Listen to the beat

A cross-linguistic perspective on the use of stress in segmentation

The research reported here was part of the project 'Parsing and Metrical Structure: Where Phonology Meets Processing' supported by the Netherlands Organisation for Scientific Research under project number 360-89-030.

Published by

LOT
Trans 10
3512 JK Utrecht
The Netherlands

phone: +31 30 253 6111
e-mail: lot@uu.nl
http://www.lotschool.nl

ISBN: 978-94-6093-202-1
NUR: 616

Listen to the beat

A cross-linguistic perspective on the use of stress in segmentation

Luister naar de beat

Een taaloverschrijdend perspectief op het gebruik van klemtoon in segmentatie

(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. Bert van der Zwaan,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op
vrijdag 4 maart 2016
des middags te 4.15 uur

door

Sandrien van Ommen

geboren 13 augustus 1985
te Apeldoorn, Nederland

Promotoren: Prof. dr. R. W. J. Kager
Prof. dr. W. Zonneveld

Voor mijn moedige ouders

Ik juich je sterrelings
Paul Rodenko

Contents

List of Tables

List of Figures

Thank you

Gratitude and an awareness of how lucky I am, if anyone was wondering what inspired this piece of text. I knew I had a lot of support doing this research, but trying to put everyone who was important during these years on one or maybe.. four pages made me truly aware of it, and dizzy.

René Kager and Wim Zonneveld were my great supervisors and I am very grateful for their energy, availability and support and their tireless rereading in the toughest part of the writing process. René, thank you for being kind, present, knowledgeable and approachable and for picking up on what I intend, I honestly don't know how you do it. Traveling with you was very relaxed and I am grateful for meeting so many interesting people. Oh, and sorry for the confusion in Konstanz. Thank you for the good talks to go with Belgian beer and good food. Wim, thank you for asking the right questions and for having writing and text structure as a second nature. Turns out writing a book is pretty difficult, thank you for your help. I was lucky to have your guidance when in the last years, as emeritus, you could just as well have gone mountain biking. Even better that you did both. Thank you for challenging me.

Brigitta Keij, my project-buddy and paranymph. I was lucky to have you by my side. Thank you for being there, always willing to talk. Even willing to listen. And giggle. Thank you for making puns at every opportunity. Silent disco and desserts. Thank you for being my friend.

Anna Sara Romøren, my other paranymph. Thank you for pointing out the beauty wherever you go, for the many conversations and the absurd jokes and the sudden inspired discussions about either language or your pink onesie. Thank you for burnt coffee. Takk for takk, takk for alt.

I was lucky to be in contact with and have so much input from researchers at the Utrecht Institute of Linguistics OTS and many institutes all over the world. I have learned more than I can remember but hopefully also more than I can forget in the Experimental Phonology Group (EPG) meetings, Experimental Linguistics Talks in Utrecht (ELiTU), Utrecht-Konstanz meetings, Utrecht-Nijmegen meetings, linguistics colloquia, LOT schools and various conferences

and workshops.

Thank you all: Tom Lentz, postdoc of the project group i.e. groupie. Thank you for the many discussions and for your help with statistics. And one should never underestimate the importance of Friday drinks and group eatings. Ania Kijak, thank you for being so welcoming. Dagmar Schadler, for not minding my asking all of those questions when I was new in the office. Şule Kurtcebe, for being so up-beat, for transcribing my corpus. Nihan Soycan, for being the second transcriber. Rocio Romero, for being 'the voice', thanks for not breaking your tongue on my stimuli. Ana Aguilar Guevara, for almost breaking your tongue on those. Alexia Guerra Rivera, for just escaping that destiny. Theo Veenker, for your help with the experiment. Maud Verlinde and Tom Lamers, for your input during meetings and Maud, for your great work on the development experiment and article. Huub van den Bergh for your input on statistics. Hugo Quené, Sieb Nooteboom and Vincent van Heuven for your input on phonetics. Jakub Dotlačil and Hayo Terband, for bearing with me while I didn't understand anything. Giorgio Magri, for your helpful comments. Shalom Zuckerman and Aoju Chen, thank you for having me teach.

I also had many helping hands outside of UiL OTS: Güliz Güneş, thank you for your help on many aspects of the Turkish experiment, including the non-words and the pilot experiment in Groningen. Beyza Sümer, for helping with the Turkish experiment and for your guest lecture. Beste Eris, for my Turkish pilot. Mária Gósy, for your advice on the transcription-by-participants of the corpus. Krisztina Polgárdi, for judging the Hungarian stimuli. Bettina Braun and Barış Kabak, coming to the Konstanz workshop was lovely, thank you for inviting us and thank you, Barış, for your time discussing the design of the research project in those early stages, and for your help throughout the process. Gizem Köksal and Ebru Ger, for collecting the optimization data and the stress transcriptions in Istanbul.

UiL OTS and NWO made it possible for me to travel for data collection and conference attendance. Doing this research would have been impossible without it and I am very grateful for the wonderful experience. Thank you, Maaike Schoorlemmer and Yvonne van Adrichem. In Adana, Turkey, at Çukurova University, I met Meryem Akçayoğlu. Thank you for welcoming me in Adana, for your kindness and help and for asking Melike Sökmen to be my lovely assistant and Çağlar Başol for showing me around. Melike, for being kind, creative and relaxed in problem solving. Çağlar for the laughs and for your face when you tried 'drop'. It has been a lovely stay. In Wrocław, Poland, I was lucky to be helped by Dorota Klimek-Jankowska, Katarzyna Czarnecka and Wojciech Witkowski. Thank you, Ania Kijak, for putting me into contact. In Budapest, Hungary, I had two very kind and easy-going assistants: Judit Süveges and Ádám Bicsérdi. Thank you so much and good luck traveling the world. Thank you Frida Kramm for putting me into contact and for helping me with several Hungarian-related questions. At Pompeu Fabra University, Barcelona, I was welcomed by Juan Manuel Toro Soto. Thank you for being so kind, attentive and open, I felt very welcome and I enjoyed taking part in the

lab meetings. Paola Crespo, for all the testing and for showing me your other lab. Sorry for upsetting your rat. Luca Bonatti, for showing me the lab, for your input on my experiment and for the lovely lunch together with Juanma. Don't forget to drink milk, they say it's good for you. Thank you Núria Sebastián Gallés, for taking your time to discuss my work. Silvia Blanch and Fofi, I had so much fun. Thank you for having me in your apartment. Frans Adriaans, you welcomed me in New York. Thank you, for the discussions on my segmentation model, a work in progress. Together with Ilke Adriaans, for the jazz and the burgers and remember: 'shockingly clean' is still a good slogan.

I had a fun office in a beautiful attic with a view on the blossoming trees. Without balloon folding, juggling, bubble-blowing, watching someone jump to earth from space, origami, discussions on Russian literature and some more juggling, this dissertation would have been.. different. I don't know how, but I'm sure. Thank you, Heidi Klockmann, kindred spirit, dancing mind. Alexia Guerra Rivera, Misha Knyazev, Loes Koring, Daria Bahtina Jantsikene, Sophia Manika and Liv Persson for everything and correspondence chess. Thanks for bringing cookies, for standing in the kitchen with your jokes, hugs and avocado salads: Björn 't Hart, for brilliant humor and for making Anna Sara blush, Stavroula Alexandropoulou, thanks for cheating in all the games you can think of, Maartje Schulpen, for enjoying blurry vision and games, for your face when you are pleasantly insulted. Marta Castella, for great cooking and impressive perseverance, Marko Simonović, thanks for being my lovely dreadful adorable mentor and thanks for good hugs and bad movies, may no branch or twig be in the way of your biking, ever. Thank you for your tango demo, Anna Volkova. General thanks and good memories of Friday drinks and games of pool, parties and dancing. Hannah de Mulder, for making me laugh, Iris Mulders, for the way you laugh, Marijn Struiksma, for good stories, Assaf Toledo for your photography course and your help with LaTeX, for kindness and ice-cream days, Sander van der Harst, thanks for all the coffee to go with your jokes. Carolien van den Hazelkamp, I like your down to earth and cheerful attitude. Natalie Boll-Avetisyan, thank you for all the laughs and your hospitality, thank you Artjom for chicken soup. Liquan Liu, for chaos and inappropriateness and for presenting 70 slides in 10 minutes. Nadya Goldberg, for Russian auction, who knew that was my new favorite thing. Sophia Manika, for your help on my last writing day: important things were done. Thank you for muffins, movie reviews and your temperament. Thank you all: Hanna de Vries, Rob Zwitserlood, Desiree Capel, Marko Hladnik, Marjolein van Egmond, Franca Wesseling, Jolien Scholten, Anne van Leeuwen, Eva Poortman, Choonkyu Lee, Zenghui Liu, Fang Li, Anqi Yang, Mirjam Hachem, Alexis Dimitriadis, Bert le Bruyn, Xiaoli Dong, Cora Pots, Arnout Koornneef, Antonio Baroni, Ileana Grama, Hans Rutger Bosker, Jennifer Cabrelli, Ao Chen, Mengru Han, Anne-France Pinget, and anyone I may have forgotten, because it happens.

I would never have been in Utrecht if it weren't for Dicky Gilbers, Petra Hendriks and Charlotte Gooskens, my Master's thesis supervisors. Thank you for everything you taught me about linguistics and scientific research. Thank you, Dicky, for everything you taught me about phonology, for your contagious enthusiasm.

My friends outside the office have to be thanked for listening and giving advice on abstract and poorly formulated issues, for letting me blow off steam or accepting me to grumpily state that I do not wish to talk about it. For dancing, singing and laughing, which is eternally important, anyone who disagrees is wrong (although no, Ewout, I did not do any research on that topic). Elsje and Katinka, thanks for being the best neighbors, for postcards and coffee and many, many cats, and dancing and singing contests. Rick and Annika, thank you for looking good in those dancing contests. Rick, it's a good thing you already know and besides, leaving things half unspoken is your specialty. Marten, thank you for all and beyond, which sounds abstract enough but surely there's good music playing and some kind of statistics flying around. Marieke and Joselien, for speaking and listening exactly the way you do. Carla and Arno, for games, talks, jokes and extreme home makeover. Alies, for racquetball and beer and fries. Thank you, Christa, Corianne, Marielle. Sabine, for being casual ('but really, are you ok?'). Nienke S. for sudden jokes and Nienke de J. for sudden revival of our friendship. Stolzies, for shifting and breaking my frame of reference constantly, for what only you can do with literature and furniture. Peter 'I feign surprisal', Simone, stop making me laugh, no keep going, Else, for self-deprecating humor, Marjolein, for stories, Matthieu, for your dancing, Niek, for making the unreasonable sound reasonable, Michiel, willing to question or defend, whichever is less expected, Henri, come back already.

My family. Papa, mama, Mathilde and Coos, Ewout and Hui Min. We had quite the time to be grateful of each other, I don't think we need many words for it anymore. Thank you for being so supportive, throughout. You never mirrored my lack of confidence, because you were sure that I could do it, thank you for that.

Thank you Gaaike, for making me look around, instead of ahead.

Kijk lieve oma: punt erachter.

Part I

Introducing

CHAPTER 1

General Introduction

The purpose of this thesis is to investigate the relation of word stress to word segmentation in a cross-linguistic perspective. While many studies have addressed this issue before, the current one takes a typologically broad cross-linguistic approach to the use of edge-aligned word stress in processing. The investigation is concerned with language-specificity, the direction of processing and the language-specific abstract nature of stress as a leading beat. The thesis concludes with an excursion into first language acquisition, regarding the issue of whether word stress can be inferred from the distribution of stress patterns in continuous speech.

Word segmentation is the division of continuous speech into words by listeners. This is a non-trivial task, since, for example for English with about 5 syllables per second (Huggins, 1964; Barik, 1977), spoken language is fast, and words are not divided by silences (Klatt and Stevens, 1973; Cole and Jakimik, 1980b), despite listeners' strong intuition to the contrary. This strong intuition may originate from the fact that language is equipped with many cues to word boundaries, most of them language-specific. That most of these cues are language-specific becomes appreciable when listening to an unfamiliar language. Suddenly, speech sounds like a long, uninterrupted sequence of alien sounds, rather than a concatenation of words. Phonology, the structure of sounds, rhythms and melodies, has, more and more so in the past decades, been identified as an interaction between the form of a word as stored in one's 'mental lexicon' and processes facilitating articulation and processing, respectively (see e.g. Kaye, 1989). Its commonalities and differences can therefore inform us about the universal and language-specific nature of speech processing, in some cases hopefully identifying interesting tension between the two.

The idea that stress can function to mark word boundaries dates back to a distinction made by Trubetzkoy (1939/1969).

> In addition to the phonological means serving to distinguish individual units of meaning (sememes), each language has a number of means that effect the delimitation of such individual units of meaning. These two functions of sound, that is, the distinctive and the delimitative function, must be carefully distinguished. (Trubetzkoy, 1939/1969, :273)

One of these means that languages have at their disposal is word stress. In some languages, word stress differentiates meaning between two words, for example in English, where the word *permit* is a noun when it has initial stress (/ˈpɜrmɪt/, 'official document granting permission to do something') and a verb when it has final stress (/pərˈmɪt/, 'to allow (something) to happen'). In other languages, word stress is fixed to a certain position in the word, as in Hungarian. Here, word stress consistently falls on the first syllable and it cannot therefore have a distinctive function. It clearly has, however, a delimitative function: the edge-prominence marks boundaries.

> Finally, the so-called "non-free" or "fixed" accent is also a non-phonemic boundary signal. Since all words with the same number of syllables (or morae) always carry this accent on the same syllable (or mora), the position of the accent cannot differentiate the meaning of words. However, it always indicates the relationship of the accented prosodeme to the word boundary. (Trubetzkoy, 1939/1969, :277)

Irrespective of whether a language has initial, final or even penultimate or antepenultimate stress, stress has a delimitative function whenever it is fixed. Interestingly, typological evidence underlines the tendency of stress to align to the word edge even in languages in which stress is influenced by other phonological and also morphological and lexical factors (Hyman, 1977; Prince, 1983; Hayes, 1995; Kager, 1995; Gordon, 2002; Goedemans and van der Hulst, 2013). In the remainder of this thesis, fixed stress is contrasted with positionally variable stress, which is an overarching term for languages with positionally variable, but phonologically or morphologically predictable stress and languages with lexically prespecified stress. Section 2.1.2 gives an overview on the typological tendencies of word stress.

The idea of word stress (and other phonological phenomena) serving as a boundary marker is, of course, not the same as saying that fixed edge-aligned stress, or any other boundary demarcator, would be useful in segmentation. Trubetzkoy realises the difference and in fact seems hesitant to take the step from one to the other, as the quote below attests.

> The external delimitation of meaningful complexes of sound, on the other hand, is not at all absolutely necessary. In an uninterrupted speech flow these complexes can occur in succession, without

> their boundaries needing to be indicated. Whether a particular one of these 'phonic marks' (= realized phonemes) occurs at the end of a meaningful complex of sound (= word or morpheme), or at the beginning of the immediately following complex of sound, can in most cases be surmised from the entire context. The possibility of a misunderstanding is in most cases very slight, especially as usually when one hears a linguistic utterance one is already attuned to a specific and very limited conceptual sphere, and one needs only to consider the lexical elements that pertain to this sphere. Still, each language possesses specific, phonological means that signal the presence or absence of a sentence, word, or morpheme boundary at a specific point in the sound continuum.(Trubetzkoy, 1939/1969, :273-274)

Why would each language possess specific, phonological means that signal the presence or absence of a boundary if there were no need for them? It was a few decades later when researchers started to unfold the difficulty of the seemingly simple task of speech segmentation, finding that boundary demarcation is indeed necessary, supposedly increasingly so under circumstances of noise (Mattys et al., 2005; White et al., 2010). Lexicon and context are not sufficient to be effective in segmentation, let alone efficient, but fortunately there is phonology. Kaye, for instance, in his 'Cognitive Phonology' argues exactly in favor of the hypothesis called into question by Trubetzkoy. In fact, like many others he presents the importance of stress for segmentation as a given.

> Let me propose the following thought experiment: Suppose we removed the effects of all phonological processes from a stretch of speech. This would mean that we take some archetypical pronunciation for each segment. This pronunciation would remain invariant in all contexts. Boundary phenomena such as final devoicing would be turned off. Stress effects, a key to the recognition of word divisions, would be leveled. It is my belief that such a stretch of speech would be incomprehensible if played back at normal speed. If comprehensible at all, it would be necessary to slow down drastically the rate of transmission. If I am correct, Phonological processes serve to facilitate parsing. [...] Many phonological processes have obvious delimitative effects, that is, they give information about domain boundaries (word, phrase, sentence). Stress systems are clear examples. (Kaye, 1989, :50-51).

Acknowledging that there is a wide variety of boundary markers, the reason that stress may stand out in the crowd, as also argued by one of the first studies on the use of stress in segmentation: Cutler and Norris (1988), is that a stressed syllable is a phonetically salient portion of speech. Stressed syllables are expanded in time and spectrum, as well as in relative amplitude and pitch[1] and prosodic information is robust as opposed to segmental information, which suffers loss under conditions of noise or channel degradation.

Considering these characteristics of stress, it seems that stress will be an effective segmentation cue. There have been many empirical studies in which a role of stress in segmentation was itself the topic of investigation. These studies are surveyed in Chapter 2, where the evidence in favor of the use of stress in segmentation turns out often to be re-analysable as confounded with phonetic edge salience. In Part II, the current thesis aims to investigate the phonological role of word stress as both an edge-aligned segmentation cue (Chapter 4) and a perceptual grouping cue (Chapter 5). This is done by taking a cross-linguistic perspective on phonological word stress. The thesis links these two potential roles of edge-aligned stress to two potential segmentation mechanisms, of which one is anticipation, and the other prosodic optimization. The first hypothesis will be discussed in Chapter 4, but a quick glance forward follows below.

The first hypothesis predicts a role in processing for the delimitative function of stress. Listeners use word stress as a beat, like a language-specific metronome, in processing.

> **Language-specific Metronome Hypothesis**
> Listeners have language-specific knowledge of the relation of the stressed syllable to the word boundary. During processing, the stressed syllable triggers expectations on the position of the word boundary, based on this knowledge.

For this hypothesis, evidence will be presented in Chapter 4. The non-word-spotting experiment described in that chapter is designed to compare the use of stress in segmentation in Hungarian, Polish, Turkish and Dutch. This comparison of languages is made because of the typological differences between them. Not only does the canonical position of stress differ between the languages, covering the three most common word positions in stress typology, the stress systems of these languages also differ on two other dimensions. On the one hand, there is a difference between languages with fixed and positionally variable stress and on the other hand there is a difference in how close canonical word stress is to the word boundary. Both of these dimensions together potentially influence how 'reliable' stress is in segmentation, that is how likely it is that segmentation predictions on the basis of word stress turn out to be correct, and thus how well listeners can rely on word stress in segmentation. Further relevant details of these languages are given in Chapter 3.

[1] These phonetic correlates are generally at the speaker's disposal to realize stress, although languages differ in their phonetic realization of stress, and these phonetic correlates are used for other purposes, too. Examples of these other purposes are the use of duration for phonemic vowel length and the use of pitch for sentence intonation. The phonetics of word stress will be further discussed in Section 2.1.1.

The Language-specific Metronome is the simple prediction that word stress is a marker and potentially predictor of word boundaries. However, word stress is part of a hierarchical prosodic system and as such does not only relate to boundaries, but also to units, in this case words. In languages with positionally variable stress, words have different stress patterns which differ in how metrically optimal they are. It is therefore hypothesized that listeners, when given time to evaluate, use word stress as a perceptual grouping cue.

> **Language-specific Metrical Grouping**
> When given time, listeners use the stressed syllable as a grouping cue. Their perception of the position of word boundaries is influenced by optimal metrical units in their language and, because of the bidirectional nature of grouping, differs from the expectations based on the Language-specific Metronome.

To investigate whether this supposed difference between the Language-specific Metronome and Language-specific Metrical Grouping has empirical ground, the two positionally variable stress languages Turkish and Dutch are compared in an offline experiment in Chapter 5 and this offline experiment is compared to the online experiment of Chapter 4. The offline segmentation experiment tests how Turkish and Dutch listeners group syllables when word stress is the only factor constraining the listeners' options. The expectation is that listeners in this case take the entire stimulus into consideration to metrically optimize their segmentation, where in the previous experiment they were likely to only look forward. This difference is visualized below.

fast, one-directional processing

w S w S S w w

$\ldots \rightarrow] \rightarrow$!

bidirectional optimization of segmentation

$\swarrow \downarrow \searrow \downarrow \swarrow \downarrow \searrow$

w S w S S w w

In this experiment, too, the cross-linguistic perspective makes it possible to compare general and language-specific preferences. Additionally, the hypothesized unidirectional and bidirectional mechanism are compared, which is possible since the two experiments were designed with comparable materials.

In Part III, the focus of the thesis switches to the information available in the unsegmented stress distribution of languages, specifically the information available to the language learning child in first language acquisition. This perspective is relevant under the hypothesis that word stress would be useful in infant word segmentation. Infant word segmentation is different from adult word segmentation because it pertains to the discovery of words, rather than their recognition. Is it possible to distill the word stress system from an unsegmented speech stream?

In Chapter 6, a simple bigram distribution analysis is performed. The stress distribution of a small corpus is compared between the four different languages described in Chapter 3 and probability measures are compared on their ability to make the word stress pattern of a language stand out from continuous speech. The analysis is subsequently repeated on a corpus of spontaneous unsegmented speech, both in Turkish and in Dutch, in which the position of stress is based on a perception survey.

In Chapter 7, an analysis of boundary patterns is hypothesized to improve the distinctiveness of language-specific word stress patterns in language acquisition. It is investigated whether the relation of perceived stress to the phrase boundary, again computed over the Turkish and Dutch spontaneous speech corpora, is useful for the acquisition of word stress, both for singling out the correct edge and for inferring the correct pattern.

Lastly, Part II and Part III are discussed together after which the thesis concludes in Chapter 8.

CHAPTER 2

Literature Review

2.1 Word stress

2.1.1 What is word stress

This thesis builds on the hypothesis that word stress is used in spoken word processing to find word boundaries and to structure speech. To be able to inquire into the role of stress in continuous speech perception, we first need to define what stress is.

Stress is a term used for the prominence of syllables. For example, in the word ‘agenda’, the middle syllable is more prominent than the first and the third. This prominence has both a phonetic and an abstract phonological dimension.

The phonetics of stress

Phonetically, there are conspiring acoustic properties that make the stressed syllable more salient than the unstressed syllable(s) and this salience is intuitively perceived as ‘strength’ or ‘loudness’. This salience was initially proposed to be a result of expiratory effort. As reviewed by Hayes (1995), stress was thought to be directly reflected in the respiratory muscles which supposedly expelled a ‘breath pulse’ for each syllable. The stressed syllable was in this view enforced with extra effort (Stetson, 1928). This idea was soon refuted, but the idea that stress could be measured by examining expiratory effort remained (a.o. Fonagy, 1966; Ladefoged, 1967). Later work showed that even this claim could not be empirically corroborated, showing that the ‘breath pulse’ was only found in some, especially emphatically stressed, syllables (e.g. Ohala, 1977), deeming the idea of respiratory muscles as the basis for a phonological stress contrast as outdated. However, perceptual intuitions to attribute stress to effort and loudness remained an important subject for phonetic studies.

Studying the perception of stress, Fry (1955, 1958) varied pitch, loudness and duration on disyllabic English words, after which participants judged the position of stress. He found that loudness contributed least to perceived prominence, and duration and pitch were much more reliable. These results were corroborated by Morton and Jassem (1965) in a similar experiment with non-words and English listeners. The authors also found that duration and loudness could be varied together, leading to a cumulative effect and the conclusion that loudness on its own appears to be insufficient. The psychological dimension of loudness is not well measured by a peak amplitude, but rather by a temporal summation of loudness, that is the combination of loudness and duration (Beckman, 1986). Another source of perceived loudness may be sought in spectral balance (Sluijter and van Heuven, 1996b; Sluijter et al., 1997), which is the finding that in Dutch stressed syllables, higher frequencies increase more in amplitude than lower frequencies do. This tilted amplitude balance was already reported by Fonagy (1966), for Hungarian, in a 1.8 and 2.3 dB increase in the second and third formants of stressed vowels, and described by Fant (1960) as a correlate of voice effort. Other languages for which spectral balance was found to be a phonetic correlate of stress are Polish, Macedonian and Bulgarian (Crosswhite, 2003) and American English (Sluijter and van Heuven, 1996a,b).

It looks as though loudness and duration, or spectral balance and duration, could together form a measure of stress through articulatory effort. Phonetic studies into stress additionally consistently find a richness of cues which conspire in the phonetics of stress. What is generally found is that the vowel quality of stressed syllables is higher (unstressed syllables are reduced), their duration is longer, onset consonants are articulated more carefully or more forcefully and pitch is either higher or has clear changing contour. All but pitch can be lead back to articulatory effort, and pitch has long been known to be the correlate of phrase-stress (a.o. Bolinger, 1958; Beckman, 1986). Whether or not word stress is indicated with pitch in its own right has been addressed in studies on among other languages Dutch (Sluijter and van Heuven, 1996b), Spanish (Ortega-Llebaria and Prieto, 2007) and Polish (Dogil and Williams, 1999), thereby tackling the methodological issue that previous studies collapsed the word and phrase domain by producing words in isolation. For Dutch and Spanish it was found that pitch was not a correlate of word stress, but there are also languages in which pitch but not duration is in fact the main correlate of word stress, such as Turkish (Levi, 2005). This is what Hayes (1995) calls the ‘parasitic’ nature of stress, being the idea that stress invokes phonetic resources which are additionally used for other phonological ends and which therefore necessarily differ between languages. Under this hypothesis, languages with phonemic vowel length would avoid using duration for perception of the stress contrast (Berinstein, 1979), although durational differences have been reported in production for example for Hungarian (Fonagy, 1966). A more extensive review of the literature can be consulted in Hayes (1995).

What remains is that there is an arsenal of phonetic cues available for the realization of word stress and that languages differ in their employment of these cues. What these cues have in common is that the phonologically stressed syllable is phonetically salient in the speech signal. But what is the phonological reason for prominence relations?

The phonology of stress

As already alluded to above, there are different levels at which roles of stress can be distinguished. Words have word stress, and phrase stress docks onto this pattern of word stresses (metrical pattern) to add information structure. Phrase stress can fall on any of the word stresses in the phrase, thereby changing meaning without changing the metrical pattern.

Word stress, on the other hand, can be understood to have two different roles, harking back to Trubetzkoy. In languages with lexical stress, such as Russian, words are stored with their specific prominence patterns. In this language stress can be understood as a phonological means to distinguish individual units of meaning (see Chapter 1) and it is in that sense comparable to phonemes. In languages with fixed stress, on the other hand, stress always falls on a specific position of the word. An example of such a language is Hungarian, which always has word-initial stress. In these languages, it is less likely to assume stress to be a meaning differentiator, it would instead be more functional as a boundary demarcator. Lexical stress, in contrast, is less reliable as a boundary demarcator, because it does not have a fixed or predictable relation to the word boundary.

There are many languages which fall somewhere in between these two extremes of fixed stress and lexical stress. Dutch, for example, has words which differ minimally in stress (see examples 2.1 and 2.2) and structurally similar words with differing stress patterns (see example 2.3 and 2.4).

(2.1) *canon*
/ˈkanɔn/
‘antiphon’

(2.2) *kanon*
/kaˈnɔn/
‘cannon’

(2.3) *mecenas*
/meˈsenɑs/
‘Maecenas’

Table 2.1: Frequencies of metrical systems

		Left	**Right**
Fixed	True Edge	92	51
	Pen.	16	110
	Antepen.	1	12
Weight-sensitive	True Edge and Pen.	37	65
	Antepen. involved	2	27
Other	Unbounded	54	
	Combined	8	
	Not predictable	26	

True Edge = peripheral, Pen. = penultimate, Antepen. = antepenultimate
Frequencies from StressTyp Database (Goedemans and van der Hulst, 2009)

(2.4) *ananas*
/ˈɑnanɑs/
'pineapple'

This example shows that in Dutch, stress is used to differentiate meaning and that there is relative variability in stress position. However, there is also strong regularity in the stress system of Dutch. In particular, the penultimate syllable in Dutch is the preferred place for stress in monomorphemic, syllabically simple words. Stress may therefore, in this language, have a dual role, although it is unclear whether boundary demarcation through stress in a language such as Dutch is clear or regular enough for it to be functional. Interestingly, typological research has found that stress in languages with positionally variable stress is often edge-aligned, which suggests a potential demarcative role for stress even when it is not truly fixed.

2.1.2 Typology

Typologically, the three most frequent stress positions are initial, penultimate and final. This can be seen in Table 2.1, where the number of languages for each main stress position are given. The table indicates the syllabic distance of stress to the word edge and it differentiates between fixed and weight-sensitive stress languages. The latter distinction indicates whether stress is influenced by internal syllable structure, depending on a difference between light and heavy syllables. In languages in which this is the case, stress may be phonologically completely regular, but positionally variable. This weight-sensitivity (also called quantity sensitivity) is further explained in Section 2.1.3.

The table makes clear that stress is naturally edge-aligned: in a majority of languages the stressed syllable is either directly aligned to the left or right edge, or aligned to the edge while separated by one or two unstressed syllables. The most common position for stress is the penultimate syllable.

2.1.3 Metrical phonology

Metrical phonology essentially formalizes the typology of stress systems, dealing with similarities and differences between stress systems. What is more, it is a field that developed around the central claim that stress is the linguistic manifestation of rhythmic structure (Liberman, 1975; Liberman and Prince, 1977).

> If the equation of stress and rhythmic structure is valid, then we automatically account for why there is no invariant physical realization for stress. The reason is that rhythm in general is not tied to any particular physical realization; one can detect and recognize the same rhythm irrespective of whether it is realized by (for example) drumbeats, musical notes, or speech. because of this independence. we are not bound to the prediction that any particular phonetic correlate will invariably realize stress in any particular language. (Hayes, 1995, p.8-9)

To be able to look at stress patterns as an overlaying rhythm at the syllabic level, it is necessary to see stress as separate from linear segments. The difference between the 'segmental' and 'suprasegmental' level was already debated around the 1940s (Haugen, 1949), but in the influential Sound Pattern of English (SPE) (Chomsky and Halle, 1968), stress is still an individual property (distinctive feature (Jakobson et al., 1951)) of the segment, in particular the vowel, which is the segment most apt to carrying the acoustics of stress because of its sonorant qualities. In SPE, the segment is a unit in a linear relation to its surroundings, undergoing change according to numerous context-specific rules. This system inaptly conceptualizes stress as a feature that is bound to the vowel. Reconceptualizing, since the 1970s phonology evolves to allow space for suprasegmentals, moving towards nonlinear phonological theory. This view is fruitful for research on tone, stress and vowel harmony and it can initially roughly be divided into autosegmental phonology (e.g. Goldsmith, 1976) and metrical phonology (e.g. Liberman, 1975; Prince, 1976; Liberman and Prince, 1977; Selkirk, 1980; Hayes, 1980; Prince, 1983; Halle and Vergnaud, 1987; Hayes, 1995; McCarthy and Prince, 1993). Lifting stress to the suprasegmental level gave space to conceptualize both the relative prominence relations between syllables and words and the temporal aspect of a prominence pattern. Prosodic categories such as the syllable, the foot and the prosodic word were introduced (Prince, 1976; Selkirk, 1980) as well as the interaction of the internal syllable structure with stress (Selkirk, 1980; Hayes, 1980). The metrical foot groups stressed and unstressed syllables into (binary (Hayes, 1980)) rhythmic units and this proved to be indispensable in metrical typology but also in segmental

Table 2.2: Iambic-Trochaic Law stress typology by Hayes (1985)

	Iambic	**Trochaic**
Quantity-Insensitive	5	32
Quantity-Sensitive	15	0

phonology. This development changed stress from a property of vowels to an organizing rhythmic structure. What is won is that the concatenation of stressed and unstressed syllables behaves according to a system that is autonomous but not independent of segmental structure. Its autonomy is reflected in the operation of stress patterns on the foot- and word domain, rather than on the segmental domain. The fact that it is not independent is reflected in weight sensitivity, that is the sensitivity of stress to internal syllable structure. Coupling this weight sensitivity to the introduction of rhythmic feet, Hayes (1985) discovered a striking asymmetry in stress typology. He found that linguistic rhythm conforms to the general law of rhythm that duration contrasts are perceived as iambic, while intensity contrasts are perceived as trochaic. This is referred to as the Iambic Trochaic Law and Table 2.2 shows the typology by Hayes (1985). What is important about this finding is that his earlier theoretical framework of foot structure turned out to be excessively symmetrical, thus underlining that stress cannot be seen as independent of syllable structure.

With the framework of Optimality Theory (OT) (Prince and Smolensky, 1993; McCarthy and Prince, 1993; Boersma, 1997), the succession of context-dependent rules in metrical phonology was replaced with universal violable 'constraints' which are ordered in a language-specific way, specifically 'arranged in a language-specific hierarchy' in the process of language acquisition. The theory is concerned with the interaction of the optimality of the surface form in terms of 'markedness' and the faithfulness of this surface form to the underlying form, rather than with context rules locally and sequentially affecting features. Because all constraints, which operate at different levels such as the segmental and the suprasegmental level, are ordered in one system, the autonomous but inter-dependent nature of metrical stress and segmental phonology are captured well.

The development of metrical theory and stress typology has made it more insightful that linguistic rhythm is on the one hand closely connected to segmental phonology, but is on the other hand necessarily recognized as an autonomous system. OT has however also made another aspect of phonology more insightful, which is the role of the listener. The theory allows for perceptually grounded constraints to interact with articulatorily grounded constraints, a minimization of perceptual confusion and a minimization of effort for the speaker and other linguistic constraints which may not be categorized in this way. The rhythmic structure of language as a temporally organized system needs to be seen in the context of other rhythms, as does the asymmetry in stress systems. The Iambic Trochaic Law has shown how perceptual rhythmic grouping explains a typological asymmetry. It would be expected that a perceptual explanation for the edge-alignment of word stress would be more simply explained, but the following section shows that the current empirical evidence is not unambiguous.

2.2 Segmentation

2.2.1 Lexical access

Early on in psycholinguistic literature, the segmentation problem was compared to the problem of reading a written piece of text without spaces (Kaye, 1989). This analogy implies that if word segmentation were a segmental problem, because comparing spoken language to written text collapses and simplifies continuous speech to delimited segments, it would be a very difficult task. This is reminiscent of the study of delimited segments in segmental phonology and just as the development of word stress from a segmental feature to a suprasegmental system, the study of word segmentation needs enrichment with an added layer of processing.

Phonetically, it can be assumed that word boundaries are not marked reliably, because words are not typically separated by pauses (Klatt and Stevens, 1973; Cole and Jakimik, 1980b). If there would be no other way in which boundaries were marked, this would mean that embedded words and boundary-straddling words were problematic for segmentation. In early models of word segmentation namely, a word's segment sequence activates multiple lexical units in parallel (e.g. Marslen-Wilson and Welsh, 1978; Cole and Jakimik, 1980a; McClelland and Elman, 1986), eventually leading to higher activation for the unit that is the best match with the input. For polysyllabic words, this point of highest activation is met at the 'uniqueness point' of the word. A problem with this approach is that for some words, there is no uniqueness point, especially for many monosyllabic words, which have been found to be recognized much later than their offset, that is as late as the occurrence of subsequent words (Grosjean, 1985). Furthermore, embeddings and straddling words somehow need to be inhibited, and, during the process, discarded as potential targets. Later research also found that competing lexical elements are not just those that have the same initial segments; effects have been found of global similarity, meaning that, upon hearing a word, competing lexical candidates are activated not just when they share the same onset (Luce and Pisoni, 1998). Later work found that early in the recognition process lexical competitors which match in onset are activated, whereas later in the recognition process, activated competitors are more globally similar (Magnuson et al., 2007). In short, there are many obvious reasons why it would be efficient for the listener to have a non-lexical way to identify word onsets and in fact, there is evidence from ERP research that listeners pay selective attention to word onsets as compared to other time points in the signal (Astheimer and Sanders, 2009), also when it concerns newly learned non-words (Sanders et al., 2002).

In fact, the phonology of languages contains myriads of delimitative cues. First of all, there are sublexical segmentation cues in segmental information. An example is phonotactics, which is the study of a languages' phonological restrictions on syllable structures and words. These restrictions can be as universal as needing a sonorant, in most languages a vowel, in the syllable nucleus, and they also include the language-specific permissibility of consonant clusters at word edges. Clusters which are not permissible or highly improbable within a word, such as for example the sequence /mr/ for English, can be recognized as straddling a word- or morpheme boundary, which makes phonotactics a useful cue for word segmentation (McQueen, 1998), even already for infants of 9 months old (Mattys et al., 1999).

Smaller than the segment, there is a role for the phonetic detail in segmentation. An example is allophonic variation and assimilation, providing the listener with positional cues (Nakatani and Dukes, 1977; Lahiri and Marslen-Wilson, 1991). Assimilation between segments shows that there is no boundary between them, or a weak boundary, such as a morpheme boundary. Different allophones within a language, in turn, provide cues about their position in the word. In aspiration languages such as English or German, for example, a stop sound such as /t/ is pronounced with clear audible aspiration at the onset of a stressed syllable, whereas the same phoneme will be pronounced without aspiration, or even with a different articulatory gesture such as a flap, when placed before an unstressed syllable, or in a syllable coda. There is evidence from American English and Dutch (Wightman et al., 1992; Fougeron and Keating, 1997; Cho and McQueen, 2005) that the consonantal strengthening which happens at the onset of the stressed syllable actually increases in magnitude in higher domains, meaning that, for instance in American English, there is more aspiration at the beginning of a word and respectively a phrase than at the beginning of a metrical foot. Combined with final lengthening, which is the increased phonetic duration of domain-final syllables, domain boundaries are phonetically cued.

Word stress is another cue which is likely to be useful in the identification of word boundaries. As described in the previous section, word stress is typologically commonly aligned with word boundaries, which makes it a useful cue for segmentation. Additionally, stressed syllables are phonetically salient and stress renders the segmental identity of the syllable more salient. Taft (1984) therefore tested two hypotheses on the use of stress in segmentation, starting in English. The first hypothesis was the Salience to Onset strategy, which predicted that languages mark onsets for them to be salient to listeners during processing, and that in English, this salience is brought about by word stress. Evidence from the segmentation of phonetically ambiguous items such as 'lettuce'/'let us' showed that listeners do not take unstressed syllables to be word onsets, unless they are utterance-initial, as predicted by the Salience to Onset Strategy. Furthermore, still in line with the Salience to Onset Strategy, participants were slower to recognize words with initial stress such as 'cactus' when stress was incorrectly placed on the second syllable, but a similar inhibitory effect was not found for words with final stress such as 'suspense' when stress was incorrectly placed on the initial syllable. Taft (1984) additionally considered the Prosodic Domain Strategy, which predicted that words would be more easily segmented when the word boundary coincided with the boundary of a metrical foot, but there was no statistical evidence to support this hypothesis.

Cutler and Norris (1988), around the same time, posited the Metrical Segmentation Strategy (MSS), which was later renamed to the Rhythmic Segmentation Hypothesis (Cutler and Butterfield, 1992). The hypothesis under both names is that a languages' rhythmic unit influences word segmentation, meaning for instance that in stress-timed languages a stressed syllable triggers segmentation, while for instance in syllable-timed languages segmentation is guided by syllables. For instance, French listeners find it easier to spot /bɑl/ in *balcon* ('balcony') than in *balance* ('balance') (Mehler et al., 1981), supposedly because /l/ straddles the syllable boundary in the latter, but not in the former. At the same time, no such difference was found in a similar study with English listeners (Cutler et al., 1986), (but see Mattys and Melhorn, 2005), which was explained to be because in English the rhythmic unit is not the syllable, but the foot.

The concept of rhythm class was originally defined as the difference between stress-timed, syllable-timed and mora-timed languages. This categorization was based on the idea that languages differ in the isochrony of linguistic elements. In syllable-timed languages such as Spanish and Turkish, syllables would be of approximately equal duration, whereas in stress-timed languages such as English and Dutch, interstress intervals would be of equal duration (Pike, 1945). The idea of isochrony was soon found not to hold. Dauer (1983), among others, shows that there is no difference between Spanish and English in isochrony of the interstress interval. She suggests that the impressionistic difference between classes can be found in onset clusters and vowel reduction in unstressed syllables for stress-timed languages and regularity of stress for syllable-timed languages. Even with these parameters, not all languages could be classified and for instance Polish and Catalan were proposed to be 'intermediate languages'. More recent studies have provided evidence for an acoustic and psychological reality of rhythm class (Ramus et al., 1999; Grabe and Low, 2002; Ramus et al., 2003; Nazzi et al., 1998; Toro et al., 2003; White et al., 2012; Arvaniti, 2012) and the variety of durational measures show that languages are scattered along a continuum and the original tripartite division does not seem to hold. Polish, for instance, seems to display unclassifiable behavior, both in production and perception (Ramus et al., 2003).

In English, the language that was tested on the MSS, 85-90% of lexical words start with strong syllables (Cutler and Carter, 1987). Strong syllables are syllables in metrically strong positions and here, it is crucial to note the difference between strong syllables and word stress. In English, word stress largely coincides with vowel quality (c.f. Hammond, 1984; Gussenhoven, 1991), meaning that the large majority of unstressed vowels is fully reduced, to such an extent that listeners of English are insensitive to any other acoustic stress cues than vowel quality in lexical recognition (Cutler, 1986). English is exceptional in this respect; even a closely related language such as Dutch differentiates between stress and vowel quality (Cutler and Van Donselaar, 2001). Dutch has stressed full vowels, unstressed full vowels, and unstressed reduced vowels (Kager, 1989; Trommelen and Zonneveld, 1989). Languages which lack significant vowel reduction but do have word stress are too numerous to list here. Evidence from the word-spotting task in Cutler and Norris (1988), where participants were asked to spot 'mint' in /mɪntejv/) and /mɪntəf/, showed that indeed strong syllables trigger segmentation[1], which was predicted by the Metrical Segmentation Strategy (MSS). However, it is unclear whether these results can be seen as evidence for the use of language-specific word stress in segmentation. Due to the segmental nature of stress in English, the results could alternatively be seen as lexical activation through segmental quality. This is indeed what was suggested by Vroomen and de Gelder (1995), when they controlled for cohort size of the target word in a replication of the study, this time with Dutch participants. Quené and Koster (1998), too, translated the study by Cutler and Norris (1988) to Dutch, and because they found the effects of vowel quality both in their experiment and in their control experiment, where targets were presented in isolation, they concluded that the reaction time differences could not be attributed to the MSS. The lack of an effect was explained by the difference between the two languages in how stress is realized, as discussed above. For English, the MSS was replicated in Mattys et al. (2005), but only for degraded speech. Under circumstances of clear speech, they argue, listeners increasingly prefer lexical strategies. In White et al. (2010), English and Hungarian word recognition was not influenced by the stress pattern of the prime, even regardless of Hungarian L2 English proficiency and this lack of an effect was again attributed to the fact that clear speech invokes lexical rather than sublexical strategies.

In the artificial language learning (ALL) study of Vroomen et al. (1998), the MSS was tested on different languages at the same time, and pitted against vowel harmony, another language-specific segmentation cue. The study tested Finnish, Dutch and French participants and initial stress was realized as a pitch peak on the first syllable. It was reported that Dutch and Finnish listeners use word-initial stress to segment words from the stream, and French listeners do not, which is in line with the fact that French does not have word-initial stress, or perhaps not even word stress at all (Dupoux et al., 1997; Jun and Fougeron, 2000), and is therefore not expected to use initial stress in segmen-

[1] Important in this discussion is the phonological principle that a stressed syllable attracts surrounding consonants, which leads to maximal onset in stressed syllables (Kahn, 1976). Evidence for this has been found in perception studies, too (Treiman and Danis, 1988). This syllabification rule also applies to intervocalic consonants, and the resyllabification of /t/ may hamper the recognition of the target word, because the coda is missing, without being, strictly speaking, a *word* segmentation effect (but see Nakatani and Schaffer 1978).

tation. Furthermore, French participants were not expected to use stress in segmentation, since French is a syllable-timed language. However, when Tyler (2006) replicated the study, using Vroomen et al. (1998)'s materials, he found the same effect for French as was found for Dutch and Finnish, and concluded that listeners are not bound to native cues in foreign language segmentation, arguing that the segmentation cue in the study by Vroomen et al. (1998) was in fact quite a salient prosodic grouping cue. Later, Hanulikova et al. (2010) found inhibition in segmentation for Slovak listeners when the target in a word-spotting experiment did not carry stress versus when it did. The results were taken as evidence for the use of fixed initial stress in segmentation. They may, however, also be taken as evidence that the presence of a stressed syllable in itself facilitates segmentation, without specification of the language of the listener or the position of stress.

The above leads to the conclusion that segmentation studies on the use of word-initial stress alone do not provide evidence for a language-specific use of stress in segmentation. Rather, the effects can be seen as task-specific, since different experiments with the same languages render different results, such as for example French and English. The effects may alternatively be interpreted as universal because no different results between languages have been found within one study, apart from Vroomen et al. (1998) which Tyler (2006) failed to replicate, and no other effect than a facilitation effect of initial stress was found. It is therefore furthermore unclear whether the results can be attributed to word stress. The effect of initial stress may alternatively be interpreted as a mere boost to lexical access (and implemented exactly that way in the model of Norris et al. (1995)) because the stressed syllable is phonetically salient, or simply because it is acoustically more recognizable (Vroomen and de Gelder, 1995; Quené and Koster, 1998). Phonologically there is, however, on top of this lexical effect, a potential of the 'beat' as a guide for the listener during processing. Rather than making the word easier to recognize, or rather than attracting attention to the word onset, word stress can be expected to have a language-specific suprasegmental role in processing. In that role, right-edge-aligned word stress may even serve as an anticipatory cue, which is arguably more efficient in prelexical segmentation than left-edge-aligned stress.

2.2.2 Anticipation

Earlier work on the use of prosody in English had already shown that a preceding intonation contour facilitated phoneme monitoring if this preceding contour predicted that the target syllable containing the phoneme was stressed (Cutler, 1976), that is, without there actually being an acoustic difference between the target syllables in the two conditions. This work provides evidence for the idea that listeners focus their attention on stressed syllables and that this happens in a forward-looking manner. Because the listener's attention is directed toward the place where a stressed syllable is likely to appear, they recognize the stressed syllable faster. A similar predictive use of word stress had already been envisioned by Martin (1972), who proposed that a listener 'locks into' the rhythm of the speaker, to, like in music, use the rhythm of speech predictively. Like any predictive mechanism, the hierarchical structure of prosody likely facilitates processing because it frees up resources to evaluate input during moments of lower information value. A study on the effect of rhythm on reaction time to an auditory target showed that a rhythmic expectation of a target, in this case a tone, lowered reaction times, indicating that rhythm helps the listener to orient temporal attention (Sanabria et al., 2011). Bohn et al. (2013) found that, in language, non-alternating stress patterns such as clashes and lapses result in processing costs for German listeners, even though these patterns are legal and common in German. Conversely, a study on the effect of rhythmic expectation in German showed that the ERP component which is typically associated with impeded lexical retrieval was smaller when the word that elicited this component fitted in the rhythm of the preceding sentence (Rothermich et al., 2012). The results of Quené and Port (2005) with American English listeners, however, suggested that the regular timing that is supposed to direct listeners' attention is better captured by temporal regularity than by metrical regularity. In sum, the idea that rhythmic regularity can be used for listeners to 'lock in' on the rhythm of the speaker seems to be justified, although the specifics are yet to be studied. The above studies suggest that the perception of a stressed syllable helps in the prediction of where the next stressed syllable will be. However, stressed syllables do not only stand in relation to each other, but also to the word boundary, since the domain of word stress is the word, as discussed in Section 2.1. It is therefore arguable that word stress is not only likely to predict the next stressed syllable, but also the position of the word boundary. In prediction, right-edge-aligned stress could even be very effective.

Typologically, there are many languages with right-edge-aligned stress, including languages with final, penultimate and antepenultimate stress (see Table 2.1). The potential of right-edge-aligned stress in segmentation is very interesting: rather than using left-edge-aligned stress as a word-beginning, right-edge-aligned stress may be used *anticipatorily*[2], as a cue that a word will follow. An example of a language which use this cue is Turkish, a language with word-final stress.

[2] A left-edge-aligned stress cue can even be understood to be regressive: the acoustic realization of stress is on the vowel, so the listener needs to regressively segment from any preceding consonants that are part of the same syllable. In the case of peninitial stress, stress would definitely cue segmentation regressively, which makes peninitial stress unpractical in this respect. Stress typology interestingly underlines this point: peninitial stress is very rare. There are 16 fixed peninitial stress languages out of 218 fixed stress languages in WALS and 39 positionally variable left-edge-aligned languages (including initial stress and antepeninitial stress languages) out of 219 positionally variable stress languages. The rarity of peninitial stress sharply contrasts the regularity of penultimate stress, of which there are 110 fixed penultimate stress languages and 92 positionally variable right edge aligned stress languages (including final stress and antepenultimate stress languages) in WALS (Goedemans and van der Hulst, 2013).

Turkish does not have fixed stress, but rather a positionally variable word-final[3] stress pattern with morphological and lexical deviations, which cause the non-canonical word stress to fall on either the penultimate or antepenultimate syllable in monomorphemic words, and even further from the right edge in polymorphemic words (Sezer, 1981; Kabak and Vogel, 2001). Kabak et al. (2010) designed a non-word-spotting experiment testing the effect of vowel harmony and stress on segmentation. Turkish and French participants were asked to spot a disyllabic non-word in an auditorily presented pentasyllabic stimulus. The non-word was presented orthographically 100-500 ms. before presentation of the auditory stimulus. Stimuli were generated using MBROLA (Dutoit et al., 1996) and stress was realized as a pitch rise and fall on the stressed vowel, while unstressed vowels had a lower pitch which slowly declined over the vowel. Turkish and French listeners both used final stress as a segmentation cue, and Turkish listeners additionally used vowel harmony. However, the use of stress in this experiment cannot be straightforwardly interpreted. Apart from the domain of stress (word vs. phrase) there was no difference between the tested languages in their canonical stress position. Since listeners of both languages used the same stress cue in the experiment, it cannot be ruled out that the cue is universal, or an artifact of the experiment. Hence, in this case an interpretation in which the segmentation cue is a language-specific strategy is confounded with other possible interpretations. Evidence from studies with penultimate stress underline this point. Any evidence for the anticipatory use of penultimate stress would disentangle the universal use of word-edge salience from a language-specific use of word stress in segmentation: If listeners of a language with penultimate stress use their canonical stress pattern in segmentation, this would be unambiguous evidence for language-specificity. Penultimate stress, unlike truly peripheral stress, does not give an immediate cue to segmentation by marking word-edges but rather prompts the native listener of a penultimate stress language to expect a word boundary further along in the signal. However, evidence from studies on the use of penultimate stress seems to make a case against the language-specific use of stress in segmentation. In Toro-Soto et al. (2007), each participant was familiarized with one out of five different artificial languages which each contained a different stress cue. Participants learning the language with penultimate stress, the canonical Spanish pattern, dropped below chance level in their performance on the subsequent 2AFC task. This performance was even below the performance of participants who learned the language with random stress cues. These results were interpreted as indicating that penultimate stress actually inhibits segmentation, shifting salience away from the edges and thereby conflicting with statistical cues. Similar conclusions were drawn from a comparable study in which English, French and Spanish listeners took part. English and Spanish listeners benefited from edge-aligned stress, but not from penultimate stress (Toro et al., 2009). Cunillera et al.

[3] Although it has to be noted that the word in Turkish is a difficult concept, due to the agglutinative nature of the language.

(2008) also tested Spanish participants in a similar design, finding that the participants learning the final-stress language outperformed the participants learning the penultimate and initial stress language. Rather than providing evidence for the use of penultimate stress in segmentation, these studies on Spanish seem to suggest that penultimate stress inhibits segmentation, even in a language in which the canonical stress pattern is penultimate. If this is the case, then earlier results on the use of peripheral stress in segmentation in other languages are open to the interpretation that they are caused by a universal facilitative effect of salience at the true edge. Any language-specific stress effects are blurred by this alternative explanation.

The interpretation that penultimate stress inhibits segmentation is not at all unreasonable, because the use of penultimate stress may be seen as a distance effect, and because penultimate stress is potentially less reliable than peripheral stress when a language has monosyllables. However, the results on Spanish may also be attributed to two other factors. First, there is the fact that the Spanish stress-system is a statistical hybrid of antepenultimate, penultimate and final stress. In the LEXESP database, 73.52% of trisyllabic Spanish words have penultimate stress (Sebastián-Gallés et al., 2000). It is possible that the penultimate pattern is not fixed or reliable enough in Spanish for it to be an effective cue in segmentation. Second, it is possible that the 20 Hz pitch increase on the stressed syllable, which is used in many of the segmentation studies, is simply not sufficient to be perceived as stress. Tyler and Cutler (2009) underlines this point. In this cross-linguistic ALL study, a 50 Hz parabolic pitch cue does not mimic word stress, but it is studied for its use as a boundary marker in segmentation with listeners of French, English and Dutch[4]. Dutch and English listeners used the pitch cue in segmentation on the left edge, and Dutch and French listeners used the same cue on the right edge. This use of the pitch movement strongly indicates a possible confound in previous studies on the use of stress in segmentation, as was for example suggested by the finding of Tyler (2006) that French listeners can use word-initial pitch in segmentation. Similarly, Cunillera et al. (2006) found enhanced attentional increase for Spanish listeners at the initial-stressed syllable in trisyllabic words, compared to non-stressed words. In this study, again, stress was realized as a 20Hz pitch increase on an otherwise flat signal.

Summarizing prior evidence for the use of 'stress' in segmentation, at least some of it may be analyzed as the use of boundary salience in segmentation. First of all, at the left edge, the use of initial stress in segmentation was not language-specific, since no counter-evidence was found for languages with non-

[4]A lengthening cue at the left and right edge was also tested. For all languages it was found that lengthening was used at the right edge but not the left. This is as expected: final lengthening and its role in boundary perception has since long been assumed to be universal (Lindblom, 1978), although Vaissière (1983) reviews evidence that final lengthening is nor universal nor innate behavior. Evidence for the languages above was however available early on (Delattre, 1966; Klatt, 1976; de Rooij, 1976) as well as more recently (e.g. Salverda et al., 2003).

initial stress; languages with non-initial stress were in fact found to use initial stress in segmentation as well, and the stress-effect could not be appreciated separately from a potential lexical effect. Second of all, at the right edge, only evidence for the use of final stress was found and languages with penultimate stress were found to be inhibited in segmentation by penultimate stress as compared to initial and final stress. Third of all, what has consistently been referred to as 'stress' has been phonetically ill-suited for the languages tested.

The question of whether the language-specific relation of word stress to the word boundary is used in segmentation is therefore clearly still wide-open. As the core chapter of this thesis, Chapter 4 will therefore describe an experiment in which four languages are compared in the same experiment, aimed at identifying whether listeners make use of their native canonical stress pattern in the anticipation of word boundaries. Chapter 3 first describes the languages that were tested. As already mentioned above, these are Hungarian, Polish, Turkish and Dutch, selected on the basis of typological criteria. Chapter 5 describes a follow-up experiment aimed at differentiating online expectations based on word stress and offline preferred segmentations based on word stress. The difference between these two hypothesized mechanisms is largely rooted in the difference between the unidirectional relation of word stress to the word boundary in Chapter 4 and the overall prosodic shape of a word in Chapter 5. In Part III, the focus of the thesis shifts to stress distributions and their relation to first language acquisition.

CHAPTER 3

Languages

3.1 Introduction

The previous chapters outlined the goal of the current thesis and the findings from the relevant literature on the subject of word stress and word segmentation. To be able to make a cross-linguistic comparison of the use of stress in segmentation, an experimental approach is taken in this thesis to test native speakers of four different languages in their use of word stress in the processing of spoken language. The linguistic characteristics of these languages are described in the current chapter, since awareness of these characteristics is necessary both for the development of predictions based on the main hypothesis and for control over possible confounding factors.

As was summarized in 2.1.2, it is typologically common for word stress to be aligned to either the left or right edge of the word. The majority of languages with edge-aligned stress have word stress on either the initial, penultimate or final syllable. Furthermore, while many languages have fixed stress, many other languages allow variability in stress position, influenced by syllable weight and morphology. In the current thesis, the main focus is on the relation of stress to the word edge. Below, a mini-typology is given on three dimensions likely to influence the use of stress in segmentation.

1) edge-alignment: left or right edge
2) peripherality: peripheral or proximate
3) regularity: fixed or positionally variable

Table 3.1: Schematic typology of edge-aligned stress with example languages

Regularity	Peripherality	Edge			
		Left edge		Right edge	
		N	Example	N	Example
Fixed	Peripheral	92	Hungarian	51	Weri
N = 282	Proximate	17	Mapudungun	122	Polish
Positionally variable	Peripheral	37	Mayalayam	65	Turkish
N = 131	Proximate	2	Archi	27	Dutch

Examples from the StressTyp Database (Goedemans and van der Hulst, 2009)

In this thesis, the hypothesis is that edge-alignment is important because of the direction of the hypothesized segmentation cue. While many studies have focused on the idea that initial stress facilitates lexical access and some have hypothesized that final stress facilitates segmentation (Kabak et al., 2010), the comparison of left- and right-edge-aligned stress has not been made and especially the forward-looking potential of right-edge marking has not been investigated to its full potential. Left-edge aligned stress cues the beginning of a word, which can be seen as immediately facilitatory for lexical access, whereas right-edge aligned stress cues the end of the word, which can be seen as a progressive cue; a cue which helps listeners anticipate an upcoming boundary. Peripherality and regularity, in turn, both influence the reliability of stress as a segmentation cue, meaning that the stressed syllable cues a word boundary with a relative certainty. When stress is fixed and peripheral, as is the case in for instance Hungarian, it is highly certain that a stressed syllable is the beginning of a word, whereas positionally variable stress at a distance from the word-edge cues the word boundary with a lower certainty. A typology with these parameters can be represented in eight cells, as in Table 3.1.

The explanatory potential of targeting a language from each cell of this typology would be impressive. However, simply to make an interesting start, the current study limits its reach to Hungarian, Turkish[1], Polish and Dutch. As can also be seen in Table 2.1, fixed initial stress, fixed penultimate stress, and positionally variable right-edge-aligned stress, that is two- or three-syllable window stress, are the most common systems within the bounded stress systems[2]. This

[1] According to the StressTyp Database, Turkish is an unbounded stress system. However, stress assignment to the final syllable is largely regular and exceptions are well-defined. See Section 3.3.5 for a description.

[2] Bounded stress systems are systems in which stress falls at a fixed distance from a boundary or another stress. In unbounded stress systems, stress can fall at an unlimited distance from a boundary or another stress, although there are other conditions which have to be met. An example of an unbounded system is that the rightmost heavy syllable should be stressed, unless there are no heavy syllables, in which case the leftmost syllable should be

Table 3.2: Selection of languages in the current study

Language	**Regularity**	**Peripherality**	**Edge**
Hungarian	Fixed	Peripheral	Left
Turkish	Positionally variable	Peripheral	Right
Polish	Fixed	Distant (Pen.)	Right
Dutch	Positionally variable	Distant (Pen.)	Right

Pen. = one-syllable distance from edge.

makes the relation of these systems to perceptual cognition very interesting: if these stress systems are common, there may be a cognitive advantage for stress to fall on these positions. This is why these metrical systems are the first suspects in a study on the use of a language-specific phonological system in a processing task.

As said, the current chapter describes the different major linguistic characteristics of the languages which are tested in the current thesis. Hungarian, Turkish, Polish and Dutch are described in terms of their segment inventory, syllable structure, word stress, phrase stress and phonetics of stress as well as other potentially relevant information, such as vowel harmony for Turkish and Hungarian.

3.2 Hungarian

Hungarian[3] (Magyar) is a Finno-Ugric language spoken by approximately 14 million speakers, of whom approximately 10 million live in Hungary. The two other most spoken Finno-Ugric languages are Finnish and Estonian and together with the Samoyedic languages, the Finno-Ugric languages belong to the Uralic language family, a language family which is not part of the Indo-European language family.

Hungarian is an agglutinative language, building words by juxtaposing suffixes. It has non-configurational syntax, meaning that it has a relatively free word order and phrase structure. The word is formed by adding derivational suffixes, and then inflectional suffixes to the stem morpheme. Each suffix represents a single morphological function and each suffix changes the previous absolute or relative stem into a new relative stem. While suffixes are important building blocks in Hungarian, prefixes are highly uncommon. Besides *leg-*, which is the superlative prefix , there are some loanprefixes such as *anti-*

stressed (Hayes, 1995).

[3]The majority of the information as found in this section is taken from Siptár and Törkenczy (2007).

and *extra-*. The so-called verbal prefixes or preverbs are syntactically better analysed as separate words, because they can be postposed, and phonologically as compound members, because they adopt primary stress.

Apart from the standard language, there are different types of non-standard speech in Hungary, including traditional rural dialects. Educated speakers from these rural areas typically speak a mixture of their dialect and the standard, referred to as the regional standard of the area. For the current study, the reference language is standard educated Hungarian.

3.2.1 Segment inventory

Hungarian has fourteen vowels, including length distinctions: /ɔ, aː, ɛ, eː, i, iː, o, oː, ø, øː, u, uː, y, yː/. The /ɔ/ is referred to as /a/ in Szende (1994), where it is, however, described as slightly rounded. A further four vowels are sometimes discussed in the literature: [a, e, ɔː, ɛː], but these are considered to be surface forms, due to lengthening and shortening processes. The language does not have any underlying diphthongs.

Hungarian has twenty-four consonants, excluding length distinctions: /p, b, t, d, c, ɟ, k, ɡ, f, v, s, z, ʃ, ʒ, x, ts, tʃ, dʒ, m ,n, ɲ, l, r, j/. All consonants can be realized as short or long, but the long ones are generally not seen as underlying geminates. Most occur as derived segments; as products of concatenation or assimilation. Genuine geminates are relatively infrequent and restricted to marginal lexical classes, such as proper names, interjections, recent loans and onomatopoeic vocabulary. There is a voicing contrast in obstruents, with voiceless stops unaspirated and voiced stops prevoiced. Voiceless stops have a strong release in the form of affrication (Göksel and Kerslake, 2005). In regressive iterative voicing assimilation all obstruents, except /v/, trigger this process. Hungarian, atypically for voicing languages, but like French and Yiddish, has no final devoicing (Jansen, 2004).

3.2.2 Syllable structure

In Hungarian words, not just any combination of well-formed syllables can make up a well-formed word. There are constraints that apply between adjacent segments belonging to different syllables, and words can be monosyllabic, but they are minimally bimoraic. This latter restriction amounts to the fact that words consisting of solely one open syllable need to contain a long vowel. This restriction holds for all stems, but affixes, function words and interjections, as well as onomatopoeic words, are exempt.

In Hungarian syllables, onsets are not obligatory, and codas are allowed. Long and short vowels are allowed both in open and in closed syllables and words can begin and end with consonant clusters. Whether or not these clusters can be seen as true onsets or codas is controversial, which is a discussion that is outside the scope of the current work. Templatic syllable structures for Hungarian are V, CV, VV, VC, VVC, CVV, CVC, CVVC, CVCC CVVCC.

Statistics of the Hungarian lexicon, as given by Siptár and Törkenczy (2007), show that in monomorphemic words, within-word clusters can contain up to four consonants, where clusters at the word-beginning and word-end reach three. All of these larger clusters are rare and most of them are found in loanwords or straddling a constituent boundary.

3.2.3 Vowel harmony

Hungarian, like many other Uralic languages, has vowel harmony. Vowel harmony is the constraint for vowels in the same word to agree on a specific dimension, such as backness, or rounding. It has been found in several studies that this phonological phenomenon, which as its domain has the prosodic word, has a role in segmentation for Finnish and Turkish (Suomi et al., 1997; Vroomen et al., 1998; Kabak et al., 2010).

The domain of vowel harmony in Hungarian is the stem and any number of suffixes. Hungarian vowel harmony is stem-controlled and directional, starting at the stem and going from left to right. Vowels agree in backness within the harmonic domain and compounds contain as many harmonic domains as they do stem morphemes. Examples of regular vowel harmony with front vowels and back vowels are found below[4].

(3.1) *perd-ül-és-etek-től*
/pɛrdyleːsɛtɛktøːl/
'from your (pl.) twirling round'

(3.2) *ford-ul-ás-otok-tól*
/fordulaːsotoktoːl/
'from your (pl.) turning round'

There are some cases in which a vowel does not harmonize with the preceding vowel. Usually, these non-harmonizing vowels are front, non-round vowels /i, iː, ɛ, eː/, and, additionally to not harmonizing themselves, they behave as transparent, meaning that they do not affect the harmony of following suffixes. This results in the examples below.

(3.3) /røvidɛn/ (*rövid-en*, 'briefly')
/hɔmisɔn/ (*hamis-an*, 'falsely')

In the former case, the transparency of the vowel is not visible, since both the suffix and the stem have front vowels, but in the latter case the vowel of the suffix does not harmonize with the vowel in the stem-final syllable, but with one to the left of that. The transparent nature of these vowels warrants the classification of these vowels as neutral.

[4] All examples were taken from Siptár and Törkenczy (2007).

Just as in the example above, some Hungarian stems have vowels differing in backness. These mixed stems can have a combination of a non-neutral and a neutral vowel, as in the example above, but they can also have two non-neutral vowels, for example a front and a back vowel. Examples of this kind are given below.

(3.4) /ʃoførnɛk/ (*sofőr-nek*, 'chauffeur'-dat.)
/nyɔnsnɔk/ (*nüansz-nak*, 'nuance'-dat.)

In this case harmony is triggered by the rightmost vowel of the stem, while still moving from left to right. Additionally, there are some stems which end in neutral vowels for which the neutral vowel selects the vowel of the suffix, and some for which the suffix can harmonize with either the front or the back vowel of the stem. These latter so-called vacillating stems appear to show sensitivity to the larger context. For example, *ezzel a pulóverell* ('with this jumper') and *azzal a pulóverall* ('with that jumper') are more frequent than *ezzel a pulóverall* and *azzal a pulóverell'*.

Non-harmonizing suffixes either contain a front non-round (neutral) vowel, or a back vowel. When the latter, the suffix will start a harmonic domain of its own. This can be called an opaque or domain-external suffix. There are also suffixes with a three-way distinction, such as for example the allative suffix *-hoz/hez/höz*[5]. These suffixes harmonize with vowel roundness, as well as with backness, resulting in *tűz-hez*('fire'-all), *víz-hez* ('water'-all), *ház-hoz* ('house'-all).

3.2.4 Rhythm type

Hungarian is considered to be a syllable-timed language. In the view of Dauer (1983), this is because it has fixed stress, a fairly simple syllable structure with few clusters, and no vowel reduction. Unpublished results of Hungarian according to the durational typology of Ramus et al. (1999) placed Hungarian in a widely variable cloud of syllable-timed and mora-timed languages (Nespor and Mehler, 2011) and more recent work found Hungarian durational differences to pattern with those in Italian and French, known as syllable-timed languages, based on the variability of vocalic intervals (White et al., 2012).

3.2.5 Stress

Word stress

Hungarian is a language with fixed word-initial stress[6]. The initial syllable of the word has primary stress, irrespective of whether the word is monomorphemic, derived, or a compound. Exceptions to this rule are very rare: there is a

[5]The allative suffix is a case marker indicating movement towards something, or in its adjacency

[6]Hungarian is the only language in the current selection of languages that has left-edge-aligned stress

class of compounds of mostly, but not exclusively, numerals, which retain stress on both members. There are some interjections with non-initial stress such as *aˈha* and some exclamations have secondary stress, such as *ˈpontoˌsan* (’Exactly’). In some cases, contrastive stress can be non-initial and some monomorphemic function words are always unstressed, such as the definite article *a/az*.

Hungarian has morpho-syntactic secondary stress which falls on the second element of a compound, on the first syllable of a deaccented word and on the first syllable of a polysyllabic word. Although there have been several different accounts additionally postulating a role for phonological secondary stress and foot structure (Szinnyei, 1912; Kerek, 1971; Hammond, 1987; Hayes, 1995; Varga, 2002), the issue remains controversial. In Siptár and Törkenczy (2007, p.22) a decided rejection of phonological secondary stress can be found:

> [...] Native intuition does not support the assumption of such regular physical patterning superimposed on Hungarian strings. At any rate, this putative rhythmic intensity alternation is phonologically irrelevant as it does not interact in any way with the rest of the phonology. [in footnote:] This is of course the reason why native intuition is unaware of the pattern even if it indeed exists (which we claim is not the case).

Kálmán and Nádasdy (1994) also take the position that there is no phonological basis for secondary stress in Hungarian and they argue that even on the phrase level there is only a difference between stressed and stressless syllables. Blaho and Szeredi (2011) found no acoustic evidence for this type of phonological secondary stress in their pilot study.

Phrase stress

Hungarian is a language with ’free’ word order. In phrases with neutral prosody, word order is rigid, whereas in phrases with non-neutral prosody, constituents can move freely around the focused part of the phrase. In a neutral phrase, as in Example 3.5[7], all words have primary stress, except for words which are lexically prespecified as enclitic. These words undergo destressing in the sentence context and join the stress domain of the previous word.

(3.5)	*ˈGéza*	*bácsi*	*ˈtáncolni*	*akar*	*a*	*ˈmagas*	*ˈfekete*	*ˈlánnyal.*
	Géza	uncle	to-dance	wants	the	tall	black	girl-with.

’Uncle Géza wants to dance with the tall black-haired girl.’

In a non-neutral phrase, complements which follow the focused constituent are destressed and this focus is therefore sometimes referred to as ‘eradicating stress’. Varga (1983) assumes some level of stress to remain after this ‘eradicating stress’, in which view this process can be seen as stress reduction. An example of such destressing or stress reduction is found in 3.6 and this type of sentence is perceived to have a contrastive, emphatic interpretation.

[7] These examples were taken from Siptár and Törkenczy (2007).

(3.6) ˈ*Jenő* ˈ*táncolni* *akar* *a* *magas* *fekete* *lánnyal.*
Jenő to-dance wants the tall black girl-with.
'It is to DANCE that Jenő wants with the tall black-haired girl'.

Hungarian has many different intonation contours and these contours can have different phonetic forms, depending on the number of syllables they spread over. A detailed analysis of the Hungarian intonational system can be found in Varga (1996).

Phonetics of stress

An important early work on the phonetics of Hungarian stress is that of Fonagy (1966). It will be explained in some detail, because the phonetic correlates of stress discussed here are important for the other languages, as well.

Fonagy (1966) reports on stress correlates, advocating stress to be an acoustic projection of articulatory effort. Parallel electrophysiological and acoustic measurements were made of Hungarian one-word sentences which were produced 20 times each by one female and one male speaker. Different acoustic correlates were isolated from the results: the formants of the vowels in stressed syllables had higher amplitudes and a broader bandwidth, especially in the higher frequency ranges. This is now referred to as 'spectral tilt' (Sluijter and van Heuven, 1996b; Sluijter et al., 1997) and it was previously noted by Fant (1960, p. 21).

> A reduction of voice effort leads to a decrease of the level of harmonics which is more prominent in the higher frequencies [...] due to a more steeply falling slope of the source spectrum envelope normally accompanying the lowering of the voice level. (Fant, 1960, p. 21)

In the same experiment, the stressed syllables were of better segmental quality, were longer and had a higher pitch. The shape of the pitch curve and the amplitude curve also seemed to be important acoustic correlates of the acoustic projection of stress. In further experiments in the same study, Italian and Romanian recordings containing short sentences were presented to 21 Hungarian speaking subjects. They misjudged stress in 5.6% of the cases. This error rate could be considered to be surprisingly low in the light of the 'stress deafness' literature (Dupoux et al., 1997, 2001). 'Stress deafness' is a term which concerns the lowered ability to lexically encode stress for speakers of languages with fixed stress. The argument is that in these languages, and mostly so in languages with fixed peripheral stress, it is lexically unnecessary to encode the position of stress. Hungarian is therefore the perfect example of a language in which listeners are expected to be 'stress deaf' and this is what Peperkamp and Dupoux (2002) found. The difference with the results of Peperkamp and Dupoux (2002) and the stress judgment task of Fant (1960) is that in the latter, it is not necessary to lexically encode stress, but merely to acoustically perceive it. Most of the students perceived the position of stress correctly even in cases where stress was produced with low pitch. When making errors, the participants typically perceived the first syllable to be stressed.

3.3 Turkish

Turkish [8] (Türkçe) is a Turkic language, spoken by approximately 65 million speakers, 60 million of whom live in Turkey. Within the Turkic family, which includes such languages as Uighur, Uzbek, Tatar and Kazakh, Turkish forms part of the southwestern or Oghuz branch, with as its closest relatives Gaguaz, Azerbaijanian and Turkmen.

Like Hungarian, Turkish is an agglutinative language, building words by adding suffixes and clitics to the stem morpheme. The same rules of affixation apply to both suffixes and clitics, but their roles in prosody are distinct. Turkish has monosyllabic words.

There are several dialects in Turkish. Modern standard Turkish is based on the dialect of Istanbul, and there is a leveling influence of the standard through the educational system and mass media. Despite this influence, much variation persists. For the current study, the reference language is standard Turkish.

3.3.1 Segment inventory

Turkish has eight vowels: /a, e, i, o, u, ɯ, œ, y/. These vowels differ in height, roundedness, and backness. This is important to realize, considering vowel harmony processes.

The language has twenty-four consonants: /p, b, t, d, k, ɡ, c, ɟ, tʃ, dʒ, f, v, s, z, ʃ, ʒ, h, ɣ, m, n, ɫ, l, r, j/. Like Hungarian, Turkish juxtaposes surface consonants as long consonants. Aspiration contrasts in prevocalic stops, with voiceless stops aspirated and voiced stops unaspirated and tenuis, meaning that voicing starts right after the release of the stop. Prevoicing, however, has also been reported for word-initial stops (Kallestinova, 2004). The velar consonants /k, ɡ, ɫ/ and the glottal consonant /h/ undergo palatalization in front-vowel environments. Stops neutralize to devoiced in coda position.

3.3.2 Syllable structure

Most Turkish stems are monosyllabic. The templatic syllable structures are, in order of frequency, VC, CV, CVC, VCC and CVCC. There is only one stem containing only a vowel (*o*, the third singular pronoun) and this word requires a consonant to be able to combine with a suffix. Consonant clusters in the onset only occur word-initially, and only in loanwords. Some speakers insert [i] or [ɯ] to adjust the cluster to the language's phonotactics.

[8] The larger part of the information in this chapter is taken from Underhill (1976); Göksel and Kerslake (2005).

3.3.3 Vowel harmony

Like Hungarian, Turkish has vowel harmony which is stem-controlled and directional, going from left to right. It is characterized by agreement in backness. In Turkish, unlike in Hungarian, suffixes with high vowels also agree in roundness with the stem.

Disharmonic stems have vowels of different backness, as for example *insan* ('person'). The suffix will harmonize with the last syllable of the word. There are, however, also exceptions to this rule. Some loanwords select the suffix with the front vowel even though the last syllable of the word contains a back vowel. These latter words often, but not always, end in /l/. An example of this kind is *kabul-ü* ('reception'-obj.).

Regular suffixes can be divided in E-suffixes, suffixes alternating between /e/ and /a/, and I-suffixes, suffixes that alternate between /i/, /ɯ/, /u/ and /y/. Invariable suffixes are of both native and foreign origin. They are opaque in that the following suffix will harmonize with its last syllable, rather than harmonizing with the quality of the stem. An example is *-iyor* in *gel-iyor-sunuz* ('you are coming'), which also shows that such a suffix can be internally disharmonic. Here, the stem has front vowels, but the last syllable of *-iyor* has a back vowel, affecting the vowel selection of the following suffixes.

3.3.4 Rhythm type

Turkish rhythm type is comparable to that of Hungarian. According to the typology of Dauer (1983) it fits because the language has relatively simple syllable structures and no vowel reduction, although according to this typology a typical syllable-timed language has fixed stress. Turkish, like Hungarian, falls within the widely variable cloud of syllable-timed and mora-timed languages of the durational variability typology of Ramus et al. (1999), according to unpublished results which are reported in Nespor and Mehler (2011).

3.3.5 Stress

Word stress

Turkish has canonical word-final stress. This applies to monomorphemic words as well as suffixed words, as can be seen below [9]

(3.7)	/oˈda/	(*oda*, 'room')
	/odaˈlar/	(*odalar*, 'rooms')
	/odalarˈda/	(*odalarda*, 'in the rooms')

However, this system also has classes of exceptions. There are, for example, stems which have lexically specified stress. In this case, the suffixes do not take on final stress, as can be seen below.

[9]The examples in this section were taken from Underhill (1976)

(3.8) /ˈmasa/ (*masa*, ‘table’)
/ˈmasalar/ (*masalar*, ‘tables’)
/ˈmasalarda/ (*masalarda*, ‘on the tables’)

Besides frequent words as *masa* (‘table’), *evet* (‘yes’) and *hayır* (‘no’), there is a class of exceptionally stressed polysyllabic words, which are typically loan words and place names. These exceptions have been described and analyzed by Sezer (1981), resulting in the Sezer stress rule: ‘If the antepenultimate syllable is heavy and the penultimate syllable is light, stress the antepenultimate syllable. Otherwise, stress the penultimate’. Examples are given below.

(3.9) /ˈaŋkara/ (*Ankara*)
/anˈtalja/ (*Antalya*)
/aˈdana/ (*Adana*)

Interestingly, this pattern is said to be productive in place names, as can be seen below.

(3.10) /sirkeˈdʒi/ (*sirkeci*, ‘vinegar seller’)
/ˈsirkedʒi/ (*Sirkeci*)
/kulakˈsɯz/ (*kulaksız*, ‘without ears’)
/kuˈlaksɯz/ (*Kulaksız*)

Subsequent debate about the nature of exceptions followed Sezer’s work. According to some, these exceptions are a combination of lexical specification and morphosyntax (Kabak and Vogel, 2001; Revithiadou et al., 2006), while according to others, this is a class of words with a separate ‘cophonology’ (Inkelas, 1999; Inkelas and Orgun, 2003). A statistical analysis of two corpora, by Revithiadou et al. (2006), additionally shows that 55% of the data that should follow the Sezer stress rule, do so, concluding that there is ‘some’ stress attraction by syllable weight, but another 35% of the data that should follow the Sezer pattern in fact defaults to final stress, and the remaining 10% deviates from the Sezer pattern in a different way. The authors conclude:

> These findings suggest that, at least for monomorphemic words, the Sezer cophonology is a rather shaky reality.

What is most important about the exceptional patterns is that these stems, like words with lexically specified stress, do not lose stress when suffixes are added.

Besides stems with exceptional stress, there are also suffixes with exceptional stress, such as *-Iyor* (prog.) and *-mA* (neg.). The former is always stressed on the penultimate syllable and the latter pre-stresses, that is, the syllable preceding it will be stressed. When these suffixes are followed by additional suffixes they retain stress. Examples are given below

(3.11) /gel/ (*gel*, 'come')
/gelˈijor/ (*geliyor*, 'come'-progr.)
/ˈgelmesin/ (*gelmesin*, 'come'-neg-2sg)

It is outside the scope of the current work to discuss the nature of exceptional stress on words and affixes, but a study by Domahs et al. (2012a) does suggest that final stress is processed in a different way than exceptional stress, interpreted by this investigator as evidence for lexical prespecification of Turkish exceptional stress.

Phrase stress

Turkish compound stress and Turkish phrase stress can be discussed together, because they are only minimally different. Both in compounds and in phrases, the main prominence falls on the left-most member. Stress on the rightmost member is reduced, more so in compounds than in phrases. This difference is contrastive (Kabak and Vogel, 2001).

(3.12) Phrase[10]

ˈsüt-beˈyaz-dır
milk-white-[epistemic copula]
'Milk is white.'

(3.13) Compound

ˈsüt-beyaz-dır
milk-white-[epistemic copula]
'(It) is milk-white'

Turkish intonation is characterized by the existence of boundary tones, which means that the end of a domain is marked by a high tone. These high boundary tones are either preceded by a low tone, or a high plateau. Apart from these boundary tones, the stressed syllables carry a high tone, which is called a nuclear accent. When words have exceptional stress, the domain will therefore have two high tones: one nuclear pitch accent and one boundary tone. When Turkish words carry canonical final stress, however, the stressed syllable and the boundary coincide, and the tone surfaces as a nuclear pitch accent according to some (Levi, 2002; Ipek and Jun, 2013), but as a boundary tone according to others (Kamali, 2011).

[10] These examples were taken from Kabak and Vogel (2001)

Phonetics of stress

Phonetically, Turkish final stress is expressed through a modest rise in pitch, and Turkish non-final stress is expressed through a more salient pitch fall. Both pitch movements are accompanied by a slight rise or fall, respectively, of intensity(Levi, 2005). In her measures, Levi (2005) also found duration to differ between stressed and unstressed syllables, but an additional discriminant analysis found this difference to be unlikely to be perceptually relevant.

3.4 Polish

Polish[11] (Polszczyzna) is a West-Slavic language of Central Europe. There are approximately 40 million speakers in Poland and about 10 million outside. The language is closely related to Czech, Slovak and the Sorbian languages of Eastern Germany.

Polish is a highly inflected language with relatively free word order. It has both prefixes and suffixes and allows compounding, as well as allowing monosyllabic words.

The Polish language has become very homogeneous in the 20th century, due to mass migration and authoritarian politics. Roughly four dialect regions are differentiated, although regional differences are small. A separate language variety, which is officially recognized as a minority language in Poland, is Kashubian. This variety is spoken in Pomerania, which is a region in the North of Poland, bordering the Baltic sea. An among West Slavic languages distinct linguistic feature of Kashubian is phonemic word stress. For the current study, the reference language is standard Polish.

3.4.1 Segment inventory

The Polish vowel inventory consists of six vowels: /i, a, u, o, e, ɯɪ/. There are no diphthongs in Polish and combinations of a vowel with a semivowel are regarded as concatenations of segments in which the vowel and the semivowel are independent units.

Unlike the vowel system, the consonantal system is quite complex. The inventory consists of 31 consonants: /p, b, t, d, c, ɟ, k, ɡ, f, v, s, z, ʃ, ʒ, ɕ, ʑ, x, ts, dz, tʃ, dʒ, tɕ, dʑ, m, n, ɲ, ŋ, r, l, j, w/. Consonants palatalize before high vowels and complex consonant clusters occur, which will be discussed below.

In a voicing contrast, voiceless obstruents are not aspirated and voiced obstruents are prevoiced. Obstruent clusters share voicing properties and most voicing assimilation in Polish is regressive. There is progressive voicing too, involving /ʃ, ʒ, f, v/. Polish obstruents undergo word-final devoicing. Further-

[11] The majority of the information in this chapter is from Gussman (2007) and Jassem (2003).

more, sonorants flanked by voiceless obstruents or by a voiced obstruent and a pause lose their voicing.

3.4.2 Syllable structure

A special characteristic of the Polish syllable structure is the possibility to have heavy consonant clusters of up to five consonants. An example of this is *skąpstwo* /skomstpfo/ ('meanness'). These clusters are especially frequent in the word-onset, where even a double affricate can occur, as in for example /dʒdʒɯstɯ/ ('rainy'). The initial C-clusters are not simplified in fluent, spontaneous speech.

Kijak (2009) reports in percentages the intuitions on syllable divisions of native speakers participating in a earlier questionnaire by Rubach and Booij (1990). She concludes that in consonant clusters consisting of two consonants, speakers either divide the cluster, having one consonant in the coda and one in the onset, or place the two consonants in the onset of the second syllable, if the sonority properties allow this. However, when more consonants are involved, speakers mostly prefer to syllabify one consonant in the coda and the others in the onset. The percentages show that speakers always have more than one option, unless the parsing of two consonants as an onset would constitute a word-internal violation involving nasals and liquids of the Sonority Sequencing Hierarchy. For example, the word *kontakt* ('contact') would always be syllabified as *kon-takt* and never as *ko-ntakt*. Note that these are dispreferred even though in Polish there are cases in which violations of the Sonority Sequencing Hierarchy are allowed, as for example *mgwa* ('fog') and *p'osnka* ('song'). Rubach and Booij (1990)'s data additionally show that a word like *Tatry* (/tatrɯ/) is syllabified as *ta-try* 60% of the time and as *tat-ry* 40% of the time. A word like *listwa* (/listfa/) is syllabified as *lis-twa* 68% of the time, as *list-wa* 19% of the time and as *li-stwa* 13% of the time.

3.4.3 Rhythm type

Polish poses a problem for rhythm classes. According to the diagnosis of Dauer (1983), it is a mix of a syllable-timed and stress-timed language, because it allows for very complex clusters, but it does not have vowel reduction, unless the speech rate is very high, and it does have fixed stress. Nespor (1990) analyzes Polish as an intermediate language between stress-timing and syllable-timing and according to the acoustic relative measures of Ramus et al. (1999) and Grabe and Low (2002), the language does, indeed, not clearly fall in one group or the other.

3.4.4 Stress

Word stress

Polish has fixed penultimate word stress (Gussman, 2007) and virtually no exceptions occur. Inflected and derived words will always carry penultimate stress, and the stress-system is quantity-insensitive, which means that the structure of the syllable does not affect stress-assignment. In compounds, the head of the compound has main stress, which is generally the final constituent (Rubach and Booij, 1985).

Even though penultimate stress is phonologically fixed, the high degree of monosyllables in Polish, which carry stress, makes the stress pattern distributionally less regular than the fixed initial stress of Hungarian. Furthermore, there are rare cases of non-penultimate stress. An example is final stress in loan words such as *menú* and *déjà vú*, and antepenultimate stress in loanwords such as *uniˈwersytet* (‘university’) and ˈ*opera* (‘opera’). However, these words can alternatively be pronounced with regular penultimate stress. Oliver and Grice (2003) and Kijak (2003) show that subjects choose the canonical regular pattern in production in about 70% of the cases which should or could receive antepenultimate stress. This shows that the re-analysis of loanwords as native, results in the production of the regular penultimate pattern, leaving the antepenultimate pattern as highly exceptional and lexically specified.

Derived and inflected words have penultimate stress, but there are so-called enclitics in Polish which cannot receive stress. These enclitics are *-by* /bɯɪ/ (conditional), *-śmy* /ʃmɯɪ/ (1st person plural past) and *-ście* /ʃtsie/ (2nd person plural past). Consider the examples below [12].

(3.14)	/roˈbili/	(*robili*, ‘make’ 3rd plural past masculine)
	/roˈbiliʃmɯɪ/	(*robiliśmy*, ‘make’ 1st pl. past masc.)
	/roˈbilibɯɪʃtsie/	(*robilibyście*, ‘make’ 2nd pl. past masc. cond.)

The enclitic is analyzed as being outside of the phonological word and part of the ’clitic group’ by Nespor and Vogel (1989), a similar analysis of which was also proposed for the behavior of Turkish clitics (Kabak and Vogel, 2001).

Polish secondary stress is a subject of controversy. There is some work advocating rhythmic or alternating secondary stress. For example, Hayes and Puppel (1985) consider Polish to behave eurythmically according to the principles proposed by Hayes (1984), that is, two prominent syllables are ideally separated by three syllables. Dogil et al. (1999) assume an alternating pattern of secondary stresses counted from the left, without finding phonetic evidence for this assumption. Rubach and Booij (1985) explain most of the examples as stress shift under focus. On phrase level, namely, Polish primary and secondary stress switch position under emphasis. This is important to discuss here, because this alternative position has been reported to be chosen as the primary

[12] These examples are taken from Dogil et al. (1999).

stress position by (then) young urban speakers (Wierzchowska, 1971), even when the word was not emphasized. More recently, Newlin-Łukowicz (2012) made a phonetic analysis of Polish stress, only finding evidence for main stress on the penultimate syllable. No evidence was found for an alternating pattern of secondary stresses, and the phonetic evidence for secondary stress on the first syllable was high intensity and F0, which they interpreted as part of declination over the utterance. Furthermore, Domahs et al. (2012b) investigated the electrophysiological response to the incorrect placement of primary stress on quadrisyllabic words. Their participants were young speakers of Polish, with a mean age of 23 years old. Crucially, placing initial primary stress on a quadrisyllabic word with canonical penultimate stress invokes a neural reaction which is typically associated with surprise. This result suggests that initial stress in polysyllabic words is not (yet) on a par with canonical penultimate stress.

Phrase stress

In Polish, the right-most head of the phrase receives the phrase accent. This means that even at the phrase level, stress is mostly penultimate. However, if the phrase-final word is monosyllabic, phrase stress will be final. Similarly, when the phrase ends with a clitic, the phrase accent will not be penultimate, since enclitics are not part of the metrical system. Furthermore, and as described above, a word with primary penultimate stress and secondary initial stress will switch prominence relations when placed in narrow focus.

Phonetics of stress

In an analysis of the phonetic correlates of Polish stress, Dogil and Williams (1999) report a higher pitch and sharper slope for primary stress, and an increased duration and expanded spectrum for secondary stress. This phonetic evidence for secondary stress was only found on the initial syllable, not on alternating syllables. The study did not manage to separate word stress from phrase stress, because no correlates for stress were found when the word was placed in a broadly focused constituent. Newlin-Łukowicz (2012) did not separate word stress from phrase stress either, making it difficult to appreciate the acoustic cues that can be attributed to word stress. The measured words were placed in focus in the recorded sentences and the study reported higher pitch, longer duration, higher intensity and greater pitch change for primary stress and it reported higher pitch, higher intensity and longer onset for the initial syllable. Crosswhite (2003) studied the presence of spectral tilt differences between stressed and unstressed syllables in various languages including Polish. The experimental words were placed outside of focus in a carrier sentence, and the results show that the important correlates of Polish word stress are spectral tilt and duration.

3.5 Dutch

Dutch[13] (Nederlands) is a West-Germanic language, closely related to German and English. There are approximately 23 million native speakers, of which about 16 million live in the Netherlands and about 6 million in Belgium. The language is also spoken in Surinam and the Dutch Antilles.

Dutch is an SOV language (word order subject-object-verb) with a verb second rule for the main clause. Otherwise it has a relatively free word order. Dutch has many monosyllabic words and it has both prefixes and suffixes and allows extensive compounding. The way in which the morphological system affects word stress will be discussed below.

There are two standard varieties: Belgian Dutch and Netherlandic Dutch. The differences between these two standards are slight and largely segmental and lexical. There is a wealth of dialects and regiolects, but at the same time there is a high degree of standardization. For the current study, the reference language is standard Netherlandic Dutch.

3.5.1 Segment inventory

The vowel inventory of Dutch consists of 13 monophthongs /a, i, u, o, ø, y, e, ɑ, ɪ, ʏ, ɔ, ɛ, ə/. There are tense and lax vowels which behave as short and long vowels in the metrical system. Tense non-high vowels are long when they are stressed (Gussenhoven, 2009). Otherwise the contrast is expressed through spectral quality. Vowels can optionally be reduced to schwa when they are part of an unstressed syllable (Kager, 1989; Trommelen and Zonneveld, 1989), but there are also underlying schwas in the vowel system. There are diphthongs /ɑu, ɛɪ, œy/ and some long vowels are realized as diphthongs in standard Netherlandic Duch [e^{j}, o^{u}, ø^{y}]. There are some loan-vowels: /ɔː, ɛː, ʏː, uː/ and some semi-diphthongs are better analyzed as concatenated segments: /ew, aj, oj, iw, uj/.

Dutch has 18 consonants /p, b, t, d, k, f, v, s, z, χ, h, m, n, ŋ, r, ʋ, j, l/. Unlike most Germanic languages, there is true voicing contrast in which voiced obstruents are prevoiced and voiceless obstruents are not aspirated. Regressive voicing asssimilation occurs in obstruent clusters and progressive voicing in stop-fricative clusters. The language furthermore has word-final devoicing.

3.5.2 Syllable structure

The language has many different syllable structures. An onset is not obligatory, and a coda is allowed. The onset and coda can be a cluster of two consonants, and /s/ and /t/ and occasionally other consonants, such as in the word

[13]The majority of the information in this chapter is from Gussenhoven (1992),Gussenhoven (2009) and Trommelen and Zonneveld (1989). The examples are taken from Trommelen and Zonneveld (1989).

kiosk, can behave as satellites, which means that they can attach to the word-beginning and word-end. This can amount to a word-beginning of maximally 3 and a word-end of maximally 4 consonants, as can be seen in the example below.

(3.15) /strɑks/ (*straks*, 'later')
/hɛrfst/ (*herfst*, 'autumn')

The Dutch syllable does not have lax vowels in hiatus (i.e. when it precedes another vowel), or word-finally.

3.5.3 Rhythm type

Dutch is seen as a typical stress-timed language, like English and German. It has complex clusters, vowel reduction and positionally variable stress (Dauer, 1983). Durational and perceptual measurements, too, place Dutch in a cluster with English and German (Ramus et al., 1999; Grabe and Low, 2002; Ramus et al., 2003; White et al., 2012).

3.5.4 Stress

Word stress

Dutch is a language with canonical penultimate word stress. Besides that, there are many rules that affect the position of primary word stress and there is lexically prespecified stress. There are stress-affecting affixes and in compounding the left-most member receives primary stress. As a result, and because of the high number of monosyllabic words and bi- and trisyllabic words with initial stress, Dutch is a statistical hybrid of basic penultimate and initial stress.

In monomorphemic words, stress is restricted to a three-syllable window, counting from the right side of the word. There are systematic ways in which Dutch word stress moves away from the penultimate syllable, under influence of syllable structure. Dutch word stress is quantity sensitive, in what seems to be a marked way, because 'long' vowels are light, but diphthongs are heavy, closed syllables are heavy, and closed syllables with long vowels are 'superheavy'(see Kager, 1989; van Oostendorp, 1995). However, Gussenhoven (2009) explains that tense non-high vowels become long only once they are stressed and under this interpretation the system of stress assignment is unexceptional.

If the penultimate syllable is light and the final syllable is heavy, stress is often antepenultimate, as can be seen in 3.16. If the final syllable is superheavy, stress will be final, as can be seen in 3.17. Cases like 3.16, among others, are in some proposals regarded as 'extrametrical' (e.g. Kager et al., 1987; Trommelen and Zonneveld, 1989), to account for the fact that this right-most heavy syllable does not receive main stress. Proposals challenging extrametricality include Gussenhoven (1993); van Oostendorp (1995); Gussenhoven (2009).

(3.16) /ˈɑlmaˌnɑk/ (*almanak*, ‘almanac’)
/ˈmɑraˌtɔn/ (*marathon*, ‘marathon’)

(3.17) /eleˈχɑnt/ (*elegant*, ‘elegant’)
/sʏroˈχaːt/ (*surrogaat*, ‘surrogate’)

There are additional exceptions to this systematic pattern, such as the difference between Example 3.18 and Example 3.19, the latter of which has the default penultimate pattern. These exceptions are generally attributed to lexical prespecification of either the main stress or the foot.

(3.18) /ˈkɑnaˌda/ (*Canada*, ‘Canada’)

(3.19) /kɔˈleːχa/ (*collega*, ‘colleague’)

The role of schwa in the Dutch stress-system is that the syllable preceding an underlying final schwa syllable will always receive stress (Kager and Zonneveld, 1986). The only exception to this rule is when the schwa is immediately preceded by a vowel, as in /ˈbɛlχiə/ (*België*, ‘Belgium’) and /ˈʋedyə/ (*weduwe*, ‘widow’).

After placing the primary stress, alternating secondary stress is applied from left to right, avoiding clash. Dutch words are assumed to be footed exhaustively and there is a tendency for a secondary stress to occupy the initial syllable, resulting in a so-called ’hammock’ (van Zonneveld, 1985). Subsequently, vowels in the ’weak’ part of the foot, that is the right-hand element of the trochee, can optionally be reduced to schwa. It is generally easier for non-high vowels to be reduced than it is for high vowels. Similarly, it is easier for vowels in open syllables to be reduced than it is for vowels in closed syllables. Word-final syllables do not have vowel reduction.

In Dutch, the leftmost part of a compound receives primary stress, which means that primary stress can fall on a syllable that falls outside of the three-syllable window on the right edge[14]. The original stresses of each member of the compound are not eradicated, but they are less prominent than the primary stress.

Dutch morphology has an influence on stress placement. First of all, there are stress-sensitive suffixes, which are suffixes which undergo stress-placement as if the resulting word were monomorphemic. Many of these suffixes are syllabically superheavy. The second class of suffixes is the stress-neutral class, in which the suffix does not receive stress even though it would if the word were

[14]This pattern also surfaces in so-called pseudo-compounds. The left or right part of the compound, or both parts, are not semantically interpretable, but the stress pattern suggests a compound status.

(3.20) /ˈalˌsχolvər/ (*aalscholver*, ‘cormorant’)
/ˈspɛrsiˌbon/(*sperzieboon*, ‘butter bean’)
/ˈuləˌwɑpər/ (’oelewapper’, ’nincompoop’)

monomorphemic. This also means that stress in a derived or inflected stem can fall outside of the three-syllable window. Thirdly, there are affixes that seem to attract stress, such as *-lijk*, *-ig*, *-baar*, *-zaam*, *-isch*, of which a few examples are given below.

(3.21)	/ˈχestdrɪft/	(*geestdrift*, 'zeal')
	/χestˈdrɪftəχ/	(*geestdriftig*, 'zealous')
	/ˈɔpmɛrk/	(*opmerk*,'observe')
	/ɔpˈmɛrkzam/	(*opmerkzaam*, 'observant')

In these examples, stress moves to the syllable directly preceding the affix. It is outside the scope of the current thesis to exhaustively describe the interaction of morphology and phonology in Dutch stress-placement, and for now it suffices to say that most of these debatably (Trommelen and Zonneveld, 1989) stress-attracting suffixes place stress on the penultimate syllable of the derived word.

Phrase stress

In Dutch, like in Polish, the right-most head of the phrase receives the phrase accent. The intonation and relative prominence in Dutch phrases does not eradicate underlying stress patterns. Stress patterns are maintained but the relative prominence of words in phrases decides on the prominence of the main word stresses on phrase level. Dutch furthermore has bidirectional rhythmic adjustments on the phrase level (Kager and Visch, 1988), which are the Hammock-principled stress-shifts that have been mentioned above.

Phonetics of stress

Sluijter and van Heuven (1996b) performed an extensive and convincing analysis of the acoustic correlates of Dutch word stress. Stress stimuli were pronounced in carrier sentences, for it to be possible to disentangle word stress from sentence accent. Stress and accent were both studied in lexical words and reiterant speech. The latter is a method to study the acoustics of stress separately from the segmental information. An example of a lexical comparison is /ˈkanɔn/ vs. /kaˈnɔn/, and an example of a reiterant comparison is /ˈnana/ vs. /naˈna/. The reiterant analysis is useful, because segments have intrinsic acoustic values which interfere with the analysis. This problem would be controlled for in the lexical analysis as well, but even languages with lexical stress, as Dutch is, do not have many minimal word pairs differing solely on the dimension of stress.

The production analysis showed, first of all, that intensity is not a marker of word stress. Rather, it is an artifact of higher pitch on the accented syllable in the phrase. When this is factored out, the acoustic correlates marking a stressed syllable when in focus are duration and spectral balance. Intensity and vowel quality did not play an important role in the contrast. The results of a follow-up perception study (Sluijter et al., 1997) were completely in line with the production study, showing that duration and spectral balance are important in stress perception, and the role of overall intensity is marginal.

Part II

Behavior

CHAPTER 4

Experiment 1: Metronome

4.1 Introduction

The first experiment [1] in the context of this thesis is designed to investigate the use of stress in segmentation in different selected languages. Native speakers of Hungarian, Turkish, Dutch and Polish are expected to use their knowledge of the relation of canonical stress to the word boundary in word segmentation. The specifics of the metrical systems of these languages were discussed in Chapter 3. Section 4.2 below will go into detail about the predictions of the current experiment.

To be able to test the use of stress in segmentation cross-linguistically, the experiment discussed in this chapter is purposefully set up to be as comparable as possible between languages. By making the experiment equal for each language, a difference between the languages will be attributable to a language-specific reaction to the experimental stimuli. This is important when one recalls the issues with previous literature as discussed in Section 2. To make the experiment work for each of the selected languages, the experimental stimuli are carefully designed to be phonologically legal and acceptable for each, controlling at the same time for potential segmentation confounds and embedded words as well as statistical tendencies within each language. Acoustically, too, the stimuli are carefully designed and checked to ascertain as much as possible that the experiment tests what it is supposed to test, that is, the use of word stress in segmentation. This method is preferred over alternative methods in

[1] A part of the data of this experiment was previously presented van Ommen and Kager (2012), although the interpretation has developed further.

which participants of each language would for instance be presented with native stimuli. Although the latter method is closer to native listening, the advantage of a maximal comparison between languages is more important in the current choice of method.

For the current experiment, it is important that the stressed syllable in the experiment be perceived as a stressed syllable by listeners of each of the languages. Section 4.3.1 will go into detail about how the experiment was designed for the current selection of languages, and how I think any possible objections against this approach have largely been overcome.

As will be described in Section 4.3.1, the experiment is designed to mimic word-spotting as closely as possible, while keeping the entire procedure the same for all participants of all languages. These two constraints resulted in the design of a non-word-spotting task. Participants learn a word, which, in the relevant part of the experiment, they have to spot in the context of a five-syllabic phrase. The task is easy and reactions are fast and non-evaluative. In Section 4.6 this characteristic of the task will become relevant and hypotheses will unfold for the next experiment, which is discussed in Chapter 5.

4.2 Hypotheses

As explained in Chapter 1, I hypothesize that when the canonical stress pattern of a listener's native language is edge-aligned, it can be used to predict word boundaries, i.e. to segment speech. Speech segmentation, then, is expected to be facilitated when stress patterns near a word boundary conform to the canonical pattern. If metrical segmentation is language-specific and relies on canonical edge-aligned stress, it is expected that listeners of Turkish, Polish and Dutch can use a stressed syllable to predict an upcoming word boundary, since their canonical stress pattern is right-edge aligned. Dutch and Polish listeners, then, would be aided by their native canonical penultimate pattern and Turkish listeners would use the canonical word-final pattern. Hungarian listeners, on the other hand, would expect a stressed syllable to coincide with a word beginning. It is therefore not expected for them to use stress patterns in anticipation.

To be able to discern the use of phonetic boundary salience from the use of word stress in segmentation, evidence for a use of penultimate stress in segmentation would be instrumental, which is why the current experiment compares languages with canonical penultimate stress with languages with initial and final stress. Besides disentangling language-specific from universal cues, the anticipatory and relatively distant nature of the cue, compared to peripheral stress, gives an insight into the reach of prelexical segmentation cues. The anticipatory nature of right-edge-aligned stress contrasts with left-edge-aligned stress, which cues segmentation very locally. Importantly, penultimate stress is less reliable as a segmentation cue than final stress, since the number of monosyllabic words influences how well stress predicts a word boundary. This specific reliability factor will be subsumed under the positional dimension of 'peripherality'. Another reliability factor is positional variability, since here too, stress does not always make the right predictions regarding word boundaries. Summarized, the potentially important dimensions for the use of stress in segmentation are:

1) edge-alignment: left or right edge
2) peripherality: peripheral or proximate
3) regularity: fixed or positionally variable

With the current study the goal is to discern whether there is language-specific use of stress in segmentation, providing a direct investigation of the role of peripherality and regularity of word stress: is word stress used in segmentation in languages with fixed and peripheral stress and do even listeners of languages with proximate and/or positionally variable word stress use word stress in segmentation, despite the unreliability of the canonical pattern? The hypothesis which will be tested is:

> **Language-specific Metronome Hypothesis**
> Listeners have language-specific knowledge of the relation of the stressed syllable to the word boundary. During processing, the stressed syllable triggers expectations on the position of the word boundary, based on this knowledge.

This hypothesis is different from earlier hypotheses in the literature on the use of stress in segmentation in two ways. First, the Metronome Hypothesis predicts the use of stress in any language with word-level stress, although the question is to what extent listeners are able to use stress in segmentation based on the positional reliability of stress. This is different from the Metrical Segmentation Strategy (MSS), later the Rhythmic Segmentation Hypothesis, which originally hypothesizes a role for stress in stress-timed languages only (Cutler and Norris, 1988; Cutler et al., 1992), as was described in Chapter 2. Second, the Metronome Hypothesis predicts the use of stress according to its canonical relation to the word boundary, whereas the MSS predicts a role for metrical units, and the Salience-to-Onset Strategy (SOS) (Taft, 1984), for example, predicts a facilitating role of word-initial stress in the activation of appropriate lexical units. For English, the MSS and SOS predict the same behavior. The Metronome Hypothesis, in sum, predicts a specifically suprasegmental prelexical role of stress in segmentation, building expectations based on language-specific phonological regularities.

The non-word-spotting experiment tests anticipation as well as the use of target stress in recognition, which makes it a more inclusive study than previous work. Besides providing a broad typological view on the use of stress in segmentation, the study is designed to overcome prior methodological issues. The artificial language, which is described in Section 4.3.3, was carefully designed to be natural for listeners of each of the languages, thereby making the behavior of the listeners of different languages maximally comparable. The experiment furthermore carefully isolates word stress as the only segmentation cue at work by controlling for potentially intervening alternative segmentation cues. Finally, the word stress cues as used in the study were carefully designed to be a combination of acoustic correlates relevant for the tested languages and was piloted to be perceived as stress by listeners of each of the languages. Section will describe the experimental method and the predictions on the behavior of participants of the different languages will be given in Section 4.4.

4.3 Method

4.3.1 Experimental design

The word-spotting task (McQueen et al., 1994) is a known method in psycholinguistic segmentation studies. Listeners detect words embedded in nonce contexts with segmentation cues being manipulated. Response latencies and accuracy scores are measured. In the current study, we are dealing with listeners of four different languages. To keep the task equal across languages, a non-word-spotting task was designed based on that developed by Kabak et al. (2010). In this task, the test stimuli consisted of a trisyllabic non-word context followed by a disyllabic non-word target. The target was always presented orthographically immediately before presenting the auditory pentasyllabic stimulus. In the pentasyllabic stimulus, the target always had final stress, and the preceding context had either penultimate or final stress. The design is attractive, because it can be used cross-linguistically and because it can be used to test various different stress conditions. For the current study, some methodological changes were made. First of all, the design was adjusted to become more like a word-spotting task. To this end, the participants learned two auditorily presented non-words, associated with pictures, in the training phase, which they then had to spot in the test phase. This way, orthographic processing is eliminated as a possible confound and the task is all auditory processing. Another advantage of this adjustment is that it is possible to add the stress pattern on the target word to the experimental conditions to be able to compare wS and Sw targets. In the context, the extra stress condition of antepenultimate stress is added. The design will further be discussed in Section 4.3.3

4.3.2 Participants

38 Dutch, 42 Turkish, 45 Polish and 46 Hungarian native speakers, all of them university students, were tested at Utrecht University, the Netherlands, Çukurova Üniversitesi, Turkey, Uniwersytet Wrocławski, Poland and Budapesti Műszaki és Gazdaságtudományi Egyetem, Hungary, respectively (Age 18-27, M = 20.9). All participants received compensation. They were all monolingual and

Table 4.1: Stress conditions

Condition	**Context**	**Target**
Antepenultimate-Final	Sww	wS
Antepenultimate-Initial	Sww	Sw
Penultimate-Final	wSw	wS
Penultimate-Initial	wSw	Sw
Final-Final	wwS	wS
Final-Initial	wwS	Sw

w = unstressed, S = stressed

were tested individually. None of the participants reported any speech, reading or hearing disorders and all had normal or corrected-to-normal vision. Participants were instructed in their native language. After the task, each participant completed a questionnaire about knowledge and use of other languages, to confirm that their native language was their dominant language. No students had to be excluded on the basis of this questionnaire.

4.3.3 Material

The materials presented to the participants of this non-word-spotting task were five-syllable nonsense strings, consisting of a three-syllabic context with a CVCVCV structure followed by a disyllabic target with a CVCCVC structure. An example of an item is /badusudarnam/, in which /darnam/ is the target to be spotted. Each target had one stressed syllable, as did each context. Table 4.1 shows the six experimental conditions according to stress pattern.

The items consisted of syllables that are phonotactically legal in all four languages. Four vowels /a u e i/ and eight consonants /b d s f r l m n/ were used, which are all legal segments in all languages. Consonants did not contrast in voice, because of different voicing characteristics between languages. The syllables were controlled for overall syllable frequency and for frequency in stressed and unstressed position. Furthermore, possible interfering segmentation cues were controlled for, such as positional frequency of syllables, transitional probability and vowel harmony. The latter was controlled for by making all strings harmonic according to the harmony-rules of both Turkish and Hungarian. Stress-affecting syllables were taken out as well: syllables matching prestressing morphemes (Revithiadou et al., 2006) and stress-attracting morphemes were not used. Syllable sequences remaining after these control measures were counter-balanced across conditions. The items did not contain polysyllabic words of any of the four languages. Monosyllabic words had to be accepted, since their removal would mean all items would be rejected: each of the CV or

CVC combinations with the above segments is a word in at least one of the four languages. There were two segmentally different targets: one, /mernel/, had front vowels and the other, /darnam/, back vowels. Both targets were embedded in 20 segmentally different experimental items. Thus, in total there were 40 segmentally different experimental items. Each individual item was used with each of the different stress patterns, which means that in total there were 6 x 40 = 240 different experimental items, which were divided over 6 lists. The lists did not contain targets which minimally differed in stress, so three of the six lists only contained the target with back vowels and initial stress /ˈdarnam/ and the target with front vowels and final stress /merˈnel/ and the other three lists only contained the target with back vowels and final stress /darˈnam/ and the target with front vowels and initial stress /ˈmernel/. This was true for experimental items as well as for fillers. Each list consisted of 80 experimental items as well as 240 filler items, 160 of which did not contain a target (e.g. /biresirefi/) and 80 of which contained the target in non-final position; either in second position (e.g. /remernelbese/) or third position (e.g. /resemernelbe/), to avoid a bias for position of the target in the string as well as a bias for a positive response. The filler items had the same stress patterns as the experimental items.

4.3.4 Phonetic resynthesis

The items were recorded integrally by a female native speaker of Spanish. This means that the context and target were recorded together, as if one word. This ensures that there are no phonetic segmentation cues between context and target. The choice for a Spanish speaker was made to make the resulting language equally non-native to the Turkish, Dutch, Polish and Hungarian listeners. Each segmentally different item was recorded minimally four times: two times with stress on the first, and two times with stress on the second syllable of the target, the context always being flat. To have a natural rhythm on the items (i.e. to avoid list intonation), items were recorded in a carrier sentence, in which they received phrase accent on the embedded target. The best token was selected by the researcher; the item was recorded again in cases in which none of the tokens was of sufficient quality. Next, the items were phonetically adjusted through resynthesis to give each condition a different stress pattern while keeping all other phonetic factors constant across conditions.

An important consideration when using the same experiment in multiple languages is language-specific acoustic perception. Stress is known to be realized differently across languages (Hayes, 1995), and the difference between the languages in the current experiment is reviewed in Chapter 3. The resynthesized ‘stress’ needs to be processed as stress for all of the participants, in order to draw conclusions on the use of stress in segmentation. This is a phonetic consideration which is often overlooked in previous studies. To make sure that all participants perceived the experimental ‘stress’ as word stress, the stress cues on the stimuli were a combination of all stress cues of each of the languages. Furthermore, a

Table 4.2: Stress judgment task

		Hungarian	**Polish**	**Turkish**	**Dutch**
Naturalness language	Own	0	2	0	0
	Other	5	4	4	4
	Computer	0	0	1	1
Mean accuracy		87.4%	84.0%	76.6%	94.8%

'Naturalness' was a multiple choice question.
'Accuracy' was the percentage correct of stress vs. no stress per syllable

pretest was run on five listeners per language. This will be discussed below.

Literature reports that important phonetic correlates for primary stress in Turkish, Dutch, Polish and Hungarian are F0, duration and spectral tilt (relative amplitude of the higher frequencies) (Sluijter and van Heuven, 1996b; Levi, 2005; Crosswhite, 2003; Fonagy, 1966). The languages differ in which phonetic correlates contribute principally to word stress. It was therefore decided to maximize the perception of stress for all participants by using each of the available cues in the resynthesized stress. The correlates were adjusted in the rhymes of each syllable, using the program Praat (Boersma and Weenink, 2011). First, the three vowels of the context were made isochronous by giving them the mean duration of the three syllables. Then the stressed syllable was lengthened by a ratio of 1.5. The pitch of the context was made flat (therefore assuming the pitch of the target) and then the stressed syllable received a boost of 8 semitones, with the peak at the first quarter of the overall duration of the vowel. Lastly, the spectral tilt of the stressed syllable was adjusted. This latter operation, however, led to a synthetic-sounding result, after which it was decided to give the overall amplitude of the syllable a boost of 8 dB instead, because it was important to have a natural result after resynthesis.

To test the acceptability of the stress patterns achieved by resynthesis, native speaker judgments were collected of five speakers of each language. For Polish, there were, by mistake, six respondents, which means there was a total of 21 participants. None of the participants took part in the non-word-spotting experiment. Participants were asked to judge which syllable was stressed and whether the language they were listening to sounded like 1) their own language, 2) another language 3) a computer language. Each participant judged a total of 40 stimuli. Each of the 40 experimental items was surveyed with one of the six stress patterns. All six conditions were therefore judged with multiple different segmental items. The results of the survey can be found in Table 4.2

4.3.5 Procedure

Each participant was tested individually on a laptop in a quiet room. In Utrecht, the Netherlands, and in Wroclaw, Poland, this quiet room was a sound-proof booth at the local university phonetic laboratory. In Adana, Turkey and Budapest, Hungary, this was a regular, but quiet, room. For running and controlling the experiment, the program ZEP was used (Veenker, 2011). The items were presented via Beyerdynamic DT 250 headphones. Each participant was randomly assigned to one of six presentation lists. The experiment started with two short animations of creatures called /darnam/ (a red and round creature) and /mernel/ (a yellow triangular creature). During the animations, the participants heard four different tokens of the name of the corresponding creature. This was followed by a static picture of the creature and its name, and a training phase in which the participants had to indicate which creature corresponded to which name. The red creature (/darnam/) was always displayed on the left of the screen, the yellow creature (/mernel/) on the right. Participants reacted with the button on the corresponding side of the button box. This introduction and training phase is intended to create a lexical entry of the non-word.

Anecdotally, participants reported that they perceived the language as a proper language (e.g. "Arabic", "some African language", "Avatar-language") and the items as utterances relating to the creatures.

Each item in the practice phase and test phase started with a fixation cross, followed by the auditory presentation of the item. Items in the test phase were presented in a random order, which was different for each participant. The participant was asked to react as quickly as possible whenever they heard the name of one of the creatures. As soon as the participant recognized one of the words, they had to hit any button of the button box. If and when they did so, a screen appeared with a three-way choice: the red creature on the left, the yellow creature on the right and a false alarm button in the middle. Here they chose which name they heard, or used the false alarm button to correct themselves if they had reacted mistakenly. Throughout the test an inter- stimulus interval of 1 second was used. When asked afterwards, the participants reported that they thought this was an appropriate speed.

4.4 Predictions

The Language-specific Metronome Hypothesis predicts listeners to use their knowledge of the relation of the stressed syllable to the word boundary in processing. Listeners with right-edge-aligned stress are expected to use their native canonical pattern superimposed on the context as a cue that a word boundary will follow, and listeners of all languages are expected to be able to recognize a target faster when it has the native canonical stress pattern than when it does not.

Concretely, this means that Hungarian listeners are expected to take a stressed syllable as a word beginning, but do not have any expectations when the target will come based on the context. As a language with left-edge-aligned stress, stress does not provide any context-induced expectations on the advent of a boundary, but it does give a clear indication of a word-beginning. Furthermore, Hungarian is the language in the selection which has the most reliable stress position; it is fixed and peripheral.

Polish is also a language with fixed stress, but because stress is not peripheral (it is penultimate), it is unclear whether listeners will use penultimate stress anticipatorily. Due to monosyllables in the language, furthermore, stress position is less reliable in languages with proximate stress. If stress need not be reliable for it to be useful in segmentation, listeners are expected to take stress as a cue that an unstressed syllable and then a word boundary will follow. For Polish, the disyllabic target with initial (thus penultimate) stress is the target with the native canonical stress pattern, which is expected to facilitate recognition.

The same is true for Dutch: listeners are expected to find it easier to recognize a target with initial stress, and to anticipate a target when the context has penultimate stress. Dutch, however, has more unreliable stress than Polish, because of the positional variability of stress in the language. Again, if reliability is a dimension in the usefulness of stress for segmentation, it is less likely for Dutch listeners to use penultimate stress anticipatorily than it would be for Polish listeners.

Turkish, lastly, has positionally variable right-edge-aligned peripheral stress. Concretely, the prediction is that listeners interpret the stressed syllable as a cue for a following word boundary if, indeed, word stress need not be fixed for it to be used in segmentation. The facilitative effect of final stress is expected both on the target and on the context preceding the target.

4.5 Results

In the analysis of the test results, the accuracy rate across participants was very high. Overall, in the experimental items and the fillers, the mean accuracy rate was 96.8% with a false-alarm rate of 0.4%. Four participants in the Turkish group and one participant in the Hungarian group were excluded on the basis of an accuracy rate of more than 2.5 SD below the mean. Because of the high accuracy scores, only response latencies were used as dependent variable, since an error analysis would not have enough data points. Response latencies were measured from the onset of the target item. In the latency-analysis, only correct answers were included. Response latencies were log-transformed, because of their naturally skewed distribution.

For this type of data, the most suitable statistical procedure is that of Mixed-Effects modeling. This technique has as the advantage over ANOVA that it is able to take several random factors into account (Baayen et al., 2008;

Table 4.3: Coding for canonicity of stress conditions

Condition		**Hungarian**	**Polish**	**Turkish**	**Dutch**
Context	**Target**	**Canonicity**			
Sww	**Sw**	1 1	0 1	0 0	0 1
Sww	**wS**	1 0	0 0	0 1	0 0
wSw	**Sw**	0 1	1 1	0 0	1 1
wSw	**wS**	0 0	1 0	0 1	0 0
wwS	**Sw**	0 1	0 1	1 0	0 1
wwS	**wS**	0 0	0 0	1 1	0 0

0 = non-canonical, 1 = canonical, w = unstressed, S = stressed

Quené and van den Bergh, 2008). The models were built incrementally in R (R Core Team, 2013), using the lme4 package (Bates et al., 2014).

An incremental build starts from a basic model with the fixed effects under consideration, and a random slope for participant and item. Step by step, covariates and random effects were added and for each added effect, it was tested whether the fit of the model significantly improved. If so, the random effect was kept in the model, if not, it was removed again, to reduce the risk of overfitting.

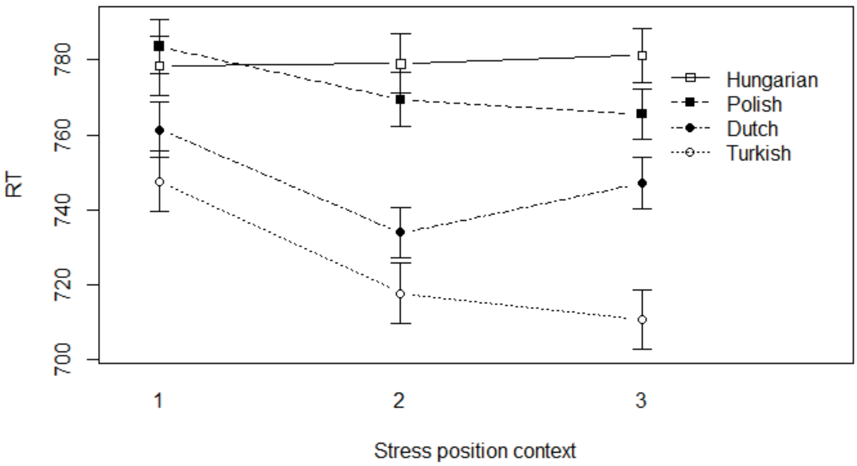
4.5.1 Language-specific effects

To assess the effect of the canonical stress pattern on response latencies, the six conditions given in the experiment were coded on canonicity (yes = 1/no = 0). For each language, each condition was coded on whether the context had a canonical stress pattern (1/0) and whether the target had a canonical stress pattern (1/0). For example, the condition SwwSw is coded as canonical context(1) and canonical target(1) for Hungarian, while the exact same condition is coded as canonical context(0) and canonical target(1) for Polish. The reader is referred to Table 4.3 for the canonicity of each condition for each language. Important to note about this table is that the conditions differentiate the stress patterns of the languages, and a canonical condition for one language is always a non-canonical condition for at least one of the other languages.

Overall model

The model assesses the effects of the canonical stress pattern on the context and the canonical stress pattern on the target, while modeling out variation caused by language, participant, trial order, creature type (/darnam/ or /mernel/) and the segmental form of the item. The results show that there is a strong

Figure 4.1: Effect of stress pattern on the context on RT per language



1 = initial, 2 = penultimate, 3 = final

effect of trial order (t(39) = -27.0, p = .000) and of creature type ((t(39) = -10.0, p = .000). Effects of language are not significant. Most importantly, the model shows that there is a convincing facilitating effect of the canonical stress pattern both on the context (t(39) = -4.1, p = .000) and the target (t(39) = -5.1, p = .000). The interaction of context and target is not significant (t(39) = -.8, p = .4).

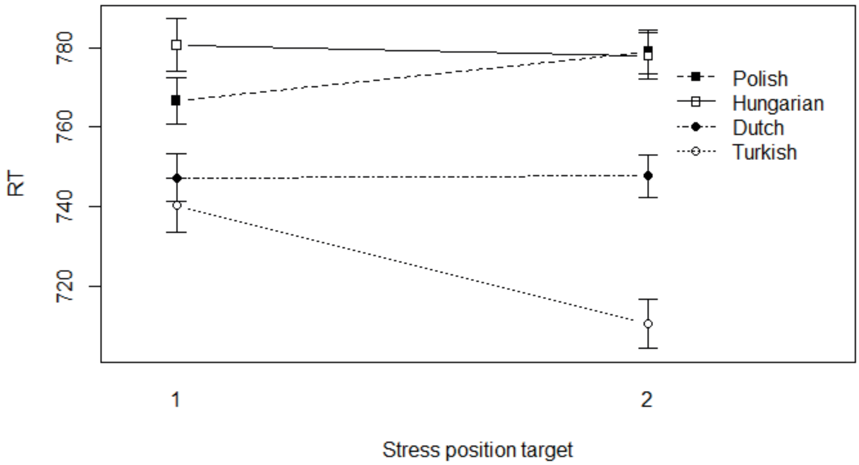
From this model we can conclude that there is an overarching facilitating effect of stress pattern on the context as well as on the target. Both factors independently facilitate segmentation, and no added interaction effect was found. This is the result when all the languages are taken together. Next, an analysis per language assesses whether the overall effect of canonical stress on segmentation is one that has the same form per language.

Separating languages

In Figures 4.1 and 4.2 the effect of stress position per language per factor is visible. In Table 4.4 the means and standard errors per language are given for each of the six experimental conditions.

The structure of the models per language is the same as the above described

Figure 4.2: Effect of stress pattern on the target on RT per language



1 = initial, 2 = final

Table 4.4: Mean and SE of the log of the reaction times per condition per language

	Hungarian		**Polish**	
Condition	**Mean Log RT**	**SE**	**Mean Log RT**	**SE**
SwwSw	6.602	.015	6.612	.013
wSwSw	6.599	.015	6.575	.013
wwSSw	6.609	.013	6.590	.012
SwwwS	6.605	.013	6.627	.012
wSwwS	6.608	.013	6.620	.013
wwSwS	6.621	.013	6.603	.012

	Turkish		**Dutch**	
Condition	**Mean Log RT**	**SE**	**Mean Log RT**	**SE**
SwwSw	6.602	.015	6.592	.014
wSwSw	6.530	.016	6.546	.013
wwSSw	6.517	.016	6.575	.013
SwwwS	6.526	.015	6.592	.012
wSwwS	6.509	.015	6.571	.012
wwSwS	6.505	.015	6.582	.012

w = unstressed, S = stressed

model for the languages overall, with the difference that this time language is not a covariate, but the data is a subset of the data in the first model, based on language group. The covariates of trial order and creature type are significant in each of the models. The fixed factors are the effects of interest, for which the values will be given in the separate language sections. In Section 4.6, the implications of the different findings per language will be discussed jointly.

Hungarian The model of the Hungarian group revealed a facilitating trend of canonical stress on the target [Sw] (t(39) = -2.01, p =.06, r = .016) but no significant effects for either canonical stress on the context [Sww] (t(39) = -1.14, p = .2, r = .015) or the interaction (t(39) = -.1, p = .4). This is an expected result from the Language-specific Metronome. It suggests, although as a trend, that the fixed initial stress on the target in Hungarian is a cue used in segmentation.

On the other hand, this finding is not as straightforward as it may sound; from several other perspectives this may be considered an unexpected effect. For example, a probabilistic perspective might hypothesize listeners to predict an upcoming boundary anticipatorily, too, based on statistical knowledge about the average distance between the right edge and the main stress. This would mean that knowledge about the average number of syllables of a word would be of assistance to a listener during fast processing, because the listener would know after how many unstressed syllables it is most probable to encounter a word boundary. This is not what was found. Another possible hypothesis which might predict an effect of the context is the Possible Word Constraint (Norris et al., 1997), which states that a word is more difficult to spot when the stretch of speech preceding that word cannot be a word in itself according to the constraints of that language. An example from English is spotting the word *apple* in the stretch *fapple* as opposed to *vuffapple*. In Hungarian, a word with non-initial stress is illegal, due to the fixed-stress nature of the language. This means that a context with non-initial stress may be expected to slow down segmentation, unless the preceding unstressed syllables are perceived as function words. The context with final stress can be interpreted as an impossible word, because Hungarian content words are minimally bimoraic (cf. Hayes, 1995; Siptár and Törkenczy, 2007), and the syllables in the context are all monomoraic.

Polish

The model of the Polish group reveals a significant facilitating effect of canonical stress on the target [Sw] (t(39) = -3.95, p = .000, r= .049) but the effects of context [wSw] (t(39) = -1.33, p = .2, r = .016) and the interaction of context and target (t(39) = -1.66, p = .1) are not significant.

The Language-specific Metronome would predict different results. We do find that Polish participants use canonical stress in segmentation, this is shown by a strong effect of canonical stress on the target. However, the lack of an effect of canonical stress on the context is unexpected. Polish has fixed right-edge aligned stress, for which the Language-specific Metronome would predict an anticipation of the upcoming word boundary. This expected effect of canonical context is not found. Reaction times are fastest when the canonical context is followed by a canonical target, but there is no significant interaction of context and target. The condition with the canonical context, but non-canonical target, however, is not faster than other conditions with non-canonical targets. Speculatively, the inhibition of the segmentation of the non-canonical target buries an anticipation effect. This strength of inhibition would surface in Polish because it is a fixed stress language, meaning that a wS target is illegal. 'Stress deafness' is unlikely to be a factor in the results, since first of all, Polish speakers were found not to be 'stress-deaf' (Peperkamp and Dupoux, 2002) and second of all, a target effect would not be expected in a 'stress-deaf' population. Third of all, 'stress-deafness' may be doubted to surface in the current experiment at all, since the task, and especially the context, do not necessitate lexical encoding of stress.

Turkish

The model of the Turkish group reveals a significant facilitating effect of canonical stress on the context [wwS] ($t(39) = -3.16$, $p = .005$, $r = .045$), a significant facilitating effect of canonical stress on the target [wS] ($t(39) = -2.06$, $p = .05$, $r = .060$) and no significant effect on the interaction ($t(39) = 1.50$, $p = .1$). This means Turkish behaves perfectly as expected: the context with final stress helps listeners anticipate a word boundary, and the target with final stress is recognized faster than the target with initial stress.

The Turkish results show anticipation in segmentation for a right-edge aligned language. Unlike Polish, Turkish listeners benefit from the beat of the word stress in directional processing. This is true in both the conditions with the canonical and the non-canonical target. The difference between Polish and Turkish is interesting. Apparently Turkish listeners do recognize the canonical target faster than the non-canonical target, but the inhibition which was speculated upon in the Polish case, does not seem to bury any anticipation effects. If this speculation would be right, it is possible that the difference between Turkish and Polish stems from the fact that Turkish is a positionally variable stress language which accepts more deviations from the standard stress position than the fixed stress language Polish. A disyllabic word with initial stress is not canonical, but not illegal in Turkish either.

When we take a closer look at the non-canonical conditions in Turkish, we see that reactions to the non-canonical condition of penultimate stress on the context are not as slow as those to the non-canonical condition with antepenultimate stress on the context. This suggests that the use of stress in segmentation is a gradual rather than a categorical process in Turkish. The default position of Turkish stress is on the final syllable, but both the penultimate and antepenultimate position are legal docking sites for stress in

Turkish. The reaction times seem to suggest that the closer the beat is to the right edge of the word, the more useful it is in segmentation. Another interpretation would be a probabilistic interpretation, where for instance the penultimate stress position is more frequent than the antepenultimate position. The current experiment does not show which of these interpretations is correct, but this gradual difference is something we thus far only see in the Turkish group. The Hungarian results, which did show an effect of context, argued against a probabilistic interpretation of the use of stress in segmentation.

Dutch

The model of the Dutch group reveals a significant facilitating effect of canonical stress on the context [wSw] ($t(37) = -2.82$, $p = .01$, $r = .053$) and a facilitating trend for canonical stress on the target [Sw] ($t(37) = -1.90$, $p =.07$, $r = .018$) as well as a facilitating trend on the interaction of context and target ($t(37) = -1.75$, $p = .09$).

Like Turkish, Dutch seems to behave perfectly according to the Language-specific Metronome. There is an effect of context, indicating that this right-edge-aligned language with penultimate stress seems to make anticipatory use of stress cues in segmentation. This is the first time evidence can be presented for the use of stress in segmentation in a language with penultimate stress. The effect not only disentangles a universal use of boundary salience from a language-specific use of word stress, but it also shows that the canonical pattern does not need to be either peripheral or regular for it to be useful in segmentation. Dutch is the one language of the four selected languages of which such a result was least expected according to the reliability hypothesis, due to its non-fixed and non-peripheral main stress.

The overall model, taking all languages together, showed an overall effect of canonical stress on the target. In Dutch, however, there is only a trend of a target effect. A possible explanation is that Dutch word stress is positionally most variable and disyllabic words with final stress are highly common in the language. The six conditions in Table 4.4 show that the RTs to the canonical target are fastest after the canonical context. In fact, for all languages with right-edge-aligned stress it is true that the condition with a canonical context combined with a canonical target is the fastest condition of all. This is an effect we will discuss in Section 4.5.1.

Interim conclusion

The results of this experiment give evidence for the use of stress in segmentation. When all languages are taken together, there is an independent effect of context and target.

In the language-specific models a facilitative effect of canonical stress on segmentation was found for each of the languages, although only suggested by a trend for Hungarian. Not only listeners of languages with peripheral stress used their canonical pattern, but Polish and Dutch listeners, too, used canonical stress in segmentation. This is an effect not previously found in comparable studies for the penultimate stress language Spanish (Toro-Soto et al., 2007; Cunillera et al., 2008). Both having the penultimate syllable as the fixed or canonical, respectively, docking site, it is interesting to see that the use of stress between Polish and Dutch differs in direction. In Dutch, we find that this effect is mostly carried by the context, but an effect of target and an interaction effect are suggested by statistical trends. In Polish, the effect is carried by the target and no other effects were found.

For Turkish, an independent facilitative effect for both the context and the target were found. These results show that Turkish participants do not only use stress in anticipation, as previously found by Kabak et al. (2010), but on the target as well.

4.5.2 Universal effects

The overall model of language specific effects, as well as the separate models for each language have given evidence for a language-specific use of canonical word stress in segmentation. To check whether these results cannot otherwise be attributed to a universal use of boundary salience in segmentation, a universal model is compared to the language-specific model, to see which of these models captures the effects better.

Model

For the universal model, the six different conditions were recoded to universal boundary cues: stress patterns that cross a word boundary. These four boundary patterns were coded in two variables with two levels: final stress (1/0), initial stress (1/0). These two variables create four different combinations, which are the cues of interest. The conditions are: 0][0, 1][1, 0][1 and 1][0. Different plausible universal segmentation cues may be suggested. Previous research has shown a facilitative effect of initial stress on segmentation in some cases, even in non-initial stress languages. If this is a universal strategy and if the current experiment inadvertently measured this strategy, it has to surface in the current analysis as a facilitative effect of initial stress. On the other hand, previous research has also shown facilitative effects of final stress or salience on segmentation, in some cases even in non-final stress languages. If this is a universal strategy, again, the current analysis should make this clear. In a case in which no evidence is found for an effect of salience on either of these two boundaries, a third potential strategy would be to use the knowledge that two stresses cannot be part of the same word (clash). In that case, an interaction effect has to surface, showing facilitation of the condition 1][1. The mirror image of facilitation of segmentation in the clash condition is the inhibition of segmentation in the lapse condition, that is 0][0.

A model was built with as the dependent variable a log of the reaction time as measured from the onset of the target. The independent variables are final stress (1/0) and initial stress (1/0). Covariates are the trial order (with consistently faster reaction times through the course of the experiment) and the type of target (/darnam/ is consistently slower to respond to than /mernel/). Random slopes were added for participant and item.

The model shows that there is a trend of initial stress on the target (t(39) = -2.04, p = .05)). The effect of final stress on the context is not significant (t(39) = -1.04, p = .23) and neither is the interaction effect (t(39) = -.07, p = .40).

The universal model and the language-specific model are not comparable since the fixed effects are different. The models are therefore compared to an reduced model, which is a model with the same random variables and covariates as the two models, but with the fixed factors taken out. An ANOVA comparison of each model to the reduced model shows that the universal model is not significantly different from the reduced model (X^2(3,9)=5.65, p =.13)), while the language-specific model is (X^2(3,9)=43.54, p=.000). In other words, the model testing the effect of language-specific canonical stress patterns is a better fit of the data than the model testing the effect of universal salience cues.

Interim discussion

The current results do not support the view that final stress universally facilitates segmentation, but the statistics suggest an effect of initial stress. However, the 'universal cues' in the current model are inevitably confounded with canonical cues and a target with initial stress is a target with canonical stress for three of the four languages (Hungarian, Dutch and Polish). In the language-specific models it became clear that in each of these three languages the canonical target was recognized faster than the non-canonical target but the language-specific models also showed the use of canonical final stress on the target for Turkish participants. This strongly suggests that the 'universal' initial-stress effect originates from the power of the majority, not from a universal effect. This power-of-majority effect does not surface in the context, because the context with final stress is canonical solely for one of the four languages (Turkish). The fact that the language-specific model is a better fit of the data underlines this point.

No universal facilitative effect for 1][1 was found in the current data. Also, no inhibiting effect was found for 0][0. Even though universally no two neighboring stresses can be part of the same word and adult speakers are expected to have this knowledge, the effect was not found. From the perspective of the Language-specific Metronome this lack of an effect is not necessarily surprising, since in the clash condition one part of the stimulus always has non-canonical stress. This non-canonical stress is not facilitative, and can even inhibit segmentation. Furthermore, strictly speaking knowledge of the relation of word stress to the word boundary is one-directional. This means it is forward-looking for right-edge-aligned stress languages and backward-looking for left-edge-aligned stress languages. A facilitative effect of stress would imply a bidirectional effect, and this is not found.

4.6 General discussion

The overarching hypothesis of the current study is that language-specific canonical word stress is used in word segmentation. There have already been ample studies claiming to give evidence for this hypothesis, but the nature of many of the found effects turned out to be unclear, when considering alternative explanations. In many studies, the effect of stress could just as well be regarded as universal salience-to-boundary principles, an interpretation for which the results of these studies do not provide counter-evidence. Similarly, not finding a language-specific use of stress within one experiment keeps the interpretation of method-specific effects open. In the current experiment, the same task was given to listeners of different languages and the results showed language-specific behavior. While providing evidence for language-specific differences in the use of stress in segmentation, the results also show that there is a language-specific *anticipatory* use of stress in segmentation for Dutch and Turkish listeners.

The effect of the use of canonical stress in segmentation, in light of the fact that ‘canonical stress’ is a different condition for each of the languages, confirms the Language-specific Metronome hypothesis. This can be concluded from the fact that all languages show a different language-specific pattern of results and the results turn out to be better explained by language-specific than by universal strategies. Furthermore, the fact that each of the languages uses stress in segmentation, despite being generally assigned to different rhythm classes, is a finding which would not be predicted by the Metrical Segmentation Strategy (Cutler and Norris, 1988) (or: Rhythm Class Hypothesis (Cutler and Butterfield, 1992)). The Prosodic Domain Hypothesis (Taft, 1984) would predict an interaction of context and target, which was not found, despite some slight trends. Lastly, the Salience to Onset Strategy (Taft, 1984) is in principle not tested; it predicts that languages have a language-specific way to mark the onset, and that this is used in segmentation. In the current experiment, the SOS only has predictions for Hungarian, where onsets are marked with stress. Here, facilitation in segmentation is suggested with a statistical trend.

A new finding of the current study is the use of penultimate stress in anticipation of the word boundary, a result found for Dutch. This effect was not found in previous studies and it serves to disentangle effects of the Language-specific Metronome from effects of a universal reliance on boundary salience, which was suggested to be the basis for some earlier results in studies on Spanish (Toro-Soto et al., 2007; Cunillera et al., 2008) and possibly other studies, too. Additional to disentangling the universal use of boundary salience from a language-specific use of word stress, this finding on Dutch shows that there is no adverse effect of reliability of stress on the use of stress in segmentation in this particular task, possibly even on the contrary. All of the languages seem to make some use of stress. This is interesting for several reasons. First of all, the dimensions of peripherality and reliability that were differentiated in Chapter 3 do not seem to have an influence on the reliance of listeners on stress. As unreliable as penultimate stress in Dutch may be, because there are many exceptions to this position, it is still used in segmentation. This shows that the canonical pattern gives the listener expectations on word boundaries, which may in some cases not be right. Relating to the latter property of positionally variable stress languages, it is quite striking that the two positionally variable stress languages in this study are the two languages that seem to make use of anticipatory strategies, visible in the context effects, whereas the two fixed-stress languages seem to only make use of target stress. Judging on reliability, it is surprising that Polish listeners do not use the penultimate pattern to anticipate an upcoming boundary, while Dutch listeners, for whom the penultimate pattern is less reliable, do rely on the penultimate stress pattern in segmentation. One hypothesis is that the non-canonical target was so much more difficult to spot than the canonical target, due to the fixed-stress nature of the language, that any facilitative effect of context was annulled. This would not be found for Hungarian, because no anticipatory cues are expected in that language. A future experiment would have to test whether this hypothesis on Polish has any ground.

While Polish listeners do not use penultimate stress anticipatorily, Dutch listeners do, even though Dutch penultimate stress would statistically be an unreliable segmentation cue. Statistically, Dutch stress is even predominantly initial but phonologically the canonical pattern is penultimate. The finding that Dutch listeners use penultimate stress therefore suggests that the Language-specific Metronome is an abstracted, phonological strategy in segmentation. The finding that in Turkish the reaction times were faster after penultimate stress as compared to initial stress, next to the finding that canonical stress was most facilitative, however, does invite follow-up research. This difference between the penultimate and initial stress context, namely, cannot be explained by the default pattern. Doing the same type of experiment with native Turkish speech is an interesting way to clarify whether the phonetics of stress influences the pattern of results. It is known that default Turkish final stress has a different phonetic realization from Turkish penultimate and antepenultimate stress (Levi, 2005) and it would therefore be interesting to see whether the difference between these two types of stress triggers different results in a segmentation task like the current.

It is important to address the difference between the context and the target, which is that the context is heteroform, so the stress pattern overlaying the context cannot be interpreted as a lexical effect. The use of stress on the context, therefore, is a prelexical segmentation strategy. The target, in contrast, is a lexical element which is repeated throughout the experiment. This makes the effect of stress on the target either a lexical or a prelexical effect: a prelexical effect of canonical stress makes it easier to segment the target from the context whereas a lexical effect makes it easier to recognize the target. The current experiment does not disentangle these two interpretations. However, the experiment does show that listeners of all languages are faster in recognizing a target with the canonical stress pattern (although in Dutch and Hungarian this is only indicated by a statistical trend) while a context-effect was only found in listeners of the languages with positionally variable stress.

The overall facilitative effect on the recognition of the target with the canonical stress pattern can be explained as a lexical effect, in the sense that the canonical target is more similar to other words in the listener's lexicon. This is especially likely because the effect is also found in Turkish, even though the canonical pattern in that language is final stress and the target therefore needs to be heard in its entirety. The effect of salience to the onset of a word, predicted by earlier accounts on the use of stress in segmentation (Taft, 1984; Cutler and Norris, 1988) is therefore confirmed to be language-specific, since the use of initial stress does not extend to Turkish listeners. A lexical effect of stress in the current selection of languages could be seen as surprising, since studies on 'stress-deafness' predict word stress not to be part of the lexical representation of the word in Hungarian, Polish and Turkish. The current results cast doubt on this interpretation and a similar conclusion can be drawn from ERP results in studies on Polish and Turkish (Domahs et al., 2012a,b), which show that placing 'incorrect' stress on words with the canonical stress pattern prompts a response consistent with a higher cost on lexical processing. The current results show that the recognition of a word is facilitated when it carries the canonical stress pattern as compared to the word carrying the non-canonical stress pattern, even when the latter is consistent with how it was learned. This effect can still be attributed to either faster recognition or easier word learning when the word is more consistent with the language's phonology, but the current experiment does not differentiate between these two interpretations. The lexical segmentation effect does not seem to depend on regularity, peripherality or edge-alignment.

The current task was very easy for the participants: reactions were fast and almost no errors were made. It is unclear whether an increase in difficulty would make the effect of stress on segmentation stand out more. On the one hand, the Metronome is a fast strategy which may disappear when the task becomes more difficult. On the other hand, stress, being a salient property of speech, may be specifically useful when a task is more difficult, especially when speech is for example perceived in circumstances with background noise, or an inferior signal, as was found in Mattys et al. (2005). Interestingly, Mattys et al. (2005) and White et al. (2010) would not predict the use of stress in a situation of clear speech such as the current experiment, since sufficient segmental and 'lexical' (the two newly acquired words) information is available to the listeners for them to not use sublexical cues. Apparently the isolation of stress as the only differentiating factor between conditions does yield the possibility to measure the use of this phonological strategy in segmentation, despite the fact that speech is clear.

Regarding universal cues, for which the comparison of the two models showed that they are an inferior predictor of the data as compared to language-specific cues, it is only moderately surprising that no evidence was found for a universal clash effect. In principle, two stresses in a row should be separated by a boundary. However, if the stressed syllable only triggers expectations on one boundary, clash is not facilitative.

Summarizing, the non-word-spotting experiment gives evidence for the use of language-specific stress-based strategies in word segmentation. Listeners of Dutch, Turkish, Polish and Hungarian all show evidence of the use of stress in segmentation in a language-specific way. This use of stress turned out to be available both for the anticipation of a target, as well as for the recognition of the target itself. The hypothesis was confirmed by the context-effects found in the current experiment, but the target-effects seem to be more difficult to explain. The experiment in Chapter 5 attempts to disentangle lexical from prelexical effects with an offline experiment. Meanwhile, the difference between the results on Turkish and Dutch are still puzzling. Both Turkish and Dutch are languages with positionally variable stress, so why would expectations not be gradual? However, while the results for Turkish tentatively suggest this possibility, the Dutch results do not. Therefore, the next section describes the results of the experiment done post-hoc with an additional positionally variable stress language: Spanish.

4.7 Spanish

While Spanish stress is positionally variable, it has fewer lexical exceptions than Dutch and its rules of assignment are more regular. Most importantly, the two main positions of stress are penultimate and final and this is mostly attributable to syllable structure. It should be noted that the experiment is not designed for Spanish, meaning that the materials were not controlled for Spanish phonology, morphology nor the lexicon and the fact that a Spanish speaker has pronounced the materials makes the non-word-spotting task easier for Spanish listeners than it was for the other language groups, since the former are more used to the way the materials sound than the latter.

4.7.1 Language characteristics

Spanish[2] (Castellano) is a Romance language spoken by over 400 million speakers all over the world, except for Antarctica. The majority of Spanish speakers live in Mexico, Columbia, Argentina, the United States and Spain and Spanish is the official language in each of the named countries except for the United States. Standard Spanish is also known as Castilian, the dialect from which it developed. Other regional varieties spoken in the Iberian peninsula include Catalan and Galician.

[2] Among the sources in this section are Martínez-Celdrán et al. (2003), Roca (2006), Colina (2009), and Hualde et al. (2012).

As a Romance language, Spanish has relatively free word order and it has prefixes and suffixes for inflection and derivation. There are some productive compounding patterns, but overall compounding is less common than in Germanic languages.

The large number of Spanish speakers of course entails considerable social and geographic variation within the language. For the current experiment the reference language is standard Iberian Spanish.

Segment inventory

Spanish has five vowels, which may occur both in stressed and unstressed syllables: /i, e, a, o, u/. There are no length distinctions. In addition, Spanish vowels can be preceded or followed by a glide, that is /w/ or /j/. In some cases this is even combined, such as in /bwej/ (*buey*, 'ox') . The phonemic status of these glides and vowel-glide combinations is debated, hence I do not refer to them as diphthongs or triphthongs.

Spanish has nineteen consonants and it has no length distinctions: /p, b, t, d, k, g, f, θ, s, x, tʃ ɟj, m, n, ɲ, l, ʎ, ɾ, r/. The sounds /b, d, g/ are complete stops after a pause or a nasal consonant, and in the case of /d/ also after a lateral. In other positions, these stops are their approximant counterparts [β, ð, ɣ], where [ð] is interdental. Spanish has a true voicing contrast; in the onset unvoiced stops are unaspirated and voiced stops are prevoiced. Unvoiced stops can be voiced in the syllable coda due to regressive voicing assimilation, and in this position voiced stops can be realized as stops or approximants or they can even be elided. /s/ and /ð/ become voiced before a voiced consonant, including sonorants, and /x/ can become uvular before a back vowel or a [w]. The trill contrasts with the tap in intervocalic position. It also appears in the onset and after /l, n, s/, but in other cases the tap is more common. The nasal in the coda is not contrastive on place of articulation; it shares this place with the following consonant.

Dialectal consonantal variation is most apparent in sibilants. /s/ is pronounced apical in most of the Iberian peninsula, while it is usually laminal in other varieties. In these, it can become [h] in coda position. In some varieties, [ð] does not exist, and it is replaced by [s]. The opposite pattern also occurs, in some areas of Andalusia. /ʎ/ is realized as [j] in all large cities of the Iberian peninsula, while /ɟj, ʎ, j/ are realized as [ʒ] in some countries in South America. Lastly, some varieties realize /x/ as [h].

Syllable structure

In Spanish syllable structure, onsets are not obligatory, and codas are allowed. The nucleus can consist of a vowel or a glide-vowel combination and syllables can begin and end with clusters of a consonant or a glide and a consonant. A satellite /s/ can be attached to this coda. Templatic syllable structures for

Spanish[3] are V, CV, VG, VC, VGC, CVG, CVC, CCV, CCVG, CCVC, CVGC, CCVGC.

Rhythm type

In the initial classification, Pike (1945) describes Spanish as the prototype of a syllable-timed language. In the classification of Dauer (1983), the perceptual illusion of syllabic isochrony can be attributed to the simple syllable structure and the lack of vowel reduction in Spanish. In the literature studying durational measures, Spanish patterns with other languages which are believed to be syllable-timed, such as French and Italian (Ramus et al., 1999), although as said in Chapter 2, the traditional classes exhibit considerable overlap on the different measures (Grabe and Low, 2002).

Stress

Word stress Spanish stress is distinctive and its contrasts are highly productive in verbs. An example of distinctive stress is given below.

(4.1) /ˈlimite/ ‘boundary’

(4.2) /liˈmite/ ‘limit-IMP’

(4.3) /limiˈte/ ‘I limited’

Spanish stress is placed on the three-syllable window at the right edge of the word. In verbs, enclitics can move stress out of this window, to the fourth syllable to the right, such as in the word /ˈkwentaselo/ (*cuentaselo*, ‘tell that to him/her’). In non-verbals, the three-syllable-window is well documented (Hualde, 2005). Lexico-statistically, Hooper and Terrell (1976) have shown that in non-verbals approximately 95% of words ending in a vowel receive penultimate stress and 95% of words ending in a consonant receive final stress. They ascribe this difference to the fact that in most words ending in a vowel, this vowel is thematic and thus has a grammatical function. By placing this vowel outside of the scope of stress assignment, Spanish stress is final in the domain of the morphological stem, leaving aside exceptions. Roca (2005) summarizes the pattern in a table which is (with adjusted terminology) given in Table 4.5.

Den Os and Kager (1986) maintain the word as the domain of stress assignment and they refer to this thematic vowel as a Class Marker (CM) which is marked as extrametrical. Their generalization is that in words with a CM, stress is never final, it will be penultimate when the penultimate syllable is closed or contains a glide and otherwise it may, but does not necessarily, move to the antepenultimate. In words without a CM, stress is likely to be final, penultimate is also found, but antepenultimate is rare. Roca (2005) distinguishes unmarked, marked and supermarked stress assignment in Spanish, departing from the

[3] ‘G’ refers to a glide, following Colina (2009).

Table 4.5: Spanish stress patterns

	Majority		Minority
	Unmarked	**Marked**	
. . . V]	Penultimate	Antepenultimate	Final
. . . C]	Final	Penultimate	Antepenultimate

Summary of Spanish stress patterns from Roca (2005)

assumption that right edge of the stem, rather than of the word, is the edge of the stress domain. In the unmarked case, the last syllable of the stem is stressed, in the marked case, a trochee is aligned to the right edge of the stem, and in the supermarked case, there is a non-final trochee, or the edge of the stress domain moves to the word edge.

According to among others Harris (1983), Den Os and Kager (1986), Harris (1992) and the empirical evidence of Shelton et al. (2012), Spanish stress is Quantity Sensitive. However, the effect of quantity seems to be clearer in vowel-glide combinations than in closed syllables (Roca, 2006). Many other accounts such as Aske (1990), Bárkányi (2002), Waltermire (2004) and Face (2006) hypothesize that speakers' and listeners' generalizations are based on regularities in the lexicon rather than on phonological rules. Evidence in favor of this view was presented in Aske (1990), where speakers had a differential stress placement in non-words ending in -'en' as opposed to non-words with any other vowel closed by -'n', where the former receive penultimate stress in around 43% of the cases and the latter receive final stress in around 96% of the cases.

Phrase stress On the phrase level, the main pitch accent is placed on the last stressed syllable in the intonation group. The accent spreads over the remaining unstressed syllables. This main pitch accent is characterized by the peak of the rise on this last stressed syllable, but in stressed syllables in prenuclear position, that is the portion of the phrase preceding the main pitch accent, stress is characterized by the pitch rise, of which the peak typically falls on the following unstressed syllable. In spontaneous speech, these prenuclear pitch rises often fail to occur and pitch peaks on prenuclear stresses only occur in emphatic speech(Face, 2003). The prenuclear portion of the phrase is furthermore characterized by downstepping, both in lab speech and in spontaneous speech (Face, 2003). Spanish compounds receive the main stress on the last member and the stresses of the other members of the prosodic word are not realized (Hualde, 2007).

Phonetics of stress Following the example of Sluijter and van Heuven (1996b), Ortega-Llebaria and Prieto (2007) studied the independent acoustic cues of unaccented stressed syllables. Their results show that Spanish word stress is mostly characterized by duration and spectral tilt and to a lesser extent by vowel quality. In a later study, Ortega-Llebaria and Prieto (2011) controlled the effect of spectral tilt by formant frequency, to find out whether the effect of spectral tilt can be disentangled from vowel quality and they found that in Catalan, these two correlates could not be separated from one another and in Spanish, there was no effect of spectral tilt. In sum, the main acoustic correlate of Spanish word stress is duration and the correlate of phrase stress is the pitch rise in prenuclear position and the pitch peak in nuclear position.

4.7.2 Hypotheses

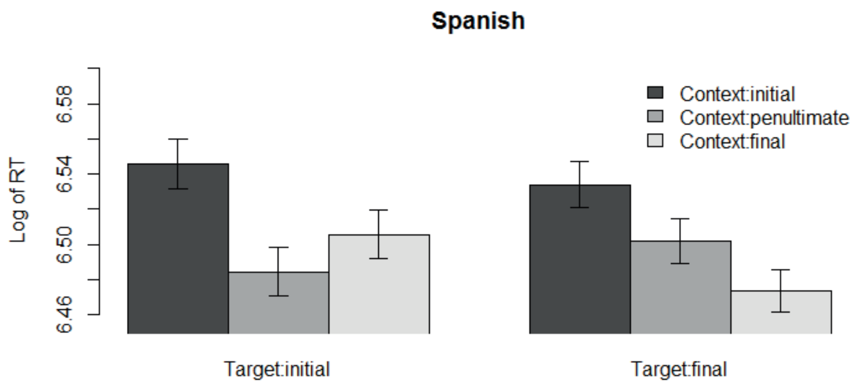
Because the Spanish stress system is linked to syllable structure and morphology, the predictions for the current non-word-spotting experiment depend on some assumptions. The previous results for Dutch and Turkish prompt to expect that Spanish listeners would expect a word boundary after penultimate stress. First of all, penultimate stress is the most frequent position for stress in Spanish even disregarding syllable structure. Second of all, the preceding context in the current experiment is a trisyllabic string with open syllables and the default stress pattern for Spanish on a similar sequence is penultimate[4]. This prediction, however, bears on the assumption that stress-based segmentation is phonological in the sense that the syllable structure, irrespective of morphological structure, is leading. After all, if the open-syllable words were to be perceived as *not* having a thematic vowel, an interpretation which is not inconceivable, they would expect the stressed syllable to indicate a word boundary, that is, final stress on the preceding context would be facilitative in segmentation if the listeners' predictions exclude the thematic vowel.

Since the materials are not designed especially for Spanish listeners, the target non-words are not native. Especially the word /darnam/ is likely to sound foreign to Spanish listeners, since the word-final nasal is non-contrastive in place of articulation, defaulting to [n][5]. The syllable structure of both non-words is such that both would have canonical final stress in Spanish. If the effect of stress on target recognition does not take syllable structure into account, the expectation is that the penultimate pattern is more facilitative in segmentation than the final pattern. If, however, syllable structure has an effect on target recognition, the expectation is that the target non-words with final stress are spotted faster than the target non-words with initial stress.

[4]About 88% of Spanish words ending in a vowel has penultimate stress (Bárkányi, 2002)

[5]This fact took extra effort and care when recording the stimuli in the first place, but with careful instruction and supervision, as well as a surplus of recordings to select from, the segmental quality of the materials was good

Figure 4.3: Effect of stress pattern on logRT for Spanish



4.7.3 Results and discussion

40 Spanish students were tested at the university of Pompeu Fabra, Barcelona, Spain. The participants spoke Spanish as their first language and they met the same criteria as the participants of the other languages. One participant had to be excluded on the basis of an error percentage of 50.6%. The accuracy of the remaining 39 participants was 98.7% and the false alarm rate was 1.2%, which is higher than that of the other languages. The Spanish reaction time data were modeled in the same way as the data of the other languages. The model assesses the effects of the canonical stress pattern on the context and the canonical stress pattern on the target, while modeling out variation caused by participant, trial order, creature type (/darnam/ or /mernel/) and the segmental form of the item. The results show that there is a strong effect of trial order (t(38) = -11.04, p = .000) and of creature type (t(38) = -5.11, p = .000). The Spanish model again shows that there is a facilitating effect of the canonical stress pattern on the context (t(38) = -2.57, p = .02), but not so on the target (t(38) = -0.16, p = .38). A glance at the interaction points at the source of this lack of an effect (t(38) = 2.08, p = .05).

In Figure 4.3, the log reaction times of the participants in the different conditions can be found and here it can be seen what the interaction of context and target is. The figure shows that the target with initial stress is spotted faster after the canonical penultimate stress context than after the other two contexts. At the same time, the target with final stress is spotted faster after the context with final stress than after the other two contexts. It is difficult to explain these results based on canonical stress and syllable structure, since the hypotheses as laid out in Section 4.7.2 do not predict an interaction between the stress pattern on the context and that on the target. The predictions were that penultimate stress would be facilitative if syllable structure is not taken into account, and that final stress on the target would be facilitative if it were. Furthermore, final stress on the context may be expected to be facilitative if the final vowel is not perceived as a thematic vowel, a prediction which is conceivable, but at the same time complicated since it involves morphology while the experiment is not explicitly designed to test the difference between these two interpretations. The main effect of penultimate stress on the context suggests that the phonological interpretation is closer to the truth, but the interaction effect shows that this is not the whole story. It is interesting to see that both final stress and penultimate stress are facilitative in segmentation, but only when both the context and the target agree.

This pattern of results may be interpreted as a harmonizing pattern: the penultimate-stress pattern makes it easier to spot the initial-stress target and the final-stress pattern makes it easier to spot the final-stress target. Besides harmonizing, the effect can be seen as a facilitation of segmentation by alternation. This explanation concurs with the evidence of Bohn et al. (2013), who found, in an ERP experiment with German listeners, that metrical clashes and lapses invoke a cost on processing as compared to alternating stress patterns, despite the abundance of clashes and lapses in German continuous speech. Irregular timing was also tested by Quené and Port (2005), though in this case irregularity was not implemented with numbers of syllables but with seconds. In the current experiment, both conditions which were facilitative for Spanish listeners have one unstressed syllable in between two stressed syllables.

Interpreting the effect as a general cognitive rhythmic pattern is not in line with the results that were found for the other four languages. Figures 4.4, 4.5, 4.6, 4.7 and 4.8 show the logRTs of the different languages, but this time normalized to z-scores. This way, the overall reaction time differences between languages do not distract from the pattern of results.

Bars rising above the middle line represent mean reaction times slower than the overall mean, and those under the middle line represent mean reaction times which are faster than the overall mean. The graph does not add a new statistic but is merely intended to visualize the effects in the different languages. As can be seen, the only language in which the pattern of results as found for Spanish can tentatively be recognized is Polish. There is only a facilitative effect for target in Polish, as can also be recognized by the direction of the bars, but the context in which the final target was spotted with the shortest RTs is the final-stress context. Unlike in Spanish, this facilitation is not statistically significant. The pattern of results found for Spanish are, then, different from the results of the other languages, but also different from what was expected on the basis of Spanish canonical stress. One might argue that both penultimate and final stress are sufficiently canonical for listeners to use in segmentation,

Figure 4.4: Zscores of logRTs per condition for Hungarian

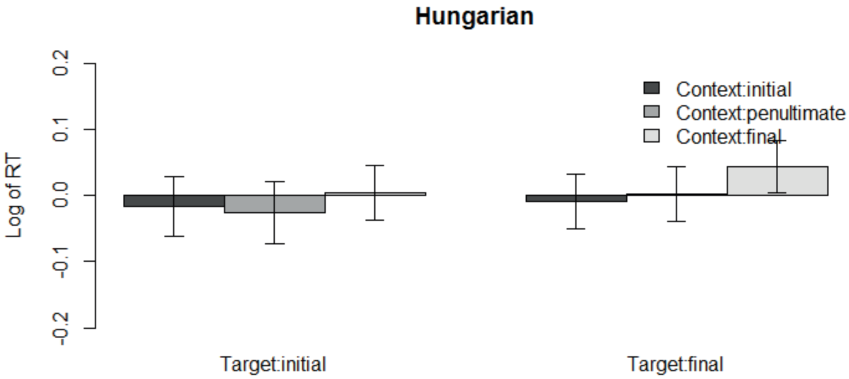


Figure 4.5: Zscores of logRTs per condition for Polish

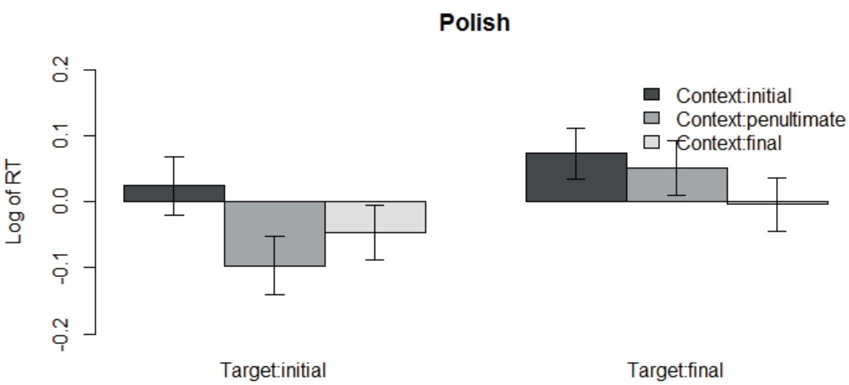
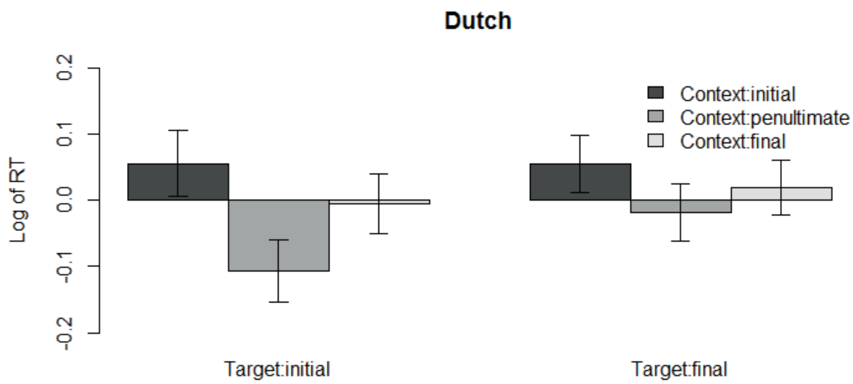


Figure 4.6: Zscores of logRTs per condition for Dutch



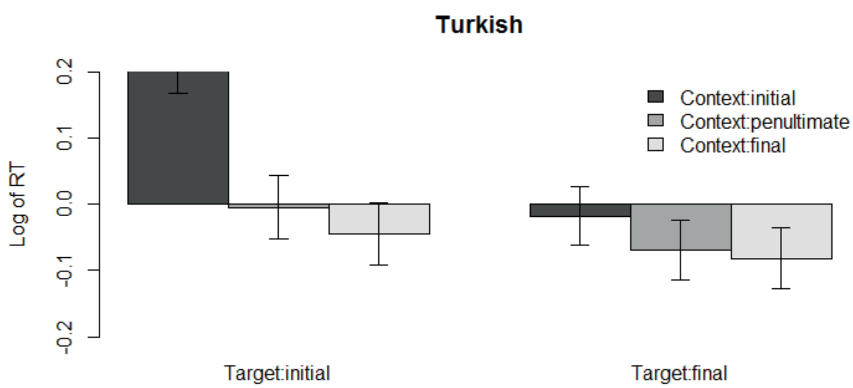
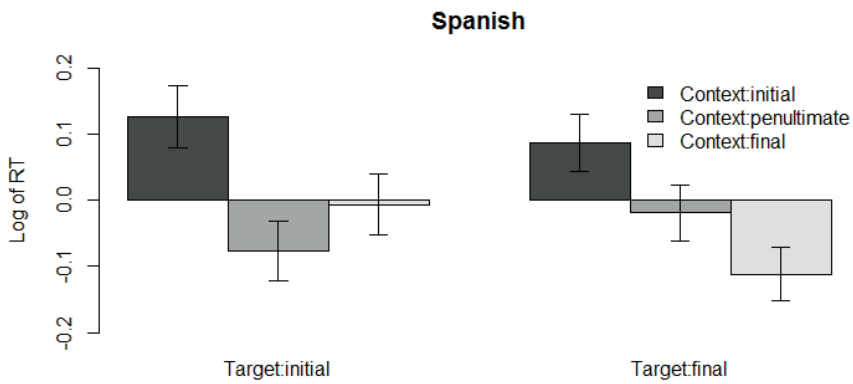


Figure 4.7: Zscores of logRTs per condition for Turkish

Figure 4.8: Zscores of logRTs per condition for Spanish



and that listeners disregard syllable structure. The fact that antepenultimate stress renders longer reaction times than both penultimate and final stress points in this direction. An additional effect of rhythmic alternation or harmony may then be responsible for the interaction effect. The question still remains, however, why this interaction is not discernable in the Dutch data. Dutch, too, has a three-syllable window of which antepenultimate stress is the least preferred position for stress, and both penultimate and final stress are allowed. It is furthermore a language with rhythmic stress alternation, which would prompt the expectation that rhythmic facilitation would surface in the Dutch results. Follow-up research should shed light on the matter. First of all, the Spanish data raise interesting questions about the interaction of alternating rhythmic stress and language-specific edge-aligned stress. Second of all, an experiment specifically targeting a potential interaction of syllable structure and stress in segmentation should clarify whether stress is independent in processing, or whether here too, it is autonomous but not independent of segmental phonology.

4.8 Conclusion

The addition of Spanish to the segmentation data raised more questions than it answered. Although the results showed that Spanish listeners use stress in segmentation, the way in which the facilitation of segmentation is to be led back to the Spanish stress system is not straightforward. The main effect of canonical penultimate stress on the context suggests that segmentation is facilitated by penultimate stress, but the interaction shows that this is only the case when the target has initial stress. Final stress on the context facilitates segmentation of a target with final stress. Although this facilitation may be attributed to rhythmic facilitation, either rooted in the harmony of context and target, or in a strict alternating stress pattern, the same pattern was not found in other languages for which it would be expected, such as Dutch. Follow-up experiments targeted at these questions are needed to disentangle these hypothesized origins of the effects.

CHAPTER 5

Experiment 2: Optimization

5.1 Introduction

In the previous chapter the results of an experiment tapping into a fast, one-directional segmentation strategy called the Language-specific Metronome were described. This strategy facilitates word segmentation by use of word stress and its language-specific relation to the word boundary. While raising new questions, the results of the experiment were to some extent compatible with the hypothesis that in languages with right-edge-aligned stress, the canonical stress pattern activates expectations of the position of word boundaries, thereby making segmentation faster and easier. For languages with positionally variable stress, however, a listener's expectations of upcoming word boundaries do not always coincide with the actual word boundaries. For example, Dutch listeners expect a stressed syllable to be followed by an unstressed syllable, but when the stressed syllable is followed by another stressed syllable, this expectation turns out to be mistaken. The question of the current chapter is whether there is another role of stress in segmentation: a role of perceptual grouping. Alongside the forward-looking, anticipatory role of stress, allowing the listener to 'lock in' on the rhythm of the speaker (Martin, 1972) and expect both the position of the next stressed syllable (e.g. Cutler, 1976; Rothermich et al., 2012; Bohn et al., 2013) and the word boundary, stress may be used as a grouping cue with which the listener has preferences on the position of both word boundaries, to arrive at metrically optimal words. This idea is visualized below.

fast, one-directional parsing
w S w S S w w
. . . →]→ !

bidirectional optimization of segmentation
↙↓↘ ↓ ↙↓↘
w S w S S w w

This view of stress as a bidirectional grouping cue is potentially interesting and an addition to what is already known about stress in segmentation. Besides aligning to word boundaries, word stress is part of a hierarchical prosodic system in which units are grouped together by use of prominence differences. Rather than seeing stress as a feature of a segment (see Chapter 2) or seeing it purely as a cue to one edge of the word (as in Chapter 4), the grouping potential of word stress can add to the knowledge about the role of this abstract and autonomous phonological system in processing.

To investigate whether listeners can make a prelexical bidirectional use of stress, i.e. as a perceptual grouping mechanism, the current chapter outlines an offline segmentation task. Where the online segmentation task in the previous chapter measured listeners' expectations, an offline segmentation task measures the segmentation that listeners arrive at, leaving space for consolidation of what they consider the best segmentation. First, the difference between the unidirectional and bidirectional use of stress in segmentation is hypothesized, after which the design of the experiment will be laid out in Section 5.3.1 and predictions are made in Section 5.2. The results are discussed separately per language, after which they are compared in Section 5.4.7. Lastly, Section 5.5 is concerned with the comparison of the offline task to the results of the online task of Chapter 4.

5.1.1 Stress as perceptual grouping

Anticipation triggered by stress is not a new idea and it has as a logical basis that speech is bound by time and rhythm should help to structure the incoming, fleeting signal, shape expectations and facilitate processing (Martin, 1972; Darwin, 1975). There is behavioral as well as ERP evidence for the prediction of upcoming stressed syllables (e.g. Cutler, 1976; Rothermich et al., 2012; Bohn et al., 2013) and recently, there is increasing attention for the relationship between neural oscillations and linguistic rhythmicity (Peelle and Davis, 2012), including metrical expectancy (Zanto et al., 2006). These studies commonly endorse the view that the listener attends to the stressed syllable to predict where the next stressed syllable can be expected, basically synchronizing with the speaker (Martin, 1972). Since the current thesis is occupied with segmentation, the expectation of stressed syllables is only one part of the story. The experiment in Chapter 4 has shown that the stressed syllable triggers expectations about upcoming word boundaries. However, in languages with variable stress, this use of stress for segmentation is limited in its success, since the prediction of where the word boundary may be is often wrong. A strict unidirectional use of stress is therefore not enough, although a processing advantage of the rhythmicity of speech remains important for speech segmentation. Next to the anticipatory use of stress, a bidirectional grouping effect of stress may be hypothesized. This bidirectional grouping allows for more positional variability of stress and may even be a check of whether anticipation on the basis of stress yields acceptable results, or whether the segmentation can be better.

In the experiment in Chapter 4, the recognition of the segmental shape of the onset of the target is in principle already enough for the listener to place a boundary before it and in that sense segment according to this new information, when it is incongruent with their initial expectations. However, the latency effects of canonical stress on the target additionally provided evidence for the use of stress in segmentation after the onset of the target. As explained before, the target effect found in the previous experiment likely comprises both lexical and prelexical effects. The use of stress in target recognition may be comparable to the anticipatory use of context stress, that is prelexical, but part of the effect may also be that it is easier to recognize a target that conforms to the canonical word stress pattern. In fact, the latter effect is known to exist, as already discussed in Section 2.2. The former, however, is still an open question.

In metrical phonology (Hayes, 1995), word stress is characterized as a hierarchical system where weak units are combined with strong units, on the syllable level, foot level and so on. Because words are units which are better characterized as chunks than as segments in a sequential relation, this use of stress is better characterized as a grouping cue than as a unidirectional guiding cue. Next to the forward-looking use of stress as related to a boundary, the prosodic shape of words as a unit may be hypothesized to be used in segmentation. The Language-specific Metronome was measured with an online task, intended to measure where listeners expect a boundary. The rapidity of the task makes it possible to measure differences in early processing, and therefore effects of anticipation. It is an open question whether a bidirectional strategy in metrical segmentation can be measured online as well, but it seems fairly simple to measure prosodic grouping in an offline task. The offline task would give the listeners enough time to settle on an optimal grouping of syllables and the comparison of this optimal grouping to the outcome of the online task in Chapter 4 may give an insight into whether listeners decide on a different segmentation when given more time. When there is no segmental or lexical information prompting listeners towards a specific segmentation, do listeners try to optimize the stress pattern of their segmentation to make metrically optimal words?

5.2 Predictions

The experiment in the current chapter is sufficiently similar to the Metronome experiment in Chapter 4 on the level of stress patterns, to allow for a comparison of the results. The difference between the experiments is first of all that the current experiment is an offline experiment, while the Metronome experiment was an online experiment. Furthermore, the stimuli in the current experiment are completely heterogeneous and do not contain the non-words that participants in the Metronome experiment were asked to spot. Lastly, listeners are allowed to segment the sequences freely, according to their preferences. Listeners are native speakers of Dutch and Turkish, since these languages have positionally variable stress and therefore in principle allow for gradual differences in segmentation preferences.

Different outcomes may be expected. The null hypothesis is that the listener does not use stress at all in the segmentation of the stimuli. To be able to find out whether this is the case, segmentations are compared across conditions. This is described in Section 5.4.2. Since conditions differ on the basis of stress alone, differences between conditions would be evidence that stress has an effect on segmentation preferences. If so, the question arises where these preferences are rooted. On the one hand, segmentations may resemble the listener's lexicon. In other words, the listener may have an idea about the optimality of stress patterns of words on the basis of the words they know. To investigate whether this is a consideration in segmentation, the stress patterns of the words that are segmented in the experiment are compared to the stress patterns in the listener's native lexicon on the basis of frequency. On the other hand, segmentations may be made according to metrical considerations. In this case, the listener has metrical considerations about optimal word forms, which they will segment where possible. This hypothesis is given below.

> **Language-specific Metrical Grouping**
> When given time, listeners use the stressed syllable as a grouping cue. Their perception of the position of word boundaries is influenced by optimal metrical units in their language and, because of the bidirectional nature of grouping, differs from the expectations based on the Language-specific Metronome.

To be able to investigate whether this is the case, the segmentations have to be regarded in context, to see what drives the listener to make certain segmentations. For example, for Dutch the trochee, that is, a disyllabic unit with initial stress, is an optimal metrical unit. However, the materials of the current experiment do not allow for only disyllabic trochaic units to be segmented, since this segmentation strategy would automatically lead to residual syllables. This can be seen in Example 5.2.

(5.1) [S w] w [S w]

(5.2) w [S w] w S

The choices the listener makes with regards to these 'residual' syllables should be informative toward the concessions listeners are willing to make in their segmentations. Do Turkish listeners prefer to include extra unstressed syllables to make longer words where Dutch listeners would want to keep unstressed syllables as separate (function) words? Would a Dutch listener prefer a monosyllabic stressed word over a longer word with final stress? These decisions would amount to an insight in metrical optimization in segmentation.

5.3 Method

5.3.1 Experimental design

The current experiment[1] was a self-paced, non-timed task, aimed at tapping into a relatively late segmentation strategy, in the sense that the results are the segmentations the listener eventually settles on. The prediction is that the results reflect a difference between optimization and expectation (cf. Chapter 4).

5.3.2 Participants

32 Dutch and 32 Turkish native speakers, all of them university students, took part in the current experiment. The Dutch participants were tested in the phonetic lab of Utrecht University, Utrecht, The Netherlands, and the Turkish participants were tested on the campus of Koç University and Boğaziçi University, Istanbul, Turkey. All participants received compensation and none of the participants took part in the Metronome experiment described in Chapter 4. They were were tested individually. None of the participants reported any speech, reading or hearing disorders and all had normal or corrected-to-normal vision. Participants were instructed in their native language and they were preselected on the criterion that their native language was their dominant language.

5.3.3 Materials

This Optimization experiment directly links to the Metronome experiment with its materials and conditions. The materials are five-syllabic strings of open syllables, which are harmonious according to the rules of Turkish vowel harmony.

[1] The idea for this design arose during the presentation of the Metronome experiment, at a meeting of the Phonology Circle at the Massachusetts Institute of Technology on December 13, 2013. At the time, it was meant to be a control experiment, but upon later thinking about the implications of the task, interesting hypotheses developed. I want to thank Michael Kenstowicz, Donca Steriade, Edward Flemming, Adam Albright and all the other members of the audience for their helpful suggestions and comments.

The syllables are phonotactically legal and consist of legal segments in both Dutch and Turkish, and they are controlled for frequency, positional frequency and interaction with stress. They are, namely, the target-less filler items of the Metronome experiment. The items were formed by random assembly of the selected syllables described in Section 4.3.3 and, as described there, they were pronounced by a female native speaker of Spanish. Phonetic resynthesis was done as it was on the experimental items in the Metronome experiment.

The fact that the items do not contain a target as they did in Chapter 4 leaves the participant free to segment at any given syllable. All items were recorded integrally in a carrier phrase, so there are no acoustic cues in the item to where a word boundary would be expected. All 160 items are heterogeneous and each participant heard and transcribed 56 randomly selected items. The heterogeneity and the random selection together minimize the chance of word-learning through accidental co-occurrence of syllables as well as the chance of boredom in the task. This lack of segmental co-occurrence increases the likelihood that segmentation is based on stress patterns.

The experimental conditions were the same as in the Metronome experiment, but this time, the conditions cannot be seen as contexts and targets, because there is no embedded target and no learning phase. Rather, there are six conditions, given in Table 5.1, each of which has two stressed syllables and three unstressed syllables. The decision to repeat the conditions of the first experiment in the second experiment makes it possible to make a direct comparison between the two experiments. However, a limitation of asymmetry that is intrinsic to this set-up should be noted. This asymmetry can be found in in Table 5.1, in which all logically possible stress conditions with two stressed syllables and three unstressed syllables are given, and each condition is presented with their mirror image directly below, if applicable (e.g. wSwSw and SwwwS are symmetrical). Next to the stress patterns, the relevant metrical conditions which are part of these patterns are listed. The trochee is Sw, the iamb is wS, the clash is SS and the lapse is ww. Both the trochee and the iamb are metrical feet, clash is a dispreferred sequence of stresses, in which a boundary may be assumed and lapse is a sequence of unstressed syllables, which is dispreferred in languages which have rhythmic alternation (such as Dutch), but less so in languages which do not (such as Turkish). The logically possible conditions which are not used in the experiment are marked with grey. The conditions in white are the stress conditions which were part of the experiment in Chapter 4, to which the current experiment is modeled exactly. The relevant metrical conditions show that the test items allow for the segmentation of, proportionally, one clash, six lapses, nine iambs and eight trochees.

The asymmetry of the setup will result in an asymmetrical frequency of boundary placement at different positions in the phrases, because there are two conditions in which there are two stressed syllables in the three-syllable window to the right (wwSwS and wwSSw), but zero conditions in which there are two stressed syllables in the three-syllable window to the left (wSSww and SwSww). Table 5.1 shows that all relevant metrical features are tested despite

Table 5.1: Stress conditions

Cond.	**Exp.**	**Relevant features**			
		Trochee	**Iamb**	**Lapse**	**Clash**
SSwww		1		2	1
wwwSS			1	2	1
SwSww		2	1	1	
wwSwS	ν	1	2	1	
SwwSw	ν	2	1	1	
wSwwS	ν	1	2	1	
wSSww		1	1	1	1
wwSSw	ν	1	1	1	1
wSwSw	ν	2	2		
SwwwS	ν	1	1	2	

w = unstressed, S = stressed, Cond. = condition
Exp. = used in experiment, grey = not part of experiment

this asymmetry. Metrically, the only asymmetry is the fact that there are eight trochees and nine iambs in the input.

5.3.4 Procedure

The participants were told that they would listen to sentences in an unknown language. They were asked to type out each sentence after hearing it completely, separating the words by spaces. Next, they heard the sentence again and were asked whether they wanted to revise their answer. If they did, they retyped their answer, and if they chose not to do so they would proceed to the next item. Participants wore headphones and they were seated in a quiet room. The experiment was written in Python. It was self-paced, not timed and could be used on any computer. Both in Utrecht, the Netherlands, and in Istanbul, Turkey, a laptop was used.

Participants reported that the language sounded ‘weird’, or it was like ‘some African language’, ‘Asian-flavored’, or ’Basque’, while some reported it sounded like Spanish, Italian or French. Furthermore, their reports indicated their preoccupation with the segmental shape of the stimuli rather than with melody or rhythm. This is understood from the fact that they reported on which sounds they thought should be written with which graphemes, and the revisions participants made during the experiment confirm this observation: in most cases participants revised segments, but not word boundaries. When asked about how they grouped the syllables or formed the words, most participants did not initially know, but they did report that their intuitions were strong. Eventually some participants concluded that this was because of ‘melody’, ‘intonation’ or ‘speed’ or ‘pauses’. As described in Section 4.3.4, there were no speed differences or pauses in the materials.

Results consisted of the literal responses of the participants. These were analyzed to be simplified to syllables and word-boundaries, and linked to the metrical form of the experimental item. See an example of stimulus, response and result below:

(5.3)	experimental item:	[ˈfelisiˈfefi]
	response participant:	‘fe lisif feffi’
	result:	[S][ww][Sw]

The result is a sequence of stressed and unstressed syllables for each case. The exact segmentation of each stress condition gives information about the effect of context on segmentation, an analysis which will be further explained in Section 5.4.1. Additionally, the separate ‘words’, henceforth referred to as units, that are segmented from the stress condition, in this case ‘S’, ‘ww’ and ‘Sw’ will be compiled in the output lexicon, which is also further explained in Section 5.4.1. For now it is most important to realize that this output lexicon will therefore consist of metrical units and their frequency.

5.4 Results

5.4.1 Analyses

In this section the results of the Optimization experiment are discussed. First, the general bias of results is examined. This is done is Section 5.4.2. The bias is the tendency of participants to display segmentation behavior irrespective of the experimental condition. An example of a bias is for participants to prefer words with fewer syllables to precede words with more syllables. This is a tendency in frozen expressions sometimes referred to as Panini’s Law and it was among other languages tested for English, French and Hungarian (Pinker and Birdsong, 1979; Pordány, 1986; Ryan, 2013). The bias found in the current results will serve as a baseline in the lexicon analysis, in the sense that the condition-specific behavior of participants will be compared to their overall behavior. Rather than testing a hypothesis, this bias is meant as a control condition to make the results of the actual analysis interpretable. Note that the bias which will be used is still language-specific, because it serves to learn about the effect of condition and it therefore has to be controlled for all other factors, including language.

Second, there will be a metrical analysis of the output lexicon. The output lexicon is a result of the segmentations of the sequences by the participants. This lexicon consists of units and frequencies and these frequencies are evaluated by their relation to frequencies in the native lexicon as well as to frequencies in an output lexicon which would emerge from a random segmentation of the stimuli[2]. This analysis will make clear whether the participants segmented at random, whether their segmentations are guided by metrical choices, whether the output lexicon is much like the native lexicon, or whether there are other factors guiding segmentation.

Third, the segmentations will be regarded per condition. When interpreting the results of the output lexicon it is important to remember that the units are not stand-alone units, but they are parts of a five-syllable sequence. For example, when segmenting Sw out of wSwwS, it is unavoidable that the initial w is segmented as a separate unit, and a choice has to be made about the final wS, too. The output lexicon is therefore not a measure that can stand by itself, but it needs to be complemented by a contextual analysis, which is the third of the analyses and it can be found in Section 5.4.6. In that section, the influence of the context is the main subject of interest. The output lexicon per condition gives an insight into the preferences of listeners and how these are influenced by the context. This role of context makes it an analysis of optimization: choices are made *given* the context. It will be possible to find out whether participants for example prefer to segment Sw even though this would mean that there is a single w left, or whether participants will solve the potential problem of this single w by segmenting wSw.

The condition-general bias is discussed for both languages in Section 5.4.2. Then, the measures used are described in Section 5.4.3 after which the lexicon analysis is discussed for Dutch and Turkish in Section 5.4.5. Next, the results of the contextual analysis of both languages is discussed in Section 5.4.6. Lastly, the experiment is compared to the Metronome experiment in Section 5.5.

5.4.2 Bias

Two Dutch participants had to be excluded from the analysis because they did not seem to have understood the instructions; one Dutch participant did not place any boundaries, i.e. they just typed five-syllable units for each stimulus; one Dutch participant separated all syllables with boundaries, rendering solely monosyllabic units. Data from all other Dutch and Turkish participants were included and the overall bias is the preference to divide the five-syllabic items into two or three units.

The overall probability for participants to place a boundary at any given position is 33% when native language is not taken into account. These numbers change when the boundary position and language are taken into account. This can be seen in Table 5.2.

The overall most likely position for boundaries across conditions is after the second or third syllable, of which that after the third is more likely than that after the second. This is the case in Dutch as well as in Turkish. At

[2]Note that this random segmentation will still be subject to the general bias found in Section 5.4.2.

Table 5.2: Overall probability of boundary placement

position	after 1st σ	after 2nd σ	after 3rd σ	after 4th σ	**overall**
Dutch	0.12	0.53	0.74	0.15	0.39
Turkish	0.03	0.42	0.57	0.04	0.26

first sight this may be counterintuitive because Panini's Law predicts words with fewer syllables to precede longer words. Seemingly therefore the current experiment gives evidence contrary to Panini's law. However, Jespersen (1905) already calls attention to the fact that frozen expressions are often ordered according to rhythmic principles, which can correlate with word length, such as: 'bread and butter' and 'rough and ready'. Only in the case that both word orders would render a rhythmic result would Panini's law decide, is what Pinker and Birdsong (1979) suggest. The reason for the overall bias for the boundary after the third syllable cannot be sought in Panini's law, but it can be found in the asymmetrical design of the study, as discussed in Section 5.3.3. There are, namely, two conditions in which there are two stressed syllables in the three-syllable window on the right, and zero conditions in which there are two stressed syllables in the three-syllable window on the left. This means that the conditions in themselves are more likely to have boundaries in the three-syllable window on the right-hand side of the five-syllable item than on the left-hand side, provided of course that it is true that a boundary is likely to be placed to prevent two main stresses to be part of the same word.

Besides the asymmetry between left and right, it can be observed that the probability of a word boundary after the first and fourth syllable is much lower than after the second or third syllable. This probability would have been higher if the two conditions with edge-aligned clash would have been included in the conditions, but it is also possible that it would be a more general effect of preferred word size. The overall probability of boundary placement is around 33%, meaning that the average stimulus consists of one or two words, which suggests a preference for disyllabic and trisyllabic units over monosyllabic units[3]. This issue will be discussed in detail in the context-dependent analysis.

[3]Even though the phonetics of stress are not a subject of research in the current thesis, I am aware of the role of phonetics in perception. The stimuli are designed to be phonetically natural and the stress in the stimuli is designed to be as acceptable as possible for all listeners, and to be interpreted as stress in a polysyllabic word. In the current experiment it becomes relevant to realize that a monosyllabic word is relatively longer than the same stressed syllable in a disyllabic word (Lehiste, 1972). The phonetics of the stimuli may influence the perception of single syllables as monosyllabic words, because they may not typically have the duration to be interpreted as stand-alone words. That being said, the results do show that participants segment monosyllabic words ($n = 1176$ for Dutch and $n = 272$ for Turkish), which refutes the idea that the phonetics of the current stimuli categorically disallow a monosyllabic interpretation.

An interesting first difference between Turkish and Dutch which can already be discerned in this overall analysis is the suggested preferred word size difference between Turkish and Dutch. In Dutch, the overall probability of boundary placement is 39%, whereas that of Turkish is 26%. This higher probability of boundary placement in Dutch is bound to result in shorter words. The distribution of probabilities over positions, meanwhile, is virtually equal, with a difference of 17% between the languages after the third syllable, and 11% at each other position. This comparability in behavior overall makes any found differences between conditions and languages more meaningfully attributable to stress than if the languages would differ overall. It furthermore strengthens the idea, although one can never be certain with only two languages, that the overall bias is attributable to the experimental design.

5.4.3 Measure

For the two remaining segmentation measures, the responses of the participants are collected in an output lexicon, which is a frequency list of the segmented lexemes irrespective of condition, and an output lexicon per condition for the contextual analysis. Both the lexicon analysis and the contextual analysis are done separately for each language and in both cases the frequency is the important measure in the analysis. The frequencies of units start to gain their value when compared to the frequencies of the same units in both a native lexicon and a random lexicon. This comparison is done with a very simple measure, called the Observed over Expected ratio (O/E), given in Example 5.4.

(5.4) O/E: $\frac{\mathrm{f}(xy)}{XY} \;/\; (\frac{\mathrm{f}(x)}{n} \times \frac{\mathrm{f}(y)}{n})$

In the comparison with the native lexicon, the Expected frequency is the frequency of the same item in the native lexicon. Because the corpora are of different sizes, the calculation is the relative frequency of the unit in the output lexicon of the experiment divided by the relative frequency of the unit in the native lexicon (implemented as a corpus). In the comparison with a random lexicon, on the other hand, the calculation is slightly different. In this case, the experimental items are segmented according to the overall bias as found in 5.4.2. So, dividing the same number of stimuli as the participant was presented with, according to the overall bias of all participants of that language, results in an output lexicon with a certain frequency for each unit. The relative frequency of the unit in this randomly generated lexicon is the expected frequency in the comparison.

Because it is a simple ratio, an O/E of 1 means that the unit is as frequent in the current experiment as would be expected on the basis of either the lexicon or the randomly generated lexicon. If it is the former, the output is close to the native lexicon, which may be interpreted as the lexicon driving segmentation decisions. If it is the latter, the listener may make random segmentations, not minding the metrical shape of the output. If the unit is overrepresented or underrepresented compared to either lexicon, the listener is making choices according to different constraints, which are likely to be metrical, considering the design of the study.

The native lexicon is modeled by a frequency count of metrical units in a self-compiled corpus. This corpus was compiled for the computational study in Chapter 6 and 7 and the design of the corpus will be discussed in detail in Chapter 6. For now it suffices to say that the corpus consists of spontaneous conversations between different speakers and on different subjects. The corpus consists of 1553 words for Dutch and 804 for Turkish. It is a very small corpus, but it should be kept in mind that it is not the words, but the stress patterns of words that matter for the current comparison. For the current analysis, the phonological[4] transcription of stress in that corpus is used, which means that the position of word stress in the corpus is transcribed by the author (Dutch) and a Turkish linguist (Turkish) and the transcriptions are checked by one native speaker and trained phonologist in Dutch, and three native speakers in Turkish.

Another important and sensible comparison is that of the output lexicon to a random lexicon. It is, after all, possible that the participants segment the items at random, i.e. without any stress- or lexicon-based system. To see whether this is the case, a random lexicon is used in the comparison. This lexicon is constructed by use of the segmentation bias of the participants when condition is not taken into account. This bias was discussed in Section 5.4.4, where it was given as the probability at each position in the five-syllabic stimulus that a participant will segment there. The random frequency of each segmentation is a product of these probabilities and the number of times this stress pattern was presented in the experiment. The formula of this calculation is given below, in which fSP is the frequency of the stress pattern and pσ is the probability of the presence or absence of a boundary at that position in the sequence. fSegm is, then, the frequency of that particular segmentation. The frequency of segmentation is the information which is used in the contextual analysis, and for the output lexicon analysis, the segmentation is divided into separate lexemes to add to the output lexicon.

(5.5) $fSP * (p\sigma 1 * p\sigma 2 * p\sigma 3 * p\sigma 4) = fSegm$

pσ is a number which depends on whether or not a boundary is placed in that position. When a boundary is placed, the number is pσ, as given in Table 5.2, and when a boundary is not placed, the number is 1-pσ. For example[5], the probability of the segmentation S][ww][Sw by Dutch listeners:

[4] In the analyses in Chapter 6 and 7 the position of stress is rendered by a perception survey with Dutch and Turkish participants.

[5] All numbers in these equations were rounded off in the text, but used unrounded in the calculation, meaning that the outcomes diverge slightly from the outcome with rounded off numbers

(5.6)	Probability of $\sigma][\sigma\sigma][\sigma\sigma$:	.12*(1-.53)*.74*(1-.15)=.036
	Frequency of SwwSw:	276
	Expected frequency of S][ww][Sw:	.036*276=10.00
	Observed frequency of S][ww][Sw:	14
	O/E S][ww][Sw	1.399

The frequencies of the randomly generated lexemes serve as Expected frequencies in the comparison of the output lexicon to the random lexicon.

5.4.4 Overall findings

5.4.5 Output lexicon analysis

The results will first be discussed on the basis of the output lexicon, after which a context analysis is discussed in which the circumstances for suboptimal choices are studied, and the arrangement of preferences are discussed.

Dutch output lexicon

Table 5.3 shows the output lexicon that emerges from the Dutch segmentation of all stress patterns taken together. Column 1 contains units arranged by their frequency in the emerging lexicon, being the stress pattern that was segmented most frequently. Column 2 contains the output frequency of that unit. In column 3, the Random Frequency is given, which, as explained in Section 5.4.3, is based on a random segmentation of the stimuli according to the general overall bias of Dutch participants. In column 4, the O/E is given on the basis of this random frequency, and in column 5 and 6, the Input Frequency and Input O/E are given, based on a phonological transcription of stress in a corpus of spontaneous speech.

Table 5.3 shows that participants most frequently segment the trochee (Sw). This unit is more frequent than all other units, and it is also more frequent than what would be expected from a random segmentation of the experimental items (O/E = 1.34) and more frequent than what would be expected from the native lexicon (O/E = 2.17). The next most frequent unit is the Iamb (wS). This unit, however, is slightly less frequent than what would be expected on the basis of a random segmentation of experimental stimuli (O/E = 0.89). It is far more frequent than it is in the native lexicon (O/E = 4.23). Both monosyllables (w and S) are slightly less frequent than in the random lexicon (O/E = 0.81 and O/E = 0.92), but compared to the native lexicon, the stressed monosyllable is highly underrepresented (O/E = 0.22). The unstressed monosyllable is as frequent as expected on the basis of the native lexicon (O/E = 1.03). The two most frequent trisyllables (wwS and wSw) are, in turn, overrepresented. The random lexicon does not contain as many of these units (O/E = 1.75 and O/E = 1.61), even though the probability of segmentation at each separate position was carefully taken into account. Compared to the native lexicon, the

Table 5.3: Complete Dutch output lexicon

1	2	3	4	5	6
Unit	**Freq.**	**R-Freq.**	**R-O/E**	**I-Freq.**	**I-O/E**
Sw	1080	804.01	1.34	178	2.17
wS	702	792.26	0.89	68	4.23
w	684	841.24	0.81	248	1.03
S	492	533.07	0.92	851	0.22
wwS	410	233.96	1.75	26	5.87
wSw	348	216.58	1.61	57	2.30
ww	265	294.86	0.90	0	-
Sww	84	171.01	0.49	51	0.63
wwSw	17	7.39	2.30	19	0.33
SwS	10	30.74	0.33	0	-
SSw	10	34.40	0.29	0	-
wwSwS	9	23.08	0.39	0	-
wSwSw	6	21.84	0.27	0	-
SwwwS	6	25.85	0.23	0	-
Swww	5	4.49	1.11	23	0.10
wwSSw	4	26.29	0.15	0	-
wSwwS	4	23.86	0.17	0	-
SS	3	5.98	0.50	0	-
SwwSw	3	24.15	0.12	0	-
wSww	3	4.15	0.72	9	0.13
www	2	0.63	3.18	0	-
wwSS	1	4.57	0.22	0	-
SwwS	1	7.54	0.13	0	-
wwwS	1	3.52	0.28	7	0.05
Total	4150			1553	

w = unstressed, S = stressed

units are even more highly overrepresented (O/E = 5.87 and O/E = 2.30). The corpus which functions as a model for the Dutch native lexicon does not contain unstressed disyllables, which is a choice of transcription. Comparing the frequency of ww to the frequency of that unit in the random lexicon, it is slightly underrepresented (O/E = 0.90). The next trisyllable (Sww) is, opposite to the earlier two trisyllables, highly underrepresented, even when compared to the random lexicon (O/E = 0.49). In the native lexicon, too, this unit is much more frequent (O/E = 0.63). The last possibly meaningful unit is wwSw. This unit is highly overrepresented in comparison to the random lexicon (O/E = 2.30), but highly underrepresented in comparison to the native lexicon (O/E = 0.33). The units which are less frequent than wwSw (N=17) can be considered to be noise. Taking into account that there are 30 participants and 56 items per participant, units with a frequency of less than 15 can be assumed to be accidental or exceptional. That being said, it should be noted that the random lexicon would expect some of these units to be segmented more frequently than they have been, such as for example the unit SwwwS, which would be expected almost 26 times, but has been segmented only 6 times. The exceptionality of these underrepresented units may therefore still be taken to imply that they are metrically dispreferred.

Some observations can be made based on the results. First, there is a difference between the comparison of the output lexicon to the random lexicon and the comparison of the output lexicon to the native lexicon. The numbers in the comparison to the random lexicon are more moderate, that is, the differences are smaller. It seems reasonable to assume that if a unit is overrepresented both in comparison to the random lexicon and in comparison to the native lexicon, the unit is metrically or contextually favored by the listeners. After all, it is a combination of a constrained context and stress as the only freedom of choice. If a unit is overrepresented as compared to the native lexicon, this either means that there was a high probability that the unit is segmented from the constrained context of the experiment, since there are not too many options, or that the unit is metrically favored, or both. In turn, the only difference between the random segmentation and the listener's segmentation is the listeners' choice to segment using stress, so this choice must be metrically favored, despite the contextual constraints. Combining the two reasons, it must be true that a unit which is overrepresented compared to both lexicons is metrically favored. In contrast, if a unit is overrepresented as compared to the native lexicon and underrepresented as compared to the random lexicon, the segmentation of this unit is likely to be an effect of the context of the experiment rather than a metrical preference, since there are relatively many opportunities to segment the unit and the listener has avoided segmenting it where possible, hence the underrepresentation as compared to the random lexicon.

Sw, wwS and wSw are the units which are overrepresented as compared to both lexicons. Sw and wSw are clear examples of metrically well-formed units in Dutch, bearing penultimate stress. The preference for wwS is more surprising. Metrically, this unit would not be expected to be frequent. It is legal (e.g. ʃoko'la, *chocola*, ‘chocolate’), but not optimal. wwS could alternatively be segmented as w][wS or ww][S. These units, however, are all dispreferred, especially in comparison to the random lexicon. Compared to the native lexicon, all of the units into which wwS could be segmented (wwS, ww, w, wS) are overrepresented or as frequent as expected, except for the stressed monosyllable (S), which is highly underrepresented. This result suggests that listeners prefer segmenting a suboptimal trisyllable over segmenting multiple smaller units, even if the stressed monosyllable is highly frequent in the native lexicon. This is a contextual argument, regarding the listeners’ choices in suboptimal contexts, and it will therefore be discussed further in the next section. Another unit which is underrepresented as compared to the native lexicon is Sww. This unit is also underrepresented as compared to the random lexicon, which indicates that it is categorically dispreferred by listeners because despite the availability in the materials, listeners avoid segmenting this unit.

When a unit is overrepresented as compared to the native lexicon, but underrepresented as compared to the random lexicon, there likely is a methodological effect. This difference in representation between the random lexicon and the input lexicon is the case for wS. The underrepresentation of wS as compared to the random lexicon suggests that listeners disprefer the metrical shape of this unit, but the fact that the unit is still overrepresented as compared to the native lexicon suggests that this dispreference for the unit is not enough for listeners to avoid it altogether. Again, the analysis of the context in the next section may shed some light on this matter.

The unit wwSw has its ratios the other way around; it is overrepresented compared to the random segmentation and underrepresented compared to the native lexicon. This particular quadrisyllable is highly overrepresented in the output lexicon compared to the random lexicon, even though the absolute frequency is low. The random lexicon indirectly takes the size preferences of the participants into account, and quadrisyllables are rare in the output, though not so rare in the native lexicon. The participants seem to make an exception for wwSw, however, which is likely to be due to its metrical shape, having penultimate stress.

The least frequent and most underrepresented units are those with two main stresses. This is confirmation that the experiment functioned in the sense that the stresses in the test items were perceived as main word stress and participants’ segmentations were stress-based. It also explains why participants do not segment long words: there are few possibilities of long units containing only one main stress in the materials (namely Swww, wSww, wwSw, wwwS). Taft (1984) also mentions listeners’ general preference for dividing the string in half and/or into disyllabic units in an experiment with comparable materials. She found this preference in an experiment inquiring into the Salience to Onset Strategy in Japanese.

Turkish output lexicon

Table 5.4 contains the output lexicon that emerges from the Turkish segmentation of all stress patterns taken together. Column 1 contains units arranged by their frequency in the emerging lexicon, being the stress pattern that was segmented most frequently. Column 2 contains the output frequency of that unit. In column 3, the Random Frequency is given, which is, as explained in Section 5.4.3, the calculated frequency on the basis of the general overall probabilities of boundary placement by Turkish participants. In column 4, the O/E on the basis of this random frequency is given and in column 5 and 6 the Input frequency and the Input O/E are given, based on a phonological transcription of stress in a spontaneous speech corpus.

In Table 5.4 we see, first of all, a surprisingly high number of trochees, overrepresented both compared to the random segmentation and compared to the input corpus. In the discussion of the Dutch results, an overrepresentation on both comparisons was interpreted to be led by metrical considerations. In this Turkish output lexicon, other units which are overrepresented in both comparisons are wwS, wS and wSw. In Dutch, the three overrepresented units were the same, except for wS, which is not very convincingly overrepresented in Turkish either, with O/Es 1.00 and 1.11. Monosyllables are underrepresented as compared to both the random lexicon and the input lexicon, which suggests a dispreference against monosyllables. This was also found in the Dutch results, although there the O/E compared to the random lexicon was much closer to 1, In sum, this first look at the lexicon surprisingly suggest that the segmentations do not differ greatly between languages.

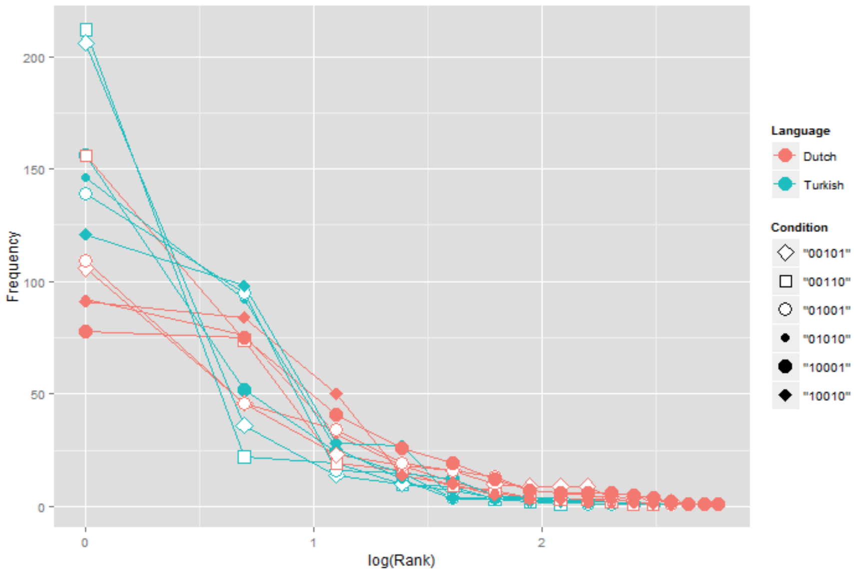
Important for the interpretation of the results of the lexicon analysis is the fact that the segmented units are not stand-alone units. The comparison of the output lexicon to the random lexicon is very useful, because the random lexicon is segmented from the materials and therefore equally limited by the context as the listeners were. The native lexicon is not segmented from a similar context, and therefore serves as a second comparison: if units are more frequent in the output lexicon than in both the random and the native lexicon, this strongly suggests that participants have a metrical preference for this unit. Similarly, if a unit is less frequent than would be expected from both the random and the native lexicon, the unit may be dispreferred for metrical reasons. The current section, in sum, shows that the output lexicon of the segmentation experiment is not directly predicted by the native input lexicon. While it is true that the least frequent segmentations in the output lexicon are also infrequent or unattested in the native lexicon, it is also true that the frequency distribution of the more frequent segmentations is very different from that of the native lexicon. This is the case for both Turkish and Dutch, although the deviations are bigger for Dutch than for Turkish, judging from the O/Es. The next section investigates the role of context.

Table 5.4: Complete Turkish output lexicon

1	2	3	4	5	6
Unit	**Freq.**	**R-Freq.**	**R-O/E**	**I-Freq.**	**I-O/E**
Sw	876	704.16	1.24	112	1.79
wwS	730	276.96	2.64	101	1.61
wS	699	700.07	1.00	144	1.11
wSw	463	283.58	1.63	50	1.89
w	156	371.45	0.42	108	0.32
Sww	153	183.01	0.84	32	1.10
S	116	211.35	0.55	58	0.47
ww	108	236.64	0.46	15	1.55
SwS	36	48.67	0.74	0	-
SwwSw	27	68.14	0.40	0	-
SwwwS	24	65.10	0.37	0	-
SSw	22	48.06	0.46	0	-
wwSSw	19	65.10	0.29	0	-
wSwwS	15	66.50	0.23	0	-
wwSwS	14	65.80	0.21	0	-
wSwSw	12	67.20	0.18	0	-
wwSw	10	4.85	2.06	29	0.07
wSww	4	2.94	1.36	15	0.06
Swww	3	2.88	1.04	17	0.04
wwwS	2	1.85	1.08	57	0.01
www	2	0.08	24.40	1	0.46
wwSS	2	2.88	0.69	0	-
Total	3493			804	

w = unstressed, S = stressed

Figure 5.1: Steepness of frequencies in Turkish and Dutch segmentations per condition



5.4.6 Contextual analysis

In this section I will discuss the circumstances which influence the participants' placement of boundaries. For both Turkish and Dutch, the participants' segmentation preferences will be examined at the phrase level, so the participants' preferred segmentations are compared per condition. First, the pattern of frequencies of segmentations can be found in Figure 5.1. The frequencies of each segmentation are ranked and the ranks are given on a log-scale, to visualize how the conditions differ in the highest ranks. It can be seen that in both languages, clash (condition 3) is the condition on which there is the highest agreement on the most optimal segmentation. The next in line in Turkish is condition 6 (wwSwS), and in Dutch it is condition 5 (wSwwS) and 6 (wwSwS). For Dutch, the remaining three conditions seem to have two optimal segmentations, while this is never the case for Turkish.

This difference in the distribution of segmentation frequencies suggests that Dutch listeners are more lenient and different segmentations can be optimal. The Dutch stress system is positionally more variable than that of Turkish, which is a potential explanation for this cross-linguistic difference.

Table 5.5: Probability of boundary placement per position for Dutch and Turkish

Language	Condition	1st σ]	2nd σ]	3rd σ]	4th σ]	N.
Dutch	**SwwSw**	0.13	0.72	0.64	0.05	271
	wSwSw	0.14	0.50	0.83	0.03	245
	wwSSw	0.13	0.37	0.94	0.05	295
	SwwwS	0.12	0.78	0.52	0.29	290
	wSwwS	0.13	0.42	0.76	0.17	268
	wwSwS	0.08	0.37	0.76	0.29	259
	overall	0.12	0.53	0.74	0.15	1628
Turkish	**SwwSw**	0.05	0.54	0.46	0.02	291
	wSwSw	0.01	0.43	0.63	0.03	287
	wwSSw	0.01	0.12	0.85	0.04	278
	SwwwS	0.05	0.68	0.27	0.07	278
	wSwwS	0.02	0.57	0.42	0.05	284
	wwSwS	0.01	0.17	0.79	0.05	281
	overall	0.03	0.42	0.57	0.04	1699

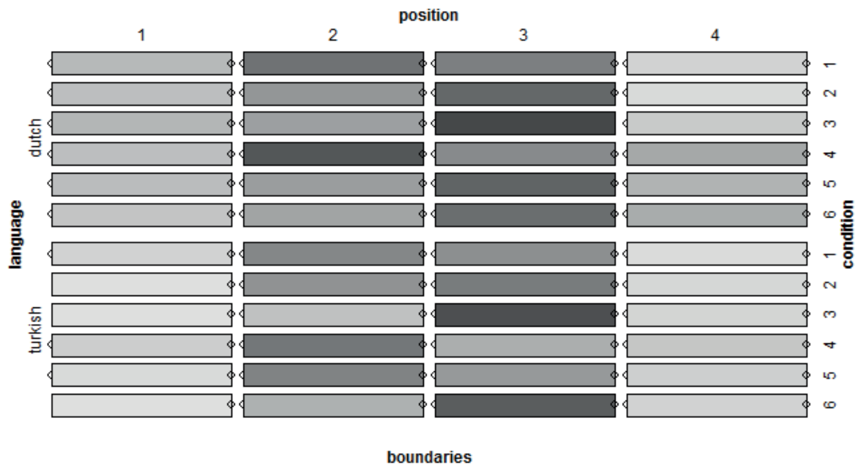
w = unstressed, S = stressed, σ = syllable, N. = number of items

The steepness of the distribution of segmentation frequencies does not clarify yet whether there are statistical differences between different stress conditions. This can be done by an analysis of the frequency of boundary placement in different positions in the phrase.

According to a Pearson's Chi-square test over the two languages, their distributions are different ($X^2(38) = 745$, p=.000). To see whether the distributions within languages, between conditions are different, a X-squared is performed for Dutch ($X^2(15) = 253.26$, p = .000) and Turkish ($X^2(15) = 326.09$, p = .000). If we normalize the frequencies over conditions, for example for Dutch by multiplying the probabilities with the average number of presentations per condition, which is 1628/6=271.33, the X-squared stays approximately the same ($X^2(15) = 250.52$, p = .000). Normalizing the data for frequency of presentation allows a more straightforward comparison of conditions, which is why the probabilities of boundary placement rather than the raw frequencies are compared. Table 5.5 gives these probabilities.

The table gives the probability within condition, meaning that the number of boundaries in that position in that condition is divided by the total frequency of presentation of that condition. For example: the condition SwwSw was presented 271 times and a boundary was placed after the second syllable 195 times. This means the probability that participants will place a boundary in that position is 195/271 = .72. The overall numbers are the same as those in

Figure 5.2: Probability of boundary placement by language, condition and position



grey scales: a darker grey means a higher probability of boundary placement

Table 5.2. Figure 5.2 visualizes what is given in Table 5.5 with darker grey areas indicating a higher probability.

It should be clear that in principle, that is, disregarding linguistic choices, the probabilities per position are independent: it is possible to have multiple boundaries within a condition and the placement of a boundary in for instance Position 1 will not prohibit the placement of one in Position 4.

The position with the highest probability of boundary placement is the clash position in both languages, with a probability of .94 for Dutch and .85 for Turkish.

In the Dutch results, a clear pattern emerges in the 4th position, where there is a divide of conditions between Sw-final conditions and wS-final conditions. For the former, the probability of splitting S and w is around between .03 and .05. For the latter, the probability of splitting w and S is between 0.17 and 0.29. Clearly, Dutch participants are more reluctant to divide Sw than wS. An alternative explanation would be that participants may be more inclined to overcome their aversion against monosyllables at the edge when these monosyllables are stressed, as in the wS-final conditions. This divide is not discernable in the Turkish results. Looking at the left edge, the Dutch probability of segmentation of monosyllables does not seem to be guided by metrical considerations, with as the only exception the probability of 0.08 in condition wwSwS, which for unknown reasons is different from the wwSSw condition (0.13). In the Turkish results, in contrast, there is a higher probability of segmentation of stressed monosyllables (0.05) than of unstressed monosyllables (0.01). A metrical difference at the left edge for Dutch is visible at the 2nd position, with higher probabilities after Sw (0.72 and 0.78) than after wS or ww (ranging from 0.37 to 0.50). Lastly, the high probability of boundary placement in Position 3 is least high in conditions starting with Sww, for both languages. In Turkish, there is a difference between a boundary in Position 3 after wSw (0.63 and 0.42) and wwS (0.85 and 0.79) which is not found for Dutch, besides the clash effect (wSw: 0.83 and 0.76, wwS: 0.94 and 0.76).

The slight differences between the languages in the overall lexicon as found in the previous section, then, seem to stem from different conditions, rather than being an effect of similar segmentation preferences. This can be concluded from the fact that the probabilities of boundary placement differ per condition per language. To better investigate the metrical preferences of Dutch and Turkish listeners it is therefore informative to look at the preferences ordered by O/E. First the Dutch results are discussed, then the Turkish, and then a conclusion will follow.

Dutch optimization

The probabilities of boundary placement result in preferences for segmentations, which clearly differ per condition. Tables 5.6, 5.7 and 5.9 show all segmentations. The order of frequency is given, and the O/E again shows how these frequencies relate to a random segmentation with the overall bias of segmentation per language.

The absolute frequencies again show that some conditions have a steeper distribution of preferences than others, as was seen in Figure 5.1. Furthermore, segmentations starting with a monosyllable are consistently rare. The highest absolute frequency for a segmentation starting with a monosyllable is 19. The pattern of probabilities and the underrepresentation of monosyllables already showed that this was the case, but including the context we can now see that in Dutch this underrepresentation is not caused by a general dispreference for monosyllables, but a dispreference for monosyllables at the edge of the phrase. This distinction is for example shown in Condition 1 (SwwSw), in which the preferred segmentation is Sw][w][Sw. The segmentations show that there is a bigger dispreference for monosyllabic unstressed syllable than for stressed syllables, and it is clearer at the left edge than at the right edge.

Despite some segmentations with monosyllables, there is a general preference for a division of the sequences into two or three parts. The dispreference for monosyllables is not condition-specific (but is sometimes overruled by other preferences), but it has an effect on O/E, because the random lexicon was segmented with the same overall bias against segmentations at different positions in the phrase. Segmentations with monosyllables at the edge therefore have a higher O/E than segmentations with only a boundary after the third

Table 5.6: Dutch preferences for segmentation by frequency, conditions 1,2

Cond.	**Segm.**	**Freq.**	**Random Freq.**	**Random O/E**
SwwSw	Sw][w][Sw	91	80.50	**1.13**
	Sw][wSw	84	27.72	**3.03**
	Sww][Sw	50	70.14	0.71
	S][ww][Sw	14	9.82	**1.43**
	S][w][w][Sw	10	11.28	0.89
	S][w][wSw	5	3.88	**1.29**
	Sw][wS][w	3	4.82	0.62
	S][ww][S][w	3	1.71	**1.75**
	S][w][w][S][w	3	1.96	**1.53**
	SwwSw	3	24.15	0.12
	Sww][S][w	2	12.18	0.16
	SwwS][w	1	4.20	0.24
	Sw][w][S][w	1	13.99	0.07
	S][wwSw	1	3.38	0.30
	S][w][wS][w	0	0.67	-
	S][wwS][w	0	0.59	-
wSwSw	wSw][Sw	92	63.41	**1.45**
	wS][w][Sw	76	72.78	1.04
	wS][wSw	32	25.06	**1.28**
	w][Sw][Sw	18	8.88	**2.03**
	w][S][w][Sw	11	10.19	**1.08**
	wSwSw	6	21.84	0.27
	wSw][S][w	4	11.02	0.36
	w][S][wSw	3	3.51	0.85
	w][S][w][S][w	1	1.77	0.56
	w][Sw][S][w	1	1.54	0.65
	wS][wS][w	1	4.35	0.23
	wS][w][S][w	0	12.65	-
	wSwS][w	0	3.79	-
	w][S][wS][w	0	0.61	-
	w][SwS][w	0	0.53	-
	w][SwSw	0	3.06	-

Table 5.7: Dutch preferences for segmentation by frequency, conditions 3,4

Cond.	Segm.	Freq.	Random Freq.	Random O/E
wwSSw	wwS][Sw	156	76.35	**2.04**
	ww][S][Sw	74	87.63	0.84
	w][wS][Sw	19	10.69	**1.78**
	w][w][S][Sw	16	12.27	**1.30**
	ww][SSw	9	30.18	0.30
	wwS][S][w	7	13.27	0.67
	wwSSw	4	26.29	0.15
	w][w][S][S][w	3	2.13	**1.41**
	ww][SS][w	3	5.24	0.57
	ww][S][S][w	2	15.22	0.13
	w][w][SSw	1	4.23	0.24
	wwSS][w	1	4.57	0.22
	w][w][SS][w	0	0.73	-
	w][wSS][w	0	0.64	-
	w][wSSw	0	3.68	-
	w][wS][S][w	0	1.86	-
SwwwS	Sw][w][wS	78	86.14	0.91
	Sw][wwS	75	29.66	**2.53**
	Sw][ww][S	41	5.15	**7.95**
	Sww][wS	26	75.06	0.35
	Sw][w][w][S	19	14.97	**1.27**
	S][ww][wS	12	10.51	**1.14**
	S][w][wwS	7	4.15	**1.68**
	S][w][w][w][S	6	2.10	**2.86**
	Sww][w][S	6	13.04	0.46
	SwwwS	6	25.85	0.23
	Swww][S	5	4.49	**1.11**
	S][ww][w][S	4	1.83	**2.19**
	S][www][S	2	0.63	**3.18**
	S][w][w][wS	1	12.07	0.08
	S][w][ww][S	1	0.72	**1.39**
	S][wwwS	1	3.61	0.28

Table 5.8: Dutch preferences for segmentation by frequency, conditions 5,6

Cond.	**Segm.**	**Freq.**	**Random Freq.**	**Random O/E**
wSwwS	wSw][wS	109	69.36	**1.57**
	wS][w][wS	46	79.61	0.58
	wS][wwS	34	27.41	**1.24**
	wSw][w][S	19	12.05	**1.58**
	w][Sw][wS	16	9.71	**1.65**
	wS][ww][S	13	4.76	**2.73**
	w][S][wwS	7	3.84	**1.82**
	w][S][w][wS	6	11.15	0.54
	wS][w][w][S	5	13.83	0.36
	w][S][w][w][S	4	1.94	**2.06**
	wSwwS	4	23.89	**0.17**
	wSww][S	3	4.15	0.72
	w][S][ww][S	1	0.67	**1.50**
	w][Sw][w][S	1	1.69	0.59
	w][Sww][S	0	0.58	-
	w][SwwS	0	3.35	-
wwSwS	wwS][wS	106	67.03	**1.58**
	ww][S][wS	46	76.94	0.60
	ww][Sw][S	23	4.60	**5.00**
	wwS][w][S	18	11.65	**1.55**
	wwSw][S	16	4.01	**3.99**
	ww][SwS	10	26.49	0.38
	ww][S][w][S	9	13.37	0.67
	w][wS][wS	9	9.39	0.96
	wwSwS	9	23.08	0.39
	w][wS][w][S	4	1.63	**2.45**
	w][w][S][wS	4	10.78	0.37
	w][w][S][w][S	3	1.87	**1.60**
	w][w][Sw][S	2	0.64	**3.10**
	w][w][SwS	0	3.71	-
	w][wSw][S	0	0.56	-
	w][wS][wS	0	3.23	-

syllable. This can for example be observed from the high O/E, despite low absolute frequency, for any segmentation containing initial [S] in the conditions SwwSw and SwwwS. In Table 5.9, the segmentations with a frequency higher than 15 (15 is more or less 5% of the phrases in each condition) are ordered by O/E. This ordering by O/E gives a slightly different view on preferences. When the overall bias against monosyllables at the edge is normalized by use of O/E, the picture emerges that choices are made by metrical optimization. The segmentations with the highest O/E are are likely driven by Sw or wSw. The overrepresentation of wwS seems to be contextual, since it only appears in Condition 3, to prevent segmenting a monosyllabic w. This can be concluded because in Condition 4, 5 and 6, where wwS can also be segmented, participants seem to prefer the segmentation of ww][S (in condition 4), Sw][wS (in condition 5) and ww][Sw (in condition 6).

In addition to comparing metrical optimization within condition, optimization can be judged by comparing syllabically similar segmentations between stress conditions. For example, $\sigma\sigma][\sigma][\sigma\sigma$ is preferred over $\sigma\sigma\sigma][\sigma\sigma$ in Condition SwwSw, but it is the other way around in condition wSwSw, while these conditions only differ in the stress pattern of the first two syllables. In Table 5.10, this analysis is done for the three most common segmentations which are most according to the segmentation probability per position, that is $\sigma\sigma\sigma][\sigma\sigma$ (33%), $\sigma\sigma][\sigma][\sigma\sigma$(25%), $\sigma\sigma][\sigma\sigma\sigma$ (15%).

Comparing conditions which differ minimally, Table 5.10, column 1 and 2, show that Dutch listeners prefer wSw and wwS over Sww. Conditions containing the latter unit are underrepresented in segmentation $\sigma\sigma\sigma][\sigma\sigma$, while conditions containing the former two units are overrepresented. In this segmentation there is also a slight preference of Sw over wS, but it is not very clear. Looking at segmentation $\sigma\sigma][\sigma\sigma\sigma$, this difference is somewhat clearer, since Sw][wSw and Sw][wwS are highly overrepresented, while the O/Es for wS][wSw and wS][wwS are somewhat lower, though still overrepresented. In this segmentation it is also clear that there is a clear preference against units without stress (ww) and units with two stresses (SSw and SwS). In the last segmentation condition it becomes clear that Dutch listeners have a slight preference against monosyllabic units, although it is still acceptable when a monosyllabic unit is flanked by two trochees or a trochee and iamb. As soon as there is a unit without stress, or two iambs, the segmentation is dispreferred.

Table 5.11 gives an overview of the segmented units per condition. The table is therefore like Table 5.4.5, which gave an overview of the overall lexicon, but this time the lexicon is kept within each condition in which it was segmented. Here, some units are overrepresented in every possible condition, where other units are overrepresented in some conditions and underrepresented in others. The latter may point towards a context-specific preference, which, rather than indicating a preference for that specific unit, may indicate a dispreference against another unit or segmentation.

Table 5.11 suggests that Dutch listeners have an overall preference for the segmentation of trochees; the frequency of trochees is higher in every condition

Table 5.9: Dutch preferences for segmentation by O/E

Cond.	**Segm.**	**Freq.**	**Random Freq.**	**Random O/E**
SwwSw	Sw][wSw	84	72.27	3.03
	Sw][w][Sw	91	80.50	1.13
	Sww][Sw	50	70.14	0.71
wSwSw	w][Sw][Sw	18	8.88	2.03
	wSw][Sw	92	63.41	1.45
	wS][wSw	32	25.06	1.28
	wS][w][Sw	76	72.78	1.04
wwSSw	wwS][Sw	156	76.35	2.04
	w][wS][Sw	19	10.69	1.78
	w][w][S][Sw	16	12.27	1.30
	ww][S][Sw	74	87.63	0.84
SwwwS	Sw][ww][S	41	5.15	7.95
	Sw][wwS	75	29.66	2.53
	Sw][w][w][S	19	14.97	1.27
	Sw][w][wS	78	86.14	0.91
	Sww][wS	26	75.06	0.35
wSwwS	w][Sw][wS	16	9.71	1.65
	wSw][w][S	19	12.05	1.58
	wSw][wS	109	69.36	1.57
	wS][wwS	34	27.41	1.24
	wS][w][wS	46	97.61	0.58
wwSwS	ww][Sw][S	23	4.60	5.00
	wwSw][S	16	4.01	3.99
	wwS][wS	106	67.03	1.58
	wwS][w][S	18	11.65	1.55
	ww][S][wS	46	76.94	0.60

w = unstressed, S = stressed

than would be expected on the basis of the general bias (calculated over the same data). The unit wSw, too, is overrepresented in every condition in which it is possible to segment it. A single unstressed syllable, on the other hand, is underrepresented in every condition. These three units give an insight into the segmentation preferences of Dutch listeners, since they are unnegotiable. All other possible units behave more context-dependent. The single stressed syllable, for example, only emerges as overrepresented when it is found at the right edge of the phrase. This indicates that the single stressed syllable is preferred over wS, but it does not indicate a preference for the

Table 5.10: Dutch preferences for segmentation by O/E

$\sigma\sigma\sigma$][$\sigma\sigma$	O/E	$\sigma\sigma$][$\sigma\sigma\sigma$	O/E	$\sigma\sigma$][σ][$\sigma\sigma$	O/E
Sww][Sw	0.71	Sw][wSw	3.03	Sw][w][Sw	1.13
wSw][Sw	1.45	wS][wSw	1.28	wS][w][Sw	1.04
wwS][Sw	2.04	ww][SSw	0.30	ww][S][Sw	0.84
Sww][wS	0.35	Sw][wwS	2.53	Sw][w][wS	0.91
wSw][wS	1.57	wS][wwS	1.24	wS][w][wS	0.58
wwS][wS	1.58	ww][SwS	0.00	ww][S][wS	0.60

w = unstressed, S = stressed

single stressed syllable overall, since it does not prevail in the other conditions. Another example is wS. Despite its being quite frequent in the output, the unit does not become as frequent as the overall bias would predict in random segmentation. Its O/E hovers underneath and around 1.00 in all conditions except for wwSSw, where wS is slightly overrepresented, possibly to prevent too many single syllables and stress clashes.

To be able to say whether these considerations are driven by the language-specific metrical system, however, the results have to be compared to the choices of participants with a different native language, in this case Turkish. In the next section, the Turkish results will be discussed and compared to the Dutch results.

Turkish optimization

The tables with all Turkish segmentations again show that the Turkish distribution of segmentation is steep, and as Figure 5.1 showed, it is steeper than the Dutch distribution.

Table 5.15 confirms the native preference of Turkish listeners to segment units with final stress. In Condition 5 and 6 the listeners prefer wS][wwS and wwS][wS, a preference which is obvious both in absolute frequency and O/E. To compare: in Dutch, for Condition 5 wSw][wS was the favorite segmentation in absolute frequency and w][Sw][wS in O/E and for Condition 6 wwS][wS was the favorite in absolute frequency and ww][Sw][S in O/E. However, the Turkish preference for final stress does have exceptions. For example, in Condition 1, the segmentation S][wwS][w is not made at all. Similarly, a potential preference for wS seemingly cannot overrule the dispreference of monosyllables: Sw][wS][w is rare (its frequency is 2). This dispreference against monosyllables can be seen overall; all segmentations with an O/E over 1 and a frequency over 15 consist of a combination of a trisyllable and a disyllable, both of which have one stressed syllable. There is an order of preference for final stress within these bounds, but the preference for final stress never overrules the need to have at least one

Table 5.11: Dutch output lexicon per condition

Unit	O/E (Freq.)					
	SwwSw	wSwSw	wwSSw	SwwwS	wSwwS	wwSwS
Sw	1.15 (344)	1.30 (216)	1.42 (265)	1.57 (213)	1.49 (17)	4.76 (25)
wS	0.55 (3)	0.92 (110)	1.51 (19)	0.64 (117)	0.93 (275)	1.02 (178)
w	0.82 (136)	0.79 (129)	0.76 (75)	0.93 (154)	0.74 (125)	0.84 (65)
S	0.71 (45)	0.49 (21)	0.71 (107)	1.50 (118)	1.12 (64)	0.97 (137)
wwS	0.00(0)	-	1.82 (163)	2.42(82)	1.31 (41)	1.58 (124)
wSw	2.82 (89)	1.27 (131)	-	-	1.57 (128)	0.00 (0)
ww	1.47 (17)	-	0.64 (88)	3.18 (58)	2.58 (14)	0.72 (88)
Sww	0.63 (52)	-	-	0.36 (32)	0.00 (0)	-
wwSw	0.30 (1)	-	-	-	-	3.99 (16)
SwS	-	0.00 (0)	-	-	-	0.33 (10)
SSw	-	-	0.29 (10)	-	-	-
wwSwS	-	-	-	-	-	0.39 (9)
wSwSw	-	0.26 (6)	-	-	-	-
SwwwS	-	-	-	0.23(6)	-	-
Swww	-	-	-	1.11 (5)	-	-
wwSSw	-	-	0.15 (4)	-	-	-
wSwwS	-	-	-	-	0.17 (4)	-
SS	-	-	0.50 (3)	-	-	-
SwwSw	0.12 (3)	-	-	-	-	-
wSww	-	-	-	-	0.72 (3)	-
www	-	-	-	3.18(2)	-	-
wwSS	-	-	0.22 (1)	-	-	-
SwwS	0.24 (1)	-	-	-	0.00 (0)	-
wwwS	-	-	-	0.28 (1)	-	-

w = unstressed, S = stressed

stressed syllable in each of the segmented units. This is very different from the Dutch behavior, where unstressed monosyllables are in the top ranks of each condition, in most cases to accommodate the segmentation of a trochee. In sum, it looks like different considerations guide the segmentation of units in Turkish and Dutch.

Again, a comparison of conditions within segmentations can be used to investigate the preference for metrical units. In Table 5.16, a comparison of conditions within segmentations, again for the three segmentations which are most common according to probability per position is done. These conditions are $\sigma\sigma\sigma][\sigma\sigma$ (48%), $\sigma\sigma][\sigma][\sigma\sigma$(6%), $\sigma\sigma][\sigma\sigma\sigma$ (33%). Actually, the third most

Table 5.12: Turkish preferences for segmentation by frequency, conditions 1,2

Cond.	**Segm.**	**Freq.**	**R-Freq.**	**R-O/E**
SwwSw	Sw][wSw	121	48.92	**2.47**
	Sww][Sw	98	89.59	**1.09**
	Sw][w][Sw	28	64.31	0.44
	SwwSw	27	68.14	0.40
	S][ww][Sw	4	2.55	**1.57**
	S][w][wSw	3	1.39	**2.16**
	S][wwSw	3	1.94	**1.55**
	Sww][S][w	2	3.96	0.50
	Sw][wS][w	2	2.16	0.92
	S][w][w][Sw	2	1.83	**1.09**
	S][ww][S][w	1	0.11	**8.87**
	S][w][wS][w	0	0.06	-
	SwwS][w	0	3.02	-
	S][wwS][w	0	0.09	-
	S][w][w][S][w	0	0.08	-
	Sw][w][S][w	0	2.85	-
wSwSw	wSw][Sw	146	88.35	**1.65**
	wS][wSw	92	48.25	**1.91**
	wS][w][Sw	26	63.43	0.41
	wSwSw	12	67.20	0.18
	wS][w][S][w	3	2.81	**1.07**
	w][Sw][Sw	3	2.51	**1.19**
	wS][wS][w	2	2.14	0.94
	wSw][S][w	2	3.91	0.51
	w][S][wS][w	1	0.06	**16.46**
	w][SwSw	0	1.91	-
	w][S][wSw	0	1.37	-
	w][S][w][Sw	0	1.80	-
	w][S][w][S][w	0	0.08	-
	w][SwS][w	0	0.08	-
	w][Sw][S][w	0	0.11	-
	wSwS][w	0	2.97	-

Table 5.13: Turkish preferences for segmentation by frequency, conditions 3,4

Cond.	**Segm.**	**Freq.**	**R-Freq.**	**R-O/E**
wwSSw	wwS][Sw	212	85.58	**2.48**
	ww][SSw	22	46.73	0.47
	wwSSw	19	65.10	0.29
	ww][S][Sw	10	61.44	0.16
	wwS][S][w	9	3.79	**2.38**
	w][wS][Sw	3	2.43	**1.23**
	wwSS][w	2	2.88	0.69
	w][w][S][S][w	1	0.08	**12.93**
	w][wS][S][w	0	0.11	-
	w][w][SSw	0	1.33	-
	w][wSS][w	0	0.08	-
	ww][S][S][w	0	2.72	-
	w][w][SS][w	0	0.06	-
	w][w][S][Sw	0	1.75	-
	w][wSSw	0	1.85	-
	ww][SS][w	0	2.07	-
SwwwS	Sw][wwS	156	46.73	**3.34**
	Sww][wS	52	85.58	0.61
	SwwwS	24	65.10	0.37
	Sw][w][wS	15	61.44	0.24
	Sw][ww][S	12	2.07	**5.80**
	S][ww][wS	5	2.43	**2.05**
	S][w][wwS	4	1.33	**3.01**
	Swww][S	3	2.88	**1.04**
	S][www][S	2	0.08	**24.40**
	S][wwwS	2	1.85	**1.08**
	S][w][ww][S	1	0.06	**17.00**
	Sww][w][S	1	3.79	0.26
	S][ww][w][S	1	0.11	**9.28**
	S][w][w][wS	0	1.75	-
	Sw][w][w][S	0	2.72	-
	S][w][w][w][S	0	0.08	-

Table 5.14: Turkish preferences for segmentation by frequency, conditions 5,6

Cond.	Segm.	Freq.	R-Freq.	R-O/E
wSwwS	wS][wwS	139	47.74	**2.91**
	wSw][wS	95	87.43	**1.09**
	wS][w][wS	16	62.77	0.25
	wSwwS	15	66.50	0.23
	w][Sw][wS	4	2.49	**1.61**
	wS][ww][S	4	2.11	**1.89**
	wSww][S	4	2.94	**1.36**
	wSw][w][S	4	3.87	**1.03**
	w][S][ww][S	1	0.06	**16.64**
	w][S][w][w][S	1	0.08	**12.65**
	w][S][wwS	1	1.36	0.74
	wS][w][w][S	0	2.78	-
	w][Sw][w][S	0	0.11	-
	w][SwwS	0	1.89	-
	w][S][w][wS	0	1.79	-
	w][Sww][S	0	0.08	-
wwSwS	wwS][wS	206	86.51	**2.38**
	ww][SwS	36	47.24	0.76
	wwSwS	14	65.80	0.21
	ww][S][wS	8	62.10	0.13
	wwSw][S	7	2.91	**2.40**
	w][wS][wS	4	2.46	**1.63**
	ww][Sw][S	3	2.09	**1.44**
	wwS][w][S	3	3.83	0.78
	w][wSwS	0	1.87	-
	w][w][SwS	0	1.34	-
	ww][S][w][S	0	2.75	-
	w][wS][w][S	0	0.11	-
	w][wSw][S	0	0.08	-
	w][w][Sw][S	0	0.06	-
	w][w][S][wS	0	1.77	-
	w][w][S][w][S	0	0.08	-

Table 5.15: Turkish preferences for segmentation by O/E

Cond.	**Segm.**	**Freq.**	**Random Freq.**	**Random O/E**
SwwSw	Sw][wSw	121	48.92	**2.47**
	Sww][Sw	98	89.59	**1.09**
	Sw][w][Sw	28	64.31	0.44
	SwwSw	27	68.14	0.40
wSwSw	wS][wSw	92	48.25	**1.91**
	wSw][Sw	146	88.35	**1.65**
	wS][w][Sw	26	63.43	0.41
wwSSw	wwS][Sw	212	85.58	**2.48**
	ww][SSw	22	46.73	0.47
	wwSSw	19	65.10	0.29
SwwwS	Sw][wwS	156	46.73	**3.34**
	Sww][wS	52	85.58	0.61
	SwwwS	24	65.10	0.37
wSwwS	wS][wwS	139	47.74	**2.91**
	wSw][wS	95	87.43	**1.09**
	wS][w][wS	16	62.77	0.25
wwSwS	wwS][wS	206	86.51	**2.38**
	ww][SwS	36	47.24	0.76

w = unstressed, S = stressed

frequent segmentation in Turkish is $\sigma\sigma\sigma\sigma\sigma$, i.e. the unsegmented sequence (7%). However, the unsegmented sequence is uninformative about segmentation decisions, more so because their frequencies do not differ much across conditions, ranging from 12 to 27. It is indicative of the difference between Turkish and Dutch, however.

Column 1 and 2 in Table 5.16 show that Turkish listeners, like Dutch listeners, prefer wSw and wwS over Sww. This difference is most clear in the last three conditions. This dispreference is unclear, since it is not as strong in condition SwwSw. A dispreference against wS can also not be found. Unlike in Dutch, wwS is clearly preferred over wSw. In Dutch, the two units were comparable in their O/E. A preference of wS over Sw or the other way around cannot be observed in Columns 1 and 2, and in column 3 and 4 this difference is very modest, preferring wS over Sw. Like in Dutch, units without stress or with two stresses are underrepresented, but Turkish listeners seem less strictly opposed to them. The last two columns show a very clear picture of a preference against monosyllabic units in Turkish. In Dutch, monosyllabic units were more acceptable in certain conditions than in others, but in Turkish these

Table 5.16: Comparing conditions by segmentations in Turkish

$\sigma\sigma\sigma][\sigma\sigma$	O/E	$\sigma\sigma][\sigma\sigma\sigma$	O/E	$\sigma\sigma][\sigma][\sigma\sigma$	O/E
Sww][Sw	1.09	Sw][wSw	2.47	Sw][w][Sw	0.44
wSw][Sw	1.65	wS][wSw	1.91	wS][w][Sw	0.41
wwS][Sw	2.48	ww][SSw	0.47	ww][S][Sw	0.16
Sww][wS	0.61	Sw][wwS	3.34	Sw][w][wS	0.24
wSw][wS	1.09	wS][wwS	2.91	wS][w][wS	0.25
wwS][wS	2.38	ww][SwS	0.76	ww][S][wS	0.13

w = unstressed, S = stressed

segmentations are strictly dispreferred.

The overrepresented units in Table 5.17, showing the separate units potentially driving segmentation, are the same as those found in Table 5.4. What is most clear from the condition-specific lexicon is the overall dispreference of unstressed monosyllables, a constraint that seems to be leading in the Turkish segmentation of the current phrases. Furthermore, there is an overrepresentation of trochees across the board, although it should be noted that in condition 5 and 6 this O/E is based on very low frequencies. The expected frequency of trochees in this condition is very low, due to the low probability of boundary placement at the edge. This normalization by bias is slightly misguiding because the design is asymmetrical. As shown in Table 5.1, a symmetrical design would have 10 levels, while the current design is lacking 4, of which 3 stress clashes. A symmetrical design would have made the bias more balanced, which may have prevented an exaggeration of O/E for the segmentation of monosyllables at the edge, more so for Dutch than for Turkish.

In the next section, an overall summary of the segmentation considerations of Dutch and Turkish listeners is given, after which a comparison is made between the current experiment and the experiment of Chapter 4.

5.4.7 Conclusion

In the range of comparisons of metrical optimization considerations by Dutch and Turkish listeners, an overall picture of preferences emerges. First of all, Turkish listeners seem to have a stronger agreement of which segmentation is optimal than Dutch listeners. Furthermore, Turkish listeners prefer longer words than Dutch listeners and are even willing to segment words with two stressed syllables to arrive at longer words. This matches the native languages in the sense that Turkish has longer words than Dutch, but it is interesting to see that this preference for long words overrules the maximum of one stressed syllable per word. This preference can also be seen in the number of unsegmented

Table 5.17: Turkish output lexicon per condition

Unit	O/E (Freq.)					
	SwwSw	wSwSw	wwSSw	SwwwS	wSwwS	wwSwS
Sw	1.02 (283)	1.12 (178)	1.49 (225)	1.62 (183)	1.54 (4)	1.40 (3)
wS	0.90 (2)	1.06 (126)	1.18 (3)	0.48 (72)	1.02 (274)	1.43 (222)
w	0.47 (40)	0.46 (41)	0.75 (17)	0.29 (22)	0.35 (29)	0.39 (7)
S	1.06 (16)	0.59 (6)	0.29 (21)	1.80 (35)	1.11 (17)	0.27 (21)
wwS	0.00 (0)	-	2.47 (221)	3.33 (160)	2.85 (140)	2.31 (209)
wSw	2.46 (124)	1.69 (240)	-	-	1.08 (99)	0.00 (0)
ww	1.88 (5)	-	0.28 (32)	4.07 (19)	2.30 (5)	0.41 (47)
Sww	1.07 (100)	-	-	0.59 (53)	0.00 (0)	-
wwSw	1.55 (3)	-	-	-	-	2.40 (7)
SwS	-	0.00 (0)	-	-	-	0.74 (36)
SSw	-	-	0.46 (22)	-	-	-
wwSwS	-	-	-	-	-	0.21 (14)
wSwSw	-	0.18 (12)	-	-	-	-
SwwwS	-	-	-	0.37 (24)	-	-
Swww	-	-	-	1.04 (3)	-	-
wwSSw	-	-	0.29 (19)	-	-	-
wSwwS	-	-	-	-	0.23 (15)	-
SS	-	-	0.00 (0)	-	-	-
SwwSw	0.40 (27)	-	-	-	-	-
wSww	-	-	-	-	1.36 (4)	-
www	-	-	-	24.40 (2)	-	-
wwSS	-	-	0.69 (2)	-	-	-
SwwS	0.00 (0)	-	-	-	0.00 (0)	-
wwwS	-	-	-	1.08 (2)	-	-

w = unstressed, S = stressed

sequences in Turkish (111) as compared to Dutch (32). From the overall lexicon analysis, we see a preference of disyllables (2050) and monosyllables (1176) over trisyllables (864) in Dutch, whereas we see a preference of disyllables (1683) and trisyllables (1406) over monosyllables (272) in Turkish.

Comparisons within contexts show that in disyllables, participants of both languages seem to segment Sw more often than wS. In Dutch, this preference would be expected on the basis of the metrical system, but it is unclear why it also occurs in Turkish. The comparison to the input lexicon similarly shows that the current segmentations contain wS more often than expected for Dutch, and Sw more often than expected for Turkish. The Turkish dispreference against monosyllables may be a reason for the higher occurrence of Sw, since many conditions either start with S or end with w. Disallowing monosyllables makes it necessary to segment suboptimal units, which seems to be a reason for the high occurrence of Sw and maybe also wSw in Turkish. For Dutch participants, monosyllables are more acceptable, and there is a preference of segmenting w][S over S][w, which indicates a preference for Sw and S over wS.

In trisyllables, both Turkish and Dutch listeners disprefer Sww. In Turkish, the preference for wwS is very clear, while in Dutch both wSw and wwS are overrepresented in the segmentations, while Sww is strongly underrepresented, even more so than in Turkish. Judging from the metrical system and the lexicon this is again a surprising result, since the Dutch metrical system allows trisyllabic words with antepenultimate stress. Although this pattern occurs mostly in words with a closed final syllable, such as ‘marathon’ and ‘almanak’, or words with a medial schwa or an open syllable preceding final schwa, such as ‘weduwe’ and ‘België’, the pattern also occurs in words with presumably lexically prespecified stress such as ‘Canada’. Whenever the final syllable does not contain schwa, however, it usually receives secondary stress (Gussenhoven, 1993). The syllabic structure of the experiment therefore may be a reason for the underrepresentation of Sww, but the strength of the dispreference is unexpected. Judging from the Metronome experiment, however, this result is less surprising. The conditions having Sww as the context were inhibiting for target segmentation both in Dutch and in Turkish. The comparison of the Optimization experiment to the Metronome experiment will be resumed in Section 5.5.

Due to the relatively short length of the stimuli, listeners were constrained to making specific segmentations. This has as its advantage that listeners had to make choices between several suboptimal choices, but on the downside it may have also exaggerated listeners’ preferences for units which they may not have preferred in more natural segmentation. The asymmetry of the design is a limitation which should be addressed in future studies.

5.5 Comparing Optimization to the Metronome

The Optimization experiment has both similarities and differences with the Metronome experiment. Figures 5.3 and 5.4 show the probability of boundary placement in the Optimization experiment, ordered by the ‘contexts’ of the Metronome experiment. The third bar is always position 3, which is the position at which the target was placed in the Metronome experiment.

Figures 5.3 and 5.4 show that both Dutch and Turkish participants are increasingly likely to segment at position 3, the closer the stressed syllable is to that position. That is, there is a dispreference against segmenting after Sww and a preference for segmenting after wSw and even more so after wwS. This behavior is comparable to the behavior in the Metronome experiment, except that for Dutch listeners, the reaction times were faster after wSw than after

Figure 5.3: Optimization of 'contexts' for Dutch

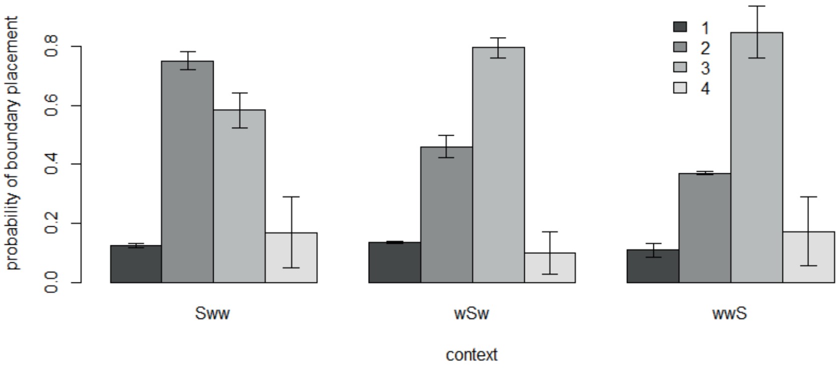


Figure 5.4: Optimization of 'contexts' for Turkish

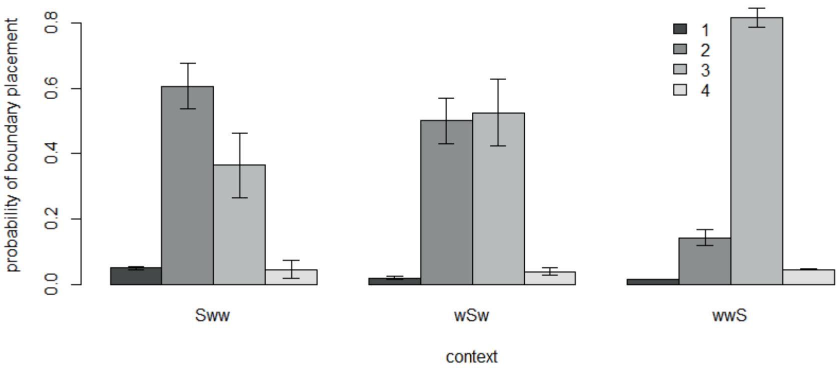
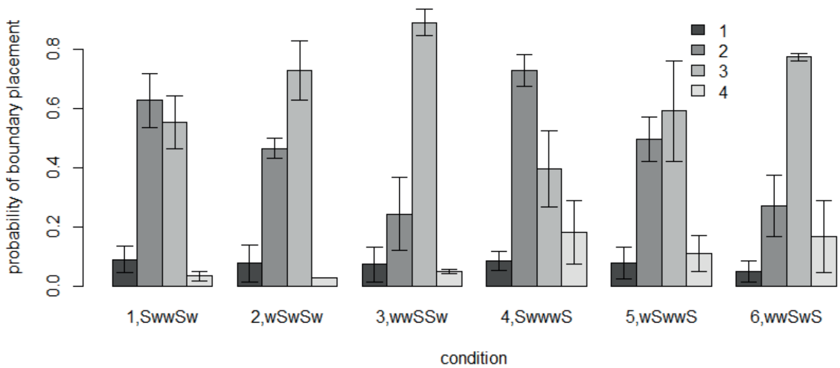


Figure 5.5: Probability of boundary placement per condition per position in Dutch and Turkish



wwS. Figure 5.5 shows that this increased probability of boundary placement at position 3 after wwS originates in the clash condition, as was already known from the probability analysis. As expected, then, the Optimization experiment showed an effect of clash which was not observed in the Metronome experiment and this difference concurs with the hypothesis that expectations are one-directional and an expectation of Sw] upon hearing S precludes a facilitative effect on segmentation of SS. The second stressed syllable is simply unexpected in online segmentation, while it is still dispreferred in offline segmentation.

Figures 5.6 and 5.7 show that in optimization, a target effect surfaces for Dutch which was not found in the metronome experiment (although there was a trend). Dutch participants are more likely to segment w][S than they are to segment S][w. For Turkish, there is no such effect, due to an overall dispreference of monosyllables. This is interesting, because in the Metronome experiment, a target effect was found for Turkish.

5.5.1 Conclusion

In Chapter 4, the Metronome experiment showed that listeners have language-specific expectations of upcoming word boundaries, triggered by stressed syllables. Because this strategy was also found for languages with positionally variable stress, a secondary strategy was hypothesized, responsible for alternative segmentations based on stress. The Optimization experiment confirmed

Figure 5.6: Optimization of 'targets' for Dutch

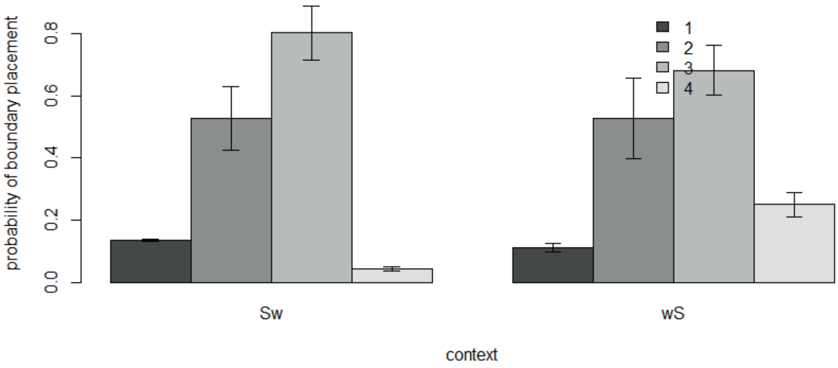
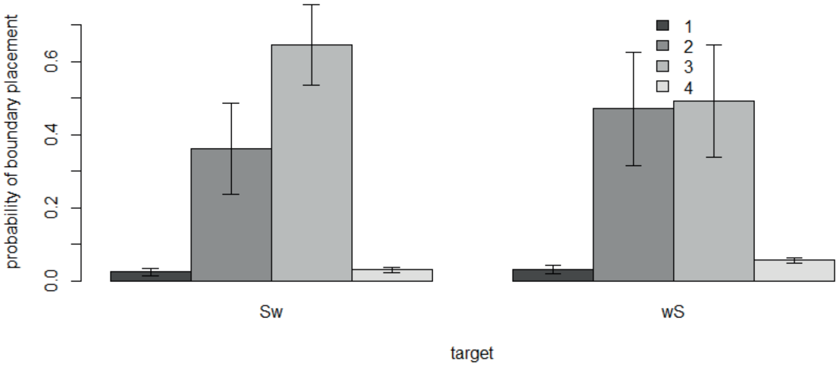


Figure 5.7: Optimization of 'targets' for Turkish



this hypothesis. The bidirectional grouping strategy led to different results from the Metronome and these differences are discussed in the current section.

The Optimization experiment gives an insight into metrical optimization in Turkish and Dutch. Because listeners have to make decisions on optimal and suboptimal units, a gradual preference of segmentation surfaces which was not observed in the Metronome experiment. Where the Metronome experiment measured online expectations triggered by stressed syllables, the Optimization experiment showed the bidirectional grouping of unstressed syllables around stressed syllables, again according to language-specific preferences. The differences between the experiments suggests that listeners use a bidirectional grouping cue in addition to a unidirectional expectation. Listeners may use this bidirectional grouping to revise their initial expectations to arrive at an optimal metrical grouping of stress patterns, although an online experiment would have to investigate what the nature and timing of such a revision strategy would be. The difference between experiments is most obvious in the clash-condition, where Metronome experiment shows no facilitation in segmentation of the target, while the Optimization experiment showed that this position had the highest probability of boundary placement. Whether these revisions can be measured online is a question which remains for future investigations.

5.6 Conclusion of Part II

This first part of the thesis set out to address behavioral differences across languages regarding the use of stress in segmentation. It was hypothesized that the typological commonality of word stress to align to word edges has a cognitive advantage in speech processing, or more specifically in word segmentation. Many languages have in common that stress aligns to the word edge, but the edge and the position of stress, as well as the positional variability differ between them. The first experiment was therefore aimed at investigating whether these typological differences would influence the use of stress in segmentation. Do listeners have language-specific expectations of word boundaries when listening to speech? The answer to this question is not straightforward. Yes, language-specific stress patterns are a good predictor of the differences in reaction times between conditions. In Turkish and Dutch, an anticipatory effect of canonical word stress is found, showing that listeners have expectations of the position of word boundaries based on canonical delimitative word stress. However, the fact that there is no anticipatory effect of stress in Polish, a right-edge-aligned fixed stress language, is puzzling enough to warrant follow up research into the potential inhibitory effect of the non-canonical target in fixed stress languages. Furthermore, the addition of Spanish to the data further complicated matters, since the results showed an interaction of context and target which was not found in any of the other languages. The interesting, potentially rhythmic, effect invites follow-up research, to investigate whether what was found in the current experiment is evidence of an interaction between a general cognitive rhythmic effect and the Language-specific Metronome, or whether there is something special about Spanish which with the current data cannot be explained.

The offline experiment that followed the online experiment gave evidence for the hypothesis that listeners make bidirectional groupings according to language-specific metrical considerations, when they are given more time. First of all, the segmentations were stress-driven, attested by the difference between the participants' segmentations and the random lexicon. Second of all, the segmentations were informed by metrical rather than lexical considerations, attested by the difference between the lexicon and the segmentations. Third of all, the segmentations were different between the languages, both in preferences for word length and stress patterns. Segmentations were influenced by metrical preferences, such as Sw and wSw for Dutch and wS and wwS for Turkish, and dispreferences, such as monosyllables for Turkish and units with two stressed syllables for Dutch.

The metrically-informed bidirectional segmentations in the offline experiment furthermore showed that the effect of canonical stress on the target in the online experiment is not necessarily a lexical effect. Participants made bidirectional groupings based on stress, despite the fact that the pentasyllabic stimuli were segmentally heterogeneous and unfamiliar. It is therefore still possible that the ease of segmentation of the canonical target in the online experiment is either a prelexical or lexical effect. A specifically designed online study is the way to disentangle the two explanations. Either way, the online and offline experiment showed that language-specific suprasegmental delimitative stress is used in segmentation; unidirectionally for anticipation, and bidirectionally for perceptual grouping.

Part III

Distribution

CHAPTER 6

Contingency

6.1 Introduction

In the experiments described in the previous chapters, it was found that listeners use stress in segmentation according to the knowledge of their native language's phonology. The Metronome experiment showed that during processing, listeners develop expectations of upcoming word boundaries based on the relation of stress to the word boundary, while the Optimization experiment showed that segmentations can also be made according to bidirectional stress-based considerations, by optimization of patterning of the whole sequence. The current chapter approaches this issue from a different perspective: how is the word stress pattern of one's native language in principle learnable. Specifically: does the distribution of stress in unsegmented, continuous speech contain enough information for the relation of stress to the word boundary to be inferable? This question is worth developing since in infant word segmentation, word stress is considered to be a potentially useful cue for the *discovery* of word boundaries, while at the same time it is unclear how infants acquire what the relation of word stress to the word boundary is. Whether the metrical system of a language is learnable for infants is typically approached from the word-domain, but whether these patterns can also be acquired before word-boundaries are known is not clear. Knowledge about information in the input can shed light on this matter, since infant language acquisition is an interplay of what information is available in the input and what the cognitive abilities of infants are at different ages. Both sides of the research are crucial, and regardless of how statistical relations of linguistic units in the corpus are acquired with the infants' cognitive abilities, research shows that statistical relations correlate with their learning behavior.

In the history of the field, the issue of stress acquisition has been approached both experimentally and computationally and it was initially seen as a chicken-and-egg problem: to segment continuous and unfamiliar input into words, word stress would be an excellent and salient measure, but to know the relation of stress to the word boundary, it is likely necessary to know words, or minimally word shapes. A third possibility, and the approach of most recent literature, is that the truth lies in the middle, and that learners simultaneously discover structure in continuous speech as well as in the developing lexicon. The current introduction will give a brief overview of the literature, after which the research question and the design of the analyses is explained. To refer to stress patterns in continuous speech, I will use the term 'distribution' and to refer to stress patterns on the word-level, I will use the term 'word stress'.

6.1.1 Infant speech segmentation

Infant speech segmentation is different from adult speech segmentation because it pertains to the *discovery* of words in speech, rather than their *recognition*. This discovery of words, or, at this stage, word forms, can be seen as the grouping of frequently co-occurring elements in a continuous stream of speech and this task is easier the more information a learner obtains about relevant linguistic patterns. The issue of how this segmentation emerges in infants is therefore a direct interplay of the information available in the input, which is what *in principle* could be grouped together or inferred from the input, and which infant cognitive abilities contribute to the discovery of this information. This is a reason why it is an interesting field for both computational and experimental work, as well as of course for theoretical work. The information available in the input is an objective maximum of what infants would be able to learn, but if a study were to approach ecological validity, this information should be obtainable without the necessity to make complicated statistical computations over large corpora. Infants, after all, acquire language incrementally and with limitations on attention and memory. Furthermore, the evaluated information units should be perceptible and discriminable by infants. Under the hypothesis that infants acquire language from a coarse-to-fine perspective, infants may be learning simple statistics over coarse distributions, as for example stress distributions, before complex abstractions over fine-grained distributions, as for example segmental phonotactics. The idea that infants 'zoom in' is supported by evidence such as the study by Bijeljac-Babic et al. (1993), who show that 4-day-old infants discriminate two- and three-syllabic utterances regardless of phonetic duration, but not disyllabic utterances that have either four or six phonemes. Bertoncini and Mehler (1981) found that newborn infants younger than 2 months old were able to discriminate a consonantal contrast within a syllable, but not within a comparable non-syllabic sequence. Stress distributions are discriminable at the syllabic level and the difference between stressed and unstressed syllables is phonetically salient, even prenatally since prosodic patterns are perceptible in utero. From the beginning of the third trimester in utero the auditory system is already functional. The womb functions as a low-pass filter which conducts the prosodic characteristics of the mother's speech (Giovanelli et al., 1999, for a review). At 35 weeks of gestational age infants are even able to discriminate vowels (Shahidullah and Hepper, 1994; Cheour et al., 1998) and French and German newborns were found to cry with native pitch contours (Mampe et al., 2009). The stress distribution is therefore one of the first types of input available to infants acquiring language.

In fact, it has been found that English infants are able to perceive the difference between trochaic and iambic units already in the first two months of life (a.o. Jusczyk and Thompson, 1978; Weber et al., 2004) and even newborns between 48 and 72 hours old were found to be able to discriminate stress patterns on disyllabic and trisyllabic words (Sansavini et al., 1997). Despite this early sensitivity, which shows that phonetic stress is acoustically perceptible and potentially linguistically interesting for infants, evidence for a language-specific knowledge of the metrical system seems to be relatively delayed as compared to their early attention to syllables and prominence differences. Preferential listening studies have shown that English infants prefer to listen to their native metrical units at 9 months of age, but not at 6 months (Jusczyk et al., 1993), that English 7-month-old infants have a preference for listening to trochaic units as opposed to the same units with post-stress pauses (Echols et al., 1997), that German, but not French learning infants have a preference for native metrical units at 6 months (Höhle et al., 2009) and that Dutch 4-8-month-olds prefer trochaic units, but Turkish 4-8-month-olds do, too (Keij and Kager, submitted). Recent work studying ERP responses to rhythmic sequences found that 5-month-old but not 4-month-old German infants could discriminate trochaic oddball items from a sequence of iambic items, while not being able to do the opposite (Weber et al., 2004). Friederici et al. (2007), however, did find an ERP response in 4-month-old French and German infants (Friederici et al., 2007) to the non-native oddball in a sequence of native disyllables, meaning that they found a language-specific effect of stress sequence processing. In fact, language-specific effects in studies on infant stress discrimination are sparse and in preference studies trochaic preferences are commonly found, even at very young ages, while preferences for the 'iambic' or word-final pattern on disyllables over the 'trochaic' pattern is sparse. It has been suggested that an infant preference for the native stress pattern is strongly linked to the native phonetic realization of stimuli in the experiment (Segal and Kishon-Rabin, 2012), although more evidence is needed (Keij and Kager, submitted).

Possible explanations for the sparsity of evidence for an iambic preference may be that the appropriate languages to provide evidence for an iambic preference have not yet been studied; most work has been done on Germanic languages and of the Romance languages the focus has been on French, which is described to be a language that does not have stress at the word level (Jun and

Fougeron, 2000). Very few other languages have been tested[1]. Furthermore, it is possible that the use of non-native speech synthesis, in particular MBROLA (Dutoit et al., 1996) voices does not invoke native listening in infants, as was suggested by infants' varying preferences according to the nativity of speech in Segal and Kishon-Rabin (2012). Another option is that an early preference for trochees is not in fact the proof of early knowledge of the metrical system, but merely evidence that infants prefer a salient word onset. An example of a salience-based approach to acquisition is that stressed syllables are stored as initial representations of content words and that unstressed syllables are recognized as necessary for natural speech, but because it is still unclear where word boundaries are, these unstressed syllables are initially not assigned meaning (Gleitman et al., 1988). Segmentation studies try to find whether these unstressed syllables are in fact assigned to words[2] and whether this is done according to the native metrical system.

A study focusing on the early use of stress in segmentation is Morgan and Saffran (1995). In this study, 6-month-old and 9-month-old English infants were familiarized with a language in which the to be learned words were either metrically consistent throughout familiarization, or segmentally, or both. In the familiarization phase, the consistent rhythm was either trochaic or iambic; the experiment was not designed to test whether infants had knowledge of the native metrical system, but whether these patterns were learned during the experiment. In the test phase, infants were expected to detect buzzes, which were word-internal or word-external, the former leading to longer latencies in detection than the latter. It was found that 6-month-old infants segmented metrically consistent words irrespective of whether these were segmentally consistent, while 9-month-old infants integrated both cues. In this study, apparently, infants relied on suprasegmental cues for segmentation, which are arguably more salient and need less fine-grained discrimination of categories, before they payed attention to the reliability of segmental cues. This finding suggests that the salience of cues may be a factor that influences the early ability of infants to discover units available in the input. Jusczyk et al. (1999) studied segmentation according to native metrical patterns and they found a use of native English stress in segmentation that precedes segmentation of units with exceptional stress. English infants of 7.5 months old, namely, missegmented the trochee 'taris' in 'the guitar is', before segmenting the correct iamb 'guitar' from the context at 10.5 months. Houston et al. (2004) continued this line of investigation with trisyllabic words and they found that infants segment words with initial strong syllables, but only when these syllables receive

[1] A hypothesis which has been put forward on the basis of children's early productions is the Trochaic Bias (Allen and Hawkins, 1979). However, this hypothesis is largely based on evidence from languages of which the rhythmic pattern is trochaic, which means that in these cases the observed bias can alternatively be led back to linguistic experience. The supposed universality of the bias has therefore met with much opposition, of which one example is the Neutral Start hypothesis (Hochberg, 1988).

[2] Unless these unstressed syllables are already recognized as function words, as was found to be possible for German 8-month-old infants in Höhle and Weissenborn (2003).

primary stress. In Polka and Sundara (2003) English and French 7.5-month-old infants were reported to segment words from native stimuli (even across dialects) but not from non-native stimuli, which was interpreted as evidence for the use of stress in segmentation, although the materials are bound to contain many other language-specific segmentation cues, which makes the evidence unclear. As said above, Morgan and Saffran (1995) investigated segmentation according to statistical cues and stress cues, and found that English 6-month-old infants only pay attention to stress, while English 9-month-old infants only segment when stress and statistics congruently group syllables together. When Thiessen and Saffran (2003) did a similar study, only this time pitting statistical cues against stress cues, it was however found that English 7-month-old infants do not pay attention to stress cues but to statistical cues, while 9-month-olds prefer paying attention to stress cues, thereby ignoring statistical cues.

The above evidence does seem to suggest a discrepancy between the initial attention to stress and the actual preference for and use in segmentation of native stress in infants. Even though stress patterns are salient and coarse, supposedly simple distributions when defined as syllables, and even though infants have the syllable as their perceptual unit and are aware of prominence patterns already as newborns, stress may not necessarily be the first cue that is used for word learning and segmentation. Why would this be the case?

It is possible that the unsegmented stress distribution is not informative enough on the relation of stress to the word boundary and is therefore not useful as a first cue on segmentation. However, the availability of a handful of word forms may be a catalyst to the hypothetical early use of stress in segmentation. It has been shown that already at 6 months infants are able to use known units, such as their name or the word 'mommy', to segment new units Bortfeld et al. (2005) and Bergelson and Swingley (2012) show that infants at 6 months have a small lexicon of nouns including their meaning. These first words can be segmented by absolute frequency or statistical clustering, or by use of isolated words, that is, single word utterances. It has been found that 9% of child directed speech (CDS) consists of isolated words (Brent and Siskind, 2001) and for example that English 8-10-month-olds are better able to learn Italian target words when the input contains isolated words (Lew-Williams et al., 2011). A first small lexicon, consisting of words or even only proxies of words, as Swingley (2005) puts it, may be necessary for infants to make an inference about the native stress system. Evidence for the ability to segment from a sentence words from such a first small lexicon has been found for English infants of 12 months old (Depaolis et al., 2014). These first words are presumably segmented by the statistical co-occurrence of their syllables, in Swingley (2005) implemented as Mutual Information[3], apparently despite the fact that the segmental distribution is more complex and fine-grained than the stress distribution, although recall the contradictory findings discussed above.

[3]Mutual information is a symmetric measure of the mutual dependence of two random variables. It expresses the reduction of uncertainty about a variable X when a variable Y is known and it is a tool of Information Theory (Shannon and Weaver, 1949).

Irrespective of the use of stress, it has been found that infants are able to use the statistical distribution of linguistic input to store their first words. At 8 months, infants have been found to segment words of which the syllables were associated by a statistical relation known as Transitional Probability (TP) (Saffran et al., 1996; Pelucchi et al., 2009). In Saffran et al. (1996) this was the probability that for example the syllable 'bu' was followed by the syllable 'sa'. The stimuli were designed for trisyllabic words to be segmented from a continuous stream, because the TP inside words was 1.0, which means 'bi' was always followed by 'da', and the TP between words was 0.33, which means that 'ku' was sometimes followed by 'pa', but more often it was not. This can be seen in Example 6.1.

(6.1) ... bidakupalotigodabubidakugodabupalotigodabubidaku...

Infants listened to a a continuous stream of syllables without any acoustic cues that would influence segmentation. After this phase, the infants listened to isolated 'words' (e.g. bidaku) and part-words, which are combinations of syllables that occur together in the input, but straddle a word boundary (e.g. dakupa). In the Headturn Preference Procedure[4] the infants showed their preference for listening to part-words, which indicates that they distinguished the two types of stimuli. In this study, frequency was a confounding factor and Aslin et al. (1998) repeated Saffran et al. (1996), this time controlling for frequency. The TP effects were replicated, which was taken to mean that the effect of TP is independent of the role of frequency.

Even though infants are not expected to make complex statistical analyses over the distribution of a corpus, the infant studies using TP showed that infants were somehow able to 'track' the statistical relations between syllables. At least, it was found that infants segmented speech into words according to the TPs between syllables. Regardless of how infants acquire this knowledge, these studies show that the statistical relations between linguistic elements in a corpus can be related to infant language acquisition. It is therefore a useful strategy to investigate the information available in the input, before speculating on how this information can be acquired with an infant brain. As said before, the discrepancy between the initial awareness of infants of prominence patterns and the relatively late preference for their native languages' stress pattern warrants the question whether it is even possible to abstract a word stress system from a continuous distribution. Below, a short introduction of the simplest statistical relations and their use in linguistic modeling is given, after which it is described how these measures relate to the stress distribution.

[4]The Headturn Preference Procedure (HPP) is a widely used method to investigate infants' preferences for auditory stimuli. An auditory stimulus is associated to a light and the infant directs and maintains their attention according to their preference (see Nelson et al., 1995). That is, the procedure is based on the assumption that when the infant loses their interest in a particular stimulus they look away.

6.1.2 Computation over input information

In Saffran et al. (1996), TP effects are the effects of forward TP. In TP the relation between x and y is asymmetric and predictive and a distinction can therefore be made between forward TP (henceforth FTP) and backward TP (henceforth BTP), given in Example 6.2, where f stands for frequency and x and y are variables in the set over which probability is calculated.

(6.2) Forward TP: $\frac{\mathrm{f}(xy)}{\mathrm{f}(x)}$

Backward TP: $\frac{\mathrm{f}(xy)}{\mathrm{f}(y)}$

In contrast, there is the Observed over Expected ratio (O/E), which is symmetric and is a simple co-occurrence measure, as can be seen in Example 6.3. Again f stands for frequency and x and y are variables in the set. Furthermore, XY stands for the total number of bigrams and n for the number of unigrams in the set.

(6.3) O/E: $\frac{\mathrm{f}(xy)}{XY} \;/\; (\frac{\mathrm{f}(x)}{n} \times \frac{\mathrm{f}(y)}{n})$

In speech segmentation, TP and O/E are both able to provide information about word boundaries. In the study of Saffran et al. (1996) and Aslin et al. (1998), the 1.0 FTP within words prompted chunking or clustering, meaning that syllables were associated with each other, while the 0.33 FTP between words prompted splitting, meaning that syllables straddling this boundary were not associated with each other and in the test phase were processed by the infants as new items. O/E is a comparable measure, with the difference that it is symmetric. It has been shown to be very useful to statistically describe phonotactics, which is the definition of permissible segmental combinations in a language. Phonotactics is described by constraints on syllable structure, consonant clusters and vowel sequences and it has been found that in many languages there is a degree of gradience in what is permissible. Pierrehumbert (1993) and among others Frisch et al. (2004) applied O/E to explain phonotactics as knowledge derived from the lexicon. Kager and Shatzman (2007) challenged this position by showing an independent role of phonological constraints in a lexical decision task, that is, the abstracted phonological rules were better able to explain subjects' behavior than pure lexical statistics. In Adriaans and Kager (2010) O/E is used to model the acquisition of phonotactics from continuous speech by phonological generalization over phonotactic distributions. The O/E ratios of segmental combinations are related to distinctive features of these segments and a generalization of features over the overrepresented combinations leads to the formation of a phonological abstraction. In turn, these abstractions were shown to improve segmentation beyond pure O/E. Interestingly, this shows that an abstraction over linguistic input may be more useful in linguistic processing than mere statistical knowledge.

Differently from O/E, the use of FTP in speech processing is a reflection of the idea that language acquisition is intrinsically a forward sequential learning task, an idea that is also behind the Simple Recurrent Networks, initially proposed by Elman (1990). The unidirectional dependency between units in these networks is an illustration of the fact that speech is bound to time. Despite this conceptual underpinning and despite findings in above-mentioned studies, Gambell and Yang (2005) show that FTP alone gives poor results in automatic corpus segmentation. An important downside of statistical learning with TP is that a word boundary is inserted at a local minimum, that is when the TP between two syllables is lower than the TP of surrounding transitions. In realistic speech there are many monosyllabic words, which are impossible to segment using these 'local minima'. Gambell and Yang (2005) add the linguistic strategy called 'the Unique Stress Constraint' to statistical learning in a segmentation model. This constraint requires a maximum of one stressed syllable per word and it thereby automatically reduces the number of word-segmentation possibilities for the model, leading to better results. In combination with lexical subtraction (segmenting by use of limited lexical knowledge) the segmentation results of this model are close to the actual input. It would be interesting to know whether this model would work comparably well on other languages, since the fact that segmentation of English CDS was so good, was partly dependent of the fact that the input given to the model contained a large number of monosyllabic words, the USC increases the number of word boundary insertions and the performance of the model was best when a boundary was inserted at a randomly chosen position at points of doubt. The assumption that infants have knowledge of a 'USC' prior to experience is furthermore something that cannot be tested. Conversely, it would be interesting to see whether this knowledge can be acquired.

In Perruchet and Peereman (2004) forward dependencies were compared to backward and bidirectional dependencies and these were all correlated to participants' word-likeness judgments of French rhymes. Interestingly, this study underlined the fact that statistical measures such as FTP, BTP and Pearson r are strongly correlated, implying that the role attributed to FTP by Saffran et al. (1996) may just as well be due to any of the correlated measures such as BTP or bidirectional measures. Perruchet and Desaulty (2008) and Pelucchi et al. (2009) indeed show that BTP is just as well relatable to word segmentation. Perruchet and Peereman (2004), too, found an important role for backward relations[5]. The study at least ventures to discuss the different statistical measures available, stressing the necessity to make a careful consideration of what a measure expresses. Crucially, Perruchet and Vinter (1998) demonstrated that the results of Saffran et al. (1996) could be replicated with a simple learning

[5]Speculatively, the word-likeness of different syllable rhymes, in this case measured by participants' judgments, is better described with BTP because arguably the phonotactic dependency is in that direction; this would be the case if in a languages' rhymes, the selection of vowels is more dependent on consonants than the other way around. An additional analysis of onsets would be needed to investigate this relation.

program called PARSER which randomly stores words of different sizes and forgets them again within a certain time-frame, or when other, overlapping, words interfere, unless they are heard again, which reinforces their position in the lexicon. What Perruchet and Vinter (1998) claimed to show is that infants need not be able to make complicated statistical computations to be able to segment words from continuous speech. By applying the PARSER model to the French data, Perruchet and Peereman (2004) make the point again that listeners need not be able to make complicated statistical computations to arrive at intuitions that can be perfectly correlated to statistical distributions. According to them, the cognitive advantage of contingency in language acquisition is the fact that it strengthens the memory trace of a hypothesized cluster without interference of similar clusters. The current chapter aims to investigate whether there are such contingency relations in the stress distribution of unsegmented speech. More specifically, the question is asked whether these contingency relations are at all informative about the native language's word stress system. This means that the contingency relations should either reflect the language's metrical unit or the relation of stress to the word boundary, or both. This furthermore entails that the relations should differ between languages, according to the differences in stress systems.

Whether or not a language learner benefits from forward, backward or bidirectional contingency in the acquisition of the metrical system can be appreciated if the uncertainty of both contingencies is calculated. For this, entropy is a useful measure. Entropy is an index which reflects the disorder in a distribution. It reflects the most efficient summary of the information in a distribution. When there is high similarity between units in a distribution, entropy is high. A simple example of high entropy, that is high similarity, is the image search of 'Where's Waldo?'. Where it would be easy to find Waldo on a blank page, or on a page where no-one else wears red-and-white clothes and umbrellas, it is nearly impossible to find him on the crowded beaches and streets where he normally resides. The latter situation is one of high entropy, because there is a large amount of information and all information is equally indistinct. When units are more distinct from each other, entropy is low. Translating this to probabilities and sets, entropy is high when the probability relations in a set resemble each other, and are therefore not distinct. For example, in a set with only trochees and iambs, if there is a high number of trochees and a low number of iambs, the entropy is low. If there is a high number of iambs and a low number of trochees, the entropy is also low. If the numbers of trochees and iambs approach each other, the entropy is high. In the current investigation, entropy is therefore interpreted as ease of acquisition, since if the canonical stress pattern is distinctive from other patterns in probability, a situation of low entropy, it is likely to be easy to acquire. The formula of Entropy can be found in 6.4, where H is Entropy, p is probability, S is the set of units, i is the index referring to units in the set[6].

(6.4) Entropy: $$H(p) = H(S) = -\sum_{i \in S} p_i \log_2 p_i)$$

Shannon (1948) conceived of entropy as the number of bits needed to encode a message. Entropy was a formula already known in statistical thermodynamics and adapted by Claude Shannon in the 1940s to use it for his theory of communication, now known as Information Theory. Shannon wanted to maximize the amount of information transmittable through a noisy channel and for this, he needed to be able to determine the theoretical maximum for data compression (Entropy) and transmission rate (Channel Capacity).

Entropy seems useful in quantifying if there is a measure of contingency which makes the canonical stress pattern noticeable from continuous speech. See an example language in Example 6.5, which is a tiny corpus consisting of only five phrases.

(6.5) 110
10
00010
010
100010

The entropy of the set of bigrams for this case, is given in Table 6.1. For comparison, the relative frequencies and entropies are given for the situation in which the syllables were randomly distributed. The entropy is the index over relative frequencies rather than over O/E, because O/E is a ratio between two probabilities and it is not a probability in itself. Instead, the entropy of the non-random distribution of relative frequencies is compared to the entropy of the set of bigrams from a random distribution of the same unigrams. Table 6.1 includes the calculation of entropy, as an example. To avoid clutter, this calculation is not spelled out for the remaining sets.

In the calculation of the relative frequency of bigrams, only the syllables are treated as units, not the boundaries. This means that the number of bigrams (14) is not the number of units (19) minus 1 (18). In contrast, in the calculation of TP as shown in Example 6.2 and 6.3, the frequency of the bigram is divided over the frequency of either x or y, meaning that in this case, the transition of x to a boundary influences TP, while in the frequency of bigrams, an edge-aligned syllable does not count as a bigram (syllable+boundary). The motivation for this choice is that in TP, the relation between units is directional, and the transition of x to a preceding (BTP) or following (FTP) element has an influence on its contingency relation. On the other hand, the relative frequency

[6]In the current investigation, the units are not random variables, but elements in a continuous distribution. The set is the set of units of this continuous distribution, that is, there is a separate set of unigrams, one of bigrams and one of trigrams in each distribution. This is elaborated below.

Table 6.1: Relative frequencies of example language in Example 6.5

ngram	**O. freq**	**O. p**	$\mathrm{p}\log_2\mathrm{p}$	**E. p**	$\mathrm{p}\log_2\mathrm{p}$
00	4	0.29	-0.52	0.40	-0.53
01	3	0.21	-0.47	0.23	-0.49
10	6	0.43	-0.52	0.23	-0.49
11	1	0.07	-0.27	0.14	-0.40
entropy			1.78		1.90

O. = Observed, E. = Expected, p = probability, in this case relative frequency
entropy = negative sum of $\mathrm{p}\log_2\mathrm{p}$ of each unit in the set

Table 6.2: FTP of example language

ngram	**Observed FTP**	**Expected FTP**
00	$4/12 = 0.33$	0.46
01	$3/12 = 0.25$	0.29
0]	$5/12 = 0.42$	0.21
10	$6/7 = 0.86$	0.50
11	$1/7 = 0.14$	0.25
1]	0	0.21
entropy	2.14	2.98

of a bigram as a unit is a symmetrical relation between two units and the frequency of unigrams at edges is not of influence on the relative frequency of bigrams. This also means that in the calculation of entropy of TPs, the probability of the transition to the boundary needs to be included.

This tiny example language has a regular right-edge-aligned trochee, as can be seen in Example 6.5, since all phrases end in 10. Because of varying word size, the relation of stress to the word boundary is regular but not certain. In the five-phrase corpus, the entropy of the set of FTPs is lowest and this is because it is quite certain what will follow a stressed syllable (FTP 10 = 0.86) and it is quite uncertain what will follow an unstressed syllable. In BTP, this relation is much less defined, because what precedes a stressed or unstressed syllable is much less predictable. This is reflected in the entropy, which is higher than that of the FTP set, and, most importantly, higher than the entropy over de set of bigrams from the random distribution. Clearly, FTP in this language with right-edge-aligned stress is a measure in which the canonical trochaic pattern is most distinctive. An exact mirror image of the example language would render the mirror image of the results, meaning that in a language with a regular left-edge-aligned trochee, the BTP would have a lower entropy than the FTP.

Table 6.3: BTP of example language

ngram	**Observed BTP**	**Expected BTP**
00	$4/12 = 0.33$	0.46
01	$3/7 = 0.43$	0.50
10	$6/12 = 0.50$	0.29
11	$1/7 = 0.14$	0.25
[0	$2/12 = 0.17$	0.21
[1	$3/7 = 0.43$	0.21
entropy	2.91	2.98

6.2 Research question

The discrepancy between the early infant attention for rhythmic alternation, found in infants of only a few days old, and the relatively late knowledge of the native stress system, of which the mere onset is only found after a few months, leads to ask whether in fact continuous speech is sufficiently informative about the relation of word stress to the word boundary or whether it is necessary to first set up a proxy-lexicon to facilitate the acquisition of the canonical stress system. Furthermore, if the canonical stress pattern is discernable in the stress distribution of continuous speech, the question rises whether the metrical system of a language is reflected on a language-specific basis in the type of contingency relation, that is whether there is a symmetric or asymmetric contingency in the stress distribution of languages. Formally, the above can be fleshed out into a research question:

Is the canonical stress pattern statistically distinctive in the stress distribution of continuous speech?

If canonical stress were to be acquired from continuous input, without any other knowledge, the stressed and unstressed syllables in continuous speech need to be in a contingency relation that reflects the word stress system. If the statistical distribution of stress would inform the learner that a stressed syllable prompts a segmentation decision, the learner would benefit more from forward contingency in a right-edge-aligned language (the probability of a boundary following a stressed syllable is high) and from backward contingency in a left-edge-aligned language (the probability of a boundary preceding a stressed syllable is high). In other words, learning the relation of word stress to the word boundary may involve the use of asymmetrical measures, such as FTP for the right boundary, and BTP for the left. On the other hand, a symmetrical contingency between stressed and unstressed syllables, such as a relatively high frequency of 10 may lead the learner to find that the trochee (or iamb in the case of 01) is the main rhythmic unit of the language. Finding an information value on language-specific stress in either TP or O/E does not necessarily indicate that these probability relations are the best possible measures for infant stress acquisition (see the discussion of PARSER), but they merely function to unearth whether continuous speech contains relevant directional or bidirectional statistical information which may be used for stress acquisition.

To answer the research question as directly as possible, the linguistic information available in a corpus will be simplified to pure stress information, being a sequence of stressed (1) and unstressed (0) syllables. As reviewed in the introduction, this is the minimal information that can be assumed to be discernable by infants, even already in utero. Realistically, infants have many more cues at their disposal in their task of language acquisition and the acquisition process is assumed to be dynamic in nature. However, to investigate whether the stress distribution of a language in itself is sufficiently informative, it is useful to eliminate all other linguistic information.

6.3 Method

The information available in stress distributions is best evaluated when compared between different languages. In order to maximize interest, the current chapter will be concerned with the four languages that were tested in Chapter 4. Recall that the summarized characteristics of these languages was that Hungarian has fixed initial stress, Polish has fixed penultimate stress, Turkish has positionally variable but canonical final stress and Dutch has positionally variable but canonical penultimate stress. In the current investigation, the fixed stress languages are expected to have the lowest entropies, since their stress patterns are most regular and predictive.

Corpora with prosodic information currently available are the Turkish Electronic Living Lexicon corpus (TELL) (Inkelas et al., 2000) for Turkish and the Corpus Gesproken Nederlands (CGN) (Oostdijk, 2000) and CELEX (Baayen et al., 1993) for Dutch. The TELL corpus is a Turkish lexicon with word stress annotation, it does not contain continuous speech. The CGN is a corpus containing spontaneous Dutch speech from a variety of sources. A part of the corpus was prosodically annotated by non-expert transcribers. The prosodic annotation does not include word stress, but only emphasis (Buhmann et al., 2002). CELEX is a large lexicon of, among other languages, Dutch, with word stress annotation, not containing continuous speech. There are no Polish and Hungarian prosodically annotated corpora of spoken language. In short, existing corpora are not suitable for the current questions, nor comparable between languages. The first evaluation of the four languages will therefore be based on the annotation of a freely available short text: 'The North Wind and the Sun'. This text is used for language descriptions by the International Phonetic Association and the annotation of the story is usually done by one or two linguist transcribers, basing their annotation on a read-aloud version of the story. For the current purpose, the only interest of the corpus is the prominence distribution of syllables in different languages. The question is whether there is information in this distribution that would be of interest for metrical acquisition, which is the reason why the annotations are simplified to contain only stressed and unstressed syllables, i.e. the corpus is translated to simply 0s and 1s. The specifics of the corpus and its adaptation for the current question will be described in Section 6.4.

To answer the question whether the canonical stress pattern is discernable from the continuous stress distribution, three measures are compared in the current chapter. These measures are selected to investigate whether there is a bidirectional or unidirectional contingency that can inform the learner about the edge-alignment of the stress system. In a language with fixed initial stress, BTP is expected to be most informative, since the stressed syllable is always *preceded* by a word boundary, while in a language with fixed final stress, the stressed syllable is always *followed* by a word boundary, which means FTP is more informative. The bidirectional measure, in this case O/E, is expected to inform the learner about the main metrical units of a language, such as the iamb and the trochee. First, these three measures of contingency are evaluated in the four different languages, after which they are evaluated on their information value.

6.4 The North Wind and the Sun

6.4.1 Corpus

The prosodically annotated translations of the short text ‘the North Wind and the Sun’ that appear in the IPA-descriptions of languages Jassem (2003); Gussenhoven (1992); Szende (1994); Zimmer and Orgun (1992) are very suitable for the current pilot. For an eventual analysis of the stress distributions in these languages, this text would be too small, but the fact that the current analysis only has two levels (stressed vs. unstressed) makes it a useful first exploration. Furthermore, the prosodic annotation and the fact that the same text is translated into all four languages that were tested in Part II, makes it an appropriate source of comparison for initial study. Even so, a careful pace is necessary, because the different transcribers have made their own transcription choices.

‘The North Wind and the Sun’ is one of Aesop’s fables and it is given in British English below.

> The North Wind and the Sun were disputing which was the stronger, when a traveler came along wrapped in a warm cloak. They agreed that the one who first succeeded in making the traveler take his cloak off should be considered stronger than the other. Then

> the North Wind blew as hard as he could, but the more he blew the more closely did the traveler fold his cloak around him; and at last the North Wind gave up the attempt. Then the Sun shined out warmly, and immediately the traveler took off his cloak. And so the North Wind was obliged to confess that the Sun was the stronger of the two.

The fable is well-known by phonologists and phoneticians because of its use as an illustration of spoken language in the *Handbook of the International Phonetic Association* (IPA). Each of the languages described in the Handbook or the Journal of the IPA include the phonetic and orthographic transcription of the spoken text, translated into that language. The languages used for the current pilot are Dutch, Turkish, Polish and Hungarian.

The differences in annotation between transcribers lead to a different type of prosodic information in the texts. In the Polish case (Jassem, 2003) the phrases are divided into accentual phrases, five levels of prominence are transcribed and spaces are used to indicate prosodic words rather than coinciding with Polish conventional spacing in writing. In the Dutch case (Gussenhoven, 1992) the phrases are additionally divided into minor intonational phrases. Only the phrase accent is transcribed, which means that not all content words carry stress, and spacing is as in writing. The Hungarian transcription (Szende, 1994) contains long phrases which stop when there is a period in the text, and two levels of stress. In this transcription, too, spacing is as in writing. In the Turkish transcription (Zimmer and Orgun, 1992), finally, word stress is indicated and spacing is as in writing. To be able to compare the stress distributions of each of the languages, it was necessary to use the same system for each language, according to the guidelines in the list below.

Conserved

Number of syllables
Stress position

Adapted

Period (.) is end of phrase (other phrase boundaries deleted)
Stress is binary (all levels of stress are now 1, and unstressed is 0)

What should be clear now is that the mini-corpora consist of long sequences of 0 and 1, there is no segmental information nor information about word boundaries. There is information about sentence boundaries, which are expected to be perceptible (just as the number of syllables and the two stress levels are). Since phrase boundaries are transcribed according to different standards depending on the transcriber, the choice is made to only conserve the orthographic period, i.e. the sentence boundary. This means that there is a relative underestimation of phrase boundaries, but this adjustment was necessary because the phrasing differed extensively between transcribers. For Polish, for example, 15 phrase boundaries were ignored in the corpus, while for Turkish only 2 phrase boundaries were ignored. The Dutch text contained minor intonational phrases, meaning that the boundaries almost completely coincided with (content) words. Furthermore, there is a difference in perceived/transcribed stressed syllables between transcribers. The choice is made to transcribe every perceived stress as 'stress' which ultimately means that in some languages, secondary stress is transcribed, while in other languages the only transcribed stress is the phrase accent.

6.4.2 Analysis

In this section, the unigram, bigram and trigram frequencies of combinations of stressed and unstressed syllables in each of the languages are given and discussed. The choice for bigrams and trigrams is related to two considerations: First of all, in metrical phonology feet are in principle binary, being iambs and trochees (Hayes, 1995), but in many languages, a third unstressed syllable has an important role in the stress system, either because the system employs so-called ternary feet or recursive feet (Prince, 1980; Martínez-Paricio and Kager, forthcoming), or because syllables can be extrametrical. In Dutch, one of the tested languages, stress in monomorphemic words is always placed within a three-syllable window at the right edge (Kager 1989; Gussenhoven 2009, c.f. van Oostendorp 2012). This possible influence of a third syllable gives reason to include trigrams in the analyses. Second of all, the addition of larger units in the analyses allows for more variation. If it were to be the case that bigrams do not contain sufficient statistical information for the learner, the extended window of analysis allows for a broader view on the available information.

Bigrams are calculated according to the formulas in Example 6.2 and 6.3. For trigrams, these calculations are comparable. For O/E, the calculation in Example 6.6 is the observed frequency divided by the frequency that would be expected on the basis of the frequencies of each separate element. In the formula, f is frequency, xyz is a trigram in the set, composed of the unigrams x, y and z, XYZ is the total number of trigrams in the set and n is the total number of unigrams. This formula conserves the symmetrical nature of the O/E.

(6.6) O/E: $\frac{\mathrm{f}(xyz)}{XYZ} \;/\; (\frac{\mathrm{f}(x)}{n} \times \frac{\mathrm{f}(y)}{n} \times \frac{\mathrm{f}(z)}{n})$

It is possible to compute the O/E of trigrams as the O/E of the two combined bigrams, too. This is a matter of increasing the window rather than increasing the size of the unit, since in that case it is measured how often two bigrams co-occur. In this first analysis, the question is whether the units as such contain information about the metrical system, so the calculation of the trigram is as given in Example 6.6. For TP, the reasoning is slightly different. The directional nature of TP should be preserved in trigram calculations, too. This means that the TP of a trigram is the probability that the third syllable follows the previous two in FTP, and the probability that the first syllable

Table 6.4: Ngram frequencies Hungarian North Wind and the Sun

N	**ngram**	**Freq**	**Relfreq**	**OE**	**FTP**	**BTP**
unigram	0	134	0.70			
	1	58	0.30			
entropy			0.88			
bigram	00	78	0.42	0.86	0.58	0.58
	01	52	0.28	1.32	0.39	0.90
	10	53	0.28	1.34	0.91	0.40
	11	4	0.02	0.23	0.07	0.07
entropy			1.67 (1.77)		1.64 (2.07)	1.65 (2.07)
trigram	000	36	0.20	0.58	0.46	0.46
	001	38	0.21	1.42	0.49	0.73
	010	47	0.26	1.76	0.90	0.89
	011	4	0.02	0.35	0.08	1.00
	100	41	0.23	1.53	0.77	0.53
	101	12	0.07	1.04	0.23	0.23
	110	4	0.02	0.35	1.00	0.08
	111	0				
entropy			2.42 (2.65)		2.55 (4.14)	2.66 (4.14)

precedes the following two in BTP. The formula is given in Example 6.7, where f is frequency, xyz is a trigram composed of the unigrams x, y and z, xy and yz are bigrams.

(6.7) FTP: $\frac{\mathrm{f}(xyz)}{\mathrm{f}(xy)}$

BTP: $\frac{\mathrm{f}(xyz)}{\mathrm{f}(yz)}$

6.4.3 Results

In Tables 6.4, 6.5, 6.6 and 6.7, the n-gram frequencies as well as O/E, FTP and BTP are given with the entropies of the sets. In each of the tables in the remainder of this chapter, the entropies are given together with the entropies of sets from a random distribution, placed between brackets. In the following subsections, the measures are discussed one by one.

Ngrams, relative frequencies and O/E

The unigram frequencies show that there are differences between the languages due to choices made by individual transcribers. The fact that Hungarian has a

Table 6.5: Ngram frequencies Polish North Wind and the Sun

N	ngram	Freq	Relfreq	OE	FTP	BTP
unigram	0	97	0.62			
	1	60	0.38			
entropy			0.96			
bigram	00	42	0.28	0.72	0.43	0.43
	01	53	0.35	1.48	0.55	0.88
	10	53	0.35	1.48	0.88	0.55
	11	4	0.03	0.18	0.07	0.07
entropy			1.73 (1.92)		1.76 (2.26)	1.76 (2.26)
trigram	000	14	0.10	0.40	0.33	0.33
	001	28	0.19	1.31	0.67	0.53
	010	47	0.32	2.19	0.89	0.89
	011	3	0.02	0.23	0.06	0.75
	100	28	0.19	1.31	0.53	0.67
	101	23	0.16	1.73	0.43	0.43
	110	4	0.03	0.30	1.00	0.08
	111	0				
entropy			2.46 (2.88)		2.72 (4.51)	3.51 (4.51)

Table 6.6: Ngram frequencies Turkish North Wind and the Sun

N	ngram	Freq	Relfreq	OE	FTP	BTP
unigram	0	112	0.62			
	1	70	0.38			
entropy			0.96			
bigram	00	56	0.32	0.84	0.50	0.50
	01	55	0.31	1.31	0.49	0.79
	10	54	0.31	1.29	0.77	0.48
	11	12	0.07	0.46	0.17	0.17
entropy			1.84 (1.92)		2.04 (2.23)	2.01 (2.23)
trigram	000	19	0.11	0.47	0.34	0.34
	001	36	0.21	1.44	0.64	0.65
	010	40	0.23	1.60	0.73	0.74
	011	11	0.06	0.70	0.20	0.92
	100	37	0.22	1.48	0.69	0.66
	101	17	0.10	1.09	0.31	0.31
	110	11	0.06	0.70	0.92	0.20
	111	1	0.01	0.10	0.08	0.08
entropy			2.68 (2.88)		3.41 (4.45)	3.47 (4.45)

Table 6.7: Ngram frequencies Dutch North Wind and the Sun

N	**ngram**	**Freq**	**Relfreq**	**OE**	**FTP**	**BTP**
unigram	0	129	0.79			
	1	34	0.21			
entropy			0.74			
bigram	00	91	0.58	0.92	0.71	0.71
	01	33	0.21	1.27	0.26	0.97
	10	33	0.21	1.27	0.97	0.26
	11	1	0.01	0.15	0.03	0.03
entropy			1.47 (1.48)		1.20 (1.83)	1.20 (1.83)
trigram	000	63	0.41	0.83	0.69	0.69
	001	25	0.16	1.25	0.27	0.76
	010	32	0.21	1.60	0.97	0.97
	011	1	0.01	0.19	0.03	1.00
	100	27	0.18	1.35	0.82	0.30
	101	4	0.03	0.76	0.12	0.12
	110	1	0.01	0.19	1.00	0.03
	111	0				
entropy			2.15 (2.22)		2.10 (3.66)	2.19 (3.66)

high proportion of unstressed syllables is as expected, but the fact that this is also the case for Dutch, a language with many short words, is due to the fact that not word stress, but phrase-stress is transcribed. In languages where the proportion of unstressed syllables in the transcription is much bigger than that of stressed syllables, the entropy of all sets is lower. This is a reminder of the importance of the comparison of the measures with sets drawn from a random distribution.

In all four languages, both 10 and 01 are overrepresented, where 00 and 11 are underrepresented, as shown by O/E. Of course, language is rhythmic and a random distribution is not. The simple analysis shows that solely the set of bigram O/Es of the current corpora do not inform a learner about the canonical stress pattern of the language. Entropy and its comparison to the entropy of a set from a random distribution is just as in Example 6.5: the entropy is not much lower, showing that the symmetrical bigram contains almost no useful information about stress.

In trigrams, too, alternation is overrepresented while trigrams containing clash (11), and a trigram of unstressed syllables (000) are underrepresented. The latter finding is interesting in a language with long sequences of unstressed syllables, such as Hungarian. In the classical use of bigram frequencies for speech segmentation, the underrepresentation of a unit would suffice for hypothesizing a word boundary, but in the distribution of stress sequences, the area around the word boundary is in fact overrepresented, both in bigrams and trigrams. Take for example 010, the most overrepresented trigram in each of the four languages: This trigram includes the canonical word boundary for Hungarian, while at the same time being the most overrepresented unit of all (O/E 1.76). In a language with regular or fixed edge-aligned stress this overrepresentation of the pattern that includes the canonical word boundary should actually not be surprising. I will return to this in the discussion. Regarding entropies: the sets of trigrams do not have much lower entropies than random. Only in Polish the entropy of the set from the actual stress distribution is somewhat lower than that from the random distribution.

TP

If not in the symmetrical measure of O/E, can a pattern be discerned in the asymmetrical measures? In the tiny example language in Example 6.5, the FTP was a more distinctive measure than BTP, attested by the lower entropy. Translating this to the current analysis, Hungarian should have a lower entropy on BTP, since it is a language with left-edge-aligned stress. However, looking at FTP and BTP does not make it clear whether Hungarian stress is oriented to the left boundary or the right boundary. The entropy for both directions is almost indistinguishable. Only in Polish, the FTP has the lowest entropy, and BTP has a higher entropy than relative frequency. Polish is like the example language in Example 6.5, with a trochee fixed to the right edge.

What does become clear is that directional TP is more informative than relative frequency in each of the languages, since the entropies over the TP sets are consistently lower than those over sets from a random distribution. A major difference between the entropy of sets of relative frequencies and that of sets of directional TPs is the inclusion of the probability that something other than a stressed or unstressed syllable will follow. This possibility obviously increases the entropy over sets from the random distribution, since more, similar, units in a set make entropy high, but it does not have much of an effect on the entropy of the actual sets. This suggests a role for phrase-edge patterns.

The individual ngrams do not seem to provide much language-specific information. Rather, the alternation of 1s and 0s translates to TPs that are higher when going from a rare unit (stressed syllable) to a more common unit (unstressed syllable) and a low TP when moving in the other direction. Instead of providing information about frequent combinations, this dressed-down distribution provides information about the importance of salience. A salient unit, i.e. the relatively infrequent stressed syllable, does not only draw attention with phonetics, but also with statistics. Compare the entropies of the set of FTP of stress-initial bigrams (Hungarian: 0.50, Polish: 0.65, Turkish: 0.97, Dutch: 0.19) to entropy of the set of FTPs with bigrams starting with unstressed syllables (Hungarian: 1.14, Polish: 1.11, Turkish: 1.07, Dutch: 1.01).

The same is true for BTP, where the entropy of the set of bigrams ending with stressed syllables is lower than that of the set of those ending with unstressed syllables.

6.4.4 Interim discussion

The results of the analysis of these four corpora suggests that no metrical information in the ngrams in continuous speech is useful to learn the canonical stress pattern. However, the term 'unsegmented' has been taken very strictly. The phrases in the corpus are very long ones which in spoken language would be broken up into smaller pieces, clearly demarcated by silence or intonational boundaries. An increased number of boundaries may make it easier to deduce the language-specific metrical system from the stress distribution. Clearly, the current analysis started from a basis with few opportunities of success, having only the one-dimensional, binary stress opposition, few boundaries and different transcription principles per language. The rationale behind this was to see whether the bare stress distribution of unsegmented speech input would be sufficiently informative for the acquisition of the metrical system of a language. Since the input now was a written text with few boundaries, this question has not been given enough opportunity to be answered successfully. In the next section, two corpora of spoken language are analyzed in the same way as 'the North Wind and the Sun', to once again investigate whether there is bigram information in spoken rhythmic distributions.

6.5 Radio Speech

To allow for the bigram analysis to have improved ground to succeed, it is, first and foremost, important to use corpora with slightly increased ecological validity. Therefore, the corpora need to contain spontaneous speech. A possible discrepancy between stress patterns in written and spontaneous language is that stress patterns in spontaneous speech are not motivated by literary and poetic choices but by processing costs. For the same reason, spontaneous speech is likely to contain shorter phrases than written language. As said in the discussion above of the results of the analysis of 'the North Wind and the Sun', and as can sensibly be expected, phrase boundaries are likely to be important in the acquisition of the relation of word stress to the word boundary.

Another way in which the analysis can be made more ecologically valid is the stress annotation. In 'the North Wind and the Sun', the stress transcription is a phonological analysis of the language, performed by trained phonologists with individual bases for decision. For the current purpose, however, the stress distribution should be based on perception, since the analysis is based on the assumption that infants have a phonetic perception of stress but are not yet making a phonological analysis of it. The stress distribution in the current corpus is therefore a result of stress perception by participants, as will be explained in Section 6.5.1.

As said at the beginning of this chapter, existing corpora are unsuitable for the current purpose, which means a corpus should be created specifically for it. Of course, this has as its downside that the created corpus will be small and only two languages will be compared. However, the analysis will again be limited to only one dimension, being the binary stressed-unstressed distinction. For now the segmental information is not relevant[7] and the corpora need not be as large as they would have to be for segmental analyses. Fewer tokens are needed for statistical regularities to stabilize in a binary system, than in a system with over 25 levels, such as a small segmental system would be, or a system with over 20,000 levels, such as a vocabulary would be.

6.5.1 Corpus

Seven fragments of each one minute of spontaneous radio conversation were recorded for both Turkish and Dutch. Seven different fragments composed a variety of topics and speakers, thereby reducing potential overrepresentation of certain words or possible speaking styles.

The Dutch corpus was compiled by the author, and checked by three trained phonologists who are also native speakers of Dutch. The manual segmental, orthographic, and boundary annotation of words and phrases was also done by the author, and checked by one trained phonologist who is also a native speaker of Dutch. The prosodic annotation was performed by native speakers who did not have knowledge of the purpose of the study. This will be discussed later in Section 6.5.1.

The Turkish corpus was compiled by a linguist who is also a native speaker of Turkish. The compiler of the Turkish corpus did the manual segmental, orthographic and boundary annotation herself. The Turkish annotation was checked by three native speakers of Turkish of whom one is a linguist, and the two others are not. Again, native speakers who did not have any knowledge of the purpose of the study were responsible for the prosodic annotation.

The radio fragments were taken from radio shows that were freely available on-line at the time of the compilation of the corpus. The website that was used for compilation is *www.tunein.com*. Audacity (Mazzoni and Dannenberg, 2013) was used for recording and preparing the audio files, and Praat (Boersma and Weenink, 2011) was used for transcription and annotation. The different broadcasts and topics are given per radio fragment in Table 6.8, together with

[7] Lexical exceptions challenge this position, however. If a certain word with irregular stress would be very frequent in the current corpus, it would have a stronger effect on the distilled metrical system than what would be desired. A more complicated analysis which combines segmental information with stress patterns, would be able to deal with this type of exception, because it would ascribe the exceptional stress pattern to the unique segmental form, setting the pattern aside as an exception. In the current analysis on this small corpus, problems of this kind may persist and caution should therefore be taken with the interpretation.

Table 6.8: Radio fragments

Lg.	**Broadcast**	**Topic**	**Sec**	**Phr**	**Wrd**	**Syll**
Du	BNR	Budget cuts army	68	65	189	319
	BNR	Malnutrition elderly	68	62	220	352
	BNN	Nostalgia	73	56	231	324
	EO	European economics	71	58	253	391
	KRO	Interview writer	71	61	184	253
	NCRV	Dutch politics	84	63	286	422
	NTR	Interview cartoonist	70	50	191	257
Tu	TRT radyo	Diet of Kazakistan	61	45	96	276
	TRT Antalya	Soccer team finances	61	51	99	285
	TGRT FM	Death penalty	61	66	124	304
	Haberturk	Turkish politics	61	53	108	266
	Alem FM	Feminist debate	61	51	94	248
	Alem FM	Cortisol	60	66	139	361
	Alem FM M.	Physical exercise	61	64	144	343

the exact time in seconds, the number of phrases, the number of words and the number of syllables.

Annotation

The annotation of the corpus was structured to include orthographic, phonetic, phonemic and boundary information. For the current analysis, the most important information is boundary information and the position of stress (see Section 6.5.1). For future use, other information was transcribed, such as the position and number of phonetic syllables (using the automatic syllable finder by De Jong and Wempe (2009)) and a broad transcription of the pronounced segments, including insertions, inversions and filled pauses, in XSAMPA. The corpus was also annotated on the dictionary position of both word and phrase stress. For Turkish, this latter task was done by two native speakers of Turkish.

The orthographic annotation of the corpus was used in the perception surveys and can be found in Appendix B.2.1. Each new line gives a new phrase. These phrase boundaries were annotated by the compilers of the corpora and checked by three native speakers each, as described in the previous section. The boundary criteria were agreed upon beforehand and checked during the process. A phrase end was defined by pauses, filled pauses, the start of a new intonation contour or the repetition of an intonation contour (list intonation). The next section describes the perception survey which was given to native listeners of the languages for the transcription of perceived stress.

Perception survey

Method 35 Dutch native speakers and 33 Turkish native speakers participated in the experiment. All participants were university students and were tested individually, receiving auditory stimuli over headphones. None of the participants were trained phoneticians or phonologists. 32 Dutch students were tested at Utrecht University, the Netherlands, and three students participated from their respective homes, using headphones as instructed. Two Dutch participants had to be excluded: one participant marked each entire word for stress instead of separate syllables, the other excluded participant reported to know from own experience and remarks from family and friends to have poor stress perception and production. The Turkish participants were tested at Koç Üniversitesi, Turkey. All participants received compensation, none reported any speech, reading or hearing disorders and all had normal or corrected-to-normal vision. Participants were instructed in their native language and they were randomly given one of the seven fragments. The instructions and small list of questions are given in Appendix B.2.2. Participants were free in the amount of time they wanted to take to finish the survey and they were free in the number of times they wanted to repeat listening to each phrase.

It is important to note that it was not specified whether participants had to indicate phrase stress or word stress. This choice was deliberate, because the participants were naive listeners. Without training, it is expected that participants are less consistent and they may be unsure about the perception of word stress. In the context of a longer phrase, word stress is less straightforwardly perceived, if at all. Sentence stress, on the other hand, indicates the position of word stress, because in most cases it is placed on a syllable that receives main word stress[8]. Phrase stress is expected to be more easily perceived than word stress, because it is produced as a salient intonation contour. To ask participants to indicate where they would perceive phrase stress, however, might prompt them to leave out certain perceived prominences, because the length of the phrase is already given in the form. Asking the participants to indicate wherever they perceive prominent syllables is a better reflection of perceived prominence overall. The number of votes for stress on a particular syllable is translated to a probability of stress perception per syllable, using Formula 6.8.

(6.8) $\text{salience} = \frac{\text{votes ‘stressed’}}{\text{no. transcribers}}$

Results In Table 6.9, the number of transcribers is given per fragment and the inter-transcriber reliability is calculated per language. The inter-transcriber reliability is based on a binary choice per syllable: for each syllable it is true that

[8]However, it can also be placed on a syllable that receives secondary stress. Both in Turkish and Dutch, the phrasal context allows word stress to shift position. For the current study, this shift is actually very interesting in the context of segmentation, since the Dutch sentence stress often moves to the left, whereas the Turkish sentence stress moves to the right of the word (see later in Section 6.5.1). This has implications for the use of stress in segmentation.

Table 6.9: Participants and inter-transcriber reliability per fragment

Fragment	**Dutch**	**Turkish**
1	5	5
2	4	5
3	5	4
4	5	5
5	6	5
6	6	4
7	4	4
	71.8%	61.5%

each transcriber regarded it as stressed or unstressed. If there are 5 transcribers that transcribed the same syllable, there are 10 combinations of transcribers, as can be seen in Example 6.9. If four of these transcribers agree, the reliability of the transcription of that syllable is $6/10 = 60\%$. This principle is made visible in Figure 6.1.

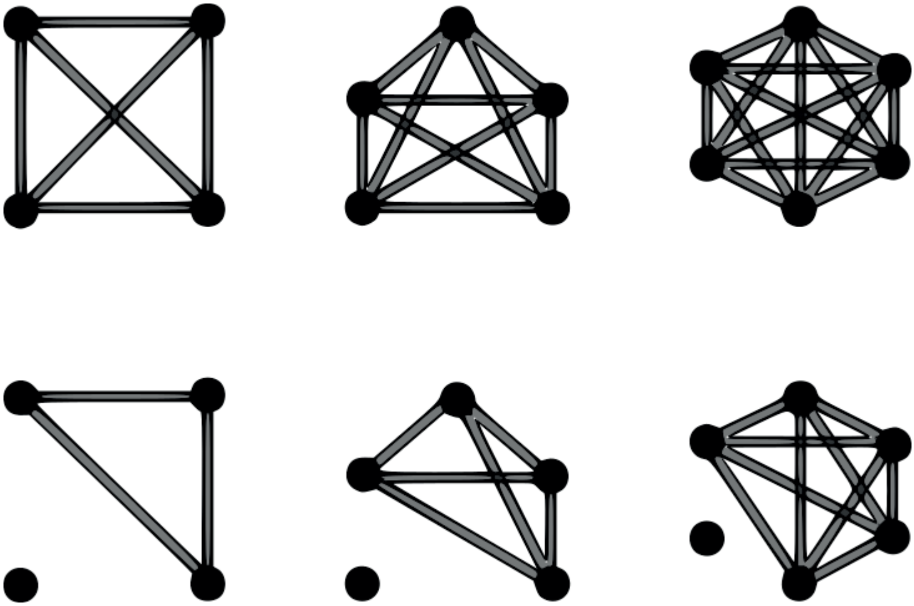
(6.9) $\binom{5}{2} = \frac{5!}{2!(5-2)!} = 10$

Table 6.9 shows that there is a clear difference in inter-transcriber reliability between Dutch and Turkish. This may be due to the assumed 'stress-deafness', reviewed in Chapter 2, found for canonical word-final stress, but not for non-canonical stress, in speakers of Turkish in a behavioral task (Domahs et al., 2012a)[9]. Furthermore, as reviewed in Chapter 3, Turkish stress is phonetically not so salient, especially in word-final position (Levi, 2005). A lower accuracy was also found in the stress judgment task of Experiment 1, as can be seen in Table 4.2. Both Dutch and Turkish inter-transcriber reliability are well above chance, however, so the transcriptions are regarded to be useful as input.

Figure 6.2 shows the first fragment as transcribed by five Dutch transcribers. Notably, the number of votes per syllable forms a rhythmic pattern: unanimity intertwined with fewer votes. This may be a sign that listeners diverge in terms of the level of stress they annotate, rather than disagreeing on the position of stress in the word. Some listeners may find phrase stress 'prominent' while skipping over word stress, while others may find every stressed syllable prominent. It is therefore an interesting option to use the divided nature of stress-judgments of the transcribers as a measure of salience: when the transcribers unanimously judge a syllable to be stressed, this is salient stress, while the judgment of one transcriber is a less salient stress. For the input of

[9] An ERP analysis of stress violations in this study showed, however, that violations of both the canonical and the non-canonical stress pattern are noticed by Turkish listeners, that is, 'stress-deafness' in Turkish is limited to behavioral tasks.

Figure 6.1: Inter-transcriber reliability



Above: possible combinations, Below: number of combinations when one transcriber disagrees

Figure 6.2: Example of the distribution of stress judgments. First part of Dutch fragment 5

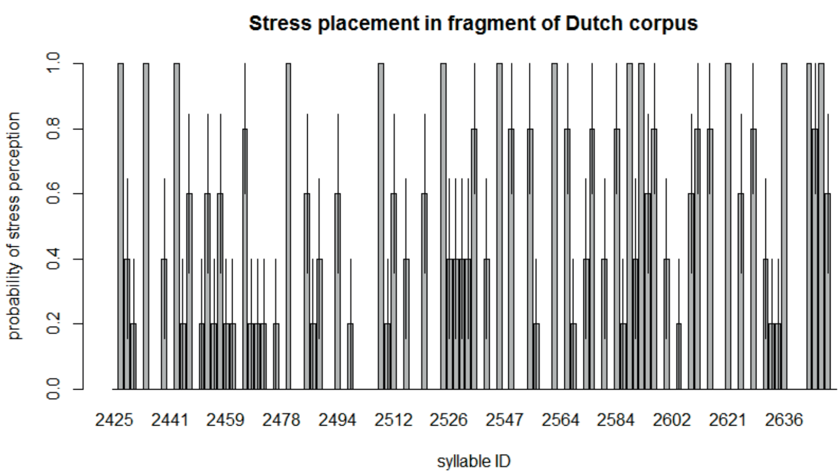


Table 6.10: Turkish radio corpus ngram frequencies, O/E and TP

N	ngram	Freq	Relfreq	OE	FTP	BTP
unigram	0	1323.65	0.64			
	1	737.35	0.36			
entropy			0.94			
bigram	00	737.35	0.44	1.06	0.56	0.56
	01	426.55	0.25	1.10	0.32	0.58
	10	368.85	0.22	0.95	0.50	0.28
	11	153.25	0.09	0.71	0.21	0.21
entropy			1.81 (1.88)		2.85 (2.83)	2.81 (2.83)
trigram	000	402.26	0.31	1.16	0.55	0.55
	001	229.55	0.18	1.19	0.31	0.54
	010	184.82	0.14	0.96	0.43	0.50
	011	73.77	0.06	0.69	0.17	0.48
	100	216.48	0.17	1.12	0.59	0.29
	101	98.16	0.07	0.92	0.26	0.23
	110	76.14	0.06	0.71	0.50	0.21
	111	29.81	0.02	0.51	0.19	0.19
entropy			2.69 (2.82)		5.72 (5.66)	5.87 (5.66)

the calculations, this measure of salience is translated to a probability of stress perception, as can be seen in Example 6.8.

Monosyllabic phrases are removed from the analysis, because they do not provide information about stress distributions. This renders 397 - 24 = 373 phrases in Turkish and 415 - 51 = 364 phrases in Dutch.

6.5.2 Analysis

Again, the unigram, bigram and trigram frequencies of combinations of stressed and unstressed syllables in both languages are given and discussed and the same calculations as in Section 6.4 are used for these calculations. These frequencies are, as said, calculated on the basis of the probability of stress perception.

6.5.3 Results

In Tables 6.10 and 6.11, the ngram frequencies as well as O/E, Forward TP (FTP) and Backward TP (BTP) are given as well as the entropy of these sets. They show that the change of corpus did not facilitate metrical acquisition through bigram and trigram contingency. If anything, the shorter phrases and

Table 6.11: Dutch radio corpus ngram frequencies, O/E and TP

N	ngram	Freq	Relfreq	OE	FTP	BTP
unigram	0	1673.98	0.74			
	1	588.02	0.26			
entropy			0.83			
bigram	00	1025.69	0.54	0.98	0.61	0.61
	01	394.69	0.21	1.08	0.24	0.67
	10	394.76	0.21	1.08	0.67	0.24
	11	87.85	0.05	0.69	0.15	0.15
entropy			1.64 (1.65)		2.58 (2.59)	2.58 (2.59)
trigram	000	628.05	0.41	1.00	0.61	0.61
	001	243.62	0.16	1.11	0.24	0.62
	010	259.42	0.17	1.18	0.66	0.66
	011	53.90	0.04	0.71	0.14	0.61
	100	228.63	0.15	1.04	0.58	0.22
	101	66.55	0.04	0.87	0.17	0.17
	110	49.67	0.03	0.65	0.57	0.13
	111	14.16	0.01	0.55	0.16	0.16
entropy			2.39 (2.48)		5.39 (5.17)	5.27 (5.17)

the perceived stress have made the metrical system more diffuse in these measures, leading to higher entropy across the board. The spectacular increase of entropy as compared to the pilot of ‘the North Wind and the Sun’ is a counterintuitive effect of the higher proportion of boundaries in the corpus. This can be seen from the increase of the entropy placed between brackets, as shown in the table. This number is the entropy of a random distribution of the same elements. When the separate elements of a random distribution approach each other in proportion, the entropy is higher than when one of the elements has a relatively low frequency. In the current analysis of directional TP, boundaries are included as units and because their relative frequency approaches that of stressed and unstressed syllables, the sets of directional TPs have high entropies, from metrical and random distributions alike.

The perceived stress patterns in spontaneous speech furthermore seem to be much less restricted to specific patterns than the phonological transcription of ‘the North Wind and the Sun’. Most languages in the latter corpus disallowed a trigram of three consecutive stressed syllables, but in the spontaneous speech corpus this is perceived, though rare. With this latter point, caution needs to be taken, since the ‘perceived’ stressed syllables are calculated by probability (based on how many participant perceived that syllable as stressed), and it is therefore not necessarily the case that a single participant perceived three stressed syllables in a row.

6.6 General Discussion

As already pointed out in the interim discussion, the bigram and trigram contingency of stressed and unstressed syllables seems to be an implausible measure for metrical acquisition. In the first corpus, the phonological transcription of stress contained little information about the metrical system, although some information did seem to be contained in forward transitional probability in Polish. The low number of phrase boundaries in that corpus is a potential reason that the direction of contingency did not surface, since the phrases were long stretches of alternating stress patterns. This was the reason to take larger corpora with shorter phrases as a basis of computation. However, the results seemed to become even less interpretable, leading to higher entropy, even compared to the entropy of a random distribution of the same unigrams. What may be the reason for this result?

First of all, it is possible that the results of Turkish and Dutch would have been easier to interpret if they were compared to a fixed stress language such as Hungarian. After all, both Turkish and Dutch are languages which allow for a large amount of positional variation of stress. However, the analysis of 'the North Wind and the Sun' did not show that Hungarian bigrams or trigrams were informative about stress, even though Hungarian has fixed initial stress. Second of all, it is possible that the choice of perceived stress, as annotated by untrained native speakers, is not an appropriate source of information. After all, it is possible that these transcriptions are less consistent than they would be if they were based on phonological choices rather than on perception. However, the intention of the use of this corpus was in fact the ecological validity of perception as compared to phonological transcription. Furthermore, the inter-transcriber reliability of the many different participants was good and any remaining disagreement has less weight, due to the translation of perception votes to probability.

Alternatively, it is possible that the current measure of contingency is unsuitable for the purpose. Imagine a trochaic language with strictly disyllabic words. This language would have an alternating stress pattern and would require segmentation before each stressed syllable. In the distribution, then, both trochees and iambs are statistically overrepresented. More generally speaking: any language that has a regular boundary marker would face this problem. Hay (2000), discussed by Pierrehumbert (2003), points out that, also in phonotactics, a low frequency phoneme transition is not necessarily correlated to word boundaries in the way that is assumed using transitional probabilities; a formal language may be invented in which words invariably begin with /t/ and end in /k/. In this language, /kt/ would probably be the most common, high frequency, phoneme transition, while being the extremely reliable marker of a word boundary. This is, in fact, precisely the situation of the stress distribution question: the more regular the stress patterns in marking word boundaries, the more common the edge combination, as was shown in Hungarian, where the most overrepresented trigram was the boundary pattern 010. This makes it clear that the contingency method is counter-intuitive for the current question: the more regular and edge-marking the rhythm, the less appropriate conclusions would be drawn, if the metrical units were to be based on high contingency. Unless, of course, phrases become so short that the pattern that crosses the boundary, /kt/ in the example, becomes less frequent than the boundary patterns /t/ and /k/ separately, that is, if one has a lexicon. The idea that phrase boundaries are important can also be approached from a perspective in which phrases do not necessarily have to be short, as long as listeners pay attention to what happens at the phrase boundary. This is the approach of Chapter 7.

Even if the analysis did not unearth how metrical patterns would become available to infant learners, the contingency measures do show that the stressed syllable is not only phonetically, but also statistically salient, in the sense that infrequent elements attract attention and probabilities departing from stressed syllables are more distinctive. The same is true for boundaries, which make their neighboring pattern salient. If we want to maintain pursuing the question whether metrical acquisition is possible on the basis of unsegmented speech, a good next step would be to focus on salience. In Chapter 7, it is investigated whether a salience-based approach on a corpus of spoken language can provide the information needed to learn the metrical system of a language. This salience-based approach investigates whether phrase-boundary regularities are informative about word boundary regularities. As found in the current analysis, the bigram does not seem suitable for the purpose. Actually, the evaluation of ngrams has the conceptually inaccurate assumption that each syllable has the same weight in the statistical calculation, that is that 0 and 1 are equivalent units. In fact, a salience-based approach to the acquisition of the metrical system should depart from the stressed syllable, which is the salient unit of speech, expected to be leading in rhythmic perception. Learning should therefore be activated by the stressed syllable, and to acquire the distance of word stress to the word boundary, the phrase boundary may be the best place to focus on. This is because, first of all, the end or beginning of a phrase is salient and it makes the pattern next to it salient because of primacy and recency constraints on memory (see Henson, 1998). Second of all, the phrase boundary is the only certain boundary on which a learner can base their hypothesis of boundary patterns. Third of all, it is known that in fact learners and listeners pay attention to boundaries, especially when they are making rule-based generalizations (Endress and Mehler, 2009). The next chapter elaborates on the relevant literature on edge-based learning and it approaches the acquisition problem of stress patterns in continuous speech from the phrase boundary: What information about the stress system of a language can be discerned from the relation of a stressed syllable to the phrase boundary?

CHAPTER 7

Salience

7.1 Introduction

In Chapter 6, it was shown that ngram statistics are an implausible tool in the acquisition of metrical patterns or the relation of word stress to the word boundary from continuous speech. The statistical relations between stressed and unstressed syllables did not differentiate languages. Furthermore, the metrical units or the relation of stress to the word boundary were not recognized when looking at the ngram contingencies and, moreover, the sets of bigrams and trigrams from actual stress distributions did not have lower entropy than sets of bigrams and trigrams from a random distribution. What did become clear from this very basic statistical exercise was that from a statistical point of view it is likely to be useful for a learner, when acquiring stress patterns from continuous speech, to pay attention to stressed syllables. Besides being phonetically salient, stressed syllables are statistically salient and metrical acquisition is the acquisition of prominence, that is salience, relations. Regardless of the regularity of the stress system of the language, the ngram stress distribution of unsegmented speech gives little information about the main metrical unit of the language and its relation to the word boundary, but what if the stressed syllable by itself gives information about the relation of stress to the word boundary? The issue of the current chapter is to investigate whether stress patterns at the phrase level in Turkish and Dutch contain sufficient information to acquire knowledge of their word stress regularities.

As discussed before, word stress is commonly right- or left edge aligned. An infant who would supposedly learn the specifics of a word stress system, such as the metrical unit, the peripherality of stress and the regularity of stress from continuous speech, would greatly benefit from knowing at which side this information is most readily learnable. Moreover, due to the salience of phrase boundaries, it would be very useful if the learner were able to infer the relation of word stress to the phrase boundary to a relation of word stress to the word boundary. Whether this inference is possible depends on how discernable the relation of word stress to word boundary is from regularities at the phrase boundary. As in Chapter 6, a measure of distinctiveness is entropy. Low entropy means low uncertainty, or similarity, and high entropy means high uncertainty, or similarity. When a set has low entropy, this can be interpreted to be a good set for learning, because there is high distinctiveness. In Chapter 6, the lack of a difference between entropies of sets from metrical distributions and random distributions showed that ngrams do not make the canonical stress pattern discernable from continuous speech. In the current chapter, the question is whether information at the phrase boundary is at all useful for the learner. For example, in a left-edge-aligned stress system, the entropy of a set at the left word edge is likely to be lower than that at the right word edge, since there is less uncertainty at that side of the word about the position of the stressed syllable. This idea will be further explained in Section 7.3.

7.2 Generalization at the edge

In language, many morphological and phonological processes are edge-aligned, among others word stress. Phonetically, too, speakers strengthen the edges of prosodic domains (Fougeron and Keating, 1997), which is an aid in perception. Taken together, there ought to be a cognitive advantage for listeners and language learners to pay attention to the edges of domains, especially when these edges are auditorily salient, to make generalizations about linguistic structure.

Endress and Mehler (2009), following Peña et al. (2002) and Endress and Bonatti (2007), argue that the first steps of human language learning can be divided into two separate mechanisms. On the one hand, there are statistical computations which lead to word learning, on the other hand there are generalizations which lead to rule abstraction. In their experiment, adult listeners were presented with a familiarization stream with pentasyllabic non-words of which the first and last syllable belonged to distinct classes and the first syllable predicted the last syllable with certainty, that is with TP 1.0: ba...de, fu...gi, mY...lo, Ri...na. In the test phase, the participants had to decide which of two words were part of the language. The critical distinction between the two options was that one of them was a part-word, that is, it had been part of the familiarization but it straddled a boundary, and the other was a word that had not been part of the familiarization, but it did conform to the artificial grammar in the sense that it started with one of the syllables from the first syllable class (ba, fu, mY, Ri) and it ended with one from the last syllable class (de, gi, lo, na), but the syllables had never appeared together in the familiarization, that is, the TP between them was 0.0. The results showed that listeners were able to learn this grammar, but only when words were separated by pauses, in this case of 1 second. Crucially, the second experiment showed that if the class-syllables were not located at word edges, but in the middle of the sequence, that is, the second and penultimate syllable belonged to predefined classes, this grammar was not learned at all, not even in the segmented condition. Here, participants performed at chance. In the unsegmented sequences, participants had a preference for part-words in both experiments. These results show that the presence of a boundary as well as a boundary-related rule helps listeners generalize the input to find that rule. Moreover, learning this rule precedes the statistical learning of words, as was found in Endress and Bonatti (2007).

A similar learning mechanism as found in Endress and Mehler (2009) was, as discussed by these authors, already found in early artificial grammar learning experiments (a.o. Braine, 1966; Smith, 1966). In Smith (1966), participants learned letter combinations of classes M (v,s,r) N (g,k,l) P (z,x,d) and Q (m,i,j). In the familiarization phase, bigrams were invariably either of type MN or of type PQ. In the test phase, however, listeners incorrectly recalled the combinations MQ and PN significantly more often than expected, which suggests that their generalization concerned positional learning, rather than learning the mutual relation of syllables. The results also suggest that both edges are considered separately, apparently the participant makes two separate positional generalizations. In Lany and Gómez (2008), 12-month-old infants did learn a similar mutual nonadjacent relation, that is, the generalization that the stimuli acX and bcY were grammatical, but not the stimuli acY and bcX, but in this case the left-hand class only consisted of two elements where the right-hand class consisted of eight. This generalization is comparable to learning article-noun relations. In principle, these nonadjacent relations are learnable and ubiquitous in language, but the studies discussed above do show that in early learning, generalizations are easier to make at edges and these edges are considered separately. Research on sequential memory also suggests that edges are important for learning, and more specifically regarding the positions of items, since positions of items are computed relative to the edge. A review on positional theory as well as on chaining and ordinal theory can be found in Henson (1998). His own work boils down tto the Start-End Model, where among others primacy and recency constraints influence short-term memory. Primacy, recency and isolation effects are well-known in psychology and they concern the fact that listeners are better able to memorize items that are placed at the beginning or end of a sequence, or items that are different from their context, such as for example a letter in a number string (see Sikström, 2006, for a joint discussion of these three effects).

To be able to make edge-based generalizations, it is necessary to recognize edges, obviously, and these edges are marked by prosody. This has for example been shown in subjective click detection, where clicks are attracted to boundaries instantiated by prosody, rather than underlying structure (Chapin et al., 1972). In their experiment, Endress and Mehler (2009) implemented edges as 1 second silences[1], which may be seen as a prosodic cue, though artificial. As mentioned in Chapter 2, however, continuous speech does not contain many meaningful silences and word boundaries are marked by linguistic cues such as coarticulation and phonotactics. If word stress would be an especially early cue in segmentation, the role of word stress ought to be inferred from continuous speech before these other cues are learned. In this scenario the infant makes generalizations at the phrase level.

The idea that infants can use prosodic phrases to learn phonology, the lexicon or even syntax, can be summarized as the Prosodic Bootstrapping Hypothesis (Gleitman and Wanner, 1982). Whether or not the perception of phrase boundaries would be useful or sufficient to acquire syntactic relations is debated (e.g. Fernald and McRoberts, 1996) and whether or not the perception of prosodic phrase boundaries is sufficient to acquire word stress remains to be investigated. In any case, empirical evidence has shown that infants already before 6 months old are able to perceive phrase boundaries (Hirsh-Pasek et al., 1987; Christophe et al., 2001; Hawthorne and Gerken, 2014; Jusczyk et al., 1992; Soderstrom et al., 2003). Mandel et al. (1994) even found that infants of as young as 2 months old benefit from prosodic clause structure in their memory of spoken language. Gout et al. (2004) found that 10-month-old infants only recognize words they were familiarized with when these words do not straddle a prosodic boundary and Shukla et al. (2011) found that infants as young as 6 months old are able to simultaneously segment and map content words to visual objects if and only if these content words are aligned with a prosodic phrase boundary, as opposed to straddling this boundary. Additionally, Seidl and Johnson (2006) and Johnson et al. (2014) show that infants are better able to segment words which are aligned with the sentence edge than those which are placed in the middle of the sentence. In this study, there is no difference between infants' ability to segment sentence-initial and sentence-final words. In sum, and this list of empirical work is not exhaustive, infants are capable of recognizing prosodic boundaries, even when prosodic constituents are not separated by silences, and words aligned with these boundaries are easier to segment. Furthermore, the fact that adults make abstract generalizations of regularities at edges prompts the hypothesis that phrase edges would contain relevant information for acquisition. The current chapter investigates whether word stress regularities at phrase edges can be captured in the statistical relation between the stressed syllable and the edge.

[1] Endress and Bonatti (2007) used .25 second silences, but Endress and Mehler (2009) mention that in their experiments, pilots showed that this short silence was not enough for participants to learn the generalization.

7.3 Method

In Chapter 6, entropy indicated distinctiveness of sets of bigrams or trigrams. In the current chapter, the sets only include units from the edges of phrases. At the left- and right edge of the phrase, patterns between the stressed syllable and the phrase edge are part of the sets. This is shown in Example 7.2 and 7.3. First, Example 7.1 repeats the five-phrase language used in Example 6.5 in Chapter 6.

(7.1) 110
10
00010
010
100010

The entropy at the edge, now, is the index of probabilities of the distance of stress to the edge. So, each instance of a first (left edge) and last (right edge) stressed syllable together with the unstressed syllables that stand between this stressed syllable and edge are taken as a pattern. The pattern thus counts from the salient stressed syllable to the salient phrase boundary. The different patterns that occur together make up a probability distribution, and the entropy of these sets indicates their distinctiveness. Example 6.1 shows the patterns of the example language at the left edge, and the patterns at the right edge are given in Example 6.2.

(7.2)

pattern	entropy
|1	$-0.6 \log_2 0.6$
|01	$-0.2 \log_2 0.2$
|0001	$-0.2 \log_2 0.2$
$\sum$	1.37

conclusion left edge: potential initial-stress system

(7.3)

pattern	entropy
10|	$-1 \log_2 1$
$\sum$	0

conclusion right edge: perfect penultimate-stress system

The example language has a fixed right-edge-aligned trochee on the phrase level and the entropies of the set of phrase edge patterns indicate that system. The left edge has a high probability of initial stress and a lower probability of other patterns, which is an indication that initial stress is a good hypothesis on the stress system of the language. However, the right edge has a probability of 1 on penultimate stress, leading to an entropy of 0. The right edge is therefore more informative on the relation of stress to the boundary.

On the word level, the clash found in the first phrase of the language would be unexpected with the current analysis of possible stress patterns in the language. The fixed penultimate stress on the phrase level suggested that two consecutive stressed syllables are illegal, while in fact it occurs in the example language.

Besides discarding uninterpretable information found in the middle of the phrase, which was found to be a disadvantage in the ngram analysis, the boundary analysis has as its additional advantage that it does not need a prespecified window. A language with fixed antepenultimate stress, for example, would be difficult to be captured in an ngram system, because that system would need a window of n=4 to be able to figure out the relation of the stressed syllable to the right edge. In principle, of course, a window of n=4 would not be difficult, but a system which would have to analyse all possible ngrams to arrive at the optimal way to capture the metrical system is not efficient. With the analysis of the phrase boundary, no such presupposed windows are needed and a language with fixed antepenultimate stress would have an entropy of 0 at the right edge, just like the example language.

The boundary analysis makes three assumptions: 1) boundaries are perceptible, 2) syllables can be counted and 3) the difference between stressed and unstressed syllables is perceptible. On the basis of these elements, the distinctiveness at the phrase boundaries is captured by entropy, and the least most distinctive edge and the most probable pattern should surface. The analysis should show whether the continuous speech in the perceived stress corpus (as used in Chapter 6) is informative enough for the learner to infer the word stress system from patterns at the phrase boundary.

7.4 Results

Edge-aligned stress patterns are analyzed over the continuous input as transcribed by the participants in Section 6.5.1. The perception of stress by these participants is translated to a probability and these probabilities are the basis for the calculations. Entropy is given in Table 7.1.

Different conclusions can be drawn from Table 7.1. First of all, the overall entropy for Turkish is lower than the overall entropy for Dutch. This means that the relation of stress to the phrase boundary is more distinctive in Turkish than in Dutch, because there is less distributional variability at the edges. This was already known from the phonological system, but the current calculation shows the same while it is based on perceived stress. That this result surfaces from the distribution of perceived stress is rather interesting, because the Turkish transcribers of the corpus had a lower inter-transcriber reliability than the Dutch transcribers. In the current analysis, as opposed to the previous analysis, this lower entropy cannot be attributed to frequency differences of stressed and unstressed syllables, since the sets are not drawn from the entire distribution, but only from edge patterns. A high proportion of unstressed syllables would not influence the results in any way other than that it would increase the proportion of units with larger distances from stress to the phrase boundary.

Table 7.1: Stress-boundary information in Dutch and Turkish

Dutch	**Left edge**		**Right edge**	
	prob	pattern	prob	pattern
	0.329	|1	0.329	1|
	0.264	|01	0.311	10|
	0.167	|001	0.147	100|
Entropy	2.537		2.466	
Turkish	**Left edge**		**Right edge**	
	prob	pattern	prob	pattern
	0.455	|1	0.621	1|
	0.285	|01	0.156	10|
	0.124	|001	0.087	100|
Entropy	2.099		1.822	

In this case, entropy would be higher rather than lower.

The second result is the fact that for both languages, entropy of the right-edge-set is lower than that of the left edge. This means that a hypothetical learner relying on entropy would choose for the right edge to infer the relation of word stress to the word boundary. Indeed, in both languages, the phonological word stress system is right-edge-aligned. The (un)certainty of where word stress falls at the phrase level may be a reasonable indication of the alignment of stress on the word level. However, the fact that the results cannot be compared to a left-edge-aligned stress language makes them interesting as an indication, but premature as a conclusion.

Moreover, and this is the third result, in Dutch, entropy is only slightly lower in the right-edge-set than in that of the left edge. Since Dutch is known to be a statistical hybrid of initial and penultimate stress on the word level, the analysis is in a way a statistical confirmation of this hybridity on the phrase level. Indeed, it is not so clear which edge of the phrase is more indicative of the relation of stress to the word boundary. At the left edge, initial stress is the most probable pattern, and at the right edge, final and penultimate stress are approximately equally probable, with a slightly higher probability for final stress. Even the peninitial syllable has a high probability of being the first stressed syllable in the phrase. The phonological word stress system of Dutch, then, seems to be hard to infer from continuous speech, since initial, penultimate, final, and even peninitial stress are all good patterns according to the statistical probabilities.

For Turkish, the entropy in the right-edge-set is notably lower than in that of the left edge, and the final stress pattern is most probable. This shows that a hypothetical learner of Turkish may rely on phrase edges to infer the word-level stress pattern. It is interesting to see that in Turkish, too, the most probable stress pattern on the left edge is initial stress. Recalling, monosyllabic phrases have already been removed from the corpus, which leaves the interpretation that speakers mark both initial and final syllables with salience, or that the continuous corpus contains many phrases which start with stress-initial, potentially monosyllabic, words. The next section will discuss the results and this possibility.

7.5 Discussion

The analysis in the current chapter shows that information regarding the relation of word stress to the word-boundary is difficult to infer from the phrase level. Although the probability relation of perceived stress to the phrase boundary in Turkish seems to give a useful indication of the relation of word stress to the word boundary, this relation is much less apparent in Dutch. Strictly speaking, also, the differences between Dutch and Turkish are not sufficient to conclude that a useful language-specific pattern can be inferred from the phrase boundary. Even though the entropies in the Turkish sets are lower than those of Dutch, the patterns are too similar to draw conclusions. In both languages, the probabilities of initial and final stress are the highest probabilities in the distributions. An analysis of a language with left-edge-aligned stress is indispensable to be able to draw conclusions from the current analysis.

The results did show that an analysis of word stress in continuous speech based on perceived stress is more informative when considering edge-alignment than at an ngram level, as shown in Chapter 6. The ngram analysis has proven to be highly insufficient to learn anything about stress in a language, while in the boundary analysis probability differences in stress placement surfaced, as can be seen in the difference between final stress ($p = 0.621$) and penultimate stress ($p = 0.157$) in Turkish. To compare, the difference between iambs (relfreq $= 0.25$) and trochees (relfreq $= 0.22$) in that language was negligible.

The fact that in both languages initial and final perceived stress on the phrase level have the highest probabilities in the distributions may be related to different causes. First of all, it is possible that many phrases start and end with monosyllabic stressed words, something that can be checked in the corpus. In Turkish, 61 of 373 phrases, that is 16.4% start with a monosyllabic word and 40 of 373, that is 10.7% end with one. These numbers are too low to be identified as the source of phrase-initial and final stress, irrespective of whether these monosyllabic words would be perceived as stressed. In Dutch, in contrast, 270 of 364 phrases, i.e. 74.2% start with a monosyllabic word and 177 of 364, i.e. 48.6% end with one. These numbers are too high to serve as a sole explanation for the high proportion of perceived phrase-initial ($p = 0.327$) and phrase-final ($p = 0.328$) stress, meaning that a proportion of these monosyllables is in fact perceived as unstressed. An informal glance at the corpora in Appendix B.2.1 suggests that at the left edge this is likely: most of the Dutch phrases start with monosyllabic function words. However, phrase-finally they do not. The high proportion of monosyllabic words at phrase ends may therefore be identified as a reason of the high proportion of perceived phrase-final stress. The high proportion of phrase-initial function words is a likely reason for the high probability of peninitial stress in Dutch. The relatively high proportion of phrase-initial stress in Turkish, on the other hand, remains to be explained. Of course, the corpus, consisting of continuous speech, contains inflected and derived words rather than merely stem morphemes. The corpus therefore contains more words with non-final, potentially initial, stress than a lexicon would. Also, in polysyllabic words, word stress tends to move to both word boundaries, according to the 'Hammock' principle (van Zonneveld, 1985) and words with right-edge-aligned stress tend to have secondary stress at the left edge, as was also mentioned for Polish in Chapter 3. Speakers may move primary stress to the position of secondary stress and the phrase context likely has an influence on this phenomenon. Furthermore, the fact that the current analysis is based on listeners' perception of stress introduces the confounding factor of phonetic salience. Crucially, this was the intention of the analysis, since it was exactly the question whether perceived stress on the phrase level would be informative for the acquisition of word stress. The results of Turkish and Dutch suggest that it is not, and further analyses would need to be typologically broader, for a firm conclusion to be warranted.

Although the current attempt and that in Chapter 6 were based on a heavily dressed-down version of continuous speech, the results do suggest that mere knowledge of syllables, prominence patterns and phrase boundaries is not sufficient to learn a language's stress system or the relation of word stress to the word boundary. It is conceivable that more information is needed, such as first words. In this light, it becomes less surprising that infants do not display knowledge of the word stress system of a language before they presumably start building a proto-lexicon. Even though newborn infants of only a few hours old already perceive the difference between di- and trisyllabic utterances (Bijeljac-Babic et al., 1993), the difference between stress patterns on disyllabic and even trisyllabic utterances (Sansavini et al., 1997) and the pitch contour on their native languages' utterances, which they transpose to their cries (Mampe et al., 2009), they do not display a preference for listening to their native languages' metrical units before around 6 months of age, an age for which among other things a use of lexical subtraction with known words such as 'mommy' and the child's own name was found (Bortfeld et al., 2005). For the acquisition of delimitative patterns it seems to be a better strategy to pay attention to boundaries than to co-occurrence statistics, and a developing lexicon is therefore likely to accelerate this acquisition.

7.6 Conclusion of Part III

Chapter 6 and 7 reported on an attempt to capture statistical regularities of the word stress system in Turkish and Dutch from continuous speech. Because Turkish and Dutch are both positionally variable stress languages, and because the corpora used were unsegmented, contained no information except for stressed and unstressed syllables as well as phrase boundaries and were based on stress perception, the attempt was ambitious. The results reflected the difficulty of inferring a word stress system and the relation of word stress to the word boundary from perceived continuous speech in these languages. The analysis in Chapter 6 showed that inferring a word stress system from bigram co-occurrences is an ineffective strategy, on the one hand because boundaries in delimitative systems such as word stress are characterized by high co-occurrence statistics, rather than the statistical troughs which are supposedly segmentation cues in phonotactics. The only way in which a delimitative system, that is, a language's regularities at boundaries, would become noticeable is when its regularities can actually be related to a boundary. Still, in Chapter 6, the high number of phrase boundaries in spontaneous continuous speech does not make the stress system stand out. In Chapter 7, therefore, the question was asked whether the assumption that listeners pay attention to boundaries would improve the potential inference listeners may make of the language's stress pattern. In this analysis, the relation of stress to the phrase boundary reflected differences in stress systems between Turkish and Dutch, which is an improvement as compared to the analysis in Chapter 6, and it suggested that the final stress-system of Turkish is more inferable from continuous speech than the Dutch stress system, although an additional analysis of a left-edge-aligned stress language is necessary to draw conclusions. The difficulty of inferring a word stress system from continuous speech, despite making a 'search light' out of phrase boundaries, makes the relatively late use of word stress in infant language acquisition more understandable.

Part IV

Closing

CHAPTER 8

General conclusion

In this thesis the role of word stress in word segmentation was reassessed. Even though a great body of literature had already been devoted to the delimitative and salient properties of word stress, the evidence was inconclusive about the way in which word stress influences processing either universally or language-specifically.

With a non-word-spotting experiment, therefore, a unidirectional anticipatory language-specific use of word stress in word segmentation was tested. By analogy, the question was whether word stress as a beat prepared listeners for the segmentation of upcoming words. To this end, a trisyllabic nonce context with antepenultimate, penultimate or final stress preceded a previously learned target non-word. Additionally, there were two disyllabic target words with different stress patterns, which made it possible to investigate a potential interaction of stress in the anticipation of a target and the recognition of the target itself. The results of the experiment gave evidence for a language-specific role of word stress in segmentation: each language group had its own pattern of results and this pattern of results matched, for the greater part, facilitation of segmentation by native canonical stress. To compare, the results were modeled with predictors of universal boundary salience marking and it was found that these predictors were an inferior fit to the results as compared to the language-specific predictors.

The prior mini-typology based on left- or right-edge-alignment, peripherality and the fixed/variable nature of stress was designed with the idea that these three factors would potentially influence the use of stress in segmentation. The results showed that edge-alignment was indeed an important factor. Reliability, expressed in regularity and peripherality, however, did not straightforwardly prove its importance for segmentation. Not some, but all languages in the typology used stress in a language-specific way in processing. The anticipatory use of penultimate stress, found for both Dutch and Spanish, are, as far as I know, new in research on stress-based segmentation. The proximate and therefore less reliable nature of the cue makes it an interesting finding.

However, the language-specific patterns also raised questions. First of all, it was not strictly the case that languages with left-edge-aligned stress (in this case represented by Hungarian) experienced facilitation in target recognition while languages with right-edge-aligned stress only experienced facilitation in anticipation. The only left-edge-aligned stress language was Hungarian, and in this language only a trend of initial-stress target facilitation was found. The other three languages, Polish, Turkish and Dutch, all have right-edge-aligned stress and for Turkish a facilitative effect of canonical stress on the context (i.e. anticipation) was found alongside a facilitative effect of canonical stress on the target (i.e. recognition). In Dutch, there was an effect of context and a trend of a target effect and an interaction. Meanwhile, an important remaining question regarding edge-alignment in the first experiment concerns the results of Polish: why is there no anticipatory effect of stress in segmentation for this right-edge-aligned stress language? It was speculated that the inhibition of the recognition of the final-stress target buried any context effect. In the pattern of results it was seen that the penultimate-stress context had the lowest reaction times in the initial-stress target condition, but not in the final-stress condition. The interaction effect, however, was not significant and the question remains open.

Second of all, word stress did not only facilitate segmentation when it was peripheral or fixed, that is, when it was a highly reliable cue, but also when it was not reliable. For Dutch, a language with positionally variable proximate stress, an effect of facilitation was found for penultimate stress on the context. This pattern is the canonical stress pattern, but statistically it is not the predominant pattern in the language. This suggests that there is an abstract phonological role of canonical stress in segmentation although this interpretation would stand firmer if the same pattern of results was found for Polish, a language with fixed penultimate stress. The Turkish results, in turn, suggested a potentially gradual effect of facilitation, since the reaction times to penultimate stress on the context were intermediate between native canonical final stress and antepenultimate stress. This pattern of results prompted the inclusion of an additional language with positionally variable right-edge-aligned stress: Spanish. Since the pattern in Spanish is roughly a combination of penultimate and final stress, facilitation in stress-based word segmentation would supposedly give a clearer insight into the Dutch and Turkish patterns. Interestingly, the Spanish pattern raised a new question. Rather than patterning with either Turkish or Dutch, Spanish stress-based word segmentation showed a harmonizing interaction of context and target. Where the penultimate stress pattern facilitated the segmentation of an initial stress target, the final stress pattern facilitated the segmentation of a final stress target. This pattern of results is reminiscent of the evidence on metrical anticipation of stressed syllables, and of literature on facilitated processing by alternating rhythm. In that sense the Spanish pattern of results is not uninterpretable. However, the fact that the pattern is not found in (any of) the other four languages does raise questions about the generalizability of the interpretation.

While the Metronome experiment provided evidence for a language-specific unidirectional use of stress in segmentation, the unidirectionality of which was attested by the lack of interaction effects in all languages except Spanish, a language-specific effect of stress on target recognition was also found. This effect may be interpreted in two ways. First of all, the effect of stress may be a prelexical segmentation effect comparable to the context effect. That is, as a beat that marks the position of the word boundary. Second of all, the target effect of stress may be a lexical effect. For example, the target word is likely to be easier to store or access when the stress pattern is canonical rather than non-canonical. In the segmentation of targets, then, a lexical effect of stress may be a bidirectional effect, one in which stress is not a delimitative mark, but a part of the word form. The second experiment was therefore aimed at disentangling this potential bidirectional effect from the unidirectional effect that was targeted in the Metronome experiment. The offline experiment, giving listeners ample time to arrive at an optimal segmentation of pentasyllabic sequences, provided evidence for this bidirectional grouping effect of stress. Listeners divided the sequences according to language-specific stress preferences, as well as according to word-length preferences. These two strategies were deduced from the fact that the segmentations by listeners of Turkish and Dutch differed from each other, from a stress-independent segmentation of the stimuli according to their general biases and from their native lexicon. Furthermore, the segmentations displayed preferences for canonical metrical units and dispreferences for those non-canonical. The distribution of frequencies of different segmentations displayed a lenience of Dutch listeners as compared to Turkish listeners in the segmentation of the stimuli: the Turkish group had a higher agreement on which segmentation was preferred than the Dutch. It is likely that this difference in agreement is caused by the fact that Dutch has a higher degree of positional variability of stress than Turkish. Across languages a difference in agreement was observed between conditions, with the highest agreement on the segmentation of the clash condition. The latter finding is a strong check that metrical considerations guide segmentation.

The stimuli in the offline experiment also differed from the facilitation effects in the online experiment. Most markedly, there was an absolute preference for a word boundary in between two adjacent stressed syllables (clash), a condition for which no facilitation of segmentation was found in the Metronome experiment. This is evidence for a bidirectional use of stress in segmentation, which can also be observed in, for instance, a differing preference for a [wSw] word over a [Sw] word. In Dutch, this preference was less pronounced than the Turkish preference of [wwS] over [wS]. The segmentations furthermore showed a gradual preference of [wwS] > [wSw] > [Sww] in Turkish, while Dutch listeners did not absolutely prefer [wSw] over [wwS], but they dispreferred [Sww]. The Turkish pattern resembles the gradual facilitation in the Metronome experiment, while the Dutch results do not resemble the Dutch Metronome results and this difference in Dutch is arguably caused by the high frequency of [wwS] due to the clash condition.

The Optimization experiment, in sum, provided evidence for a language-specific grouping strategy in segmentation which was led back to preferences for canonical word stress patterns and which furthermore differed from the use of word stress in online segmentation.

The behavioral part of this thesis, then, provided evidence for a language-specific suprasegmental use of stress in segmentation. The stimuli of the two experiments were segmentally heteroform, except for the targets in the first experiment, which made the stress pattern simply a metrical beat to which the listeners attended. This beat provided listeners with language-specific expectations and segmentation preferences and when given more time, listeners used stress as a bidirectional grouping cue. At the same time, listeners' behavior raised new questions about the interaction of word boundary expectations and rhythmic harmony, at least in Spanish, and the way in which expectations may (as suggested by the Turkish results) or may not (as suggested by the Dutch results) be gradual. Furthermore, the apparent lack of anticipation in the Polish use of word stress in segmentation invites further research on this topic in fixed stress languages.

The distributional part of this thesis took a quick glance at the available information of prosodic patterns in continuous speech. The literature suggested a discrepancy between early attention to prosodic patterns in infants and the later acquisition of the language-specific regularities of these prosodic patterns. The role of word stress in infant word segmentation is paradoxical, since some knowledge of word boundaries seems to be necessary to grasp the relation of word stress to the word boundary. Bigram statistics which have in previous literature been found to be informative about the potential way phonotactics and word forms are acquired and, respectively, stored, were in the current research shown to be uninformative about language-specific stress patterns. The analysis of ngram statistics did serve as a reminder that salience, or distinctiveness, is a statistical concept and the statistics of distinctiveness are likely to be useful in language acquisition. It appeared, indeed, to be a better, more informative, strategy for acquisition to focus on the distributions of the relation of word stress to the phrase boundary, although the difference between Dutch and Turkish in these results showed that this strategy at least in some languages would remain a guessing game. Given that research has shown that listeners attend to boundaries in learning and abstraction, however, does indicate that this line of reasoning has more potential than the current analysis was able to provide evidence for. Follow-up research should at least include left-edge-aligned stress languages.

Together, the behavioral part and the distributional part of this thesis provided an account of the use of delimitative stress in processing and, indirectly, acquisition. A crucial difference with previous research is the typological breadth of the tested languages, the comparison of unidirectional and bidirectional segmentation and, most of all, the view of stress as a suprasegmental delimitative cue.

8.1 Suggestions for future research

Following up on pointers given in the above section, the current section suggests future research.

In the area of unidirectional processing, the lack of an anticipatory effect of stress in segmentation for Polish was surprising. Fixed penultimate stress should, in theory, be a reliable cue for boundary anticipation. A study specifically designed to exclude illegal targets in Polish may give insight into whether inhibition by non-canonical stress on the target caused the lack of an effect. To have a cross-linguistic design, which is recommendable, emulating the Polish study would ideally include a fixed peninitial stress language. Peninitial stress is likely to be a poor delimitative cue in processing, on the one hand because stress is proximate, although the results of Spanish and Dutch have shown that this is not a disqualifying factor for the use of stress in segmentation. On the other hand, however, the fact that it is proximate *left-edge-aligned* stress makes it a regressive cue in the strong sense of the word. Listeners can only use this stress in segmentation when they are practically halfway or even nearing the end of a word, depending on the word lengths in the language. This cognitive disadvantage may well be a reason for its typological markedness; very few languages have peninitial stress.

The interaction effect of context and target in the Metronome experiment in Spanish is intriguing. Final context stress facilitated the recognition of targets with final stress and penultimate context stress facilitated recognition of those with penultimate (i.e. initial) stress. The experimental design did not allow for an explanation based on syllable structure, but the stress system of the language does motivate research in that direction. Is the use of final stress on the context a sign that context is processed as a 'word' *without* a thematic vowel or is the use of penultimate stress on the context a sign of the opposite, i.e. that they are processed as words ending in a thematic vowel? An experiment testing both syllable structure and rhythmic alternation, possibly of influence on the current results, in Metronome segmentation could give an interesting insight in the interaction of general cognitive and language-specific mechanisms, as well as in the interaction of suprasegmental and segmental phonology in processing. This experiment, too, would not be complete without a reference language.

Even though the results of the distribution analysis suggested that it is difficult to acquire the word stress system of a language from continuous speech, the analysis of phrase-boundary patterns of Chapter 7 is not complete without a left-edge-aligned stress language. Turkish and Dutch radio corpora were analyzed, but these languages both have right-edge-aligned stress which is furthermore positionally variable. Hungarian, a language with fixed left-edge-aligned stress, would therefore be a good candidate. Furthermore, a similar analysis of Polish could serve as a type of baseline, in the sense that entropy for Polish phrase-boundary patterns is expected to be lowest. In Polish, namely, the right-most head of the phrase receives the phrase accent, which means that even at the phrase level, stress is mostly penultimate. Results on Polish would therefore put the results of Turkish and Dutch, having positionally variable stress, in perspective.

This list has an infinite expansion potential, since research is never 'done'. The corpora and data collected for the current research will therefore be made publicly available for further investigation. The perception of Spanish segments/syllables by Turkish and Dutch listeners, for instance, may be investigated judged by their orthographic transcription of the stimuli of the offline experiment in Chapter 5. Seven minutes of transcribed radio speech in Turkish and Dutch also contain more information than what was currently used, as does everything, really. My last research suggestion is therefore to look about with curiosity.

APPENDIX A

Stimuli Part II

All the instructions found in this appendix and the next are translated to English. In every experiment and survey, the participants received the instructions in their own native language.

A.1 Stress Judgment

Figure A.1 presents the instructions and the first four items of the stress judgment task. The square at the beginning of each line was a button which could be pressed to hear the item, the radio button above each syllable could be selected to indicate where stress was perceived.

A.2 Metronome

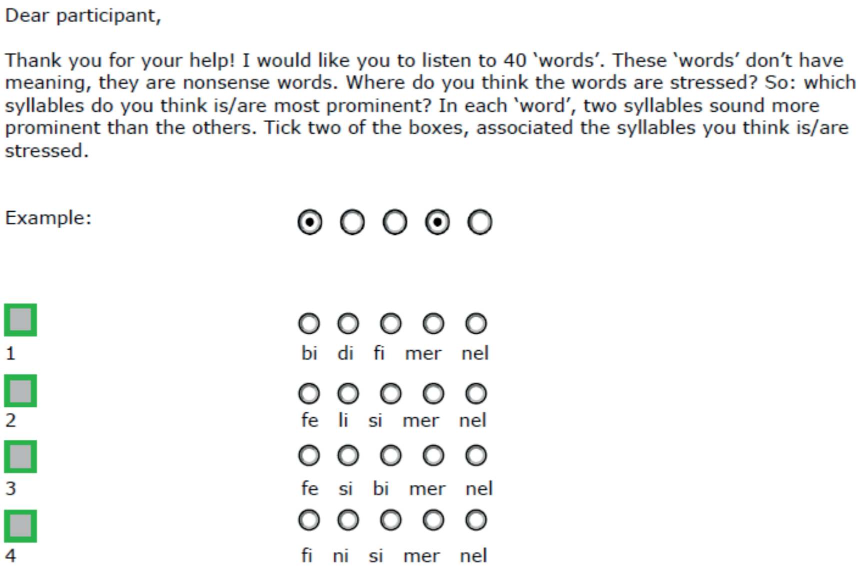
A.2.1 Instructions

Dear participant,
Thank you for participating in our experiment. You will see two animations of two different creatures. Try to remember what these creatures are called; you need their names in the rest of the experiment.

(Here, the participants go through an introduction in which they learn the names of the two creatures, as given in Figure A.2)

Do you remember?

Figure A.1: Part of stress judgment form



Dear participant,

Thank you for your help! I would like you to listen to 40 'words'. These 'words' don't have meaning, they are nonsense words. Where do you think the words are stressed? So: which syllables do you think is/are most prominent? In each 'word', two syllables sound more prominent than the others. Tick two of the boxes, associated the syllables you think is/are stressed.

Example:

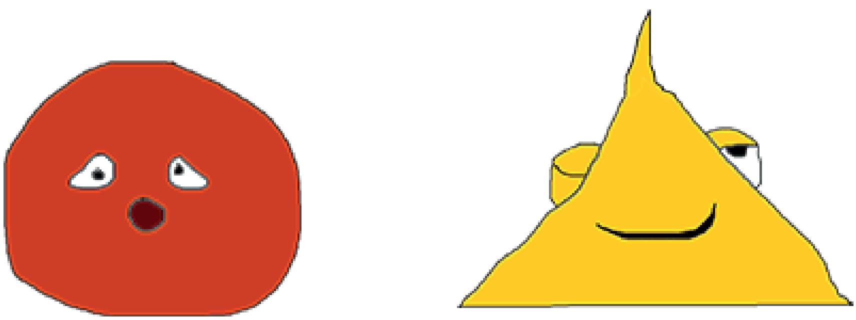
1 bi di fi mer nel

2 fe li si mer nel

3 fe si bi mer nel

4 fi ni si mer nel

Figure A.2: /darnam/ and /mernel/



We are about to check that. This time you will hear the names of the creatures. Press the button of the creature that goes with the name. Use the left button for the creature on the left, and the right button for the creature on the right.

(Here, the participants press the left and right buttons in response to the two names)

Okay, that was not so hard. Now, we are ready for the test. You will hear unknown words. Sometimes the name of one of the creatures is hidden in these words. If and when you hear one of the names, press any button **as quickly as possible**. After that, a screen will appear, in which you can choose the name of which creature you heard. If you pressed a button by accident, this is the time to correct it, by pressing the middle button. Let's practice first. Then the actual test will start.

(The participants go through a practice and a test phase. The test phase consists of four blocks, divided by self-paced breaks.)

A.2.2 Items

The items used in the metronome experiment can be found in Tables A.1, A.2, and A.3.

A.2.3 Questionnaire

Thank you for filling out this questionnaire.

All information provided in this questionnaire is confidential and will be used anonymously for scientific purposes only.

1) Name:. . . . .

2) Age:. . . . .

3) Sex:. . . . .

4) Availability (in case it is necessary to consult you about the information provided):

e-mail:. . . . .

or tel.. . . . .

5) What is your native language (i.e. the language you learned first and which feels most natural to you)?

Table A.1: Test items

Front harmony	**Back harmony**
bidifimernel	badusudarnam
felisimernel	bafududarnam
fesibimernel	bafasudarnam
finisimernel	budusudarnam
firidimernel	dunufudarnam
fisidimernel	dusufudarnam
lifisimernel	fulusudarnam
lirefimernel	lufusudarnam
nedifimernel	lusufudarnam
nefidimernel	lusafudarnam
nesefimernel	nalusudarnam
niferimernel	narufudarnam
nirefimernel	nafasudarnam
refidimernel	nufududarnam
resebimernel	ralufudarnam
ridifemernel	ranududarnam
rinibimernel	rasududarnam
rifebimernel	runududarnam
rinefimernel	runafudarnam
siredimernel	safududarnam

Table A.2: Fillers containing targets

Position target	**Front harmony**	**Back harmony**
Syllable 3 and 4	sefemernelbe	faludarnamdu
	dinimernelse	subadarnamsa
	biremernelre	lufudarnamba
	nebimernelni	duludarnamdu
	rerimernelri	sabadarnambu
	risemernelbe	lusadarnamna
	sedimernelse	rusudarnamdu
	firimernelfi	sabudarnamru
	sirimernelri	fanadarnamdu
	fefemernelsi	sasudarnambu
	belimernelsi	buludarnamfa
	senemernelfi	dubudarnamfa
	riremernelfe	fubudarnamlu
	benimernelne	rurudarnamna
	dirimernelli	ranadarnamlu
	seremernelri	duradarnamdu
	sinimernelne	fanadarnamru
	firimernelri	dunudarnamna
	bebimernelfi	burudarnamnu
	sibemernelfe	nadudarnamba
Syllable 2 and 3	nemernelnibi	budarnamnaba
	semernelbere	sadarnambasu
	simernelfibi	fudarnamlunu
	nimernelreni	nadarnamnasu
	nemernelnine	ludarnamnalu
	simernelrine	radarnamdusa
	simernelfebe	sudarnamfuna
	femernelbisi	badarnamfana
	bimernelsise	radarnamduru
	fimernelbebi	budarnamrusu
	simernelbesi	nadarnamfasa
	bemernelfibe	radarnamrusa
	fimernelsisi	badarnambulu
	semernelsene	fudarnamsusa
	semernelnibe	dudarnamsasu
	femernellire	fudarnambufu
	nimernelline	fadarnambuba
	limernelsine	dudarnamranu
	dimernelrene	dudarnamsubu
	bimernelsibe	fudarnamduru

Table A.3: Targetless fillers

Front harmony		**Back harmony**	
biresirefi	feneberefi	nanusabufu	babarabura
serenirili	sefifenedi	rufafubaru	nasunasuna
silirisifi	bininibefi	duraruruRu	barasudusu
refilirebe	nenifebidi	lusanaradu	fubanufalu
didibifebi	difisesire	sulurusulu	sunalubana
dibiselife	sinibebifi	narufusuna	sabulufusu
sirenilise	bebesirisi	dusanufuna	nunadusaba
rinenesibe	sibifiriri	budubafudu	safubususu
fesinefise	finerisefi	bururunuba	bubafalura
lirebefebe	lifesefeli	sufudunasu	ranadulusu
ninebibefi	seribibefi	lulufusara	lufarulubu
dinibifibe	nelibesene	basubabaru	farabafusu
ninirefebe	liniserili	dulululudu	fusurubasa
dibilirili	direbireri	dunadufuru	bubulufalu
dibelisene	siliberibi	susufanura	sasabufabu
bebesisili	befibeberi	rasabaduna	rusabababu
serefibibe	sifenesebi	sufanafabu	nulududura
febibiseli	finesefeli	nasuranubu	basafufafa
ribibenebi	nibenefili	buraburasu	bufafabasa
bibelisesi	silifesebi	susurubura	dufusafusu
benefifefi	sinisenili	bufafuluba	sabadurana
rerereniri	sebinisefe	burusanabu	sulubanana
sirefereni	nefibifese	surusunulu	sabarudunu
felisebiri	rirefefibe	nadububufu	sarafarana
lifesebere	bifenesife	rabubafaba	bufafanana
firisefini	silisilise	sunafunubu	rasasunudu
nififibire	disiberiri	suburufaba	bubufufasa
ririferidi	finefibebe	sarubusasa	rufafarafu
filirisese	relinilise	rarurunudu	basunabulu
bilisedifi	rereferedi	safurudufa	fubasuraba
linineberi	renesifiri	fananaluba	raluraluba
binebesiri	selisisebe	lufusubulu	nurunulufu
referidise	rifelinidi	narurafufa	rurabarunu
fifireneni	ririserisi	bulusarubu	dubururuba
neribinifi	bisinefene	ralusanafu	lunafusadu
besiniseri	beliberine	nusubudubu	fulusubunu
nenerinene	benisenene	sabafunusu	dufalubulu
nefinelifi	bedilifefi	furafubabu	ranubufasu
risedifibi	libesisifi	lufasusara	barabadubu
nerirerere	selibifese	fafubasara	nurafabadu

. . . .

6) Which region of your country did you grow up in? . . . .

7) Do the people in your region, including you, speak a dialect that is different from the standard language used in your country? If yes, please name it
. . . .

8) What is the native language of your parents?

mother. . . . .

father. . . . .

9) Which countries have you lived in during your lifetime (please indicate the corresponding years)? . . . .

10) Do you speak any other languages? If yes, indicate them in the order learned
. . . .

11) How old were you when you started to learn these other languages? Keep the same order as above
. . . .

12) What do you study?
. . . .

13) What was your language of study (i.e. what language was used for teaching your non-language lessons (e.g. mathematics, biology, religion etc.) at:

primary school . . . .

secondary school . . . .

university. . . . .

14) Please evaluate your linguistic competence in your second language (if any) (choose one of the given options):

Speaking	Near-native fluent good average bad
Comprehension	Near-native fluent good average bad
Reading	Near-native fluent good average bad
Writing	Near-native fluent good average bad
Vocabulary	Near-native fluent good average bad
Grammar	Near-native fluent good average bad
Accent	Near-native fluent good average bad
Fluency	Near-native fluent good average bad

15) What languages do you speak…

with close family (parents, siblings, grandparents)… ……………

with friends, colleagues………………………

to yourself (when you're angry, when you calculate or when you dream) ……

16) Do you have dyslexia, or any speech or hearing disorder?…

Thank you very much.

A.3 Optimization

A.3.1 Instructions

Thank you for participating in this experiment!

What is your age?

Are you male or female?

Is …[1] your native language?

What language do you speak best - after …[2]

You will hear a number of sentences, in a language you don't know.Listen carefully and then type what you think the sentence was. Separate the words by spaces. Then you will hear the sentence again. You will be asked, if you want to change your answer. You can say yes (y) or no (n). You can also wait with your answer, to take a short break. This you can do as often as you want to.

This is an example of an item:

You hear: baraligopuredo

and you type, for example: bara ligopu re do

There are no rules for how this should be done and nothing is right or wrong. It's only about what you think it sounded like.

Do you understand, until now?

[1] Turkish or Dutch.
[2] See above.

Type ’y’ to continue:

There are three blocks and in total about 50 items. We will start with three practice items

Are you ready to start?

(The participant gets three practice items)

These were the practice items.

If there is still something unclear to you, ask the experimenter now.

A.3.2 Test items

The test items of the optimization experiment can be found in Table A.3.

APPENDIX B

Stimuli Part III

All the instructions found in this appendix are translated to English. In every experiment and survey, the participants received the instructions in their own native language.

B.1 The North Wind and the Sun

B.1.1 Hungarian

Orthographic

The text below is the orthographic transcription from Szende (1994).

Egyszer az északi szél és a nap vetélkedtek, hogy melyikük az erösebb. Épp arra jött egy vándor, vastag köpönyegbe burkolódzva.
Az északi szél és a nap nyomban megegyeztek, hogy az lesz a gyöztes, aki hamarabb rábírja a vándort, hogy levegye a köpönyegét.
Akkor az északi szél elkezdett süvölteni, ahogy cask bírt. De a vendor annál szorosabban vonta maga köré a köpenyt, minél erösebben fújt.
Így aztán az északi szél el is veszítette a versenyt. A nap meg elkezdte ontani tüzö sugarait, mire a vendor egyszeriben kibújt a köpöjyegéböl.
Az északi szél tehát kénytelen volt megadni, hogy bizony a nap az erösebb.

Stress distribution

The stress distribution is based on the prosodic transcription from Szende (1994).

Phrasing corresponds to periods (.). The transcription originally had 3 stress levels. In the stress distribution, each stress level above 1 has been changed to 1 (wherever stress is perceived, it is now transcribed as ‘stress’). Hard returns are phrase boundaries, the distribution does not contain word boundaries.

1001001001100000100010010000101010000000
01001001101000001000000100100010010001000
000100110010001001001010100000010000101010001
1000100010100000100101001001010010010100010010000
010011000100010010100

B.1.2 Turkish

Orthographic

The text below is the orthographic transcription from Zimmer and Orgun (1992).

Poyrazla güneş, birbirlerinden daha kuvvetli olduklarını ileri sürerek iddialaşıyorlardı. Derken, kalın bir palot giymiş bir yolcu gördüler.
Bu yolcuya paltosunu çıkarttırabilenin daha kuvvetli olduğunu Kabul etmeye kara verdiler.
Poyraz, var gücüyle esmeye başladı. Ancak, yolcu paltosuna gitgide daha sıkı sarınıyordu.
Sonunda poyraz uğraşmaktan vazgeçti. Bu sefer güneş açti; ortalık ısınınca yolcu paltosunu hemen çıkardı.
Böylece poyraz, güneşin kendisinden daha kuvvetli olduğunu Kabul etmeye mecbur kaldı.

Stress distribution

The stress distribution is based on the prosodic transcription from Zimmer and Orgun (1992).

Phrasing corresponds to periods (.). The transcription originally had 3 stress levels. In the stress distribution, each stress level above 1 has been changed to 1 (wherever stress is perceived, it is now transcribed as ‘stress’). Hard returns are phrase boundaries, the distribution does not contain word boundaries.

010010000101001000010010100001000100111001101101
1001100101000010100100010100101001

01101000100110011001100010100100
10001000110010101010010010001100001001
100010011001010010001010010101

B.1.3 Polish

Orthographic

The text below is the orthographic transcription from Jassem (2003).

Pewnego razu Północny Wiatr i Słońce sprzeczali się, kto z nich jest silniejszy.
Właśnie przechodził drogą jakiś człowiek owinięty w cieply płaszcz.
Umówili się więc, że ten z nich, który pierwszy zmusi przechodzącego, aby zdjął okrycie, będzie uważany za silniejszego.
Północny Wiatr zaczął od razu dąć z całej siły, ale im więcej dął, tym silniej podróżny otulał się w płaszcz.
Wreszcie Północny Wiatr dał spokój. Wtedy Słońce zaczęło przygrzewać,
a w chwilę później podróżny zdjął płaszcz.
W ten sposób Północny Wiatr musiał przyznać, że Słońce jest silniejsze od niego.

Stress distribution

The stress distribution is based on the prosodic transcription from Jassem (2003).

Phrasing corresponds to periods (.). The transcription originally had 5 stress levels. In the stress distribution, each stress level above 1 has been changed to 1 (wherever stress is perceived, it is now transcribed as 'stress'). Hard returns are phrase boundaries, the distribution does not contain word boundaries.

0101001010100100100010100101000100010101
10100100010100000001000101010001000010
010110010110100010101001001001
10010101010100100100101001001
110010110100100010010

B.1.4 Dutch

Orthographic

The text below is the orthographic transcription from Gussenhoven (1992).

De Noordenwind en de zon hadden een discussie over de vraag wie van hun tweeën de sterkste was, toen er juist iemand voorbij kwam die een dikke, warme jas aanhad.

Ze spraken af dat wie de voorbijganger ertoe zou krijgen zijn jas uit te trekken de sterkste zou zijn.
De noordenwind begon uit alle macht te blazen, maar hoe harder hij blies, des te dichter de voorbijganger zijn jas om zich heen trok.
Tenslotte gaf de noordenwind het maar op. Vervolgens begon de zon krachtig te stralen, en onmiddellijk daarop trok de voorbijganger zijn jas uit.
De noordenwind kon toen slechts beamen dat de zon de sterkste was.

Stress distribution

The stress distribution is based on the prosodic transcription from Gussenhoven (1992).

Phrasing corresponds to periods (.). Wherever stress is perceived, it is now transcribed as 'stress'. Hard returns are phrase boundaries, the distribution does not contain word boundaries.

0100000100000100000100000001000000100010001010100
0001000000000001001000001 0000
01000001000100010010010000000010010
0100000000101000011001000100000000000010
010000001000101000

B.2 Radio Speech

B.2.1 Orthographic transcription

Literal orthographic transcriptions of the radio fragments. Interjections such as 'eh' are not interpreted as syllables and were therefore not included in the forms that were given to the participant nor in the analysis.

Dutch

Fragment 1: BNR, Budget cuts army

(Speaker 1)
nou kijk
eh stel je nou voor dat je
defensie zou op-
opheffen he
gewoon dat gedachte-experiment hoewel we
er
als ik echt cynisch ben
eh niet ver van afzitten zo langzamerhand
eh met eh

alles wat er eh
gaat komen
dan moet je toch constateren dat een hele hoop eh
taken die defensie op dit ogenblik eh
doet
eh gewoon over moeten gaan naar een nieuw ministerie
bijvoorbeeld eh
de grensbewaking
dat door eh
de marechaussee wordt uitgevoerd
op bijvoorbeeld schiphol
dat kost driehonderzeventigmiljoen euro
nou
dat moet worden eh
worden overgeheveld
kustwachttaken
eh de luchtverdediging
eh van nederland
eh we hebben laatst nog zo'n ehm
akkefietje gehad
met een vliegtuig dat
ongeoorloofd het nederlandse luchtruim eh
binnen zou komen eh
zetten
nou
dan eh
moeten straaljagers de lucht in
om dat eh
vliegtuig eh
te onderscheppen
te kijken of dat eh
bijvoorbeeld niet een
een soort aanslag is a la elf eh
september
de eh
elke dag
nou
niet
elke dag
maar bijna elke week
is de explosievenopruimingsdienst eh
wel in het
ehm in het nieuws
dat
kost

ja
dat is niet heel duur
maar dat kost twintigmiljoen euro
die taak zou je gewoon echt moeten eh
moeten blijven uit eh
voeren
ehm zandzakken vullen
eh als dijken doorbreken
ja
dat doet op dit ogenblik eh
defensie

Fragment 2: BNR, Malnutrition elderly

(Speaker 1)
eh vertel
wat is dat precies voor methode
(Speaker 2)
ja we zijn
we
we zijn wel een beetje aan het klagen
maar in Nederland doen we het natuurlijk wel eh
hartstikkeeh
goed he
want weeh
hebben als eh
stuurgroep ondervoeding
samen met eh
allerlei andereeh
groepen
eh simpele screeningslijsten geinfre- eh
geintroduceerd zo-
zodat je in allerlei sectoren
met een paar simpele vragen
de eh
de risicomensen eruit kan halen
en dan komt natuurlijk de
de dietist en de he-
de zorgprofessional aan bod
om die mensen dan in zeer korte tijd
te zorgen
dat ze weer op voldoende eiwit en energie komen
(Speaker 1)
ja

en heeft dat ook effect op
eh een een
ander punt wat eh eh eh
ter discussie staat
dat heel veel mensen
niet goed uit een operatie komen of althans niet goed herstellen
omdat ze last hebben van delier
ze zijn verward
(Speaker 2)
ja ja
dat is zo
dat het natuurlijk eh
ook allerlei geestelijke factoren is
als je ondervoedt raakt dan ra-
raak je in een soort eh
ja ve-
verwaarloosde toestand
en dan gaat ook dat
de
de
geestelijke
toestand eh
achteruit
en dat is wat we proberen te zeggen
je moet voorkomen dat mensen in zo'n negatieve gezondheisspiraal komen
waar allerlei klachten tegelijkertijd eh eh
naar voren komen
en wat eh eh
hè
de mijn eh
colle-
de chi- chirurg aan de telefoon ook zegt
dan valt 't kaartenhuis eh
in
dus je moet vroegtijdig eh

Fragment 3: BNN, Nostalgia

(Speaker 1)
van de andere kant
eh eh zeker eh
hè in retrospect
eh vinden mensen toch wel eh
duidelijk ook een

plezierige kant eraan
ik
ik vind dat zo wonderlijk
je ziet dat zelfs in
in Oost-Duitsland hè
waar de mensen na de val van de muur
toch ook een enorme nostalg-
nostalgische herinneringen hadden en
en
terwijl het eigenlijk vreselijk was
maar van de andere kant
de sociale binding
de sociale cohesie da
dat is iets waar men sterk naar terugverlangt
(Speaker 2)
ja want dat is ook waarom het
waarom het zo in is
dat mensen opeens zoiets hebben van
ja de jaren vijftig
ik
ik weet nu al
dat wordt een enorme hit dat programma niet omdat ik hier
bij omroep Max zit
maar je zi-
je ziet het
ik werd er zelf ook eigenlijk wel vrolijk van toen ik het net even
langs zag komen ik weet niet wat er dan gebeurt
maar je ziet iets
wat je nooit meer ziet
en je wordt er toch een beetje blij van
klopt dat
(Speaker 1)
ja
't 't
nostalgie
dat heeft eh
een
daar komen we steeds meer achter
een een
heel bijzondere functie dat eh
of je nu ehm
nou ja
je eenzaam voelt
of of of wat
wat depri-achtig

of wat eh
wat voor negatieve gesteldheid ook
van nostalgie daar word je
vrolijk van
dat
helpt je om je om te schakelen naar een meer
positieve kijk op het leven

Fragment 4: EO, European Economics

(Speaker 1)
dus dat moeten we niet doen
maar dan hebben we
nog wel een klein probleempje
namelijk met Europa
we hebben namelijk afgesproken
drie procent eh
en
en niet meer en
als u
als we niet bezuinigen
krijgen we een tekort van viereneenhalf procent
(Speakers speak simultaneously)
(Speaker 2)
we hebben met Europa afgesproken
dat het de bedoeling is
dat je je aan die drie procent houdt
en als er nou sprake is van uitzonderlijke omstandigheden
dan kun je dus
je wenden tot de Europese Commissie
(Speakers speak simultaneously)
(Speaker 2)
Nederland en Griekenland in de economie
hè dus
dus eh
daar is echt
hebben ze een behoorlijke zooi van gemaakt
en hier in Nederland
de rente is zo laag als maar kan
we krijgen bijna geld toe
ehm als we geld lenen
dus ehm
(Speaker 1)
geluiden uit Brussel komen

die zeiden
ja Nederland eh
allemaal leuk voor Nederland
maar Nederland moet zich gewoon aan die drie procent houden
net als alle andere landen
dus
ik
ik proef daar nog niet echt veel ruimte om
om
om een beetje coulant te zijn voor Nederland
(Speaker 2)
nee maar goed dat moeten Mark Rutte en
en
en De Jager die deze situatie oaakt hebben
moeten ze da daar maar eens gaan uitleggen
kijk er
er kwam net een rapport van de Europese Commissie
ik heb het hier eh
voor me liggen
die dus
eh analyseert
hoe dat
hoe het kan
dat alle andere landen
hier heb ik het
dat is van eh
februari
hoe alle andere landen in de noordelijke
eh regio van Europa
helemaal geen problemen hebben om
om zich aan die grens te houden

Fragment 5: KRO, Interview with writer

(Speaker 1)
ja nee ja ik eh
fantastisch natuurlijk
(Speaker 2)
goed geantwoord
(Speaker 1)
ja
goed he
(Speaker 2)
nou eh

ik las jouw boek enneh
ik was gelijk verkocht
en ik zette jou op mijn lijstje
eh eh uit te nodigen gasten
en ik raakte vervolgens dat lijstje kwijt
en eh
nou
toen verscheen
eh wat is het
eind vorig jaar
of is het dit jaar

(Speaker 1)
eh eind vorig jaar

(Speaker 2)
toen verscheen eh
haar eerste
roman
eh half mens
nou die las ik weer
en toen eh
hè
werd ook weer gelijk eh
genomineerd
voor allerlei prijzen
voor de BNR

(Speaker 1)
ja
voor de BNG prijs
dat is voor onbekend talent
ja
en eh
voor de Opzij literatuurprijs
en voor de
a- a- op de AKO eh
tiplijst
heet dat volgens mij

(Speaker 2)
ja precies
nou goed
eh de om maar eens even aan te geven
nou eh
dus ik
eindelijk eh
durfde ik je uit te nodigen
hè

dacht ik
oh ja
Maartje Wortel
nou
en ze vond het prima
en nou zit ze hier
in de auto van eh
van paps of mams
(Speaker 1)
ja van
van allebei
want we hebben maar één auto
(Speaker 2)
en dat is Eemnes
geboren getogen
(Speaker 1)
en ik reed hier net een stuk of vijf rondjes

Fragment 6: NCRV, Dutch politics

(Speaker 1)
er is een eh
vreselijk geval
eh eh vreselijk incident
en we duiken er met z'n allen op
om te kijken wat er nou gebeurd is
en 't is allemaal halve informatie die we hebben
eh normaal gesproken moeten dit soort dingen
moet het o- o-
Openbaar Ministerie moet dit gaan onderzoeken
moet kijken wat er aan de hand is
(Speakers speak simultaneously)
(Speaker 1)
ja natuurlijk
maar we moeten niet met z'n allen gaan zitten speculeren
hoe het nou met die jongen zit
en dan in de kamer ook nog
(unintelligible)
op te
op dat individu ingaan
ik dacht nog iets anders
toen wij de
TBS-commissie
afsloten

hebben we dat gedaan met een
belangrijke conclusie
dat het in de jeugd veel erger was
(Speakers speak simultaneously)
ja eh
dat die
er wordt nu z- geze-
steeds gezegd jeugd-TBS hè
nou er is geen jeugd-TBS
dat is g-
klinkt natuurlijk lekker op het moment
er is wel een bepaalde maatregel die erop lijkt
(Speakers speak simultaneously)
de PIJ-maatregel
en wij concludeerden eigenlijk
gedurende de rit
en hebben we toen ook tegen de kamer gezegd
dat daar een heleboel dingen in zeiden
waar de kamer zich op zou moeten storten
en de kamer heeft toen ook gezegd
hè
was ook alweer eh
zes jaar geleden volgens mij
dat gaan we doen
en ze hebben het nooit gedaan
dus als ik
deze mevrouw hoor
en een paar anderen denk ik
oh
had daar nou nog een paar jaar geleden
gedaan wat je zou doen
was er toen in gedoken
en ga nu niet
weer op een incident duiken
op één geval
hou ook
als kamer ook distantie en afstand
om het geheel te bekijken
(Speaker 2)
ja goed
tegelijkertijd
geef je zelf al aan
het was toen al mis
dus misschien
moet dit dan

maar de aanleiding zijn
(Speakers speak simultaneously)
(Speaker 1)
maar laat even
laat
de
eh eh
laat hier onderzoeken wat er werkelijk gebeurd is
door de instanties die er wel over gaan
(Speaker 2)
ja

Fragment 7: NTR, Interview cartoonist

(Speaker 1)
op dat moment dacht ik
er nog niet aan dat dat
na vijf jaar al gemeengoed was
en ik heb het zelf ook wel eens meegemaakt
dat ik in
bij een begraaf-
dus dat mijn telefoon aanstond
maar iemand had mij al een keer gezegd van
dit wordt gemeengoed
het is nou nog
een prent om
om te lachen
maar straks
begrijpt
(Speakers speak simultaneously)
nee
nee
hij is gewoon
dus de waarheid kan het
ook je
je
ideeën
die nu leuk zijn
heel snel inhalen
en daarom heb ik ook
kijk daar staan dan die politiek getinte tekeningen van de Gelderlander eh in
daarom heb ik ook een
de tekeningen gebruikt uit de Volkskrant

want door de and-
de andere eh
literaire tekeningen daar ben ik nog
over aan het nadenken
wat voor een boekvorm ik dat wil gaan uitgeven
want het zijn zevenhonderd tekeningen
waar je straks eh
uit kunt eh
kunt eh
eh god hoe zeg je dat
(Speakers speak simultaenously)
ja
eh en en daar staan
de
de tekeningen in
over de
de e-readers en
en ook dat
daar zijn we straks zo
weer zo aan gewend
dat de grappen die daar in st-
staan
nu nog wel leuk zijn

Turkish

Fragment 1: TRT, Diet of Kazakistan

(Speaker 1)
etli
böyle bol etli
bizim zaten
Kazakistan bölgesinde
çok
mal
şey tarımcılığıyla uğraştığı için
et
çok et tüketilir
hayvancı
evet
havancılıkta çok
di mi
et tüketimi
bir de
soğuk memleketlerde zaten

et tüketimi de
fazla oluyor
yani yağlı et
yeniyor
ve hayvancılık
yoğun oluyor
ondan da zannediyorum herhalde
etli yemekler
revaçta
nasıl Türkiye'de
çok
sebze meyve çok tüketilir
ve ekmek tüketilirse
bizde de orda etsiz yani
sofra oturduğunda
nasıl Türklerde
ekmek olmazsa
oturmazsa
bizde de
et olmazsa
oturmayız
memleketinizde
hiç bir şey olmasada
bizim memleketimizde mesela
bir iki yeri
yani
hatta bir tanesinin
ekmek kültürü yoktu
hangisiydi şimdi
hatırlayamıyorum tam
bir ülkeyi
bizim coğrafyamızda işledikte
bu kadar çok ülke işledik ki
onlardan bir tanesinde
ekmek kültürü yok mesela
yani
ekmek bizde diyor yenmez
ancak
bazı pastaneler var
bu pasta türü falan olur onlarda
efendime söyleyim
hani
yemeğin haricinde
yenilen şeyler
yoksa yemeğin yanında

Fragment 2: TRT Antalya, Soccer team finances

(Speaker 1)
takım
ligi
zannedersem
onikinci onüçüncü bitirdi
yine
ikinciyle
sıkıntıyla başladılar
bu bir gerçek
görecekler
ama ilk yıl
çok büyük
harcamalar yapıldı
şampiyonluk hedefiyle girildi
hadi diyelim
beşinci olsun
ama onikinci bitirdi
ya da onbirinci onüçüncü
yani
onbirinciyle onüçüncülük arasında bitirdi
tamam
ikinci yıl haklılar
büyük bi mahkeme süreci
kaosla girdiler
daha sonra mahkeme süreci işte
aralarında ne geçti bilmiyorum
Cemal Aydın ve Melih Gökçek
Cemal Aydın
açtırdığı söyleniyor
mahkemeleri kamuoyunda
kendi açmadı ama
Onun
desteklediği kişilerin açtığı söyleniyor
delegelerden
yani delegeler
delegelerde Cemal Aydın'la olduğu biliniyor
bu süreçte
sıkıntılar yaşadılar
mahkeme kararı verdi
tekrar
Melih Gökçek'te
me
pardon

Melih Gökçek değil
Ahmet Gökçek'in başkanlığı
yok
hükmünde sayıldı
bu arada yapılan
bu arada
o arada
kulübün
kırk trilyona yakın
bir geliri olmuştu
ve
seksen trilyonluk da
bir harcama olmuş
yüz yirmi trilyon
luk
yüz trilyona yakın
bir borçla bırakıldı
Ankara gücü

Fragment 3: TGRT FM, Death penalty

(Speaker 1)
idam cezasının kaldırılmasına
karşı olan
çevreler var
idam cezasının
tekrar konması için
görüşmeyen eden
çevreler var
efendim
ama
tabi bunun
politik ve hukuki
şeyleri var
evet
yani
dediğim gibi
hemen
bugünden yarına
çözüme bağlanacak
bir konu değil
fakat
sadece bizde değil
avrupanın da

bazı ülkelerinde
zaman zaman
bu türden
tartışmalar oluyor
yani
bazı ülkelerde
mesela İngiltere'de
yeniden idam cezalarının
konulmasına dair
dönem dönem
gerek siyasilerden
gerek vatandaşlardan
görüşler
ve çıkışlar
oluyor
evet
yani aslında
baktığınız zaman
bunun
çok da şaşırtıcı
bir iddiası yok
alenen
dünyanın
birçok ülkesinde
idam cezası uygulanıyor
evet
Amerika gibi
Federal

Fragment 4: Haberturk, Turkish Politics

(Speaker 1)
ihbar kavgasının
Türkiye'ye
empoze edilmesinin
(Speaker 2)
ama dediniz ya hocam
iç
siyasal hesaplar üzerinden
bir gerginliği
görüyor
sayın cumhurbaşkanı Gül
böylesiyle müdahale ediyor
sayın

cumhurbaşkanı Gül'ün
bu anlamda
bazı
siyasi hesapları
var mı acaba
ya
o yüzden mi müdahale
(Speakers speak simultaneously)
(Speaker 1)
en
fazla
isteyeceği
nihai nokta
devlet
cumhurbaşkanlığı'dır
o zaten cumhur
ya
olsa da
olmasa da
bir kere herkesin
aklından geçirdiği
en yüksek
mertebe olarak
düşündüğü yere
o zaten
varmış
böyle bir
aynı zamanda da
ama sonrasında
(Speakers speak simultaneously)
hayır aynı zamanda da
başka gene
hukuk ve anayasa
konuşmadığımızı
söyledik
benim bildiğim
yürütmenin başıdır
cumhurbaşkanı
evet
geçen şu bakan
bakanlar kurulunun
yemidir tabi canım
(Speakers speak simultaneously)
(Speaker 2)
başbakan da söyledi

(Speakers speak simultaneously)
(Speaker 1)
böyle bir şey değil
o günkü

Fragment 5: Alem FM, Feminist debate

(Speaker 1)
benden ya
yani başka bir insan
benden
rahatsızlık
duyduğunu
gördüm
buna şahit oldum
(Speaker 2)
doğru
aynen öyledir
(Speaker 1)
böyle bir durumda
hele sonuçta
diyor ki mesela
bu
bir
cinsellik ögesidir sadece
kare olsaydı
e kare olsaydı
affedersiniz
sizin de kaba etiniz
kare olsaydı
oturamazdınız
oraya
o özel
bir
üretim
yani sonuçta
(Speaker 2)
hayır
Zaten
o kareye
dikdörtgene falan
geldin mi
söylenecek çok şey var
tabi radyoda söyleyemiyoruz onu

(Speakers speak simultaneously)
o kareyi
doğru söylüyorsun
hani
oturduğun yerde
insanın iki tane
yuvarlak
organı var
biliyorsun
bir tanesi
başı beyni
bir tanesi de
oturduğu yer
pek ikisinin
yer değiştirmemesi lazım
aslında
(Speakers speak simultaneously)
(Speaker 1)
beyniyle çalışmalı
değil mi
evet
(Speakers speak simultaneously)
(Speaker 2)
doktorun
Açıklamalarındaki
dangu kısım şurada
meme
önemsizdir diye
anlatırken
memeyi
aşağılamak
başka bir şey yani
anladın mı
siz erkekleri
bununla kandırıyorsunuz
bu zaten
yağ tulumudur

Fragment 6: Alem FM, Cortisol

(Speaker 1)
altına sokmuşlar
böyle
eli

yağ oranı
düşük olanları
fit olanları
filan
onların organizmalarında
aşırı
kortizol hormonu
oluştuğunu görmüşler
halbuki
bu
balık etinde olan insanların
olayı
sizin o beyefendinin
söylediği gibi
daha yaratıcı
evet
daha savunulur hale gelerek
o
stresin
yaratmış olduğu anksiyetin
yaratmış olduğu baskıyı
azaltarak
organizmanın hormon sistemini
tahrik edecek
unsur olmaktan
şiddetinden kurtararak
evet
beyin tarafından
algılanmasını
sağlayan bir yapı
geliştirdiklerini
evet
dola
dolayısiyle
daha rint
meşrep olduklarını
olayları
çok
dramatize etmeden
değerlendirmeye aldıklarını
şey etmişler
görmüşler
ve bu sebepledir ki
böyle bir farklılık
olduğu zaman

aman onu gidereceğim diyerek de
ekstra bir stres unsuru haline
getirilmemesini

Fragment 7: Alem FM M. Physical exercise

(Speaker 1)
organizmanın
o otuz dakikalik
hareketlilik
imkanını ona vermezseniz
immun sisteminin
şey olayı
yeteri kadar
güçlü olamayacağı
doğrusunda haber yapıyorlardı
dolayısıyla siz bahçede yürürdünüz
tahmin ediyorum
ondan sonra da
devam ettireceksiniz
sevgili büyüğüm
bugün
ki programımızda istiyorsanız
biraz
neşeli
konulardan bahsedelim
absürt böyle sıradışı olaylardan
bahsedelim çünkü konuların neresinden
tutsanız gene
öyle karanlık
konular oluyor
insanın içerisini
karartacak hususlar var
trajikomik olaylar da var
da gerçi
gerçi de
mesela
ben
tam
gene
böyle
göğüs kanserini
azaltma
konusunda sporun

etkileri konusuyla yapılan araştırmalardan
filan bahsedecektim
tam bu arada
balıkesirde de
meme kanserinde
yeni yaklaşımlar konulu
bir panel
yap

B.2.2 Instructions

Dear participant,
Thank you for being willing to participate in my research by filling out the following form. With this document you also received a sequence of numbered sound files, of which each file name corresponds to the sentences you find in the form. Together, these sentences are a radio fragment. I would like to ask you to indicate where in each of these sentences you hear stress. Stress is a type of prominence which makes syllables stand out more. One could say they are the 'beat' of the language. I would like to know where you hear these prominent syllables and I would therefore like to ask you to show me with the 'highlight' function of Adobe Reader (the yellow rectangle with the pen in the toolbar; in the menu tools -> comments/markup -> highlight text tool) each syllable on which you perceive stress. It is possible that you would put stress on a different syllable if you would pronounce the word, but I would like to know where the speaker put it. In case you hear multiple stressed syllables per audio file, which is definitely possible, I would like you to indicate where you hear each of them.

A. This is an example

Please fill out the form and comment on the survey afterwards (click on the arrow with the cursor in the toolbar; in the menu it is tools -> select en zoom -> select) and save the form. Then please send it to me: s.vanommen@uu.nl .

Thank you very much!
Best,
Sandrien van Ommen

Name:
Age:
Education:
Occupation (if any):
Native language:
Place of birth:
Comments (after filling out the form):

Bibliography

Adriaans, Frans W., and René W. J. Kager. 2010. Adding generalization to statistical learning: The induction of phonotactics from continuous speech. *Journal of Memory and Language* 62:311–331.

Allen, George D, and Sarah Hawkins. 1979. Trochaic rhythm in children's speech. *Current Issues in the Phonetic Sciences* 927–933.

Arvaniti, Amalia. 2012. The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics* 40:351–373.

Aske, Jon. 1990. Disembodied rules versus patterns in the lexicon: Testing the psychological reality of spanish stress rules. In *Proceedings of the Sixteenth Annual Meeting of the Berkeley Linguistics Society*, 30–45.

Aslin, Richard N., Jenny R. Saffran, and Elissa L. Newport. 1998. Computation of conditional probability statistics by 8-month-old infants. *Psychological Science* 9:321–324.

Astheimer, Lori B., and Lisa D. Sanders. 2009. Listeners modulate temporally selective attention during natural speech processing. *Biological Psychology* 80:23–34.

Baayen, R. Harald, Richard Piepenbrock, and H. van Rijn. 1993. The CELEX lexical database. CD-ROM.

Baayen, R.H., D.J. Davidson, and D.M. Bates. 2008. Mixed effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59:390–412.

Barik, Henri C. 1977. Cross-linguistic study of temporal characteristics of different types of speech materials. *Language and Speech* 20:116–126.

Bárkányi, Zsuzsanna. 2002. Primary stress in spanish. *Amsterdam Studies in the Theory and History of Linguistic Science Series* 4:17–32.

Bates, Douglas, Martin Maechler, Ben Bolker, and Steven Walker. 2014. *lme4: Linear mixed-effects models using Eigen and S4*. URL `http://CRAN.R-project.org/package=lme4`, r package version 1.1-5.

Beckman, Mary E. 1986. *Stress and non-stress accent*. Walter de Gruyter.

Bergelson, Elika, and Daniel Swingley. 2012. At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences* 109:3253–3258.

Berinstein, Ava E. 1979. A cross-linguistic study on the perception and production of stress. Doctoral Dissertation, University of California, Los Angeles–Linguistics.

Bertoncini, Josiane, and Jacques Mehler. 1981. Syllables as units in infant speech perception. *Infant Behavior and Development* 4:247–260.

Bijeljac-Babic, Ranka, Josiane Bertoncini, and Jacques Mehler. 1993. How do 4-day-old infants categorize multisyllabic utterances? *Developmental Psychology* 29:711–721.

Blaho, Sylvia, and Dániel Szeredi. 2011. Secondary stress in Hungarian:(morpho)-syntactic, not metrical. In *Proceedings of the 28th West Coast Conference on Formal Linguistics*, 51–59.

Boersma, Paul. 1997. How we learn variation, optionality, and probability. In *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam*, ed. R.J.J.H. van Son, volume 21, 43–58.

Boersma, Paul, and David Weenink. 2011. Praat: Doing phonetics by computer. Software. Version 5.3.

Bohn, Karen, Johannes Knaus, Richard Wiese, and Ulrike Domahs. 2013. The influence of rhythmic (ir) regularities on speech processing: Evidence from an ERP study on German phrases. *Neuropsychologia* 51:760–771.

Bolinger, Dwight L. 1958. A theory of pitch accent in English. *Word* 14:109–149.

Bortfeld, Heather, James L. Morgan, Roberta Michnick Golinkoff, and Karen Rathbun. 2005. Mommy and me. familiar names help launch babies into speech-stream segmentation. *Psychological Science* 16:298–304.

Braine, Martin D. 1966. Learning the positions of words relative to a marker element. *Journal of Experimental Psychology* 72:532–540.

Brent, Michael R., and Jeffrey Mark Siskind. 2001. The role of exposure to isolated words in early vocabulary development. *Cognition* 81:B33–B44.

Buhmann, Jeska, Johanneke Caspers, Vincent J. van Heuven, Heleen Hoekstra, Jean-Pierre Martens, and Marc Swerts. 2002. Annotation of prominent words, prosodic boundaries and segmental lengthening by non-expert transcribers in the Spoken Dutch Corpus. In *Proceedings of the third International Conference on Language Resources and Evaluation*, ed. C. González Rodriguez, M Paz Suárez Araujo, 779–785.

Chapin, Paul G., Timothy S. Smith, and Adele A. Abrahamson. 1972. Two factors in perceptual segmentation of speech. *Journal of Verbal Learning and Verbal Behavior* 11:164–173.

Cheour, M, K Alho, R. Čeponiené, K. Reinikainen, K. Sainio, M. Pohjavuori, O. Aaltonen, and R. Näätänen. 1998. Maturation of mismatch negativity in infants. *International Journal of Psychophysiology* 29:217–226.

Cho, Taehong, and James M. McQueen. 2005. Prosodic influences on consonant production in Dutch: Effects of prosodic boundaries, phrasal accent and lexical stress. *Journal of Phonetics* 33:121–157.

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

Christophe, Anne, Jacques Mehler, and Núria Sebastián-Gallés. 2001. Perception of prosodic boundary correlates by newborn infants. *Infancy* 2:385–394.

Cole, Ronald A., and Jola Jakimik. 1980a. How are syllables used to recognize words? *The Journal of the Acoustical Society of America* 67:965–970.

Cole, Ronald A., and Jola Jakimik. 1980b. A model of speech perception. In *Perception and production of fluent speech*, ed. Ronald A. Cole, 136—163. Hillsdale, NJ: Erlbaum.

Colina, Sonia. 2009. *Spanish phonology: A syllabic perspective*. Georgetown University Press.

Crosswhite, Katherine. 2003. Spectral tilt as a cue to word stress in Polish, Macedonian and Bulgarian. In *Proceedings of the 15th International Congress of Phonetic Sciences*, ed. Maria-Josep. Solé, D. Recasens, and J. Romero. Barcelona: Causal Productions.

Cunillera, T., A. Gomila, and A. Rodriguez-Fornells. 2008. Beneficial effects of word final stress in segmenting a new language: evidence from ERPs. *BMC Neuroscience* 9.

Cunillera, Toni, Juan M. Toro, Nuria Sebastián-Gallés, and Antoni Rodríguez-Fornells. 2006. The effects of stress and statistical cues on continuous speech segmentation: An event-related brain potential study. *Brain Research* 1123:168–178.

Cutler, Anne. 1976. Phoneme-monitoring reaction time as a function of preceding intonation contour. *Perception and Psychophysics* 20:55–60.

Cutler, Anne. 1986. Forbear is a homophone: Lexical prosody does not constrain lexical access. *Language and Speech* 29:201–220.

Cutler, Anne, and Sally Butterfield. 1992. Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language* 31:218–236.

Cutler, Anne, and David M. Carter. 1987. The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language* 2:133–142.

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

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

Cutler, Anne, and Dennis Norris. 1988. The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance* 14:113–121.

Cutler, Anne, and Wilma Van Donselaar. 2001. Voornaam is not (really) a homophone: Lexical prosody and lexical access in Dutch. *Language and Speech* 44:171–195.

Darwin, Chris J. 1975. On the dynamic use of prosody in speech perception. In *Structure and process in speech perception*, ed. A. Cohen and S. G. Nooteboom, 178–194. Springer.

Dauer, Rebecca M. 1983. Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics* 11:51–62.

De Jong, Nivja, and Ton Wempe. 2009. Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods* 41:385–390.

Delattre, Pierre. 1966. A comparison of syllable length conditioning among languages. *International Review of Applied Linguistics in Language Teaching* 4:183–198.

Den Os, Els, and René Kager. 1986. Extrametricality and stress in Spanish and Italian. *Lingua* 69:23–48.

Depaolis, Rory A., Marilyn M. Vihman, and Tamar Keren-Portnoy. 2014. When do infants begin recognizing familiar words in sentences? *Journal of Child Language* 41:226–239.

Dogil, Grzegorz, Jadranka Gvozdanovic, and Sandro Kodzasov. 1999. Slavic languages. In *Word prosodic systems in the languages of Europe*, ed. Harry Van der Hulst, 813–876. Walter de Gruyter.

Dogil, Grzegorz, and Briony Williams. 1999. The phonetic manifestation of word stress. In *Word prosodic systems in the languages of Europe*, ed. Harry Van der Hulst, 273–334. Walter de Gruyter.

Domahs, Ulrike, Safiye Genc, Johannes Knaus, Richard Wiese, and Barış Kabak. 2012a. Processing (un-) predictable word stress: ERP evidence from Turkish. *Language and Cognitive Processes* 28:335–354.

Domahs, Ulrike, Johannes Knaus, Paula Orzechowska, and Richard Wiese. 2012b. Stress 'deafness' in a language with fixed word stress: an ERP study on Polish. *Frontiers in Psychology* 3:1–15.

Dupoux, Emmanuel, Christophe Pallier, Nuria Sebastian, and Jacques Mehler. 1997. A destressing 'deafness' in French? *Journal of Memory and Language* 36:406–421.

Dupoux, Emmanuel, Sharon Peperkamp, and Núria Sebastián-Gallés. 2001. A robust method to study stress 'deafness'. *The Journal of the Acoustical Society of America* 110:1606–1618.

Dutoit, Thierry, Vincent Pagel, Nicolas Pierret, François Bataille, and Olivier Van der Vrecken. 1996. The MBROLA project: Towards a set of high quality speech synthesizers free of use for non commercial purposes. In *Proceedings of the 4th International Conference of Spoken Language Processing*, ed. H. Timothy Bunnell and William Idsardi, 1393–1396.

Echols, Catharine H., Megan J. Crowhurst, and Jane B. Childers. 1997. The perception of rhythmic units in speech by infants and adults. *Journal of Memory and Language* 36:202–225.

Elman, Jeffrey L. 1990. Finding structure in time. *Cognitive Science* 14:179–211.

Endress, Ansgar D., and Luca L. Bonatti. 2007. Rapid learning of syllable classes from a perceptually continuous speech stream. *Cognition* 105:247–299.

Endress, Ansgar D., and Jacques Mehler. 2009. Primitive computations in speech processing. *The Quarterly Journal of Experimental Psychology* 62:2187–2209.

Face, Timothy L. 2003. Intonation in Spanish declaratives. *Catalan Journal of Linguistics* 2:115–131.

Face, Timothy L. 2006. Cognitive factors in the perception of Spanish stress placement: Implications for a model of speech perception. *Linguistics* 44:1237–1267.

Fant, Gunnar. 1960. *Acoustic theory of speech production*. Mouton, Den Haag.

Fernald, Anne, and Gerald McRoberts. 1996. Prosodic bootstrapping: A critical analysis of the argument and the evidence. In *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*, ed. James L. Morgan and Katherine Demuth, 365–388.

Fonagy, I. 1966. Electrophysiological and acoustic correlates of stress and stress perception. *Journal of Speech and Hearing Research* 9:231–244.

Fougeron, Cécile, and Patricia A. Keating. 1997. Articulatory strengthening at edges of prosodic domains. *The Journal of the Acoustical Society of America* 101:3728–3740.

Friederici, Angela D., Manuela Friedrich, and Anne Christophe. 2007. Brain responses in 4-month-old infants are already language specific. *Current Biology* 17:1208–1211.

Frisch, Stefan A., Janet B. Pierrehumbert, and Michael B. Broe. 2004. Similarity avoidance and the OCP. *Natural Language and Linguistic Theory* 22:179–228.

Fry, Dennis B. 1955. Duration and intensity as physical correlates of linguistic stress. *The Journal of the Acoustical Society of America* 27:765–768.

Fry, Dennis B. 1958. Experiments in the perception of stress. *Language and Speech* 1:126–152.

Gambell, Timothy, and Charles Yang. 2005. Mechanisms and constraints in word segmentation. Ms., Yale University.

Giovanelli, Giuliana, Alessandra Sansavini, and Alessandra Farneti. 1999. Perception of sound, rhythm and speech from pre-natal to post-natal life. In *Current issues in developmental psychology*, 137–159. Springer.

Gleitman, Lila R, Henry Gleitman, Barbara Landau, and Eric Wanner. 1988. Where learning begins: Initial representations for language learning. In *Language: Psychological and biological aspects*, ed. Frederick J. Newmeyer, 150–193. Cambridge University Press.

Gleitman, Lila R., and Eric Wanner. 1982. *Language acquisition: The state of the art*. Cambridge University Press.

Goedemans, Rob, and Harry van der Hulst. 2009. *StressTyp: a database for word accentual patterns in the world's languages*, 235–282. New York/Berlin: Mouton de Gruyter.

Goedemans, Rob, and Harry van der Hulst. 2013. *Fixed stress locations*. Leipzig: Max Planck Institute for Evolutionary Anthropology. URL `http://wals.info/chapter/14`.

Göksel, A., and C. Kerslake. 2005. *Turkish: A comprehensive grammar*. Routledge, London.

Goldsmith, John A. 1976. Autosegmental phonology. Doctoral Dissertation, Massachusetts Institute of Technology.

Gordon, Matthew. 2002. A factorial typology of quantity-insensitive stress. *Natural Language and Linguistic Theory* 20:491–552.

Gout, Ariel, Anne Christophe, and James L. Morgan. 2004. Phonological phrase boundaries constrain lexical access ii. infant data. *Journal of Memory and Language* 51:548–567.

Grabe, Esther, and Ee Ling Low. 2002. Durational variability in speech and the rhythm class hypothesis. *Papers in Laboratory Phonology* 7:515–546.

Grosjean, François. 1985. The recognition of words after their acoustic offset: Evidence and implications. *Perception and Psychophysics* 38:299–310.

Gussenhoven, Carlos. 1991. The english rhythm rule as an accent deletion rule. *Phonology* 8:1–35.

Gussenhoven, Carlos. 1992. Dutch. *Journal of the International Phonetic Association* 22:45–47.

Gussenhoven, Carlos. 1993. The Dutch foot and the chanted call. *Journal of Linguistics* 29:37–63.

Gussenhoven, Carlos. 2009. Vowel duration, syllable quantity and stress in Dutch. In *The nature of the word: Studies in honor of Paul Kiparsky*, ed. K. Hanson and S. Inkelas, 181–198. MIT Press, Cambridge.

Gussman, E. 2007. *The phonology of Polish*. The Phonology of the World's Languages. Oxford University Press, Oxford.

Halle, Morris, and Jean-Roger Vergnaud. 1987. *An essay on stress*. Mit Press Cambridge.

Hammond, Michael. 1984. Constraining metrical theory: a modular theory of stress and destressing. Doctoral Dissertation, UCLA.

Hammond, Michael. 1987. Hungarian cola. *Phonology Yearbook* 4:267–269.

Hanulikova, Adriana, James M. McQueen, and Holger Mitterer. 2010. Possible words and fixed stress in the segmentation of Slovak speech. *The Quarterly Journal of Experimental Psychology* 63:555–579.

Harris, James Wesley. 1983. *Syllable structure and stress in Spanish. a nonlinear analysis.* MIT Press, Cambridge, Massachusetts.

Harris, James Wesley. 1992. *Spanish stress: The extrametricality issue.* Bloomington, IN: Indiana University Linguistics Club.

Haugen, Einar. 1949. Phoneme or prosodeme? *Language* 25:278–282.

Hawthorne, Kara, and LouAnn Gerken. 2014. From pauses to clauses: Prosody facilitates learning of syntactic constituency. *Cognition* 133:420–428.

Hay, Jennifer. 2000. Causes and consequences of word structure. Doctoral Dissertation, Nortwestern University.

Hayes, Bruce. 1984. The phonology of rhythm in English. *Linguistic Inquiry* 15:33–74.

Hayes, Bruce. 1985. Iambic and trochaic rhythm in stress rules. In *Proceedings of the Eleventh Annual Meeting of the Berkeley Linguistics Society*, 429–446.

Hayes, Bruce. 1995. *Metrical stress theory: Principles and case studies.* Chicago: The University of Chicago Press.

Hayes, Bruce, and Stanislaw Puppel. 1985. On the rhythm rule in Polish. In *Advances in nonlinear phonology*, ed. Harry van der Hulst and Norval Smith, 59–81. Dordrecht: Foris Publications.

Hayes, Bruce Philip. 1980. A metrical theory of stress rules. Doctoral Dissertation, Massachusetts Institute of Technology.

Henson, Richard N.A. 1998. Short-term memory for serial order: The start-end model. *Cognitive psychology* 36:73–137.

Hirsh-Pasek, Kathy, Deborah G. Kemler Nelson, Peter W. Jusczyk, Kimberly Wright Cassidy, Benjamin Druss, and Lori Kennedy. 1987. Clauses are perceptual units for young infants. *Cognition* 26:269–286.

Hochberg, Judith G. 1988. First steps in the acquisition of Spanish stress. *Journal of Child Language* 15:273–292.

Höhle, Barbara, Ranka Bijeljac-Babic, Birgit Herold, Jürgen Weissenborn, and Thierry Nazzi. 2009. Language specific prosodic preferences during the first half year of life: Evidence from German and French infants. *Infant Behavior and Development* 32:262–274.

Höhle, Barbara, and Jürgen Weissenborn. 2003. German-learning infants' ability to detect unstressed closed-class elements in continuous speech. *Developmental Science* 6:122–127.

Hooper, Joan B, and Tracy Terrell. 1976. Stress assignment in Spanish: A natural generative analysis. *Glossa* 10:64–110.

Houston, Derek, Lynn Santelmann, and Peter Jusczyk. 2004. English-learning infants' segmentation of trisyllabic words from fluent speech. *Language and Cognitive Processes* 19:97–136.

Hualde, José Ignacio. 2005. *The sounds of Spanish*. Cambridge University Press New York.

Hualde, José Ignacio. 2007. Stress removal and stress addition in Spanish. *Journal of Portuguese Linguistics* 5:59–89.

Hualde, José Ignacio, Antxon Olarrea, and Erin O'Rourke. 2012. *The handbook of Hispanic linguistics*. John Wiley and Sons.

Huggins, A. W. F. 1964. Distortion of the temporal pattern of speech: Interruption and alternation. *The Journal of the Acoustical Society of America* 36:1055–1064.

Hyman, Larry M. 1977. On the nature of linguistic stress. In *Studies in stress and accent*, ed. Larry Hyman, 37–82. Los Angeles, University of Southern California.

Inkelas, Sharon. 1999. Exceptional stress-attracting suffixes in Turkish: representations versus the grammar. In *The prosody-morphology interface*, ed. René W. J. Kager, Harry van der Hulst, and Wim Zonneveld, 134–187. Cambridge University Press.

Inkelas, Sharon, Aylin Küntay, Orhan Orgun, and Ronald Sprouse. 2000. Turkish electronic living lexicon (TELL). *Turkic Languages* 4:253–275.

Inkelas, Sharon, and C.O. Orgun. 2003. Turkish stress: A review. *Phonology* 20:139–161.

Ipek, Canan, and Sun-Ah Jun. 2013. Towards a model of intonational phonology of Turkish: Neutral intonation. *Proceedings of Meetings on Acoustics* 19.

Jakobson, Roman, Morris Halle, and C. Gunnar M. Fant. 1951. *Preliminaries to speech analysis*. MIT.

Jansen, Wouter. 2004. Laryngeal contrast and phonetic voicing: A laboratory phonology approach to English, Hungarian, and Dutch. Doctoral Dissertation, Rijksuniversiteit Groningen.

Jassem, Wiktor. 2003. Polish. *Journal of the International Phonetic Association* 33:103–107.

Jespersen, Otto. 1905. *Growth and structure of the English language.* BG Teubner.

Johnson, Elizabeth K, Amanda Seidl, and Michael D. Tyler. 2014. The edge factor in early word segmentation: utterance-level prosody enables word form extraction by 6-month-olds. *PloS ONE* 9.

Jun, Sun-Ah, and Cécile Fougeron. 2000. A phonological model of French intonation. In *Intonation: Analysis, modeling and technology*, ed. A. Botinis, 209–242. Dordrecht: Kluwer Academic Publishers.

Jusczyk, Peter W., Anne Cutler, and Nancy J. Redanz. 1993. Infants' preference for the predominant stress patterns of English words. *Child Development* 64:675–687.

Jusczyk, Peter W., Kathy Hirsh-Pasek, Deborah G. Kemler Nelson, Lori J. Kennedy, Amanda Woodward, and Julie Piwoz. 1992. Perception of acoustic correlates of major phrasal units by young infants. *Cognitive Psychology* 24:252–293.

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

Jusczyk, Peter W., and Elizabeth Thompson. 1978. Perception of a phonetic contrast in multisyllabic utterances by 2-month-old infants. *Perception and Psychophysics* 23:105–109.

Kabak, Baris, Kazumi Maniwa, and Nina Kazanina. 2010. Listeners use vowel harmony and word-final stress to spot nonsense words: A study of Turkish and French. *Laboratory Phonology* 1:207–224.

Kabak, Baris, and Irene Vogel. 2001. The phonological word and stress assignment in Turkish. *Phonology* 18:315–360.

Kager, René. 1989. *A metrical theory of stress and destressing in English and Dutch.* Dordrecht: Foris Publications.

Kager, René. 1995. The metrical theory of word stress. In *A handbook of phonological theory*, ed. J. Goldsmith, 367–402. Oxford: Basil Blackwell.

Kager, René, and Keren Shatzman. 2007. Phonological constraints in speech processing. *Linguistics in the Netherlands* 24:99–111.

Kager, René, and Ellis Visch. 1988. Metrical constituency and rhythmic adjustment. *Phonology* 5:21–71.

Kager, René, and Wim Zonneveld. 1986. Schwa, syllables, and extrametricality in Dutch. *The Linguistic Review* 5:197–221.

Kager, René W. J., Ellis Visch, and Wim Zonneveld. 1987. Nederlandse woordklemtoon (hoofdklemtoon, bijklemtoon, reductie en voeten). *GLOT: Leids Taalkundig Bulletin* 10:197–226.

Kahn, Daniel. 1976. *Syllable-based generalizations in English phonology*, volume 156. Indiana University Linguistics Club Bloomington.

Kallestinova, Elena. 2004. Voice and aspiration of stops in Turkish. *Folia Linguistica* 38:117–144.

Kálmán, László, and Ádám Nádasdy. 1994. A hangsúly [stress]. *Strukturális Magyar Nyelvtan* 2:393–467.

Kamali, Beste. 2011. Topics at the PF interface of Turkish. Doctoral Dissertation, Harvard University.

Kaye, Jonathan. 1989. *Phonology: A cognitive view*. Hillsdale, NJ: Lawrence Erlbaum.

Keij, Brigitta, and René Kager. submitted. Rhythmic preferences in monolingual Dutch- and Turkish-learning infants between 4 and 8 months of age.

Kerek, Andrew. 1971. *Hungarian metrics: some linguistic aspects of iambic verse*. Indiana University Bloomington.

Kijak, Anna. 2003. Anglístyka or Anglistýka? A study of extrametricality in Polish and L2 parameter 'resetting' for metrical systems. Master's Thesis, Utrecht University.

Kijak, Anna. 2009. How stressful is L2 stress? Doctoral Dissertation, Utrecht University. Published by LOT (Landelijke Onderzoekschool Taalwetenschap).

Klatt, Dennis H. 1976. Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *The Journal of the Acoustical Society of America* 59:1208–1221.

Klatt, Dennis H, and Kenneth N Stevens. 1973. On the automatic recognition of continuous speech: Implications from a spectrogram-reading experiment. *Institute of Electrical and Electronics Engineers Transactions on Audio and Electroacoustics* 21:210–217.

Ladefoged, Peter. 1967. Stress and respiratory activity. *Three Areas of Experimental Phonetics* 1–49.

Lahiri, Aditi, and William Marslen-Wilson. 1991. The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition* 38:245–294.

Lany, Jill, and Rebecca L. Gómez. 2008. Twelve-month-old infants benefit from prior experience in statistical learning. *Psychological Science* 19:1247–1252.

Lehiste, Ilse. 1972. The timing of utterances and linguistic boundaries. *The Journal of the Acoustical Society of America* 51:2018–2024.

Levi, Susannah V. 2002. Limitations on tonal crowding in Turkish intonation. In *Proceedings of 9th International Phonology Conference*, 1–16.

Levi, Susannah V. 2005. Acoustic correlates of lexical accent in Turkish. *Journal of the Internation Phonetic Association* 35:73.

Lew-Williams, Casey, Bruna Pelucchi, and Jenny R. Saffran. 2011. Isolated words enhance statistical language learning in infancy. *Developmental Science* 14:1323–1329.

Liberman, Mark, and Alan Prince. 1977. On stress and linguistic rhythm. *Linguistic Inquiry* 8:249–336.

Liberman, Mark Yoffe. 1975. The intonational system of English. Doctoral Dissertation, Massachusetts Institute of Technology.

Lindblom, B. 1978. Final lengthening in speech and music. In *Nordic prosody*, ed. E. Gårding, G. Bruce, and R. Bennet, 85–101. Lund University: Department of Linguistics.

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

Magnuson, James S., James A. Dixon, Michael K. Tanenhaus, and Richard N. Aslin. 2007. The dynamics of lexical competition during spoken word recognition. *Cognitive Science* 31:133–156.

Mampe, Birgit, Angela D. Friederici, Anne Christophe, and Kathleen Wermke. 2009. Newborns' cry melody is shaped by their native language. *Current Biology* 19:1994–1997.

Mandel, Denise R., Peter W. Jusczyk, and Deborah G. Kemler Nelson. 1994. Does sentential prosody help infants organize and remember speech information? *Cognition* 53:155–180.

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

Martin, James G. 1972. Rhythmic (hierarchical) versus serial structure in speech and other behavior. *Psychological Review* 79:487–509.

Martínez-Celdrán, Eugenio, Ana Ma Fernández-Planas, and Josefina Carrera-Sabaté. 2003. Castilian Spanish. *Journal of the International Phonetic Association* 33:255–259.

Martínez-Paricio, Violeta, and René Kager. forthcoming. Non-intervention constraints and the binary-to-ternary rhythmic continuum. *Phonology* .

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

Mattys, Sven L., and James F. Melhorn. 2005. How do syllables contribute to the perception of spoken English? Insight from the migration paradigm. *Language and Speech* 48:223–252.

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

Mazzoni, Dominic, and Roger Dannenberg. 2013. Audacity. Software, version 2.0.6. URL `http://audacity.sourceforge.net/`.

McCarthy, John J, and Alan Prince. 1993. *Generalized alignment*. Springer.

McClelland, James L., and Jeffrey L. Elman. 1986. The TRACE model of speech perception. *Cognitive Psychology* 18:1–86.

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

McQueen, James M., Dennis Norris, and Anne Cutler. 1994. Competition in spoken word recognition: Spotting words in other words. *Journal of Experimental Psychology: Learning, Memory and Cognition* 20:621–638.

Mehler, Jacques, Jean Yves Dommergues, Uli Frauenfelder, and Juan Segui. 1981. The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior* 20:298–305.

Morgan, James L., and Jenny R. Saffran. 1995. Emerging integration of sequential and suprasegmental information in preverbal speech segmentation. *Child Development* 66:911–936.

Morton, John, and Wiktor Jassem. 1965. Acoustic correlates of stress. *Language and Speech* 8:159–181.

Nakatani, Lloyd H., and Kathleen D. Dukes. 1977. Locus of segmental cues for word juncture. *The Journal of the Acoustical Society of America* 62:714–719.

Nakatani, Lloyd H., and Judith A. Schaffer. 1978. Hearing ‘words’ without words: Prosodic cues for word perception. *The Journal of the Acoustical Society of America* 63:234–245.

Nazzi, Thierry, Josiane Bertoncini, and Jacques Mehler. 1998. Language discrimination by newborns: toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance* 24:756–766.

Nelson, Deborah G. Kemler, Peter W. Jusczyk, Denise R. Mandel, James Myers, Alice Turk, and Louann Gerken. 1995. The head-turn preference procedure for testing auditory perception. *Infant Behavior and Development* 18:111–116.

Nespor, Marina. 1990. On the rhythm parameter in phonology. In *Logical issues in language acquisition*, ed. Iggy M. Roca, 157–175. Dordrecht: Foris Publications.

Nespor, Marina, and Irene Vogel. 1989. On clashes and lapses. *Phonology* 6:69–116.

Nespor, Shukla Mohinish, Marina, and Jacques Mehler. 2011. Stress-timed vs. syllable-timed languages. In *The blackwell companion to phonology*, ed. Ewen Colin J. Hume Elizabeth Van Oostendorp, Marc and Keren Rice, 1147–1159. Wiley-Blackwell.

Newlin-Łukowicz, Luiza. 2012. Polish stress: looking for phonetic evidence of a bidirectional system. *Phonology* 29:271–329.

Norris, Dennis, James M. McQueen, and Anne Cutler. 1995. Competition and segmentation in spoken-word recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition* 21:1209–1228.

Norris, Dennis, James M. McQueen, Anne Cutler, and Sally Butterfield. 1997. The possible-word constraint in the segmentation of continuous speech. *Cognitive Psychology* 34:191–243.

Ohala, John J. 1977. The physiology of stress. In *Studies in stress and accent*, ed. Larry Hyman, 145–168. Los Angeles, CA: Department of Linguistics, University of Southern California.

Oliver, Dominika, and Martine Grice. 2003. Phonetics and phonology of lexical stress in Polish verbs. In *Proceedings of the 15th International Congress of Phonetic Science. Barcelona*, ed. Maria-Josep. Solé, D. Recasens, and J. Romero, 2027–2030.

van Ommen, Sandrien, and René Kager. 2012. Progressive use of metrical cues: A cross-linguistic study. In *Proceedings of Natural Language Processing and Cognitive Science 2012*, 74–84.

Oostdijk, Nelleke. 2000. The spoken Dutch corpus. overview and first evaluation. In *Proceedings of the second International Conference on Language Resources and Evaluation*, 887–894.

van Oostendorp, Marc. 1995. Vowel quality and phonological projection. Doctoral Dissertation, University of Tilburg (KUB).

van Oostendorp, Marc. 2012. Quantity and the three-syllable window in Dutch word stress. *Language and Linguistics Compass* 6:343–358.

Ortega-Llebaria, Marta, and Pilar Prieto. 2007. Disentangling stress from accent in Spanish: Production patterns of the stress contrast in deaccented syllables. In *Segmental and prosodic issues in Romance phonology*, ed. Pilar Prieto, Joan Mascaró, and Maria-Josep Solé, 155–176. Amsterdam; Philadelphia; John Benjamins Publishing Company.

Ortega-Llebaria, Marta, and Pilar Prieto. 2011. Acoustic correlates of stress in Central Catalan and Castilian Spanish. *Language and Speech* 54:1–25.

Peelle, Jonathan E., and Matthew H. Davis. 2012. Neural oscillations carry speech rhythm through to comprehension. *Frontiers in Psychology* 3.

Pelucchi, Bruna, Jessica F. Hay, and Jenny R. Saffran. 2009. Learning in reverse: Eight-month-old infants track backward transitional probabilities. *Cognition* 113:244–247.

Peña, Marcela, Luca L. Bonatti, Marina Nespor, and Jacques Mehler. 2002. Signal-driven computations in speech processing. *Science* 298:604–607.

Peperkamp, Sharon, and Emmanuel Dupoux. 2002. A typological study of stress 'deafness'. *Laboratory Phonology* 7:203–240.

Perruchet, Pierre, and Stéphane Desaulty. 2008. A role for backward transitional probabilities in word segmentation? *Memory and Cognition* 36:1299–1305.

Perruchet, Pierre, and Ronald Peereman. 2004. The exploitation of distributional information in syllable processing. *Journal of Neurolinguistics* 17:97–119.

Perruchet, Pierre, and Annie Vinter. 1998. PARSER: A model for word segmentation. *Journal of Memory and Language* 39:246–263.

Pierrehumbert, Janet. 1993. Dissimilarity in the Arabic verbal roots. In *Proceedings of the North East Linguistics Society*, volume 23, 367–381.

Pierrehumbert, Janet. 2003. Probabilistic phonology: Discrimination and robustness. In *Probability theory in linguistics*, ed. Rens Bod, Jennifer Hay, and Stefanie Jannedy, 177–228. The MIT Press, Cambridge MA.

Pike, Kenneth L. 1945. *The intonation of American English.* Ann Arbor, Michigan: University of Michigan Press.

Pinker, Steven, and David Birdsong. 1979. Speakers' sensitivity to rules of frozen word order. *Journal of Verbal Learning and Verbal Behavior* 18:497–508.

Polka, Linda, and Megha Sundara. 2003. Word segmentation in monolingual and bilingual infant learners of English and French. In *Proceedings of the 15th International Congress of Phonetic Sciences*, ed. Maria-Josep. Solé, D. Recasens, and J. Romero, 1021–1024. Causal Productions Barcelona.

Pordány, László. 1986. A comparison of some English and Hungarian freezes. *Papers and Studies in Contrastive Linguistics* 21:119–127.

Prince, Alan. 1976. Applying stress. Ms., University of Massachusetts, Amherst.

Prince, Alan S. 1980. A metrical theory for Estonian quantity. *Linguistic Inquiry* 11:511–562.

Prince, Alan S. 1983. Relating to the grid. *Linguistic Inquiry* 14:19–100.

Prince, Alan S., and Paul Smolensky. 1993. Optimality theory: Constraint interaction in generative grammar. Technical report, Ms, Rutgers University, New Brunswick and University of Colorado, Boulder.

Quené, Hugo, and Huub van den Bergh. 2008. Examples of mixed-efects modeling with crossed random effects and with binomial data. *Journal of Memory and Language* 59:413–425.

Quené, Hugo, and Mariëtte L. Koster. 1998. Metrical segmentation in Dutch: Vowel quality or stress? *Language and Speech* 41:185–201.

Quené, Hugo, and Robert F. Port. 2005. Effects of timing regularity and metrical expectancy on spoken-word perception. *Phonetica* 62:1–13.

R Core Team. 2013. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. URL `http://www.R-project.org/`, software.

Ramus, Franck, Emmanuel Dupoux, and Jacques Mehler. 2003. The psychological reality of rhythm classes: Perceptual studies. In *Proceedings of the 15th International Congress of Phonetic Sciences*, ed. Maria-Josep. Solé, D. Recasens, and J. Romero, 337–342. Barcelona: Causal Productions.

Ramus, Franck, Marina Nespor, and Jacques Mehler. 1999. Correlates of linguistic rhythm in the speech signal. *Cognition* 73:265–292.

Revithiadou, Anthi, Hassan Kaili, Sophia Prokou, and Maria-Anna Tiliopoulou. 2006. Turkish accentuation revisited: A compositional approach to Turkish stress. In *Advances in Turkish Linguistics: Proceedings of the 12th International Conference on Turkish Linguistics*, ed. S. Yagcioglu, A. Cem Deger, O. Kosaner, and A. Celtek, 37–50.

Roca, Iggy M. 2005. Saturation of parameter settings in Spanish stress. *Phonology* 22:345–394.

Roca, Iggy M. 2006. The Spanish stress window. In *Optimality-theoretic studies in Spanish phonology*, ed. Colina Sonia Martínez-Gil, Fernando, 239–277. John Benjamins Publishing.

de Rooij, J. J. 1976. Perception of prosodic boundaries. *IPO (Instituut voor Perceptie Onderzoek) Annual Progress Report* 11:20–24.

Rothermich, Kathrin, Maren Schmidt-Kassow, and Sonja A. Kotz. 2012. Rhythm's gonna get you: regular meter facilitates semantic sentence processing. *Neuropsychologia* 50:232–244.

Rubach, Jerzy, and Geert Booij. 1990. Syllable structure assignment in Polish. *Phonology* 7:121–158.

Rubach, Jerzy, and Geert E. Booij. 1985. A grid theory of stress in Polish. *Lingua* 66:281–320.

Ryan, Kevin M. 2013. Onset weight, word weight, and the perceptual interval.

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

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

Sanabria, Daniel, Mariagrazia Capizzi, and Ángel Correa. 2011. Rhythms that speed you up. *Journal of Experimental Psychology: Human Perception and Performance* 37:236–244.

Sanders, Lisa D., Elissa L. Newport, and Helen J. Neville. 2002. Segmenting nonsense: an event-related potential index of perceived onsets in continuous speech. *Nature Neuroscience* 5:700–703.

Sansavini, Alessandra, Josiane Bertoncini, and Giuliana Giovanelli. 1997. Newborns discriminate the rhythm of multisyllabic stressed words. *Developmental Psychology* 33:3–11.

Sebastián-Gallés, Núria, M. A. Martí, M. Carreiras, and F. Cuetos. 2000. LEXESP: Una base de datos informatizada del Español.

Segal, Osnat, and Liat Kishon-Rabin. 2012. Evidence for language-specific influence on the preference of stress patterns in infants learning an iambic language (Hebrew). *Journal of Speech, Language, and Hearing Research* 55:1329–1341.

Seidl, Amanda, and Elizabeth K. Johnson. 2006. Infant word segmentation revisited: Edge alignment facilitates target extraction. *Developmental Science* 9:565–573.

Selkirk, Elisabeth O. 1980. The role of prosodic categories in English word stress. *Linguistic Inquiry* 563–605.

Sezer, Engin. 1981. On non-final stress in Turkish. *Journal of Turkish Studies* 5:61–69.

Shahidullah, Sara, and Peter G. Hepper. 1994. Frequency discrimination by the fetus. *Early Human Development* 36:13–26.

Shannon, Claude E. 1948. A mathematical theory of communication. *Bell System Technical Journal* 27:379–423.

Shannon, Claude E., and Warren Weaver. 1949. *The mathematical theory of communication*. Urbana: University of Illinois Press.

Shelton, Michael, Chip Gerfen, and Nicolas Gutierrez Palma. 2012. The interaction of subsyllabic encoding and stress assignment: A new examination of an old problem in Spanish. *Language and Cognitive Processes* 27:1459–1478.

Shukla, Mohinish, Katherine S. White, and Richard N. Aslin. 2011. Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-month-old infants. *Proceedings of the National Academy of Sciences* 108:6038–6043.

Sikström, Sverker. 2006. The isolation, primacy, and recency effects predicted by an adaptive ltd/ltp threshold in postsynaptic cells. *Cognitive Science* 30:243–275.

Siptár, Péter, and Miklós Törkenczy. 2007. *The phonology of Hungarian*. The Phonology of the World's Languages. Oxford University Press, Oxford.

Sluijter, Agaath M. C., and Vincent J. J. P. van Heuven. 1996a. Acoustic correlates of linguistic stress and accent in Dutch and American English. In *Fourth International Conference on Spoken Language Processing*, ed. H. Timothy Bunnell and William Idsardi, 630–633.

Sluijter, Agaath M. C., and Vincent J. J. P. van Heuven. 1996b. Spectral balance as an acoustic correlate of linguistic stress. *The Journal of the Acoustical Society of America* 100:2471–2485.

Sluijter, Agaath M. C., Vincent J. J. P. Van Heuven, and Jos J. A. Pacilly. 1997. Spectral balance as a cue in the perception of linguistic stress. *The Journal of the Acoustical Society of America* 101:503–513.

Smith, Kirk H. 1966. Grammatical intrusions in the recall of structured letter pairs: Mediated transfer or position learning? *Journal of Experimental Psychology* 72:580.

Soderstrom, Melanie, Amanda Seidl, Deborah G Kemler Nelson, and Peter W. Jusczyk. 2003. The prosodic bootstrapping of phrases: Evidence from prelinguistic infants. *Journal of Memory and Language* 49:249–267.

Stetson, R. H. 1928. *Rhythm and the characteristic pronunciation of a language*, 379–382. Springer.

Suomi, Kari, James M. McQueen, and Anne Cutler. 1997. Vowel harmony and speech segmentation in Finnish. *Journal of Memory and Language* 36:422–444.

Swingley, Daniel. 2005. Statistical clustering and the contents of the infant vocabulary. *Cognitive psychology* 50:86–132.

Szende, Tamás. 1994. Hungarian. *Journal of the International Phonetic Association* 24:91–94.

Szinnyei, József. 1912. *Ungarische sprachlehre*. GJ Göschen.

Taft, Lori Alice. 1984. Prosodic constraints and lexical parsing strategies. Doctoral Dissertation, University of Massachusetts.

Thiessen, Erik D., and Jenny R. Saffran. 2003. When cues collide: use of stress and statistical cues to word boundaries by 7-to 9-month-old infants. *Developmental Psychology* 39:706–716.

Toro, Juan M., Núria Sebastián-Gallés, and Sven L. Mattys. 2009. The role of perceptual salience during the segmentation of connected speech. *European Journal of Cognitive Psychology* 21:786–800.

Toro, Juan M., Josep B. Trobalon, and Núria Sebastián-Gallés. 2003. The use of prosodic cues in language discrimination tasks by rats. *Animal Cognition* 6:131–136.

Toro-Soto, Juan M., Antoni Rodriguez-Fornells, and Núria Sebastián-Gallés. 2007. Stress placement and word segmentation by Spanish speakers. *Psicologica* 28:167–176.

Treiman, Rebecca, and Catalina Danis. 1988. Syllabification of intervocalic consonants. *Journal of Memory and Language* 27:87–104.

Trommelen, Mieke, and Wim Zonneveld. 1989. *Klemtoon en metrische fonologie*. Dick Coutinho, Muiderberg.

Trubetzkoy, Nikolaj. 1939/1969. *Principles of phonology*. University of California Press, Berkeley and Los Angeles. Translated by Christiane A.M. Baltaxe.

Tyler, Michael D. 2006. French listeners can use stress to segment words in an artificial language. In *Proceedings of the 11th Australasian International Conference on Speech Science and Technology*, ed. Paul Warren and Catherine I. Watson, 222–227.

Tyler, Michael D., and Anne Cutler. 2009. Cross-language differences in cue use for speech segmentation. *The Journal of the Acoustical Society of America* 126:367–376.

Underhill, Robert. 1976. *Turkish grammar*. The MIT Press.

Vaissière, Jacqueline. 1983. Language-independent prosodic features. In *Prosody: Models and measurements*, 53–66. Springer Verlag.

Varga, László. 1983. Hungarian sentence prosody: an outline. *Folia Linguistica* 17:117–152.

Varga, László. 1996. Hungarian intonation contours. *The Even Yearbook* 2:111–144.

Varga, László. 2002. *Intonation and stress: evidence from Hungarian*. Palgrave Macmillan.

Veenker, Theo G. J. 2011. The Zep experimental control application. Software, version 0.8. Utrecht Institute of Linguistics OTS, Utrecht University.

Vroomen, Jean, and Beatrice de Gelder. 1995. Metrical segmentation and lexical inhibition in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance* 21:98–108.

Vroomen, Jean, Jyrki Tuomainen, and Beatrice de Gelder. 1998. The roles of word stress and vowel harmony in speech segmentation. *Journal of Memory and Language* 38:133–149.

Waltermire, Mark. 2004. The effect of syllable weight on the determination of spoken stress in Spanish. In *Laboratory approaches to Spanish phonology*, ed. Timothy L. Face, 171–191. Mouton de Gruyter Berlin.

Weber, Christiane, Anja Hahne, Manuela Friedrich, and Angela D. Friederici. 2004. Discrimination of word stress in early infant perception: Electrophysiological evidence. *Cognitive Brain Research* 18:149–161.

White, Laurence, Sven L. Mattys, and Lukas Wiget. 2012. Language categorization by adults is based on sensitivity to durational cues, not rhythm class. *Journal of Memory and Language* 66:665–679.

White, Laurence, James F. Melhorn, and Sven L. Mattys. 2010. Segmentation by lexical subtraction in Hungarian speakers of second-language English. *The Quarterly Journal of Experimental Psychology* 63:544–554.

Wierzchowska, Bożena. 1971. *Wymowa Polska.* Państwowe Zakłady wydawnictw szkolnych.

Wightman, Colin W., Stefanie Shattuck-Hufnagel, Mari Ostendorf, and Patti J. Price. 1992. Segmental durations in the vicinity of prosodic phrase boundaries. *The Journal of the Acoustical Society of America* 91:1707–1717.

Zanto, Theodore P., Joel S. Snyder, and Edward W. Large. 2006. Neural correlates of rhythmic expectancy. *Advances in Cognitive Psychology* 2:221–231.

Zimmer, Karl, and Orhan Orgun. 1992. Turkish. *Journal of the International Phonetic Association* 22:43–45.

van Zonneveld, Ron. 1985. Word rhythm and the Janus syllable. In *Advances in nonlinear phonology*, ed. Harry van der Hulst and Norval Smith, 133–142. Dordrecht: Foris Publications.

Nederlandse samenvatting

Deel I Het is moeilijk voor te stellen dat gesproken taal niet uit losse woorden bestaat, maar een lange stroom van elkaar overlappende en variabele klanken is, zonder aanduiding waar woorden beginnen en eindigen. Het horen van losse woorden, alsof ze gescheiden zijn door pauzes, is een illusie die ingegeven wordt door de talige bekwaamheid van de luisteraar. Tijdens de spraakverwerking gebruikt hij structurele kenmerken van de eigen taal om spraak in losse woorden op te delen; 'woordsegmentatie', vaak hier kortweg 'segmentatie'. Deze dissertatie bepleit dat klemtoon een rol heeft in segmentatie en dat het gebruik van klemtoon bij segmentatie taalspecifiek is, dus toegespitst op de verwerking van de eigen taal.

Al in 1939 merkte Trubetzkoy op dat woordklemtoon verschillende functies heeft; die van betekenisonderscheid en die van woordgrensmarkering. In een grote groep talen staat klemtoon namelijk op een vaste afstand van de woordgrens, waarmee het deze lijkt te markeren. Enige tientallen jaren later kreeg deze veronderstelling een nieuwe psycholinguïstische interpretatie, omdat bleek dat woordgrensmarkering van groter belang is voor spraakverwerking dan eerder gedacht. Behalve grensmarkering lijkt klemtoon bovendien een handig houvast omdat klemtoondragende lettergrepen opvallen ten opzichte van onbeklemtoonde lettergrepen: ze zijn langer in duur, luider in de hogere frequenties van het spectrum, en ze worden vaak gemarkeerd door toonhoogte of toonhoogteveranderingen. Klemtoon blijft opvallend, zelfs wanneer de rest van de spraak door een storm, feestgedruis of een slechte telefoonverbinding onverstaanbaar is. Hoofdstuk 2 bespreekt de wetenschappelijke context van deze dissertatie, waarna het in Deel II een taaltypologisch karakter krijgt door het gebruik van klemtoon in segmentatie te vergelijken tussen het Hongaars (klemtoon op de eerste lettergreep), Pools (klemtoon op de voorlaatste lettergreep), Turks (klemtoon 'vaak' op de laatste lettergreep) en Nederlands (klemtoon 'vaak' op de voorlaatste lettergreep). De kenmerken van deze talen worden besproken in Hoofdstuk 3; klemtoontypologisch verschillen ze langs de volgende dimensies:

1) grensverbinding: links of rechts
2) periferaliteit: perifeer of naburig
3) positionele regelmatigheid: strikt of variabel

Deze dimensies hebben potentiëel implicaties voor het gebruik van klemtoon in segmentatie. Als klemtoon rechts verbonden is, *voorspelt* hij de aankomende woordgrens, terwijl deze relatie aan de linkerkant *terugkijkend* is. Punten 2 en 3 impliceren een mate van betrouwbaarheid; als klemtoon positioneel variabel en/of niet perifeer aan de woordgrens verbonden is, is de relatie tussen klemtoon en woordgrens niet een-op-een. De luisteraar kan in principe niet volledig vertrouwen op klemtoon als aanwijzing voor de woordgrens.

De Taalspecifieke Metronoom Hypothese, centraal in deze dissertatie, voorspelt dat luisteraars de klemtoon van hun eigen taal gebruiken in segmentatie, als een beat die woordgrenzen aanduidt, zelfs voorspelt. A priori hoeft klemtoon geen rol te spelen in segmentatie, en zelfs als hij dat wel doet (en de literatuur bevat aanwijzingen dat dat zo is), kan het zijn dat het gebruik van klemtoon in segmentatie universeel, dat wil zeggen, taalonafhankelijk is. Met dit in gedachten wordt de hypothese getest in een taaloverschrijdend experiment.

Deel II De Taalspecifieke Metronoom wordt in Hoofdstuk 4 getest met een 'nonwoordspotexperiment'. In zo'n experiment moeten de deelnemers stukjes op zich betekenisloze taal herkennen in langere eenheden. In de hier gehanteerde versie luisteren Hongaarse, Turkse, Poolse en Nederlandse deelnemers naar een taal die voor dit experiment is ontworpen. De deelnemers leren aan het begin van de taak twee nonwoorden: 'mernel' en 'darnam'. Vervolgens horen ze korte 'zinnen' die wel of niet een van deze nonwoorden bevatten, zoals 'badusudarnam' of 'biresirefi'. Ze worden geacht zo gauw ze een van de nonwoorden horen, op een knopje te drukken. Cruciaal voor dit experiment is dat een deel van de 'zinnen' het klemtoonpatroon heeft van de taal van de deelnemer. Aan de reactietijden wordt gemeten of de deelnemer het nonwoord sneller segmenteert als dit volgt op een stukje dat qua klemtoon met de eigen moedertaal overeenstemt, zodat er een woordgrens kan worden verwacht.

Uit de analyses van de reactietijden blijkt dat segmentatie door het taalspecifieke klemtoonpatroon vergemakkelijkt wordt: over het algemeen reageren deelnemers sneller als het woord het eigen standaardpatroon heeft en volgt op een context met het eigen standaardpatroon. Turkse deelnemers reageren sneller als het nonwoord volgt op een context als 'baduSU' (het Turkse standaardpatroon) dan op een context als 'BAdusu' of 'baDUsu', terwijl Nederlandse deelnemers juist sneller reageren als het nonwoord volgt op 'baDUsu' (het Nederlandse standaardpatroon). Dit volgt precies de voorspelling van de Taalspecifieke Metronoom. Bij Hongaarse deelnemers heeft de context geen effect; wat in lijn is met de verwachting dat een linksverbonden klemtoon geen verwachtingen schept voor de rechter woordgrens. Onverwacht is er geen bewijs dat Poolse luisteraars een woordgrens verwachten na 'baDUsu', terwijl klemtoon in het Pools strikt op de voorlaatste lettergreep valt. Een vervolgexperiment zou uitkomst moeten bieden.

Een vraag die naar aanleiding van de taaltypologische achtergrond van het onderzoek gesteld kan worden, is of verwachtingen in talen met positioneel variabele klemtoon strikt of gradueel zijn. Deze vraag ontstaat door een suggestie van gradatie in de Turkse groep, maar niet in de Nederlandse. Hiertoe werd aanvullend een groep sprekers van het Spaans getest, een taal waarin de voorlaatste of de laatste lettergreep beklemtoond is, afhankelijk van de lettergreepstructuur. De resultaten zijn verrassend: als de context hetzelfde patroon heeft als het nonwoord wordt het nonwoord sneller ontdekt. Een vergelijkbaar patroon is niet te zien in de andere vier talen en dit schijnbaar ritmische effect vraagt dus om meer onderzoek.

In Hoofdstuk 5 wordt de 'Taalspecifieke Metrische Groeperingshypothese' getest. Deze voorspelt dat deelnemers 'zinnen' zo segmenteren dat elk van de gesegmenteerde woorden maximaal in overeenstemming is met het taalspecifieke klemtoonpatroon. In dit geval gaat het dus niet om de relatie van klemtoon tot een van de woordgrenzen, maar om het ritmische patroon van het woord als geheel. Turkse en Nederlandse deelnemers worden vergeleken, omdat in deze talen positionele variabiliteit van klemtoon bestaat en de segmentaties dus kunnen variëren.

De 'zinnen' die de deelnemers segmenteren zijn die uit Hoofdstuk 4 zonder 'darnam' en 'mernel', zoals 'biresirefi'. Zo kunnen de resultaten van Hoofdstuk 5 en 4 worden vergeleken. Uit de resultaten blijken verschillen tussen de talen die op basis van hun kenmerken kunnen worden verwacht. Turkse deelnemers hebben bovendien meer overeenstemming over optimale segmentaties dan Nederlandse, waar meer variatie gevonden wordt, en ze segmenteren langere woorden, ook dit in overeenstemming met woordlengtes en klemtoonvariabiliteit in de respectievelijke talen. Een opvallend verschil met de faciliterende contexten in Hoofdstuk 4 is dat deelnemers een grens plaatsen tussen twee aangrenzende klemtonen en dit verschil valt samen met dat tussen de hypotheses: de Taalspecifieke Metronoom voorspelt dat deelnemers tijdens het luisteren in één richting verwachtingen hebben op basis van woordklemtoon. De Metrische Groeperingshypothese daarentegen voorspelt dat deelnemers, wanneer ze tijd gegeven is, de gehele 'zin' optimaal segmenteren, en dus beide richtingen beschouwen.

Deel III In Hoofdstuk 6 verschuift de focus van verwerken naar leren. Doel hiervan is om te zien of het in principe mogelijk is om taalspecifieke woordklemtoonpatronen te distilleren uit ongesegmenteerde spraak. Net als bij segmentatie voor volwassenen zou klemtoon de segmentatie voor taalleerders kunnen vergemakkelijken, al dreigt hier een kip-en-eiprobleem: hoe kan een leerder, met name een leerder zonder kennis van woorden, woordklemtoon gebruiken om een begin te maken met woordleren?

Een mogelijkheid is dat de ongesegmenteerde input voor zo'n vroege leerder informatie bevat om woordklemtoon te leren zonder gebruik van zoiets als een mini-woordenschat. Om dit te onderzoeken is een verzameling van spontane spraak omgezet naar reeksen van beklemtoonde en onbeklemtoonde lettergrepen, gescheiden door frasegrenzen.

Hoofdstuk 6 analyseert deze verzameling op de informatiewaarde van klemtoonpatronen op twee- en drielettergrepige 'brokken', specifiek hun onderlinge statistische verbondenheid. Deze verbondenheid wordt onder andere uitgedrukt in de relatieve frequentie van de brokken vergeleken met hun frequentie als dezelfde elementen willekeurig verspreid waren geweest, in plaats van door taal gestructureerd. Uit de analyse blijken de statistische relaties echter ongeschikt voor het ontdekken van klemtoonpatronen. De analyse suggereert dat de informatie niet binnen grenzen, maar bij de grenzen gezocht moet worden. Deze aanpak wordt in Hoofdstuk 7 gekozen.

In plaats van naar de relatie van alle samen voorkomende lettergrepen wordt gekeken naar de relatie van de beklemtoonde lettergreep tot de frasegrens. Het argument is dat zowel klemtoon als frasegrenzen in het oog vallend zijn. Uit de literatuur blijkt bovendien dat luisteraars en leerders aandacht besteden aan grenzen bij het leren van regels en patronen, en aandacht besteden aan de onderlinge statistische relaties van elementen (als in Hoofdstuk 6) bij het leren van woorden. Het leren van klemtoon, een patroon, is dus waarschijnlijk makkelijker aan een grens. De deelvraag is nu of de statistische relatie van klemtoon tot de frasegrens informatief is voor het leren van het klemtoonpatroon op woordniveau. Als een frase altijd begint met een beklemtoonde lettergreep, maar eindigt met meerdere verschillende patronen die ongeveer even vaak voorkomen, is het onderscheidingsvermogen van het patroon aan het begin van de frase hoog en aan het eind van de frase laag. De resultaten laten zien dat er voor het Nederlands meerdere patronen opvallen; het woordklemtoonpatroon onderscheidt zich niet goed op fraseniveau. In het Turks is de frasefinale klemtoon het meest opvallende patroon, overeenkomend met de woordklemtoon. Hier lijkt het klemtoonpatroon dus te onderscheiden te zijn op fraseniveau. Het feit echter dat slechts in een van de twee talen de analyse erop lijkt te wijzen dat woordklemtoon herkenbaar is op fraseniveau vraagt op zijn minst om een uitgebreidere analyse met meerdere talen.

Deel IV Deze dissertatie heeft de relatie van klemtoon tot de woordgrens nader bekeken als taalspecifiek instrument in woordsegmentatie en heeft geanalyseerd in hoeverre deze relatie uit ongesegmenteerde spraak te distilleren is. De taaltypologische neiging van klemtoon om zich te verbinden aan een van de woordgrenzen blijkt zijn nut voor segmentatie te bewijzen, en het ritmische kenmerk van klemtoon creëert structurele temporele verwachtingspatronen die het luisteren vergemakkelijken. Deze verwachtingspatronen blijken taalspecifiek te zijn, dus toegespitst op de verwerking van de eigen taal. Tegelijk blijkt het leren van de relatie van klemtoon tot de woordgrens niet eenvoudig wanneer de leerder niet over een woordenschat beschikt.

Curriculum Vitae

Sandrien van Ommen was born in 1985 in Apeldoorn, the Netherlands. Her pre-university education included the study of classical Greek and Latin and focused on science subjects. Learning Italian through immersion during a year of voluntary service in Italy strengthened her interest in general linguistics, which inspired her to enroll in the BA programme at the University of Groningen in 2004. Her bachelor's thesis on the phonetics of emphatic lengthening concluded her degree and Sandrien subsequently enrolled in the research master's program at the same university. She spent a semester at the University of Leuven studying Dutch language varieties and she worked on several different projects concerning among others language variation, processing, computation and acquisition, mostly in phonology. Sandrien concluded her research master with a thesis on the mutual intelligibility of Danish and Swedish, focusing on intervocalic plosives. In 2010, she obtained her master's degree cum laude.

This thesis is the result of work Sandrien van Ommen carried out between 2011 and 2015 as a PhD-candidate at UiL OTS, Utrecht University. She presented her work at many international conferences and workshops as well as intradepartmental meetings of different universities. As a visiting researcher she worked at New York University (NYU) in the United States and at Pompeu Fabra University in Barcelona, Spain. Sandrien currently works as a postdoc at the Université Paris Descartes, continuing her research on prosodic processing in a cross-linguistic perspective.