THE ROLES OF LEXICALITY, PRIMARY STRESS AND VOWEL HARMONY IN HUNGARIAN WORD SEGMENTATION

BA Thesis Linguistics

Domonkos Király (12131598)
Supervisor: Suki S.Y. Yiu
Submitted: June 21st, 2020

UNIVERSITEIT VAN AMSTERDAM

Abstract

This thesis explores some mechanisms of word segmentation in Hungarian. I hypothesised that lexicality, word stress and vowel harmony would serve as important cues for auditory word segmentation. These features, specifically, wordhood, target stress and prefix-target disharmony, are predicted to result in faster detection times of a prefix-stem boundary. Hungarian listeners took part in a detection task with a visual prime using Hungarian words and nonwords. Stimuli consisted of a nonsense prefix and a CVCV target. Participants were told to listen out for the word written on screen and to press a reaction button as quickly as they could when they heard it. Before the target, random fillers were played. A linear mixed effects model analysis for response times as a function of lexicality, stress and vowel harmony revealed that both lexicality and stress had significant facilitative effects on boundary detection, however, vowel disharmony did not decrease reaction times. A possible explanation is that while lexicality is the most important segmentation cue in all languages and mid-word stress has boosted saliency in leftmost-stress languages, a third and theoretically less important segmentation cue simply does not contribute to word segmentation when more robust cues are present.

Acknowledgments

First, I would like to express my sincere thanks to my supervisor, Suki Yiu, for her ongoing support during the entire thesis trajectory, as well as her flexibility and readiness to solve any methodological or other problem. Further, I owe thanks to Dirk Vet, the author of the experiment-software environment in which this study was conducted. He worked very quickly and precisely to tailor the present experiment to the exact needs of my research design. I thank the participants of my experiments, all contacts of mine, without whom I could not have completed this work. Finally, I am also grateful to my fellow students for providing a collegial environment of academic growth, as well as my partner, friends, and family for their support.

Table of contents

1. Introduction

During reading, detecting word boundaries is a straightforward task. In languages which employ the Latin or the Cyrillic script, breaks between words in written text are indicated by a physical space between characters. Alternatively, many writing systems exist where characters are not separated, e.g., in isolating languages such as Mandarin Chinese. When it comes to speaking, however, in no human language do we find a break or audible pause as clear as a space between written words (Klatt, 1979). Despite the continuous nature of spoken language, which is evident from looking at the waveform of any recording, the perception of continuous speech is that of a string of discrete words. The location of word boundaries in fluent speech is clearly far more ambiguous than noticing small gaps between words. Listeners need to rely on an array of different segmentation cues to tell them where one word ends and the next one begins, which is a topic that has been widely studied (e.g. Cutler & Norris, 1988; Davis et al., 2002; Gow & Gordon, 1995; Johnson & Jusczyk, 2001; Kabak et al., 2010; Klatt, 1979; Mattys et al., 2005; Norris et al., 1995; Saffran, 2002; Suomi et al., 1997; Vroomen et al., 1998).

2. Literature review

Prior research has provided evidence for many of the alternative aids on which listeners rely during on-line speech segmentation. Two predominant approaches to speech segmentation and cue categorisation can be identified. One perspective focuses on listeners' reliance on the statistically regular occurrence of acoustic-phonetic cues, i.e., boundaries are consistently marked by changes in the physical properties of the sound signal (e.g. Cutler & Norris, 1988; Johnson & Jusczyk, 2001; Saffran et al., 1996; Saffran, 2002). According to this view, acoustic-phonetic cues are probabilistically associated with certain parts of a word, such as increased intensity on word onsets, or regular metrical stress. Stressing the onset boosts its acoustic saliency, and in stress-initial languages, this signals a word boundary retroactively (Kabak et al., 2010). Speakers have also been shown to exploit other phonological features such as vowel harmony patterns in languages where they are present (Suomi et al., 1997, Vroomen et al., 1998, Kabak et al., 2010). Thus, according to this view, word boundaries are identified exclusively based on the presence of perceptual features such as metrical stress or other acoustic-phonetic properties (for a review, cf. Davis et al., 2002).

According to the opposing perspective, lexical processes and word identification serve as the basis for word segmentation rather than being the result of it (e.g., McClelland & Elman, 1986; Norris et al., 1995). In this view, word segmentation is a lexically driven process, where sublexical cues, both acoustic-phonetic or metrical, are secondary only to lexical cues. During listening, multiple potential parsing solutions are activated, and segmentation is reached when the recognition process settles on the lexically most acceptable candidate, while also considering the context of the utterance and pragmatics (cf. e.g., Mattys et al., 2005).

Both accounts have their own limitations, especially in providing a comprehensive description of multiple segmentation cues, both lexical and sublexical. Purely sublexical views do not include the aiding contributions of contextual and lexical cues and thus fall short of accounting for the segmentation effects of word recognition. As Mattys et al. (2005) note, these approaches generally are not suited for situations in which multiple cues compete or where the same sublexical cues can be interpreted in different ways, leaving the resolution of ambiguous strings to contextual information. The example they give is that of /nɑɪtɹeɪt/, for which 'nitrate' and 'night rate' are equally viable solutions since their pronunciation contains the same acoustic-phonetic information. To decide between the candidates, the semantic context of the utterance and the content of previous utterances must be considered (pharmacy vs. parking garage). Conversely, lexically driven views are problematic for novel or nonsense words, since they do not correspond with anything that is stored in the listener's mental lexicon. The issue gets more complicated if we consider words that are not *yet* stored in the listener's lexicon, as it makes purely lexical accounts incapable of describing (second) language acquisition or the learning of new words. Both second and first language learners face the challenge of segmentation without any lexical knowledge, and therefore all lexically-driven approaches must consider the contributions of sublexical cues.

In an effort to synthesise the sublexically and lexically driven approaches to speech segmentation, Mattys et al. (2005) developed a hierarchical model of boundary cues. The authors' starting point for creating a comprehensive hierarchical model was the assumption that a fitting and complete description of speech segmentation must contain more than just the individual contributions of different cues. Six experiments were conducted in English, the first five of which were structured similarly: they tested the segmentation effects of cues in comparison to one another, using lexical-decision tasks. For example, in their third experiment, which pitted metrical cues (stress) against lexicality, it was revealed that word recognition aided segmentation efficiently and operated independently of metrical cues. In the other experiments, the segmentation facilitation effects of other cues were measured (e.g., Experiment 1: Stress Versus Acoustic-Phonetic Cues; Experiment 2: Stress Versus Phonotactics, etc). In addition to measuring the facilitation effects of lexical and sublexical cues, the authors introduced another factor in their model: signal quality, which describes the 'quality' of the speech stream as it reaches the listener. Low signal quality would entail speech with background noise, interruptions etc., whereas a sample with high signal quality is free from any noise, artifacts or acoustic events that compromise intelligibility. Signal quality as a factor was included in Experiment 6, in the form of a lexical judgement task where signal quality gradually decreased. The inclusion of this dimension enabled the construction of a comprehensive segmentation cue hierarchy that also accounts for varying listening conditions and the intelligibility of the speech sample.

In their resulting model, all segmentation cues of English were divided into three hierarchically ordered tiers: Tier I = lexical (semantics, pragmatics, lexical knowledge), Tier II = segmental (phonotactics, acoustic-phonetics) and Tier III = metrical prosody (word stress). In optimal interpretive conditions, the sentential context and word recognition outrank all other (sublexical) cues. With decreasing signal quality, however, first segmental (Tier II), then metrical prosody (Tier III) cues start to gain more importance and having stronger facilitation effects. In sum, the hierarchy predicts that sublexical cues are always secondary to lexical cues, which they attribute to the "communicative, meaning-related nature of speech" (Mattys et al. 2005, p. 491). Following from the predominance of lexical cues, the authors, similarly to Norris (1994) and Gow & Gordon (1995), conclude that sublexical cues may only facilitate recognition or boost lexical activation, especially with poor signal quality, but are not likely to inhibit any of those processes. Furthermore, their hierarchical model predicts that the more listeners hear from an utterance, cues from Tiers 2-3 are increasingly less likely to be engaged due to the growing presence of contextual and sentential cues. Crucially for the present paper, Mattys et al. (2005) finally conclude that the segmentation cues of any language may be grouped into Tiers and ranked in a hierarchical manner, depending on the importance of each cue. This ranking varies cross-linguistically in both the specific ordering and the availability of cues in the given language. However, as Tier I lexical segmentation cues (wordhood, semantic and syntactic context) are superior to any sublexical cues, lexicality is expected to top the cue hierarchies of all languages. In this thesis, stress and vowel harmony are investigated as potential segmentation cues of Hungarian, alongside word recognition.

2.2. Hungarian segmentation cues and prior empirical studies

Hungarian has a palatal, or front-back vowel harmony system, similarly to Finnish (Suomi et al., 1997). The vowels /y, yː, ø, øː/ belong to the front harmony class, and the vowels /u, uː, o, oː, ɔ, aː/ belong to the back class (Törkenczy, 2011). In addition, there is a third, 'tolerant' group of vowels, consisting of the (phonetically) front vowels /i, iː, ɛ, eː/ (Hyllested, 2017). The latter class is often referred to as "neutral", however, as Hyllested points out, roots containing only tolerant vowels generally pattern with suffixes of front harmony. Whenever a tolerant vowel co-occurs with a non-tolerant vowel, the stem assumes the harmony class of the non-tolerant vowel. This system restricts vowels of the front and back classes from co-occurring within the same uncompounded stem. The harmony system can be illustrated with examples from loanword adaptation into Hungarian. In the centuries following the arrival of Hungarians in the Carpathian Basin in the 9th century, loanwords of German, Latin and Slavic provenance were subjected to vocalic harmonisation, by assimilating vowels in loanwords to either the front or back harmony class. This is why words like the Latin *Angelus* ("angel") became *angyal* in Hungarian (cf. Hyllested, 2017). Hungarian vowel harmony propagates from left to right. Inflectional and derivational suffixes are subject to harmony restrictions, but not prefixes. For example, the prepositional suffix meaning "in" has two variants: back *-ban* and front *-ben*, differing only in the harmony class of the vowel. Words of the back harmony class, for example *ház* ("house") are only allowed to be suffixed with back-class suffixes, so the correct derivation is *házban* ("in a house") and not **házben*. Conversely, front-class words such as *tűz* ("fire") are only allowed to take front-class suffixes: *tűzben* ("in a fire") but not **tűzban*. For adverbial prefixes, there is only one variant for either harmony class, for example, the prefix *rá-* meaning "on" or "over" does not assimilate: *rárak* (back, "put on"), vs. *rájön* (front, "realise"). Thus, since vowels from opposing harmony classes cannot occur within the same uncompounded stem, disharmony is typically associated with a word boundary (Vroomen et al., 1998), which supports the assumption that vowel (dis)harmony plays an important role in word segmentation, albeit only in languages that exhibit vowel harmony features.

Concerning stress, primary stress in Hungarian content words consistently falls on the first syllable and is associated with the left boundary of the word (Varga, 2002). Research on Finnish, another leftmost-stress language has shown that with the aid of the stress cue, listeners detected lexical targets significantly faster than in conditions where the stress cue was unavailable or did not signal any boundary (Vroomen et al. 1998). The study by Vroomen, Tuomainen and De Gelder had a similar purpose to the present paper: in the first two out of three experiments, the authors investigated the relative importance of simultaneously occurring stress and vowel harmony cues. In Experiment 1, the authors replicated the results of Suomi et al. (1997) with a target detection task. Stimuli consisted of an existing Finnish word, e.g. *hymy* "smile" with a monosyllabic nonsense prefix, resulting in, e.g., *PUhymy*, in this case with a disharmonious prefix. Since in Finnish, the leftmost syllable carries main stress, prefixes were stressed, and the target words themselves had no stress cue. Vowel disharmony between prefix and target was confirmed to significantly boost detection. In Experiment 2, the authors introduced the variable of stress alongside vowel harmony. There were four possible stimuli in this detection task: 1. harmonious prefix, first-stress (*PYhymy*), 2. harmonious prefix, second-stress (*pyHYmy*), 3. disharmonious prefix, first-stress (*PUhymy*) and 4. disharmonious prefix, second-stress (*puHYmy*). The contributions of vowel disharmony lost importance in this experiment, as they turned out to be secondary only to stress cues. Furthermore, the facilitation effect of vowel disharmony was only found in items where the stress cue did not mark a boundary (prefix-stress). Concerning the segmentation cue hierarchy proposed by Mattys et al. (2005), this would imply that the contribution of vowel harmony would have to be ranked below the contributions of word stress and only exploited when higher-tier cues are insufficient or unavailable. As for the supposed effect of lexicality, however, this design does not reveal any facilitation, since real words were used exclusively. For any such effect to be determined, segmentation times of real words would have to be evaluated against those of nonwords. Given that both Finnish and Hungarian belong to the Uralic language family, and both have word-initial stress and vowel harmony in their sublexical cue inventories, it is reasonable to expect finding similar effects in Hungarian, as the detection-boosting qualities of these cues have already been documented in Finnish.

To further investigate the roles of stress and vowel harmony in speech segmentation, Kabak et al. (2010) designed a cross-linguistic study with Turkish and French speakers who took part in a nonword segmentation task. Similarly to Finnish and Hungarian, Turkish also has a backness-based vowel harmony system, where vowels from the front class cannot co-occur with vowels from the back class within the same stem, though exceptions exist in the form of loanwords, place names etc (Kabak & Vogel, 2001). In addition, both Turkish and French are known to consistently assign word stress to the rightmost edge of the word, signalling an immediately following word boundary (Kabak et al., 2010). This progressive boundary assignment is the opposite of the Finnish/Hungarian system, where a boundary is signalled regressively by leftmost stress. The authors motivate their choice of languages with the fact that in leftmost-stress languages, it is impossible to differentiate between the facilitative effects of stress and the general prominence of word onsets. Hence, by having right-stress languages which are still predictable in their stress assignment, it is possible to attribute the potential facilitation to stress only. As French does not have any provisions for vowel co-occurrence in the way that Turkish, Finnish and Hungarian do, the authors expected French speakers to benefit from stress cues but not from vowel harmony cues. Turkish speakers, on the other hand, were expected to use both types of boundary cue in speech segmentation. This theory is in line with the hierarchical approach and underlines that cue rankings are language-specific, as it posits that speakers only benefit from features that are actually present in their language. In their experiment, Kabak et al. (2010) presented participants with 5-syllable nonsense sequences (nonwords in both French and Turkish) with varying stress position and vowel harmony relationships between syllables. Each sequence contained a target and the participant's task was to press a reaction button as soon as they heard the target, resulting in reaction time and accuracy data. The findings of this study provided strong support for the view that languages with predictable stress assignment rank word stress as an important segmentation cue. Kabak et al. (2010) also conclude that vowel disharmony significantly aided detection, resulting in higher accuracy rates and lower reaction times for Turkish but not French participants.

2.3. Hypotheses

Based on the research summarised above, we can theorise about the roles of word stress and vowel harmony in Hungarian word segmentation. As it has been shown for Finnish, Turkish and French, speakers of languages with predictable (non-lexical) stress benefit from stress cues during word segmentation. Because Hungarian too belongs to this group of languages, we can expect to find the same result in the present paper: detection will be more efficient when it is aligned with the left edge of the target and thus regressively marks a boundary, as opposed to when it is not. Turkish and Finnish have also been proven to rely on vowel harmony to at least some extent, in cases where a word boundary was marked by a switch from one harmony class to another. Being a member of the Uralic language family alongside Finnish, Hungarian too is predicted to use disharmony cues. Lastly, it has been shown that more than anything, segmentation is driven by word recognition. Nevertheless, prior segmentation research has largely employed either words or nonwords exclusively, so in order to gain insight into how much wordhood facilitates segmentation, both categories need to be investigated and compared. This would also reveal if sublexical cues only play a role in the absence of lexical cues, as the hierarchical approach of Mattys et al. (2005) predicts.

3. Methods

To investigate the contributions of stress and vowel harmony to the segmentation of words and nonwords, a target detection task with a visual prime was conducted. The experiment had a 2x2x2 design, with three binary independent variables: *lexicality* (word/nonword), *harmony* (harmony/disharmony) and *stress* (prefix/target). This design is similar to that of Kabak, Maniwa & Kazanina (2010), but instead of language, the dimension of lexicality is added, which serves to compare the effect of segmentation cues in both words and nonwords.

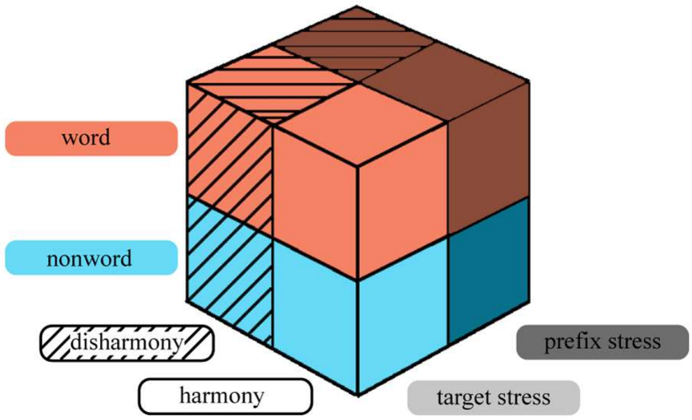

Figure 1. *Matrix of variables plotted in 3D. The 2x2x2 design results in 8 possible combinations and therefore 8 experimental groups.*

3.1. Participants

25 native Hungarian speakers were recruited for the experiment, 16 female and 9 male with a mean age of 34. Their ages ranged from 18 to 54. None of the participants reported hearing problems or visual impairments that might have compromised their participation. Subjects were recruited through contacts of the experimenter, most of them being residents of Vas County in the Western Transdanubia region of Hungary. During the collection of data, active measures against the spread of the novel SARS-CoV-2 were respected at all times.

3.2. Materials

The design of stimuli was partly based on the studies by Vroomen, Tuomainen & De Gelder (1998) and Kabak, Maniwa & Kazanina (2010). Every test item consisted of a CV nonsense prefix and a CVCV target. In total, 96 unique experimental items were spoken and recorded by the experimenter, comprising 48 words and 48 nonwords (lexicality). For half the items in the groups, the assigned prefix contained a vowel that matched the harmony of the target, in the other half, the vowel contrasted with the harmony of the target (prefix harmony). To create a balance between words of the front and back harmony classes, there was a further divisions made in the ‘mismatch’ condition: this group was made up of 6 front-back and 6 back-front items, where the two values describe the harmony class of the prefix and the target, respectively. Additionally, those groups were halved again and were assigned stress either on the first syllable, i.e., the prefix, or the second syllable, i.e., the first syllable of the target (stress). There were 12 items for every unique combination of the binary values of the variables, as illustrated in Table 1.

Table 1. *Example stimuli for every experimental group*

Transcript	Lexicality	Prefix harmony	Stress	N
TUkapu	word	harmony	prefix	12
tuKApu	word	harmony	target	12
TYkapu	word	disharmony	prefix	12
tyKApu	word	disharmony	target	12
TUbola	nonword	harmony	prefix	12
tuBOla	nonword	harmony	prefix	12
TYbola	nonword	disharmony	target	12
tyBOla	nonword	disharmony	prefix	12

To create the experimental list, the order of the 96 experimental items was randomised. Then, before each experimental item, a pseudorandom number of filler items between 1 and 5 was inserted. This randomisation was needed in order to eliminate any patterns or temporal regularities in the detection task (Swallow & Jiang, 2010), and resembles the design of Vroomen et al. (1998), who told their participants that the stimuli “sometimes” contained the target word. This way, participants focus on detecting the target word without knowing when to expect it. Filler items were either words from the list of experimental items (but not containing the target of the current trial) or words from a separate list of 96 items which did not contain any target words. 96 filler items were recorded, with the same parameters as the examples in Table 1. These items were used dilute the proportion of experimental items to filler items. All stimuli were spoken by the experimenter and recorded using a digital audio workstation, with a PreSonus M7 large-diaphragm condenser microphone, connected to a USB audio interface and set at 44.1 kHz sampling frequency and 16-bit-depth. Participants listened to the stimuli over Sony WH-1000MX3 headphones.

3.3. Procedure

Participants were seated in a quiet room in front of a computer and were instructed to wear the headphones provided by the experimenter. The experiment was built and run in ED (Vet, 2021), a software for designing and running such and similar experiments. At the start of every trial, a visual target word, e.g., *KAPU* (lexical word for "gate"), was displayed on the screen, without any prefix, spelled in accordance with the orthographic conventions of Hungarian. Subjects were told that they were going to hear words over the headphones, in one of which the target would be 'embedded'. They were instructed to press the spacebar to start the trial and stimuli would be played with a pause of 1000 ms between the offset of a word and the onset of the following. Participants' task was to press the spacebar as quickly as they could upon hearing the target but to keep listening without pressing any keys if they did not hear the target. Much like in Vroomen et al. (1998), response times were measured from the offset of each word, by subtracting the duration of each item from the response time measured from the onset. To allow subjects to become familiar with the procedure, a short practice session of 10 training items with 3 targets was provided before the start of the experiment. In total, every participant heard 374 auditory stimuli over the approximately 20-minute experiment, including the 96 targets.

3.4. Predictions

If target lexicality is the most reliable segmentation cue, as predicted by the cue hierarchy of Mattys et al. (2005), response times for words should be consistently lower than for nonwords, in four experimental conditions: prefix stress + harmony, prefix stress + disharmony, target stress + harmony and target stress + disharmony. Since leftmost stress is associated with an immediately preceding word boundary, it can be predicted that segmentation will be quicker when words stress is assigned to the first syllable of the target (second syllable overall), than when it is assigned to the prefix (first syllable overall). Furthermore, as concluded by Mattys et al. (2005), languages in which the placement of stress is entirely predictable, like Turkish, Finnish or indeed Hungarian, word stress is likely to be placed higher in the segmentation cue hierarchy than other sublexical cues. Therefore, the contribution of stress to segmentation should show in the form of significantly lower response times for target stress conditions, for both words and nonwords, as well as harmonious and disharmonious prefixes. As the rules of vowel harmony predict, if a vowel from one harmony class is followed by a vowel from the opposing harmony class, it is likely that there is a word boundary between them, as there cannot be a mix of two harmony classes within the same stem. Thus, if the prefix of a stimulus is disharmonious with the target, detection of the word boundary should be faster than in a case of harmony, resulting in lower response times for disharmonious items. That said, the presence of two hypothetically more important cues might cause the contributions of vowel harmony to become insignificant in some conditions. The theoretical ordering of the three cues can be visualised simply as in (1).

(1) **lexicality** >> stress >> vowel harmony

4. Results

From the experimental list with 374 items, response times for the 96 targets were extracted for each participant. Instances where a participant did not react to a target item (2.75%) were coded as errors and were excluded from the analysis. Since there was a 1000 ms pause between the offset of a stimulus and the onset of the next, no response could be slower than 1 second as measured from the offset and therefore no response times were treated as outliers. Exclusion criteria for items and participants were set up in the vein of Vroomen et al.'s (1998) target detection task (Experiment 1), however, no item was missed by more than half of the participants and no participant made more than 50% errors, and therefore all items and participants could be included in the analysis. Response times as the dependent variable were analysed with a linear mixed effects model, fitted in R (R Core Team, 2020) using the *lme4* package (Bates et al., 2015). Using this method, it is possible to incorporate random effects, which makes it more powerful than a traditional analysis of variance. To reveal the contributions of segmentation cues, a model was fitted for reaction time from word offset as a function of lexicality, stress, and vowel harmony, with stimuli and participants as random effects.

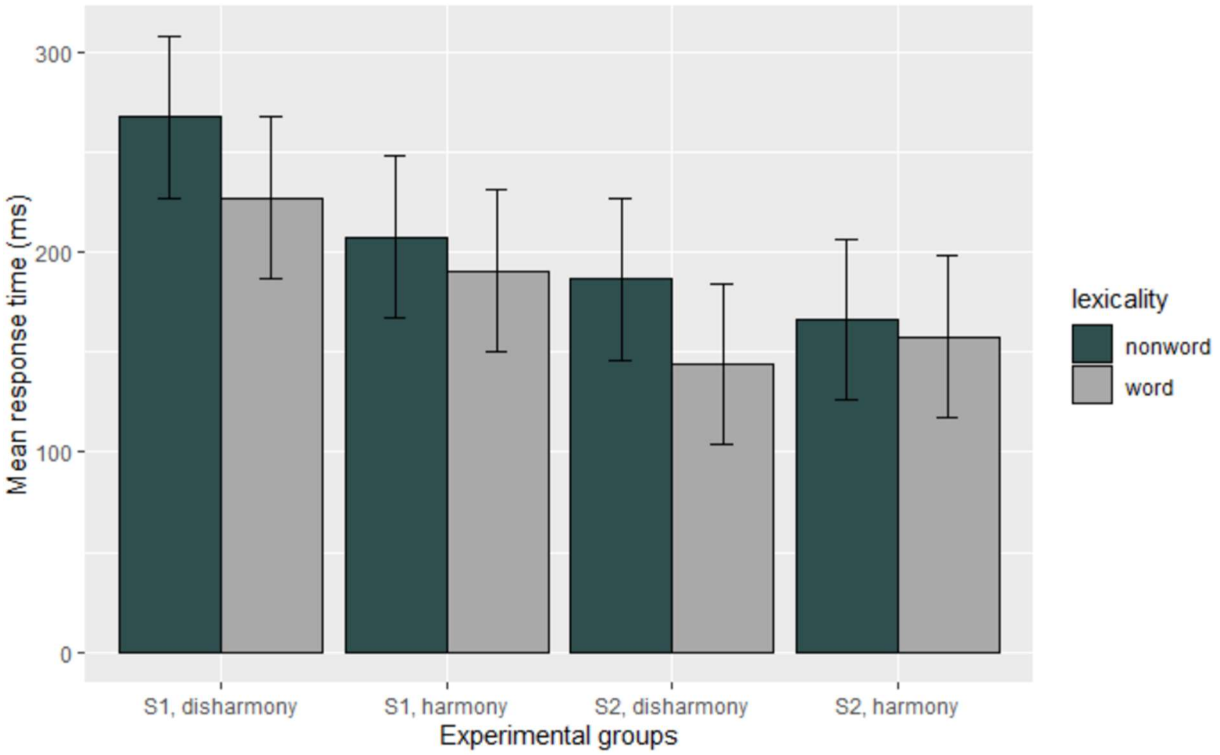

Figure 2. *Mean response times for words and nonwords in four conditions: S1 and S2 stand for items where main stress fell on either the first or the second syllable of the item respectively, harmony and disharmony refer to the relationship between the prefix vowel and the harmony class of the target. Error bars of 1 standard deviation are included.*

Mean response times for all eight experimental groups are reported in Figure 2. Concerning each variable individually, the means are shown in Table 2. LMER results revealed statistical significance for lexicality and stress, and marginal significance for vowel harmony, as follows. Lexicality: β = -28.38, 95% C.I. = -54.34 … -2.42 ms, t = -2.129, p = 0.0360, stress: β = -62.37, 95% C.I. = -88.35 … -36.39 ms, t = -4.674, p = 0.0000102 and harmony: β = -25.43, 95% C.I. = -51.41 … 0.55 ms, t = -2.443, p = 0.0598. On average, words were identified more quickly than nonwords (words: 180 ms, nonwords: 206 ms). Stimuli where the target was stressed were detected faster than leftmost-stress stimuli (target: 164 ms, nontarget: 222 ms). In the harmony class, results were the opposite of what was predicted, since stimuli with a harmonious prefix were identified quicker than stimuli with disharmony between prefix and target: harmony: 181 ms, disharmony: 204 ms.

Table 2. *Mean response times for all experimental groups*

Lexicality	Prefix harmony	Target stress	Mean RT in ms
nonword	disharmony	nontarget	267.6616
word	disharmony	nontarget	227.3898
nonword	harmony	nontarget	207.7279
word	harmony	nontarget	190.8185
nonword	disharmony	target	186.4182
word	disharmony	target	144.125
nonword	harmony	target	166.1185
word	harmony	target	157.7898

In separate analyses for the word and nonword groups, word stress revealed significance ($p < 0.001$) for both conditions, whereas there was a slight difference in the effect of vowel harmony: in the word group, vowel harmony only had a marginally significant effect (p = 0.0644), but in the nonword group, the effect was significant with p = 0.039893. Again, these results are still the exact opposite of what was predicted, as harmony boosted response times in both the word and the nonword groups. There were no significant interactions between any combination of the main effects, and no significance was found for participants or items as random effects.

5. Discussion

In this research, the facilitative effects of lexicality, words stress, and vowel harmony in Hungarian have been investigated, by manipulating those factors in experimental stimuli consisting of a nonsense prefix and a target. As the theoretically most important segmentation cue (Mattys et al., 2005), lexicality was hypothesised to contribute significantly to word boundary identification, decreasing response times in lexical words. Alongside facilitation effects from word recognition, decreased response times were expected from two sublexical cues, stress and vowel harmony. Results from the detection task with visual primes provided unequivocal evidence for the effect of word recognition, but only partly proved the prediction for sublexical cues: only stress cues signalling a word boundary were associated with faster detection, as harmony effects were only marginally significant. Furthermore, the opposite of what was expected for vowel harmony turned out to be the case, as items with a prefix that matched the harmony class of the target were detected (marginally) faster than words with a disharmonious prefix.

Lexical status of targets has proven to have an important facilitative effect on word recognition, making a difference of around 30 ms in boundary detection time. The inclusion of this factor alongside other segmentation cues is an important contribution to not only the description of Hungarian boundary cues, but to the study of word segmentation in general. Whereas other studies often aimed to investigate the contributions of sublexical cues alone (e.g. Kabak et al., 2010) and therefore only employed nonword detection tasks, this study showed the role of lexicality, but more specifically of word recognition in segmentation. As for the implications of these results for the hierarchical categorisation of Hungarian segmentation cues, it is within reason to conclude that lexicality is a crucial facilitator of boundary detection.

However, determining the exact ordering of cue weights in the vein of the hierarchical model by Mattys et al. (2005) is difficult, as word stress showed great significance for both words and nonwords. This goes slightly against the view that sublexical cues only start having a facilitative effect in the absence of higher-tier cues, in this case, lexicality. In the terms of Mattys et al. (2005), the lack of significant interactions between Tier I and Tier II cues and their simultaneous effect suggest that multiple cues may be used at the same time. An absence of higher-tier cues may also be caused by impoverished signal quality, but this also fails to explain the independent importance of stress in the results of this thesis, as it took place in optimal listening conditions with an uninterrupted signal.

An interesting point may be raised with regards to the relationship of lexicality and word stress. Vroomen et al. (1998) discussed that word stress in Finnish, while being an important segmentation cue, does not belong to the lexical representation of the word, therefore, a separation of stress into a sublexical tier is adequate. Well before the hierarchical model of Mattys et al. (2005) was proposed, Vroomen and colleagues already highlighted the important difference between lexical and sublexical segmentation cues, by likening predictable, e.g. word-initial stress to a long silence: a long break in the speech signal almost certainly marks the location of a word boundary, but the silence itself does not belong to the lexical representation of the word that either preceded or followed it. In this way, stress cues are similar to any other acoustic-phonetic cue, meaning that they belong to a different cue category (or Tier). The inclusion of stress in the lexical representation of Finnish words would simply be redundant, since it consistently falls on the first syllable of every word, just as in Hungarian. Knowing the completely predictable nature of stress in Hungarian, we might speculate that manipulating it in the present experiment and moving it one syllable further into the word boosted its saliency and therefore the stress cue's facilitative effect so much that the theoretically more important lexicality cue lost its superiority in the cue hierarchy.

As in Experiment 2 in the study by Vroomen and colleagues (1998), the presence of two strong segmentation cues seems to have overshadowed the potential effect of vowel disharmony in the present study too. Contrary to the findings of Kabak et al. (2010), who found robust evidence for the facilitation effect of vowel disharmony, the present results suggest the opposite. However, this seems to go against the findings of most segmentation studies which included vowel harmony (Suomi et al., 1997; Vroomen et al., 1998) and therefore must be treated with caution. As mentioned above, while the inclusion of three different segmentation cues (both lexical and sublexical) is useful for providing a comprehensive account of segmentation in cases where multiple cues are available, this design is not suitable for investigating the contributions of one cue in isolation. Vroomen et al. (1998) call stress simply more informative than vowel harmony, saying that vowel harmony is not used when stress cues are present. Mattys et al. (2005) attribute "foremost importance" to the lexicality cue in all languages, dominating all other Tiers. Kabak et al. (2010) found that Turkish listeners were using vowel harmony cues, however, their experiment was designed around nonword detection, lexicality as a cue being thus completely absent. In the present study, it is likely that alongside two hypothetically superior cues, the effects of vowel (dis)harmony were negated.

Another important point on the subject of vowel harmony is the distribution of tolerant vowels in the list of experimental stimuli. As described in Section 1.2 above, there is an additional 'tolerant' class of vowels alongside the mutually exclusive front and back classes in Hungarian. The vowels in this class, /i, iː, ɛ, eː/, are front vowels, phonetically speaking. Nevertheless, if they co-occur with a back-class vowel in the same word, the word stem adopts the harmony class of the non-tolerant vowel and, in this case, functions as any other stem of the back-harmony class. In the 96 experimental items of this study, there were 14 back-class items (14.58%) that had a tolerant (=front) vowel in the first syllable. Therefore, in the disharmony condition, a front prefix was followed by another phonetically front vowel, damping the potential effect of vowel disharmony. More problems arise when one considers the natural succession of same-class words. Disharmony cues are absent from such sequences, which is also pointed out by Vroomen et al. (1998), who employed an algorithm to detect disharmony-based word boundaries in two newspaper texts. Their algorithm identified 19% of the boundaries in the first text and 17.5% in the second. Clearly, when disharmony occurs, it most likely signals a word boundary, however, no disharmony is *required* for a word boundary to be marked. All of this makes vowel disharmony a weak segmentation cue and explains why, in the present experiment, items with disharmonious prefixes were not detected significantly faster than their harmonious counterparts.

6. Conclusion

To conclude, it has been shown that Hungarian listeners employ lexical and sublexical cues for word segmentation in conjunction. The understanding of boundary detection as a primarily lexically-driven process thus remains uncontested, however, since a strong facilitative effect of stress has been found, both in real words and in nonwords, the need for further research remains, on whether sublexical cues are only used when word recognition is not available, or whether the two factors can be used in conjunction, particularly with the boosted saliency that occurs when main stress is moved in a language where it is otherwise completely predictable. For future research about vowel harmony as a segmentation cue, stronger control of tolerant vowels in the stimuli is advised. Furthermore, a potential decrease in participants' level of attention should either be prevented or controlled for, as a 20-minute perception task may induce a level of fatigue that impacts reaction times. Answering the question whether stimuli should be spoken and recorded by a trained speaker or synthesised by computer goes beyond the scope of this thesis, however the issue must be mentioned, since any additional saliency of mid-word stress originating from the recorded speaker's pronunciation may have a crucial effect on the importance of the stress cue. Though this thesis has provided solid evidence for listeners' use of both lexical and sublexical segmentation cues, the need for further research on word segmentation is clearly indicated.

References

Bates D., Mächler M., Bolker B., Walker S. (2015). "Fitting Linear Mixed-Effects Models Using lme4." *Journal of Statistical Software*, 67(1), 1–48. doi: https://doi.org/10.18637/jss.v067.i01

Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(1), 113–121. https://doi.org/10.1037/0096-1523.14.1.113

Davis, M. H., Marslen-Wilson, W. D., & Gaskell, M. G. (2002). Leading up the lexical garden path: Segmentation and ambiguity in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(1), 218–244. https://doi.org/10.1037/0096-1523.28.1.218

Gow, D. W., & Gordon, P. C. (1995). Lexical and prelexical influences on word segmentation: Evidence from priming. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(2), 344–359. https://doi.org/10.1037/0096-1523.21.2.344

Hyllested, A. (2017). Loanword adaptation in Hungarian: unexpected vowel harmony in material of Latin and Slavic provenance. *Acta Linguistica Hafniensia*, *49*(1), 1–11. https://doi.org/10.1080/03740463.2017.1341811

Johnson, E. K., & Jusczyk, P. W. (2001). Word Segmentation by 8-Month-Olds: When Speech Cues Count More Than Statistics. *Journal of Memory and Language*, *44*(4), 548–567. https://doi.org/10.1006/jmla.2000.2755

Kabak, B., & Vogel, I. (2001). The phonological word and stress assignment in Turkish. *Phonology*, *18*(3), 315–360. https://doi.org/10.1017/s0952675701004201

Kabak, B., Maniwa, K., & Kazanina, N. (2010). Listeners use vowel harmony and word-final stress to spot nonsense words: A study of Turkish and French. *Laboratory Phonology*, *1*(1). https://doi.org/10.1515/labphon.2010.010

Klatt, D. H. (1979). Speech perception: a model of acoustic–phonetic analysis and lexical access. *Journal of Phonetics*, *7*(3), 279–312. https://doi.org/10.1016/s0095-4470(19)31059-9

Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of Multiple Speech Segmentation Cues: A Hierarchical Framework. *Journal of Experimental Psychology: General*, *134*(4), 477–500. https://doi.org/10.1037/0096-3445.134.4.477

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1–86. https://doi.org/10.1016/0010-0285(86)90015-0

Norris, D. (1994). Shortlist: a connectionist model of continuous speech recognition. *Cognition*, *52*(3), 189–234. https://doi.org/10.1016/0010-0277(94)90043-4

Norris, D., McQueen, J. M., & Cutler, A. (1995). Competition and segmentation in spoken-word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(5), 1209–1228. https://doi.org/10.1037/0278-7393.21.5.1209

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

Saffran, J. R. (2002). Constraints on Statistical Language Learning. *Journal of Memory and Language*, *47*(1), 172–196. https://doi.org/10.1006/jmla.2001.2839

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word Segmentation: The Role of Distributional Cues. *Journal of Memory and Language*, *35*(4), 606–621. https://doi.org/10.1006/jmla.1996.0032

Suomi, K., Mcqueen, J. M., & Cutler, A. (1997). Vowel Harmony and Speech Segmentation in Finnish. *Journal of Memory and Language*, *36*(3), 422–444. https://doi.org/10.1006/jmla.1996.2495

Swallow, K. M., & Jiang, Y. V. (2010). The Attentional Boost Effect: Transient increases in attention to one task enhance performance in a second task. *Cognition*, *115*(1), 118–132. https://doi.org/10.1016/j.cognition.2009.12.003

Törkenczy, M. (2011). Hungarian Vowel Harmony. *The Blackwell Companion to Phonology*, 1–27. https://doi.org/10.1002/9781444335262.wbctp0123

Varga, L. (2002). *Intonation and stress: evidence from Hungarian*. Palgrave Macmillan.

Vet, D.J. (2021) ED - Experiment designer [Computer program]. Beta version 1.32, https://www.fon.hum.uva.nl/dirk/ed.php

Vroomen, J., Tuomainen, J., & de Gelder, B. (1998). The Roles of Word Stress and Vowel Harmony in Speech Segmentation. *Journal of Memory and Language*, *38*(2), 133–149. https://doi.org/10.1006/jmla.1997.2548

Appendix: List of all experimental items and their translation

	Harmonious prefix		Disharmonious prefix				Harmonious prefix		Disharmonious prefix	
Stress location	Prefix	Target word	Prefix	Target word	Gloss	Stress location	Prefix	Target nonword	Prefix	Target nonword
Prefix	pa	szoba	pe	szoba	room	Prefix	pa	kola	pe	kola
	ka	csiga	ke	csiga	snail		ka	tipu	ke	tipu
	tu	liga	tü	liga	league		tu	dipa	tü	dipa
	pa	gumi	pe	gumi	gum		pa	doku	pe	doku
	pu	kocsi	pü	kocsi	car		pu	bola	pü	bola
	tu	satu	tü	satu	vise		tu	gita	tü	gita
	ke	pösze	ka	pösze	gate		pe	kitü	pa	kitü
	kü	teke	ku	teke	horn		kü	peke	ku	peke
	tö	teve	to	teve	flash		tö	nibe	to	nibe
	tü	mese	tu	mese	hoof		kü	tili	ku	tili
	tü	gebe	tu	gebe	mantle		tü	gipü	tu	gipü
	ke	süti	ka	süti	pipe		ke	biki	ka	biki
Target	tu	kapu	tü	kapu	lisping	Target	tu	lavu	tü	lavu
	ku	duda	kü	duda	roof		ku	futa	kü	futa
	to	vaku	tö	vaku	camel		to	bagu	tö	bagu
	ku	pata	kü	pata	tale		ku	taka	kü	taka
	to	suba	tö	suba	nag		to	gopa	tö	gopa
	ka	pipa	ke	pipa	cookie		ka	bibu	ke	bibu
	pe	kivi	pa	kivi	kiwi		tü	mime	tu	mime
	ke	mise	ka	mise	mass		ke	bikü	ka	bikü
	ke	csibe	ka	csibe	chick		ke	neri	ka	neri
	pe	Feri	pa	Feri	(name)		pe	zebe	pa	zebe
	pe	zene	pa	zene	music		pe	sekü	pa	sekü
	ke	fenyő	ka	fenyő	pine		kü	tepi	ku	tepi