Word final consonant extrametricality in Hungarian

Stephen Grimes
stgrimes@indiana.edu
August 2007

1 Introduction

The extrametrical status of word final consonants in Hungarian is a matter of debate in the Hungarian phonology literature. For example, Siptár and Törkenczy entertain but ultimately discard the hypothesis that word final consonants could be extrametrical (Siptár and Törkenczy, 2000:151). In this paper, I take a closer look at a variety of phonological phenomena compatible with extrametricality, and I also reanalyze data that have been interpreted as evidence against extrametricality in the literature. I also examine new data that illustrate how a minimal word constraint in Hungarian interacts with extrametricality. In the end, I conclude that classifying Hungarian as a language with word final extrametricality is compatible with the data presented.

1.1 Hungarian preliminaries

A few basic facts about the Hungarian phoneme inventory must be mentioned here at the outset of the paper. Hungarian makes a binary length distinction for all segments – vowels and consonants alike (except that there is no long counterpart to [h], and naturally segmental length contrasts are not available in all environments). Minimal pairs differing in segment length can be found for most vowels and consonants. Long segments can be specified lexically, but they can also be created by certain morphological processes. Most cases of productive gemination occur with the instrumental and translative case suffixes that double a final singleton consonant of a consonant final stem (but surfaces as -[v] after vowel-final stems). There are very few morphophonological alternations involving

1

length in the vowel system, with the exception of the so-called Low Vowel Lengthening (cf. Vago, 1980).

1.2 Hungarian vowel preliminaries

The seven short-long vowel pairs in Standard Literary Hungarian are given below in (1). Hungarian orthography marks vowel length with an acute accent or with a double acute accent in the case of the front rounded vowels; I will follow the practice of using Hungarian orthography for most examples in this paper except where doing so would be exceedingly unclear.

(1)   <u>Short Vowels</u>   <u>Long Vowels</u>
          ü i   u   ű í   ú
          ö e   o   ő é   ó
                  a   á

There is a special peculiarity about two of the vowel pairs. Although the height feature for the pair of vowels *e* [ɛ] and *é* [e] cannot be disputed to be *phonetically* mid (with the long vowel higher and more peripheral than the short vowel), these vowels seem to have a sort of dual status in the *phonological* system. Specifically, the *e/é* pair behaves phonologically as both mid and low vowels. Two Hungarian locative suffixes have three allomorphs containing mid vowels (-hez/-höz/-hoz 'towards' and -en/-ön/-on/-n 'on'). The quality of the vowel of the suffix is determined by vowel harmony – the harmony system primarily partitions stems into front and back classes, but a secondary rounding harmony is available for front vowel stems in which the final vowel is round. The existence of such a vowel alternation as exhibited by these suffixes instantiates a natural class of vowels one could call "mid" in Hungarian.

By contrast, other locative suffixes (e.g. the inessive -ban/-ben) or the dative case suffix (-nak/-nek) have only two allomorphs, and in these cases we see an alternation of a mid (front) vowel with a low (back) vowel chosen according to the harmony of the stem; the rounding harmony observed for front vowels apparently does not apply in this case because the front vowel is specified as [+low], and there are no front rounded vowels with a [+low] specification on which to realize this harmony. (See Kornai 1991 for an interesting and detailed discussion of what phonological features are necessary and sufficient to create the proper natural classes for Hungarian vowels.) In any case, I assume that the existence of suffix pairs using *e/é* and *a/á* constitutes evidence that these vowels form a natural class. I follow a somewhat standard practice by referring to this class of vowels as low vowels despite their actual phonetic values.

Having established a "low" category for the vowel system, I now wish to point out that these low vowel pairs *e/é* and *a/á* are also the only vowel pairs in the system for which there is a clear difference in *quality* between the short and long variants (not simply *quantity*). This observation may be relevant because the low vowels often are exceptional when trying to make generalizations about syllable weight, and I will return to this issue several times throughout the paper.

1.3 Hungarian consonant preliminaries

All instances of geminate consonants occur in intervocalic position or are found "wholly contained" in the coda in word final position. Underlying geminates that would appear in a cluster with another consonant due to the morphology surface as singletons due to a length reduction process applying to consonant clusters – this reduction of GC or CG to CC is likely related to restrictions on maximal syllable weight. Triconsonantal clusters are generally limited to foreign borrowings and typically involve a sibilant. Underlying

geminates are outnumbered by geminates occurring at morpheme boundaries arising through gemination processes or through simple concatenation of identical sounds. Overall it can be said that geminate consonants have a relatively small functional load in Hungarian (Obendorfer, 1975).

For Hungarian, a great deal of discussion has focused on what the language-internal representation of geminates should look like. So-called "fake" geminates are geminates comprised of a sequence of identical consonants (Hayes, 1986:326-327). Following Hayes' definition, a "true" geminate cannot be split by epenthesis nor can there exist phonological processes which only act on a single half of the geminate. Polgárdi (2005) describes the fake geminates appearing in Hungarian as those arising through concatenation where identical segments appear at the concatenation boundary. For example a stem ending in [b] precedes the suffix *-ben* in the word *habban* 'foam-INESSIVE', creating a geminate of the type Polgárdi describes.

The issue of whether a geminate is fake or not is related to the discussion about whether the representation of the geminate should use a single or double root node (for a discussion specific to Hungarian see Vago, 1992). The dominant view in the cross-linguistic literature seems to treat geminates as having single root nodes (Hayes, 1989, Hock, 1986, Hyman, 1992, McCarthy and Prince, 1986/1996); under this view the geminate acts as a heavy variant of the corresponding singleton consonant and is associated with a mora in an underlying representation (in contrast to single consonants which have no underlying moraic association). In contrast Selkirk has been the principle advocate of geminates as occupying a double root node (Selkirk, 1990). To support Selkirk, Tranel (1991) cites languages where geminates pattern as light in the coda – Selkup, Malayalam, and Tubatulabal. Ringen and Vago provide a discussion of the relevant issues and characterize the single/double root debate as involving weight versus

length interpretations of geminates (Ringen and Vago, 2006), and Davis' discussion on quantity (2007) provides an overview of the different proposals for theoretical treatments of quantity in the literature.

One reason to treat geminates as having a double root node representation would be if they behave similarly to consonant clusters, which naturally occupy two consonant "slots". In the pattern of vowel epenthesis for verbs in Hungarian shown below, the forms in (2a) do not take epenthesis while the forms in (2b) do. Presumably this demonstrates that Hungarian geminates should in fact be treated as occupying a double root node:

(2) Vowel epenthesis   (data from Vago, 1992)

	3S	2S	infinitive	gloss
a.	kap	kapsz	kapni	receive
	nő	nősz	nőni	grown
	fél	félsz	félni	fear
b.	áld	áldasz	áldani	bless
	önt	öntesz	önteni	pour
	hall	hallasz	hallani	hear
	függ	függesz	függeni	depend

Crucially, note that the word for 'fear' in (2a) with a long vowel plus single consonant does not take epenthesis. However, the word for 'pour' in (2b) with a short vowel plus consonant cluster does take epenthesis. This led Vago (1992) to conclude that the epenthesis process in not sensitive to syllable weight per se, but rather the process counts the number of consonants. Given that geminates here pattern with consonant clusters, this gives reason to believe geminates occupy two C-slots or a double root node.

For the purposes of this paper, however, I do not want to commit to a single or double root node analysis. This is not simply out of ambivalence but rather because it may be the case that both interpretations of geminates are available to the phonology of a single language. Hyman (1992) studied the patterning of morass in Bantu languages and

termed the discrepancies in interpretation of geminates as moraic mismatches. Hyman gives examples to show that a preconsonantal nasal appears to contribute to syllable weight with respect to how weight is computed in reduplication. However a preconsonantal nasal is not considered heavy when assigning high tone to the second mora of the verb stem. The conclusion is that segments can be considered moraic for one phonological process but non-moraic in another instance, and it seems reasonable to allow for this in Hungarian as well.

Finally, somewhat as an aside concerning Hungarian syllable weight, it is interesting to note that Trubetzkoy's generalization (cf. Zec, 1995) holds in Hungarian. Trubetzkoy's generalization states that if a language treats CVC syllables as heavy, then it will also treat CVV syllables as heavy.[1] In (2) above we see that the contrapositive (a logical equivalent) holds for Hungarian. For vowel epenthesis, a CVV syllable is treated as light, and it is also the case that CVC syllables are also regarded as light (no epenthesis required).

1.4 Extrametricality preliminaries

In the metrical phonology literature, the term *extrametricality* (Ito, 1989) can often be synonymous with extraprosodicity, Non-Finality constraints (McCarthy and Prince, 1993), the Peripherality Condition (Hayes, 1982) and other so-called edge effects. In this paper, I will use extrametricality to denote final consonant extrametricality, or perhaps even more precisely, consonant extramoraicity. The defining characteristic of extrametricality here is that a word final consonant is not a member of the prosodic word domain. Under the assumption that weight-by-position (Hayes, 1989) applies in Hungarian, a word final consonant is not syllabified and is not available to contribute to

---

[1] In the case of CVC, this is informally used throughout to mean C_0VC – the presence of zero or more consonants in the syllable onset.

syllable weight under weight-by-position. Indeed the consonant is not moraic, and in terms of syllable structure one can view the final consonant as linked directly as an appendix to the prosodic word with no intermediate syllable affiliation.

In a language which treats word final consonants as extrametrical, one consequence of extrametricality is that when comparing allowable syllable weights, one should observe a phonotactic asymmetry in the apparent allowable weight between word final and non-word final syllables. Word final syllables should appear to be heavier by allowing an additional segment or longer segments word finally. Researchers generally assume unified syllable weight constraints for a language, and hence this is the underlying reason to not consider the extra mora as part of the syllable. Hence, in order to show that there exists extrametricality in Hungarian, I will seek to demonstrate that an asymmetry in weight in fact exists between non-word final and word final syllables.

1.5 Organization of the remainder of the paper

This paper presents six independent justifications for why consonants in Hungarian are extrametrical. The first three arguments appear in separate subsections of Section 2. In Section 3, I introduce the Minimal Word Condition and discuss reasons to believe it is an active constraint in Hungarian. Under the assumption that Hungarian words must respect a minimum length requirement, Section 4 provides three further arguments for final consonants in Hungarian being extrametrical. Section 5 summarizes and concludes the paper.

2 A survey of arguments for extrametricality in Hungarian

The three subsections of Section 2 address three independent cases demonstrating word final extrametricality. The first case involves an examination of the lexicon to determine the range of maximum allowable syllables.

2.1 The lexicon and lexical gaps in syllable structure

In Hungarian there is an interesting pattern of lexical gaps in the distribution of syllable weight for monomorphemic words. As shown below in the column (3a), both long and short vowels are permitted word finally before a single consonant. Gaps in the lexicon occur in the columns in (3b) and (3c) – except for the low vowels, long vowels are not permitted to appear before a consonant cluster word finally or word internally. (For the data in (3) the appropriate segmental values of digraphs are given in square brackets.)

(3) Lexical gaps for three syllable patterns (data from Siptár and Törkenczy, 2000).

	(a) V(V)C#		(b) V(V)CC#		(c) V(V)C.C	
i	hit	'belief'	ring	'sway'	inger	'stimulus'
ü	sün	'hedgehog	csüng	'hang'	kürtő	'funnel'
ö	sör	'beer'	gyöngy [dʸöndʸ]	'pearl'	ördög	'devil'
e	nem	'gender'	szent	'saint'	persze [perse]	'of course'
u	fut	'run'	must	'grape juice'	undor	'disgust'
o	lop	'steal'	gyors	'fast'	boglya [bogja]	'haystack'
a	hat	'six'	tart	'hold'	apró	'tiny'
í	sír	'grave'	----		----	
ű	bűn	'sin'	----		----	
ő	bőr	'skin'	----		----	
é	kém	'spy'	érc	'ore'	érték	'value'
ú	rút	'ugly'	----		----	
ó	kór	'disease'	----		----	
á	láp	'marsh'	márt	'dip'	árpa	'barley'

I would like to direct the reader's attention to observe that a sequence of a long vowel followed by a single consonant within a syllable is permitted for a word final syllable but not in a word internal syllable (again, with the exception of the low vowels). This asymmetry in allowable weight is what I sought to find -- the asymmetry is "explained away" if the word final consonant is not moraic. In this way, the syllables would be maximally bimoraic, and the evidenced gap in the syllable typology points toward final consonant extrametricality.

Despite the asymmetry observed in (3), Siptár and Törkenczy do not consider it evidence for extrametricality because no vowel length alternations take place. Here only static forms in the lexicon are being considered, and they would prefer to see further alternations such as the nucleus of a superheavy syllable reducing as evidence for an active extrametricality process. Furthermore, Siptár and Törkenczy suggest that the data in (4), which show that superheavy syllables may surface when polymorphemic forms are considered, illustrate that the constraint on maximal syllable weight is a constraint only on the morpheme, not the syllable.

(4) Polymorphemic forms exhibiting heavy syllables not found in monomorphemic words for each of the seven long vowels

	VCC#		VC.C	
í	sír-t	'cry-ACC'	szív-t-am	'inhale-PAST-1S'
ű	bűn-t	'crime-ACC'	bűn-ben	'crime-INESSIVE'
ő	fő-bb	'main-COMP'	bőr-ben	'skin-INESSIVE'
é	kér-t	'ask-PAST'	kér-t-ék	'ask-PAST-3S'
ú	túr-t	'dig-PAST'	túr-nak	'dig-3P'
ó	kór-t	'disease-ACC'	kór-nak	'disease-DAT'
á	vár-t	'wait-PAST'	vár-t-am	'wait-PAST-1S'

The maximal syllable constraint in Hungarian, which does not apply across morpheme boundaries such as in (4), is often referred to as *VVCC in the Hungarian phonology literature. I am not familiar with any explanation as to why this constraint only applies to stems while allowing extra heavy syllables with low vowels to appear in derived forms. Note that the data in (4) are apparently problematic for an extrametricality analysis because the expected asymmetry in syllable weight between word final and non-word final positions disappears. However, to maintain an extrametricality analysis, I advance that the forms in (4) are permitted to have superheavy syllables due to a high ranked Output-to-Output or Paradigm Uniformity constraint that dominates the phonotactic

constraint concerning syllable weight. The optimality theory tableau in (5) shows how these competing constraints might interact:

(5) Tableau to illustrate proposed relationship between constraints

/sír + t/ 'cry-PAST'	Paradigm Uniformity	*μμμ
→ a. sírt		*
b. sirt	*!	

The reason to believe that paradigm uniformity is at work in such cases is because only one other word in the relevant paradigms has a superheavy syllable – the second person indefinite conjugation. This is shown in (6); one may assume that a CC consonant cluster is syllabified C.C such that the cluster straddles a syllable boundary.

(6)   Four paradigms for the verb sír 'cry'

	1S	2S	3S	1P	2P	3P
Present definite[2]	sírom	sírod	sírja	sírjuk	sírjátok	sírják
Present indefinite	sírok	sírsz	sír	sírjunk	sírjatok	sírnak
Past definite	sírtam	sírtad	sírta	sírtuk	sírtátok	sírtak
Past indefinite	sírtam	sírtál	sírt	sírtunk	sírtatok	sírták

The purpose of this digression was to convince the reader that Hungarian does not in fact freely permit superheavy syllables, as this was an argument used against final consonant extrametricality. In fact a superheavy syllable is only permitted so that the suffixed form minimally deviates from the base stem. In this way, the lexical gaps shown in (3) can in fact be used as evidence for extrametricality.

2.2 Syllable weight with respect to stress

Kerek (1971) reports facts concerning variable secondary stress and how stress interacts with syllable weight. In Hungarian, primary stress always falls on the first syllable of the word. Primary stress does not induce vowel length nor do long vowels attract primary

---

[2] The verb 'cry' in Hungarian has a definite conjugation that does not correspond to a direct object in English; Hungarian uses the accusative to indicate what is being "cried about".

stress. As for secondary stress, some accounts have it not existing at all, while others report that secondary stress is regularly assigned to odd-numbered syllables after the initial syllable (3rd syllable, 5th syllable, and so forth). Yet other researchers, including Kerek, report that weight-sensitive secondary stress shift is possible as long as doing so does not create two adjacent stressed syllables. Secondary stress may optionally shift away from a light syllable onto an adjacent heavy syllable, similar to the secondary stress shift pattern reported for Finnish and other languages. The example in (4) shows two hypothetical stress patterns for a four-syllable word σ σ σ$_L$ σ$_H$ with a light penultimate syllable and heavy final syllable.

(7)   a. (**σ σ**) (**σ$_L$ σ$_H$**)
        b. (**σ σ**) (σ$_L$ **σ$_H$**)   (boldface indicates stress)

Kerek gathered data on possible syllables types that may constitute a heavy syllable for the observed stress shift phenomenon. The data are summarized in the table in (8), where it can be seen that word internal and word final environments do not treat stress the same.

(8) Classification of syllable weight according to secondary stress shift

	Light	Heavy
Word internal	CV	CVV, CVC, CVVC
Word final	CV, CVC	CVV, CVVC

Crucially, a closed CVC syllable is considered light when it is the final syllable of the word but heavy in non-final positions. No data is given by Kerek to indicate whether words ending in a final geminate attract secondary stress – such an example would need a geminate in the final fourth syllable, and unfortunately such words are not countenanced. Incidentally, this is precisely the patterning of CVC observed by Abu-Mansour (1995) for Levantine Arabic, a language which is standardly assumed to have word final consonant extrametricality. In Levantine Arabic, CVC syllables generally attract stress, but not in

word final position; CVG syllables (i.e. syllables closed by geminates) do attract stress word finally. Similar to Arabic, adopting final consonant extrametricality Hungarian would erase the discrepancy in weight interpretation of syllables in final and non-final positions, and hence evidence from Hungarian secondary stress shift supports extrametricality.

2.3 Evidence from word final length contrasts

The third argument for extrametricality concerns the language-internal representation of geminates in word final position. Consider the consonant length minimal pairs in (9).

(9)   a. singleton   b. geminate   (some data from Vago, 1992)

        hal   'fish'   hall   'hear'
        sok   'many'   sokk   'shock'
        megy   'go'   meggy   'sour cherry'
        kel   'rise'   kell   'need'
        tol   'push'   toll   'pen'
        lap   'page'   lapp   'Lapp'

Exactly as Curtis (2003), Morén (2001), and Ham (2001) have noted, if the forms in (9a) were to receive a mora through a weight-by-position principle, their representations would be indistinguishable from the forms in (9b) where the weight-bearing unit was associated with true geminate. However, if a final consonant is underparsed, it does not receive prominence through weight-by-position. Hence the argument here is one concerning the representation of geminates – Ham goes as far as to claim that CVC syllables are always extrametrical in languages that have word final geminates. Whether this holds in all languages is not clear to me, but clearly one should have a phonological account for the phonetically real distinction that Hungarian listeners can easily detect – one in which the word final geminate consonant has a delayed burst release.

The remaining arguments for extrametricality surround its interaction with minimal word length, so we take a slight diversion to first consider minimality.

3 Minimal word length

3.1 Theoretical background

The Minimal Word Condition (Hayes, 1995, McCarthy and Prince, 1986/1996) grew out of research on metrical stress theory and templates. It assumes that prosodic words contain a hierarchical structure consisting not just of syllables but of intermediate metrical feet. Typically a prosodic word must contain at least one metrical foot, which in turn consists prototypically of two syllables (putting aside the possible attestation of ternary stress systems). Despite the above assumptions, many languages allow monosyllabic words, and in these cases a so-called degenerate foot is allowed. This is a metrical foot consisting of a single syllable which is minimally bimoraic. While there are a great number of monosyllabic words in Hungarian, one-syllable nouns and verbs are almost universally bimoriac. The minimal word constraint for Hungarian is stated in (10).

(10) $PrWd_{MIN} = \mu\mu$ A prosodic word is bimoraic.

The interaction of extrametricality and the Minimal Word Condition concerning examples such as in (9) will be treated in Section 4 of this paper. Function words are typically and typologically immune from the condition in (10) in cross-linguistic examinations. That function words survive such minimality violations is generally left unexplained, but this could perhaps be due to their clitic-like nature where the prosodic deficiency is resolved by joining a stress phrase with neighboring content words.

3.2 Minimal words in Hungarian

The minimality constraint in (10) does not apply to function words in Hungarian. Pronouns (e.g. *te* '2S', *ti* '2P'), wh-words (e.g. *ki* 'who', *mi* 'what'), and several separable verbal prefixes (e.g. *be* 'into', *ki* 'out') would violate a minimal word condition applying to function words. The minimal bimoraic constraint seems to generally hold for content words; the two content word exceptions are *fa* 'tree' and *ma* 'today'. Note that the vowel in these two words is a low vowel, and hence the apparent exceptions are not surprising given the erratic behavior of the low vowels with respect to their length interpretation in the phonological system.

Once again Törkenczy and others have pointed out that word minimality is not entirely without controversy in Hungarian. Due to a general phonotactic constraint mandating mid, rounded vowels to be long in word final position (words do not end in [o] or [ö], it is argued that this phonotactic constraint and not word minimality forces the mid, rounded vowels to be long in monosyllables and polysyllabic words alike. While it is clear that such a phonotactic constraint exists, I would like to add that the existence of only mid, rounded vowels with [+long] specification in open syllables of monosyllabic words could be due to *both* the Minimal Word Condition as well as the more general word final length phonotactic condition on these segments. Rather than using this phenomenon to argue against minimality, it is my position that this data simply imparts no crucial evidence.

One reason to believe the Minimal Word Condition is relevant in Hungarian can be shown by examining high vowel length vacillation. First off, as can be seen in (11), there are few minimal pairs to differentiate length in the high vowels, and the glosses for the existing examples show that they are not easily constructed. The vacillation of length in the high vowels may be due to the low functional load of length in the high vowels,

14

and Nádasdy and Siptár (1994:62) note that the low functional load of high vowels made it possible to interpret texts typed on typewriters lacking long, high vowel keys without difficulty.[3,4]

(11)   High vowel minimal pairs   (examples from Pintzuk et al., 1995)

      színt   'color-ACC'   szint   'floor'
      fűlnek   'they are heated'   fülnek   'to an ear'
      nyúlunk 'we grasp'   nyulunk 'our rabbit'

Speakers of Educated Colloquial Hungarian (ECH) in Budapest and elsewhere are often uncertain as to the correct value for the length of high vowels in certain syllable positions. Here I focus on the length of the nucleus of the word final syllable, and the table in (12) compares the traditional pronunciation found in Standard Literary Hungarian (SLH) with the speech found in ECH, the unmarked dialect of Budapest.

(12)   Written Form   SLH   ECH   Gloss

a. high vowel length alternation in polysyllabic words

      fiú   [fiu:]   [fiu]   'boy'
      tetű   [tetü:]   [tetü]   'louse'
      házból   [ha:zbo:l]   [ha:zbol]   'from the house'
      hegyről   [hed^{y}rö:l] [hed^{y}röl]   'down the hill'
      szomorú   [somoru:]   [somoru]   'sad'

b. no alternation in monosyllabic words

      fű   [fű]   [fű]   'grass'
      sí   [si:]   [si:]   'ski'

---

[3] Please refer to (Pintzuk et al., 1995) for an in-depth treatment of the whether the lack of typewriter keys affected the decline of long, high vowels in spoken Hungarian.
[4] The present-day equivalent of the typewriter problem involves email correspondence where vowel length is not indicated for any vowel due to issues with different character encodings or foreign keyboard-character maps. Such emails can be interpreted with some ambiguity resolution necessary on the part of the reader, but nonetheless it is the high vowels for which there is the least difficulty resolving ambiguities.

Nádasdy (1985) asserts that the direction of the change-in-progress is towards short final high vowels. I attribute the failure for high vowels to reduce in monosyllables as being due to an active Minimal Word Condition for Hungarian.

Having argued for word minimality, I can now continue by presenting additional arguments for extrametricality that are relevant to and interact with the Minimal Word Condition.

4 Minimal word length and extrametricality

The key point that will underscore each example in this section is that due to the assumption of extrametricality, a CVC word (in which the vowel is specifically short like those in 9a) *does not meet* the requirement of the Minimal Word Condition. I hope the reason for this is clear – because the final consonant is underparsed, no weight is attributed to it by weight-by-position, and hence the syllable is only monomoraic.

The three additional arguments for extrametricality are presented in this section are based on arguments concerning vowel length in monosyllabic words (4.1), evidence from German loanwords (4.2), and vowel length alternations (4.3).

4.1 Vowel length in monosyllabic words

While examining a diverse set of Hungarian texts and wordlists, one can develop the intuition that in monosyllabic, closed-syllable words ending in a single consonant, long vowels appear a surprising number of times compared to the overall long vowel frequency in all other contexts. In this subsection I seek to quantify this hypothesis empirically, and I present statistical data from corpora. If this observation can be demonstrated, the scarcity of CVC words (where the V is a short vowel) suggests the failure of such words to meet the Minimal Word Condition (and that the words in 9a are just a small group of exceptional words).

4.1.1 Frequency statistics

In order to avoid spurious word examples as well as hapax legomena (words appearing only once in a corpus), my wordlist for frequency examination is complied by considering words if and only if they appear in both of two dictionaries. The first dictionary was compiled by Ándrás Kornai and possibly others at the Hungarian Academy of Sciences (Kornai, 1986), and the second dictionary is the word frequency list from the Hungarian National Corpus (Váradi, 2002). Because the Hungarian National Corpus is tagged for part of speech, I was able to select from it only the nouns, verbs, adjectives, and adverbs – for this examination these will serve as the content words for which the Minimal Word Condition is presumably relevant. Techniques described in (Grimes, 2007) were used to derive a phonemic transcription from the orthographical form of the word using a Perl script for each word in the combined dictionary. I did this of course because my questions are not concerned with the spelling but with the pronunciation of Hungarian words. (Performing a grapheme-to-phoneme conversion allows for queries on the CV structure of each word; this particular task requires identifying each of the digraphs in the orthography where multiple letters correspond to a single sound.) Then, using regular expression searches, I identified words of the desired structure, such as CVC or CVCC.

Because I worried that only examining the presence or absence of word forms (type frequencies) might distort the characterization of the language, I also compiled word frequency data (token frequencies) for contrastive purposes. As will be seen below, the results of frequency comparisons for both type and token frequencies are somewhat similar to one another, and there is no reason apparent to me to prioritize one type of frequency data over the other in this situation. The individual word frequencies were from the Hungarian National Corpus (HNC); like corpora such as the Brown Corpus, the

HNC makes an attempt to compile a variety of written sources and genres to reflect the diversity in usage of the language.

The table in (13) illustrates data on the distribution of superheavy CVVC and CVCC syllables in mono- and disyllabic words. Superheavy here is used in a descriptive sense to mean a rhyme with a long vowel and one or more consonant or a short vowel and two or more consonants; studies of syllable typologies reveal such syllables are marked and generally dispreferred crosslinguistically.

(13) Percentages of words containing a CVVC or CVCC syllable

Two syllable word first syllable	Two syllable word final syllable	One syllable word (only syllable)
13.0% of types	34.1% of types	73.6% of types
8.8% of tokens	29.5% of tokens	46.9% of tokens

An examination of the types of two-syllable words with superheavy first syllables in the first column of (13) reveals that the 13.0% of apparent word type exceptions involve almost exclusively compounds or multimorphemic words – these exceptions are why the maximal syllable constraints are said to only apply monomorphemically. Otherwise virtually no superheavy, word-internal syllables are permitted. For the second and third columns, recall that under extrametricality these would only be considered heavy syllables, not superheavy. Finally, the higher rate of "superheavy" final syllables in monosyllabic words is presumably due to an effect of the Minimal Word Condition.

4.1.2 A closer look at the quality of the vowels

When a language ignores word final consonants for syllable weight calculation (as is assumed under extrametricality), CVC words appear then to have a tenuous status as prosodic words. Such words are interestingly not ruled out by the Hungarian grammar. For this reason, it is instructive to examine the class of possible exceptions – the CVC

words that seemingly fail to attain bimoraic status under extrametricality. In monosyllabic words, many of the apparent exceptions to the Minimal Word Condition contain one of the so-called low vowels; the structure of the word is CaC or CeC. The table in (14) shows that low vowel words constitute 87.8% of the token frequencies that violate the minimal word condition under extrametricality.

(14)   Distribution of short vowels in one syllable C_0VC words

	Types	%	Tokens	%
a	72	26.1%	3372953	23.0%
e	64	23.2%	9487668	64.8%
i	21	7.6%	427409	2.9%
o	47	17.0%	621353	4.2%
u	21	7.6%	327504	2.2%
ö	34	12.3%	329757	2.3%
ü	17	6.2%	68976	0.5%
Total	276	100.0%	14635620	100.0%

An explanation as to why low vowel CVC words are so frequent may be found in the concept of coerced weight or moraicity (Morén, 2001, 2003). For Morén, coerced weight is a restriction on minimal or maximal surface moraicity in some phonological context; this is to be differentiated from distinctive weight in which underlying moraicity is reflected in a surface contrast. An example of coerced moraicity is found in English where low lax vowels in penultimate position can receive word-level stress. The lax-tense distinction in English, which is sometimes compared to a short-long length distinction, naturally prefers that tense vowels bear word-level stress. Applied to Hungarian, it may be the case that in the CVC phonological context, minimal surface moraicity conditions require a bimoraic interpretation of a short vowel. Furthermore, as low vowels are more sonorous and more hence more prominent, a short low vowel is more likely to receive a bimoraic interpretation than non-low vowels. Zec (1994) carefully explored the relationship between syllabicity, moraicity, and sonority, and this was precisely the

relationship discovered – the more sonorous a segment is, the more likely it will be moraic (in the syllable rhyme). This view is also reinforced by Hayes (1995: Ch. 7) who claims that syllables with low vowels are both more prominent and more perceptually salient than other syllables. He also adds that although this prominence may not be relevant to the phonology, sometimes languages can take differences in prominence and phonologize them, which may be the case here.

Finally, I would like to present data on the syllable weight of monosyllabic words which do not contain a low vowel. As the table in (15) shows, 77% of the word types and 80.8% of the word tokens meet the bimoraic requirement under extrametricality when low vowels are factored out. This evidence points to the Minimal Word Constraint more or less respecting the extrametricality assumption that a final C does not constribute to weight calculations in the case of non-low vowels.

(15)   One syllable content words only; words with low vowels excluded

		Type	%	Token	%
a.	μ. C_0V	0	0%	0	0.0%
	μ. C_0VC	140	23.0%	1,774,999	19.2%
b.	$\mu\mu$.C_0VV	31	5.1%	1,187,687	12.8%
	$\mu\mu$. C_0VVC, C_0VCC, or C_0VVCC	437	71.8%	6296365	68.0%

The prevalence of monosyllabic words with bimoraic weight (under extrametricality) is explained by the Minimal Word Condition. The reader may be concerned that the minimal word effect observed here is apparently gradient and not absolute. I share this concern, and one explanation in a constraint-based theory would be that there is an IDENT-IO constraint ranked such that the minimality violation is not fatal. I return to this discussion in the final section of the paper to offer another explanation.

4.2 Germanic loan words

Nádasdy (1989) describes a particular case of "unmotivated" consonant gemination in recent borrowings (where recent here means since 1750) primarily from German into Hungarian. Some words borrowed from English and a few other languages also fit the pattern described. While Nádasdy describes the gemination as unmotivated, I believe the account I offer in this paper provides some explanatory weight.

The phenomenon is that short consonants in German, a language which apparently lacks geminates, are borrowed as a geminates into Hungarian. Some examples of this borrowing for monosyllabic words are given in (16), although the consonant lengthening process also applies to some but not all consonants in longer words, both in word final and word internal positions.

(16) Examples of consonant lengthening
lakk	(<Ger. Lack)	'lacquer'
sokk	(<Ger. Schock)	'hock'
tipp	(<Ger. Tipp)	'idea'
meccs [mečč]	(<Eng. match)	'match'
blöff	(<Eng. bluff)	'bluff'

The pattern of borrowing in monosyllabic words closed by a single consonant (C_0VC) is best described as follows:

(17)   a. If the vowel is borrowed as long, no consonant lengthening occurs.
          (To do so otherwise would create a superheavy syllable).

         b. If the vowel is borrowed as short, the consonant is always borrowed as a geminate.

Based on this generalization, Germanic loanword phonology respects word minimality under extrametricality. The gemination of the final consonant in borrowed words such as in (16) adds the required extra mora.

4.3 Vowel length "alternations"

The final piece of evidence I consider for Hungarian extrametricality involves an interesting case of vowel length alternations. There is a closed class of words that exhibit an alternation in stem vowel length between nominative and suffixed forms. These words are always monosyllabic, and examples of the so-called shortening stems are given in (18). The vowel shortening is triggered by what can be described as vowel-initial suffixes, or what are clearly not consonant-initial suffixes.

(18) | Nominative | Accusative | Gloss |
|------------|------------|-------|
| víz | vizet | 'water' |
| nyár | nyarat | 'summer' |
| kéz | kezet | 'hand' |
| jég | jeget | 'ice' |

In Hungarian grammars these words are likely described as shortening stems because the nominative, unaffixed form of the lexeme (lemma) is assumed to be basic. However, a phonological analysis might ask whether the underlying form of the vowel should be taken as long or short – whether the relevant process is actually shortening or lengthening. Indeed, more insight is available by taking the underlying form of the vowel to be short as shown when the accusative and other cases are suffixed. Under this view, vowel lengthening would be motivated for monosyllables due to the constraint on word minimality applying. To be clear, the minimality effect is seen because a final consonant does not contribute to syllable weight.

Note, however, that these words must be lexically marked; many words have stem vowels – both long and short – that do not alternate in length. Because this is a closed class of words, its descriptive ability concerning the global phonology of modern Hungarian is limited in this regard.

5 Summary and discussion

The purpose of this paper has been to survey arguments for and against extrametricality in Hungarian. I hope to have advanced the argument for the existence of extrametricality by showing both its independent effects and effects related to the minimal word condition. The table in (19) below summarizes the phenomena discussed in this paper. In the final column of (19) I raise an issue concerning diachronicity which I discuss immediately below.

(19) Summary of evidence relating to extrametricality in Hungarian

Evidence	Supports extrametricality?	Notes / Caveats	Suggestive of diachronic extrametricality?
Asymmetry in allowable size of a maximal syllable word finally vs. word internally	yes	confusion as to whether maximal syllable is a syllable or morpheme constraint	yes
Secondary stress weight sensitivities	yes		?
Geminate moraic representation	yes	other geminate representations possible	?
Minimal word length – vowel length in CVC	yes	gradient phenomenon; approximately 20% of words remain unaccounted for	yes
Gemination in Germanic loan words	yes	polysyllabic words may also have unmotivated gemination	yes
Vowel length alternations in nominal paradigm for a closed class of nouns	yes	frozen vowel length "alternations"	yes

5.1 Diachronic versus synchronic extrametricality

An issue that frequently lingers behind analyses in phonology and is an issue related to the gradience observed in Section 4.1 is the synchronic status of a phonological process.

Whether the extrametricality that has been demonstrated in this paper is a synchronic condition on lexical items in the language or whether it reflects evidence of a phonological constraint from an earlier period of the language system is unclear. In other words, I am interested in whether Hungarian possibly *was* an extrametrical language but is no longer. It is not obvious what type of evidence would be conclusive in this case; foreign borrowings would be a natural place to look, but these words are known to freely violate generalizations made about the rest of the language. The synchronic/diachronic issue is complicated by the observation that even if a lexicon absolutely respects extrametricality and word minimality absolutely, does this license the linguistic researcher to say that extrametricality is a synchronic phenomenon? Extremely recent English loanwords no longer undergo consonant gemination seen in earlier Germanic loanwords, which indicates that whatever the constraint was that caused the Germanic gemination is no longer active in the language system.

Similarly, the closed class of shortening stems presented in Section 4.3 seems to provide evidence that extrametricality was in effect at a previous point in the Hungarian language and that it interacted with word minimality. The words in this stem-shortening class are very obviously "older" words in the language – words that could appear on the Swadesh list or that refer to concepts that have been part of society for at least several hundred years. Early Hungarian religious documents from the middle ages document the longevity of the stem vowel length alternation; this information can be easily checked and verified by consulting the Hungarian language etymological dictionary (Benkő, 1967-1984). The final decision as to the synchronic status of extrametricality might depend on what happens with the length feature in the vowel system and in particular with respect to low vowels.

5.2 Low vowels in Hungarian

A few remarks remain to be made about the low vowels in Hungarian. First, it was noted

earlier that the words *fa* 'tree' and *ma* 'today' are the only CV content words violating the

Minimal Word Condition. However, one would not expect to see an alternation of short *a*

with long *á* in this context either – *á* is prohibited word finally in major lexical categories

(i.e. content words) in underived contexts (Kenesei et al., 1998:409-410). The only

exceptions to this generalization are *hajrá* 'final spurt' and *burzoá* 'bourgeois'.

To reiterate, there are also low vowel exceptions to the maximal syllable

condition referred to as *VVCC that I first described in Section 2.1. Törkenczy

(1994:343) notes that the low vowels *é* and *á* are in fact permitted in monomorphemic

syllables such as *márt* 'dip, douse' and *férc* 'seam' (and in a small handful of other

words). After observing so many exceptions and footnotes concerning the low vowels,

one may begin to think that the relevant contrast in the low vowels is not quantity but

quality (perhaps specifically vowel height, as the "long" vowels are generally higher and

perhaps more tense). In fact, there may be an emerging distinction between phonetic

length pairs and phonological length pairs in these vowels – Siptár reports four new

"minor vowels" are appearing – two short vowels to act as quantitative counterparts to *é*

and *á*, as well as two long vowels to act as counterparts to short *e* and *a* in which there is

no difference in quality (Siptár, 1991). Aside from orthography, the only basis for

considering these low vowels length pairs is the Low Vowel Lengthening process

whereby a final low vowel is lengthened after certain case endings are suffixed.

For the present, this problem seems unresolved. I can only offer that there appears

to be a unique identity length relationship for low vowels that allows them to retain their

length in contexts where the other vowels would not surface:

(20)   IDENT-LOW-VOWEL$_{\text{LENGTH}}$ >> PHONOTACTICS >> IDENT-VOWEL$_{\text{LENGTH}}$

As I suggested earlier, this may be phonetically motivated by the higher sonority and prominence of the low vowels.

5.3 Final remarks

As the weight contrast in low vowels becomes diminished, this may cause the entire system to fluctuate and rely on alternate feature cues for phonemic length disambiguation. The duration of high vowels already exhibits fluctuation in certain phonological environments. As demonstrated in the paper, long low vowels can act as short vowels with respect to syllable weight computation. However one would not really expect a merger in the case of the low vowels, as there already exists a substantial quality difference between the short-long pairs. Finally, with the mid vowels, there is a reduced number of contrasting minimal pairs available due to the phonotactic requirement that round final mid vowels be long. While decreasing the functional load of the binary length feature of the language, this also lessens the opportunity for contrast and allows the language user to predict mid vowel length based on the position of the vowel within the word. It will be interesting to continue to follow developments in how the Hungarian phonology computes syllable weight of word final syllables.

References

Abu-Mansour, Mahasen Hasan. 1995. Optimality and conspiracy in the syllable structure of Arabic. In *Papers in Optimality: University of Massachusetts Occasional Papers*, eds. Jill N. Beckman, Laura Walsh Dickey and Suzanne Urbanczyk. Amherst, MA.

Benkő, Loránd. 1967-1984. *A Magyar nyelv történeti-etimológiai szótára [The Hungarian language etymological dictionary]*. Budapest: Akadémiai Kiadó.

Curtis, Emily. 2003. Geminate weight: case studies and formal models, Department of Linguistics, University of Washington.

Davis, Stuart. 2007. Quantity. In *To appear in The Handbook of Phonological Theory*: Blackwell Publishers.

Grimes, Stephen. 2007. On the creation of a pronunciation dictionary for Hungarian: Indiana University, manuscript.

Ham, William H. 2001. *Phonetic and Phonological Aspects of Geminate Timing*: Outstanding Dissertations in Linguistics. New York: Routledge.

Hayes, Bruce. 1982. Extrametricality and English stress. *Linguistic Inquiry* 13:227-276.

Hayes, Bruce. 1986. Inalterability in CV Phonology. *Language* 62:321-351.

Hayes, Bruce. 1989. Compensatory lengthening in moraic phonology. *Linguistic Inquiry* 20:253-306.

Hayes, Bruce. 1995. *Metrical Stress Theory: Principles and Case Studies*. Chicago: University of Chicago Press.

Hock, Hans. 1986. Compensatory lengthening: in defense of the concept "mora". *Folia Linguistica* 20:431-460.

Hyman, Larry. 1992. Moraic mismatches in Bantu. *Phonology* 9:255-265.

Ito, Junko. 1989. *Syllable Theory in Prosodic Phonology*. New York: Garland Press.

Kenesei, István, Vago, Robert M., and Fenyvesi, Anna. 1998. *Hungarian*: Descriptive Grammars. New York: Routledge.

Kerek, A. 1971. *Hungarian metrics: Some linguistic aspects of iambic verse*.vol. 117: Uralic and Altaic Series. The Hague: Mouton.

Kornai, András. 1991. Hungarian vowel harmony. In *Approaches to Hungarian III*, ed. István Kenesei, 183-240. Szeged: JATE.

Kornai, Ándrás. 1986. Szótári adatbázis az akadémiai nagyszámitógépen [A dictionary database of Hungarian]. In *Working Papers*, 65-79. Budapest: Hungarian Academy of Sciences Institute of Linguistics

McCarthy, John J., and Prince, Alan. 1986/1996. Prosodic Morphology 1986. New Brunswick, NJ: Rutgers University Center for Cognitive Science.

McCarthy, John J., and Prince, Alan. 1993. Prosodic Morphology I: Constraint Interaction and Satisfaction. New Brunswick, NJ: Rutgers University Center for Cognitive Science.

Morén, Bruce. 2001. *Distinctiveness, Coercion and Sonority: A Unified Theory of Weight*: Outstanding Dissertations in Linguistics. New York & London: Routledge.

Morén, Bruce. 2003. Weight typology: An Optimality Theoretic approach. *The Linguistic Review* 20:281-304.

Nádasdy, Ádám. 1985. Segmental phonology and morphology. In *Approaches to Hungarian I*, ed. István Kenesei. Szeged: JATE.

Nádasdy, Ádám. 1989. Consonant length in recent borrowings into Hungarian. *Acta Linguistica Hungarica* 39:195-213.

Nádasdy, Ádám, and Siptár, Péter. 1994. A magánhangzók. In *Strukturális magyar nyelvtan [Structural Hungarian Grammar], volume 2*, ed. Ferenc Kiefer, 42-182. Budapest: Akadémiai Kiadó.

Obendorfer, Rudolf. 1975. The ambiguous status of Hungarian long consonants. *Lingua* 36:325-336.

Pintzuk, Susan, Kontra, Miklos, Sándor, Klára, and Borbély, Anna. 1995. The effect of the typewriter on Hungarian reading style. *Working Papers in Hungarian Sociolinguistics* 1.

Polgárdi, Krisztina. 2005. Geminates and degemination in Hungarian. In *International Conference on the Structure of Hungarian* Veszprém, Hungary.

Ringen, Catherine O., and Vago, Robert M. 2006. Geminates: Heavy or Long? Ms.

Selkirk, Elisabeth. 1990. A two-root theory of length. *University of Massachusetts Occasional Papers*.

Siptár, Péter. 1991. Marginal vowels in Hungarian. Paper presented at *The XIIth International Congress of Phonetic Sciences (ICPhS)*, Aix-en-Provence, France.

Siptár, Péter, and Törkenczy, Miklos. 2000. *The phonology of Hungarian*. Oxford: Oxford University Press.

Törkenczy, Miklos. 1994. A szótag [The syllable]. In *Strukturális Magyar Nyelvtan, 2. kötet: Fonológia [Structural Hungarian Grammar volume 2: Phonology]*, ed. Ferenc Kiefer, 272-392. Budapest: Akadémiai Kiadó.

Tranel, Bernard. 1991. CVC light syllables, geminates, and moraic theory. *Phonology* 8.

Vago, Robert M. 1980. *The Sound Pattern of Hungarian*. Washington, D.C.: Georgetown University Press.

Vago, Robert M. 1992. The root analysis of geminates in the moraic phonology of Hungarian. In *Approaches to Hungarian*, eds. István Kenesei and Csaba Pléh. Szeged: JATE.

Váradi, Tamás. 2002. The Hungarian National Corpus. Paper presented at *Third International Conference on Language Resources and Evaluation*, Las Palmas, Spain.

Zec, Draga. 1994. *Sonority Constraints on Prosodic Structure*: Outstanding dissertations in linguistics. New York: Garland.

Zec, Draga. 1995. Sonority constraints on syllable structure. *Phonology* 12:85-129.