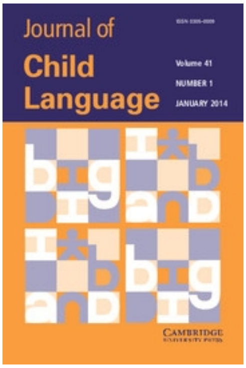

Harmonic cues for speech segmentation: a cross-linguistic corpus study on child-directed speech

F. NIHAN KETREZ

Journal of Child Language / Volume 41 / Issue 02 / March 2014, pp 439 - 461
DOI: 10.1017/S0305000912000724, Published online: 21 February 2013

J. Child Lang. **41** (2014), 439–461.
doi:10.1017/S0305000912000724

BRIEF RESEARCH REPORT

Harmonic cues for speech segmentation: a cross-linguistic corpus study on child-directed speech*

F. NIHAN KETREZ
Istanbul Bilgi University

(*Received* 5 *May* 2011 – *Revised* 26 *January* 2012 – *Accepted* 29 *November* 2012 – *First published online* 21 *February* 2013)

ABSTRACT

Previous studies on the role of vowel harmony in word segmentation are based on artificial languages where harmonic cues reliably signal word boundaries. In this corpus study run on the data available at CHILDES, we investigated whether natural languages provide a learner with reliable segmentation cues similar to the ones created artificially. We observed that in harmonic languages (child-directed speech to thirty-five Turkish and three Hungarian children), but not in non-harmonic ones (child-directed speech to one Farsi and four Polish children), harmonic vowel sequences are more likely to appear within words, and non-harmonic ones mostly appear across word boundaries, suggesting that natural harmonic languages provide a learner with regular cues that could potentially be used for word segmentation along with other cues.

INTRODUCTION

Word segmentation and vowel harmony

One of the first tasks for a child in the process of language acquisition is to learn which sound sequences correspond to words in natural speech. Speech that children hear is usually composed of long strings of words and does not necessarily provide cues regarding word boundaries (Cole & Jakimik, 1980). Despite that, children are universally successful at word learning, and they rarely produce word boundary errors. Thus the question how children learn which sound sequences correspond to words in a particular language naturally arises.

[*] I would like to thank W. U. Dressler, Fereshteh Kowssar, Agnieszka Lazorczyk, and Shadi Ganjavi for their help with Hungarian, Polish, and Farsi data, and Bariş Kabak, Toben H. Mintz, Andrew Nevins, Charles Yang, and two anonymous reviewers for their comments and suggestions on various versions of this study. Address for correspondence: Istanbul Bilgi University, Department of English Language Teacher Education, Eyüp, İstanbul 34060, Turkey. e-mail: ketrez@gmail.com

Previous research has shown that infants rely heavily on prosodic cues when they are segmenting words in a string of speech (Curtin, Mintz & Christiansen, 2005; Cutler & Norris, 1988; Jusczyk, 1997; 1999; Jusczyk, Houston & Newsome, 1999; Morgan, 1996; Thiessen & Saffran, 2003; among others). Statistical probabilities and phonotactic constraints, i.e., constraints in the order of phones, are also observed to be helpful in word segmentation (Aslin, Saffran & Newport, 1998; Brent & Cartwright, 1996; Saffran, Newport & Aslin, 1996; among others, though see also Cairns, Shillcock, Chater & Levy, 1997; Mattys, Jusczyk, Luce & Morgan, 1999; Yang, 2004). It seems natural to assume that other phonological regularities, for instance, vowel harmony, may contribute to word segmentation as well. In harmonic languages, words tend to have either all front or all back vowels. Assuming that a great majority of the words are harmonic in a language, vowel harmony may be relevant to word boundaries because shifts from one vowel type to the other (shifts from back to front or the other way around) may signal word boundaries. Indeed, studies based on Finnish show that adults can recognize words on the basis of harmony cues (Suomi, McQueen & Cutler, 1997; Vroomen, Tuomainen & Gelder, 1998). In these studies, pseudo-words were created based on Finnish vowel harmony rules where word boundaries and harmony shifts overlapped and adults were observed to be sensitive to such shifts. A similar study based on Turkish vowel harmony rules was conducted on Turkish and French speakers and the role of vowel harmony vs. word stress on word segmentation was tested. The results suggested that Turkish speakers showed sensitivity to vowel sequences in word segmentation and used them together with word stress regularities, while French speakers relied only on stress cues in word recognition (Kabak, Maniwa & Kazanina, 2010). These promising results reported in the literature based on adult data raise the question of whether or not children use vowel harmony cues in a similar way in learning words. To investigate this question, Mintz and Walker (2006) conducted a head-turn experiment in which infants aged 0;7 listened to nonsense sequences such as *detipobubeditopu* and observed that infants were sensitive to vowel shifts in such sequences and recognized harmonic sequences such as *deti* or *pobu* as words. These findings suggested that when harmonic information was available, children could use it to assign word boundaries. Similar results were reported by van Kampen, Parmaksiz, van de Vijver and Höhle (2008) based on a study where children acquiring Turkish and German were tested using artificially created words. These studies showed that children were capable of segmenting words following vowel cues, but they did not show whether they could use vowel harmony as a segmentation mechanism in word learning in real life.

It is important to note that languages created and tested in the literature are all artificial languages where harmonic cues RELIABLY point to word boundaries. In other words, when word segmentations coincide with non-harmonic vowel shifts, they are attributed to the role of vowel harmony (or disharmony) in segmentation. The question that still remains to be answered is whether NATURAL speech has such cues and thus whether it is reasonable to assume that children acquiring harmonic languages have access to reliable harmonic cues.

In a learning mechanism where vowel harmony is used, words are expected to be consistently harmonic, as a non-harmonic word would give an erroneous cue or a false alarm for a word boundary. Such false alarms are difficult to undo in a mechanism that operates on vowel harmony cues. Similarly, a mechanism that assigns word boundaries only at vowel shifts will have many 'misses' due to utterances that are composed of words that have only the same type of vowel. In such instances, a learner would treat the whole utterance as one single word unless s/he can rely on some other cues. Unlike the artificial languages that are used in experiments, natural languages have many false alarm contexts as well as misses, which make it challenging to segment words with vowel harmony cues. Therefore a learning mechanism should assume additional cues, and such cues are available in natural languages in the form of the phonotactic regularities of phonemes, stress patterns, the distributional properties of words, or some other regularities. Earlier studies have already shown that stress patterns were acknowledged as more reliable cues by speakers of Finnish, Turkish, and French, for example, and were preferred over vowel cues (Kabak *et al.*, 2010; Vroomen *et al.*, 1998). In this study we investigate the nature and the amount of vowel harmony cues that could be used WITH OTHER CUES to segment words in natural speech by contrasting the vowel distributions within and across words.

Because we investigate the relevance of vowel harmony to word segmentation, we need to examine the harmony patterns BOTH within words, and across word boundaries. The motivation for this study lies in the idea that, although harmonic languages are not fully harmonic, and word boundaries are not necessarily non-harmonic, there may still be some contrastive regularity in 'within word' vs. 'across words' contexts that could potentially complement other segmentation cues. For example, although a harmonic sequence such as /aa/ or /au/ is observed both at word boundaries and within words, it may be more frequent within words. An opposite trend may be recorded for non-harmonic sequences such as /ai/ or /ae/, i.e., they may appear across word boundaries more often. If we can document such a complementary distribution pattern, it may enable us to recognize vowel harmony as a possible segmentation tool. To this end, we look at vowel distributions in two harmonic languages, Turkish and Hungarian. We further contrast the patterns in harmonic languages with the patterns in two non-harmonic languages, Farsi and Polish, through parallel analyses run on the

TABLE 1. *Turkish and Hungarian vowels in IPA symbols and orthography*

		Front vowels in IPA	Corresponding orthography	Back vowels in IPA	Corresponding orthography
Turkish	+round	y œ	ü ö	u o	u o
	−round	i ɛ	i e	ɯ ɑ	ı a
Hungarian	+round	y ø y: ø:	ü ö ű ő	u o u: o:	u o ú ó
	−round	i e i: e:	i e í é	ɒ ɑ:	a á

four languages individually. We can assume a mechanism that can rely on vowel harmony only if within- and across-word distributions are reliably different, especially in harmonic languages.

Vowel harmony in Turkish and Hungarian

In both Turkish and Hungarian, words tend to have either all front or all back vowels. Turkish has eight vowels (four front and four back), as seen in Table 1. According to the rules of the front/back or palatal vowel harmony, a word can have either all front or all back vowels (Clements & Sezer, 1982; Demircan, 1996; among others). Turkish word formation is mainly realized through suffixation, and suffixes undergo vowel harmony in such a way that the harmonic status of the words is maintained. In the word *at-lar-ımız-dan* 'horse-PLU-POSS-ABL','from our horses' the word *at* has a back vowel and all the suffixes that are attached to this word are back (i.e., *-lar*,*-ımız*, *-dan*). In contrast, a word such as *ev* has a front vowel, and takes the front variety of the suffixes (i.e., *-ler*, *-imiz*, *-den*): *ev-ler-imiz-den* 'house-PLU-POSS-ABL'. The language has many non-harmonic words as well due to borrowings and compounding. Some of these non-harmonic words, including the word *anne* 'mother' are very common in child-directed and child speech. Despite these non-harmonic words, a great majority (90%) of words are reported to be harmonic, based on a Turkish corpus study of 601 words (Rodd, 1997). In the case of the non-harmonic words, suffixes harmonize in accordance with the last vowel of the word. Some inflections, such as the progressive marker *-Iyor*, harmonize only partially (it appears as *-üyor*, *-uyor*, *-iyor*, and *-ıyor*, with the change of the first vowel only; the last vowel remains as /o/). Besides the palatal vowel harmony, Turkish has rounding harmony that applies to high vowels as well, but it will be beyond the scope of this study.

Hungarian has seven short vowels and seven long vowels, which are longer versions of the short counterparts, with two exceptions (see Table 1). The sound represented by é is the long version of /e/, not /ɛ/ and the sound represented by á is /ɑ:/ not /ɒ:/. In addition to these fourteen vowels, nearly half of Hungarian speakers distinguish an eighth vowel, a short vowel, ë /ɛ/ as in the word *szëg* 'carpenter's nail' (Abondolo, 1987), but it was not found in the corpus, so it was not covered in the analyses. Just as is the case in Turkish, Hungarian has an internal vowel harmony that concerns the vowel sequences within word stems, and an external vowel harmony through which suffixes harmonize in word formation (e.g., *szűr-tök* 'you-PLU strain' vs. *szúr-tok* 'you-PLU pierce'). An important exception to this generalization is the words that have *i* and *í*, so-called 'neutral' vowels, which take suffixes with back vowels although they are front (e.g., *ír-tok* 'you-PLU write'). Just like Turkish, Hungarian has rounding harmony as well (e.g.,*-tok*/*-tök*, as exemplified above vs. *-tëk* after non-round vowels), but it will be beyond the scope of this study.

In general, the terms 'harmonic' and 'non-harmonic' are used to refer to relations between vowels WITHIN words. In the present study it is important to analyze harmony relationships both within words and across word boundaries. For the sake of simplicity, we use these same terms to refer to relations of vowels ACROSS word boundaries as well. Those word boundaries that have only back or only front vowels on both sides of the boundary are categorized as harmonic. In an utterance such as *Erel-cim bu ne* 'Erel-DIM this what' 'Erel-dear, what's this?' from the CHILDES Turkish corpus (File 1aa), the last vowel of the word *erelcim* is a front vowel. The vowel in the following word is back. The one in the last word is front, resulting in the sequence front–front–front # back # front, where # stands for a word boundary. So a shift from front to back and then back to front vowel corresponds to a word boundary in this particular example. In contrast, in an utterance such as *kimin teybi* 'who-GEN-3S (tape)recorder-POSS-3S' 'Whose (tape) recorder?' from the same corpus, all the vowels are front; therefore there is no word boundary cue provided by vowel shifts.

Farsi and Polish as control languages

In order to test whether the regularities that are observed are due to the harmonic nature of Hungarian and Turkish and are not just accidental, results from harmonic languages are compared to non-harmonic languages. Farsi, an Indo-European language spoken in Iran, is selected as a non-harmonic language because its vowel system is symmetrical – it has three front and three back vowels, as seen in Table 2. In regular orthographic spelling, vowels are not represented, but a corpus with phonologically transcribed utterances (including vowels) is available in CHILDES. In the table, the vowels are shown just as they are represented in the orthography of the corpus.

Polish is a West Slavonic language of central Europe. It is written in the Latin alphabet with a few additions. Its transcription is transparent in terms of the pronunciation of vowels, which are listed in Table 2 (Jassem, 2003). In addition to three front and three back vowels, it has two nasal vowels,

TABLE 2. *Farsi and Polish vowels in IPA symbols and orthography*

		Front vowels in IPA	Corresponding orthography	Back vowels in IPA	Corresponding orthography
Farsi	+round			u o	u o
	−round	i ɛ æ	i e æ	α	a
Polish	+round			u ɔ ɔ̃	u/ó o ą
	−round	i ɛ ɨ ɛ̃	i e y ę	α	a

TABLE 3. *Corpus details*

Language	Corpus	Age range of children	No. of utterances	No. of words
Turkish	Aksu	2;0–4;8	10,232	34,391
Hungarian	MacWhinney	2;3–2;10	11,478	41,514
Farsi	Family	1;11–2;10	13,325	40,472
Polish	Weist	1;7–3;2	13,258	130,778

which are marked clearly in orthography. The vowel /u/ has two different representations, u and ó. When *i* appears before vowels, it is not pronounced as /i/, it rather palatalizes the preceding consonant, so in such sequences it is not treated as a vowel. The alphabet character *y* represents a variant of /i/ rather than a consonant and is treated as such. It is not included in the vowel chart as a distinct phoneme in some sources (e.g., Stone, 1987). An important difference between Polish/Farsi and Hungarian/Turkish is that in non-harmonic languages, unlike the harmonic ones, suffixes do not undergo vowel harmony to fit the properties of the vowels in the stems.

METHOD

The corpora

All the analyses were conducted on child-directed speech from the corpora available at CHILDES. Details of the corpora are reported in Table 3. The Turkish analysis was conducted on the entire Aksu corpus of thirty-five children (Slobin, 1982). The data were collected at the children's homes by an experimenter. Family members were present during the sessions and were involved in the conversations, but the majority of the child-directed speech comes from the experimenters. All the utterances produced by all the adults were included in the analysis. A parallel analysis was conducted on the Hungarian corpus (MacWhinney, 1974). Three children were included in the analysis because the number of utterances and word tokens they provided were similar to the amount of data that was available for

Turkish. The sessions were recorded at a children's kindergarten. In addition to the experimenter, teachers and other children contributed to the conversations. All child-directed utterances were included in the analysis.

The Farsi and Polish data are from CHILDES as well (Farsi: Family, 2009; Polish: Weist & Witkowska-Stadnik, 1986; Weist, Wysocka, Witkowska-Stadnik, Buczowska & Konieczna, 1984). For Farsi, data from one child (Lilia) was sufficient to provide a comparable amount of data. Lilia's recordings were made at home by her parents. A nanny, a brother and other family members were also present in some sessions and contributed to the recordings. All the child-directed speech was included in the analysis. The Polish analysis was conducted on the Weist corpus of four children that provided a similar number of utterances. Parents, investigators and in some sessions other family members participated in the recordings and all child-directed speech was included in the analysis.

Procedure

The first analysis presents the proportion of harmonic versus non-harmonic words in the language and shows how harmonic a language is. The 200 most frequent multisyllabic words (frequency range: 20–453 in Turkish, 16–402 in Hungarian, and 23–687 in Farsi, 18–400 in Polish) were selected and coded as harmonic or non-harmonic according to the frontness–backness features of the vowels. The 200 most frequent words were selected in order to exclude words with very low frequency. In the case of Hungarian, for example, this restriction excluded words that occurred less than sixteen times in the whole corpus of over 40,000 word tokens. All multisyllabic words that have the same frequency as the 200th word were included in the analysis as well. So in Turkish, Farsi and Polish 203, and in Hungarian 212, multisyllabic words were included.

The second analysis compared the occurrence frequencies of vowel pairs within words and across word boundaries. The goal of this analysis was to see whether harmonic sequences (e.g., /aa/, /au/, /ee/) were more likely to occur within words rather than across word boundaries, and whether it was the other way around for the non-harmonic sequences (e.g., /ae/, /ai/, /uœ/). For this analysis, the same corpus was used, but in their entirety rather than limiting the analysis to the 200 most frequent multisyllabic word types. A list of all possible vowel pairs was created, categorized as harmonic or non-harmonic, and the frequency of each vowel sequence within and across word boundaries was calculated and compared. These lists of vowel pairs that were created individually for each language can be seen in the 'Appendix'.

The reason why we look at vowel pairs composed of two vowels rather than longer sequences of vowels is that word boundaries are immediately

TABLE 4. *Harmonic and non-harmonic words in types and tokens in Turkish and Hungarian*

		Harmonic (%)	Non-harmonic (%)	Total
Turkish	Types	177 (87·2)	26 (12·8)	203
	Tokens	12083 (89·47)	1421 (10·53)	13504
Hungarian	Types	151 (71·2)	61 (28·8)	212
	Tokens	7989 (77·4)	2331 (22·5)	10320

surrounded by two vowels (the last vowel of the first word and the first vowel of the second word), and we would like to compare the same vowel pairs occurring within and across word boundaries.

It is important to note that how we look at harmony within words in the second analysis is different from what we do in the first analysis, where whole words were coded as harmonic and non-harmonic regardless of the individual (co-)occurrences of vowels within a word. In the second analysis, we look at vowel pairs that occur adjacent to each other (when consonants are left out). For example, in the first analysis, a word such as *masaya* 'table-DAT' is counted as one harmonic word. In the second analysis, it is counted as two /aa/ sequences. We need the first analysis to see how harmonic a language is. We need to look at vowel pairs rather than entire words as well because it is the only way we can compare within-word contexts to across-word boundary contexts.

RESULTS

Harmonic words

In this analysis, we look at the proportion of harmonic words in harmonic and non-harmonic languages to see how harmonic a language is. The results of this analysis will provide us with a clearer idea of the vowel distribution patterns in word-size units. The analysis is restricted to the most frequent 200 multisyllabic words in each language.

The results show that, despite the presence of non-harmonic words in both Turkish and Hungarian, a majority of words are harmonic in both these harmonic languages. Table 4 displays the number of word types and tokens in frequencies in the two languages.

Table 5 reports the frequencies of harmonic and non-harmonic words in two non-harmonic languages, Farsi and Polish. As predicted for a non-harmonic language, in Farsi, harmonic and non-harmonic word types have equal proportions. Interestingly, non-harmonic words are even more frequent if we look at the frequency of words in terms of tokens. Polish presents a similar pattern, where harmonic and non-harmonic words have

TABLE 5. *Harmonic and non-harmonic words in types and tokens in Farsi and Polish*

		Harmonic (%)	Non-harmonic (%)	Total
Farsi	Types	100 (49·2)	103 (51·8)	203
	Tokens	7133 (32·4)	14924 (67·6)	22057
Polish	Types	89 (43·8)	114 (56·2)	203
	Tokens	5982 (54·6)	4961 (45·4)	10943

TABLE 6. *Harmony within (VV) versus across (V#V) word boundaries in Turkish and Hungarian*

		V#V token (%)	VV token (%)	Total
Turkish	Harmonic	10724 (48·8)	35097 (89·4)	45821
	Non-harmonic	11245 (51·1)	4140 (10·5)	15385
	Total	**21969**	**39237**	**61206**
Hungarian	Harmonic	15284 (50·5)	20465 (77·6)	35746
	Non-harmonic	14955 (49·4)	5880 (22·3)	20835
	Total	**30239**	**26345**	**56541**

almost the same proportion, in both the token and the type analysis, as seen in the table.

Overall results from the four languages showed that the great majority of words in child-directed speech were harmonic in harmonic languages. In non-harmonic languages, such a trend was not observed. Moreover, an opposite trend was recorded in the case of the word tokens in Farsi.

Harmony within versus across word boundaries

In this part of the study, token occurrences of harmonic and non-harmonic sequences of vowels within and across word boundaries (VV vs. V#V) were contrasted. The goal of this analysis is to see whether there is a contrastive pattern of occurrence of harmonic and non-harmonic vowel pairs. If we can show that harmonic and non-harmonic pairs are likely to occur with different distributional properties, the former appearing more within words, and the latter more across word boundaries, we could use this to support our hypothesis that vowel distributions can provide a learner with helpful cues for word segmentation.

Table 6 displays the total counts of Turkish and Hungarian analyses. Harmonic and non-harmonic vowel sequences that occur across (V#V) and within word boundaries (VV) are shown in the first and the second columns of the table. The within-words counts show that a great majority of harmonic vowel pairs appear within words in Turkish and Hungarian. This result is similar to the result of the first analysis (Table 4), where we saw that a great majority of words are harmonic in harmonic languages. The word boundary counts (V#V) show that about half of the word boundaries are harmonic in Turkish and Hungarian. What is interesting and more relevant for the purpose of this study, however, is the contrast between within- and across-words contexts. Although only half of the word boundaries are non-harmonic, the contrast between within-words and across-words contexts clearly shows that vowel distributions within and across word boundaries are different. Harmonic sequences mostly occur within words, while non-harmonic sequences are found at word boundaries.

In the case of Turkish, although harmonic sequences can be found both across and within words, they are more likely to appear within words, while non-harmonic sequences appear mostly at word boundaries. Examination of individual vowel pairs that are reported in the 'Appendix' provides further evidence for the contrastive distribution pattern of the vowels. For example, although a harmonic sequence such as /aa/ can be found both within words and at the two sides of a word boundary, it is more likely to be found WITHIN words. In all, 6,318 occurrences of /aa/ were recorded in the corpus and 5,133 (81·2%) of such occurrences were within word boundaries. Similarly, the /œɛ/ sequence is recorded 712 times, and 706 of them (99%) are within words, although it may appear at word boundaries as well. Not all harmonic pairs appear mostly within words though. In all, 11 out of 32 vowel pairs are found more than 50% of the time across word boundaries. However, as seen in Table A in the 'Appendix', the token frequencies of such unexpected occurrences are not very high. Therefore, they do not affect the overall tendencies.

An opposite trend is observed in the non-harmonic sequences, i.e, they are more likely to be observed at word boundaries. The sequence /ea/, for example, is recorded 2,919 times in the sessions, and 2,663 (91%) of them occur across word boundaries, although it is a possible sequence within words as well. Only 2 out of 32 non-harmonic pairs are observed within word boundaries in more than 50% of the cases, and both of them have the vowel /o/ as the second member of the VV pair and a high vowel as the first member. What contributes to this unexpected distribution is the non-harmonizing and frequent progressive marker *-Iyor*.

The overall Turkish results suggest that, although harmonic word boundaries are frequent, it is more likely for a harmonic sequence to be found within a word boundary than across word boundaries (χ^2 (1, N=61,206)=12,357·419, p <.0001). This picture suggests that when occurrences of individual sequences within and across word boundaries are contrasted, a promising picture appears in Turkish.

TABLE 7. *Harmony within (VV) versus across (V#V) word boundaries in Farsi and Polish*

		V#V token (%)	VV token (%)	Total
Farsi	Harmonic	14232 (53)	20010 (50·8)	34242
	Non-harmonic	12597 (47)	19308 (49·1)	31905
	Total	**26829**	**39318**	**66147**
Polish	Harmonic	17659 (46·1)	17520 (48·1)	35179
	Non-harmonic	20645 (53·8)	18933 (51·9)	39578
	Total	**38304**	**36453**	**74757**

A similar, but not exactly the same, distributional regularity is observed in Hungarian. In the case of the non-harmonic sequences, the Turkish results are replicated, i.e., a majority of non-harmonic sequences appear across word boundaries. The Hungarian results are even stronger, because about half of the non-harmonic vowel pairs are not observed within words at all, while in Turkish those non-harmonic pairs that only occur across word boundaries are about 20% of the total pairs. The sequence /ae/, for example, is recorded 2,176 times, and 2,022 (92·9%) of them appear across word boundaries. Although there are non-harmonic vowel pairs that are more frequent within words, they are not that many in terms of token frequencies, as seen in Table B in the 'Appendix'. What is also important is that a majority of non-harmonic sequences that are exceptionally more frequent within words contain the 'neutral' vowels that behave like both a back vowel and a front vowel.

In the case of the harmonic sequences, the distribution of vowels within and across word boundaries is not always as expected. In the case of Turkish, almost all individual harmonic sequences are more frequent within words, while in Hungarian, the occurrence of harmonic sequences are not necessarily associated with within-word contexts. The sequence /aa/ is found across a word boundary in 54% of the instances, for example. Despite that, the proportion of non-harmonic sequences that occur across word boundaries is high, and the overall results suggest that in Hungarian, just as is the case in Turkish, harmonic sequences tend to occur within words, and non-harmonic sequences across words (χ^2 (1, N=56,541)=4,424,498, $p<.0001$).

The similarity of Hungarian to Turkish becomes apparent when these two harmonic languages are compared to Farsi and Polish, two non-harmonic languages. In Farsi (Table 7), within-word and across-words contexts have equal proportions of harmonic and non-harmonic sequences. Although the proportion of harmonic sequences found within words may provide potentially useful cues for word recognition, despite the fact that Farsi is not a harmonic language, the distribution of non-harmonic sequences with

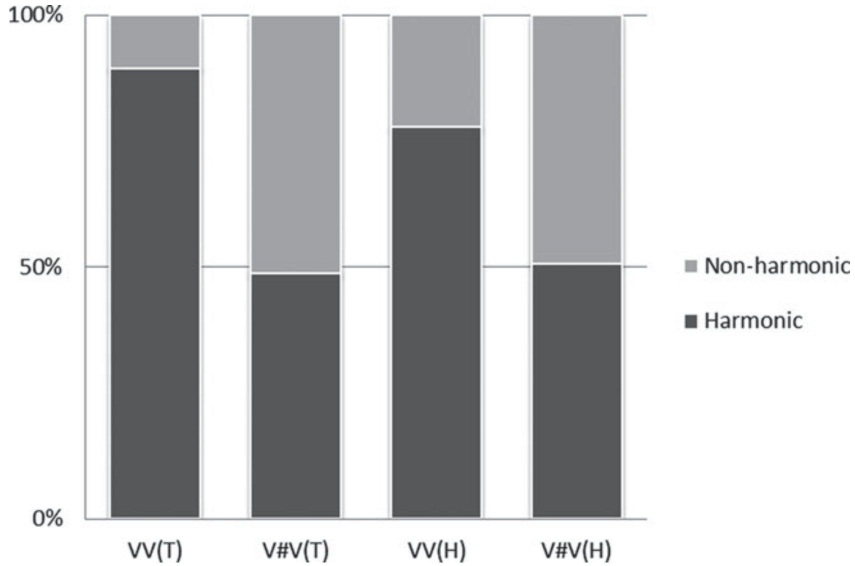

Fig. 1. Harmony within (VV) versus across (V#V) word boundaries in Turkish (T) and Hungarian (H).

an exact same pattern blurs the results because harmonic and non-harmonic sequences do not display a contrastive distribution. The Farsi results, especially when all the analyses are considered together, provide a clear contrast to Turkish and Hungarian, where harmonic sequences are more likely to occur within words, and non-harmonic sequences across word boundaries. Another contrast can be seen between Polish and the harmonic languages. Unlike the pattern in Turkish and Hungarian, in Polish both harmonic and non-harmonic sequences are equally likely to appear across word boundaries and within words (Table 7).

A closer look at the individual pairs reported in the 'Appendix', Tables C (Farsi) and D (Polish) provides a clearer contrastive pattern. Unlike harmonic languages, in neither Farsi nor Polish are there vowel pairs that exclusively appear within words or across word boundaries. A majority of vowel sequences can be found in both contexts with equal frequency, regardless of their harmony status.

The contrast between harmonic and non-harmonic languages can be clearly observed in Figures 1 and 2 as well. VV(T), VV(H), VV(F) and VV(P) bars represent the harmonic (dark) and non-harmonic sequences within word boundaries in Turkish, Hungarian, Farsi, and Polish, respectively. The numbers are the number of sequence (/aa/, /ae/, etc.) tokens. What is important to note for the purpose of this study is the

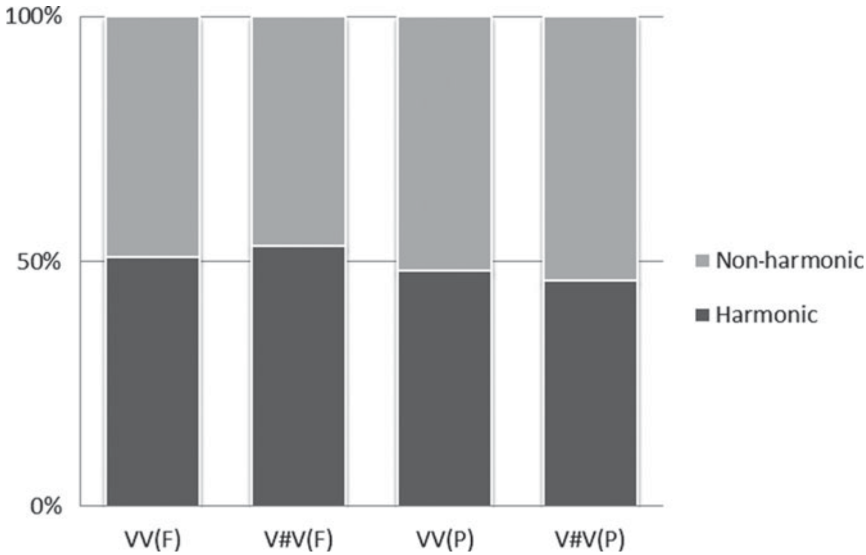

Fig. 2. Harmony within (VV) versus across (V#V) word boundaries in Farsi (F) and Polish (P).

similarity of the VV(T) and VV(H) bars, and their dissimilarity to VV(F) and VV(P).

As seen in Figures 1 and 2, in both Hungarian and Turkish, harmonic sequences are clearly more frequent within words (VV) than across word boundaries (V#V), and non-harmonic sequences are more frequent across word boundaries (V#V). In Farsi and Polish, such a contrastive pattern is not observed. Harmonic and non-harmonic sequences show the same pattern within words as across word boundaries, failing to provide any potential cues for word recognition.

DISCUSSION AND CONCLUSION

In this study, we investigated whether harmonic languages provide a learner with harmonic cues that could be used in word segmentation together with other cues. We looked at the proportion of harmonic vs. non-harmonic words in harmonic and non-harmonic languages to see whether harmonic languages really are as harmonic as one expects. We further looked at the distributional properties of vowel pairs to see whether harmonic and non-harmonic vowel pairs appear in distinct contexts (within words and across word boundaries), providing potential cues for segmentation.

The results suggested that a great majority of words are harmonic in Turkish and Hungarian child-directed speech. It was also observed that, although some harmonic sequences may occur across word boundaries as frequently as non-harmonic sequences do, they are more likely to occur WITHIN words in harmonic languages. Therefore there is some vowel harmony information embedded in the natural data, although it is not as easily accessible as it has been assumed in the literature, and it may be useful in word boundary recognition, especially when it is used together with some other cues. Newport and Aslin (2004) show that learners are capable of acquiring patterned relations among non-adjacent segments. It is likely that vowel harmony is one such non-adjacent relationship that can be learned and used in word learning together with some other cues. A comparison of vowel sequences occurring within and across word boundaries may provide a learner with helpful statistical cues to rely on in speech segmentation.

Analyses conducted on two non-harmonic languages, Farsi and Polish, further showed that the regular pattern observed in harmonic languages is not an accidental one. Rather, it is peculiar to harmonic languages, and learners of harmonic languages most probably use the language-specific regularities available to them, while speakers of non-harmonic languages learn to rely on other cues. These results find support in experimental settings. They correctly predict the findings reported by van Kampen *et al.* (2008), who show that Turkish infants aged 0;6 (but not the German infants of the same age) prefer listening to harmonic over non-harmonic pseudo-words. Moreover, Turkish infants aged 0;9 can segment words using vowel-harmony cues. The results in the present study are in line with the results of Kabak *et al.* (2010), as well, who report that adult Turkish speakers, but not the speakers of a non-harmonic language, follow harmony cues in word segmentation together with other cues. These results are not surprising considering the vowel harmony and word boundary patterns reported in this study.

The goal of this study was to examine whether there is any regularity based on vowel harmony that could potentially be used in word segmentation in the acquisition of harmonic languages. Although we examined only the sequences of vowels, without taking into consideration any other potential cue (for example, co-occurrence regularities of consonants, word stress, distributional properties of word-size units) that can be used together with harmony cues, we do not have the intention of assuming a learning mechanism that operates exclusively on vowel harmony cues. The results suggested that there is some regularity due to vowel harmony that could potentially be helpful in word segmentation, but it is also important to acknowledge that there is a need to rely on additional cues, especially in contexts where harmonic cues result in the wrong segmentation. For example, harmony across word boundaries is not problematic only if the child has access to additional cues to segment words (when one of the words is a very frequent word that the child hears in non-harmonic contexts, for example). Similarly, the difficulty due to non-harmony within words can be overcome by considering, for example, transitional probabilities. Frequent use of some suffixes in unambiguous contexts may also be helpful. For example, in the case of non-harmony within words in Turkish, one single suffix, the progressive marker *-Iyor*, has a special place, because it appears to be the reason for the only two non-harmonic pairs that appear more frequently within words. *-Iyor* is one of the earliest acquisitions in Turkish, due to its semantic salience, obligatory contexts and frequent production in child-directed speech (Aksu-Koç & Ketrez 2003). Once the learner figures out that *-Iyor* is part of the word, s/he may rely on this information for further segmentation. With multiple cues considered simultaneously, a learner can gather information that could work against erroneous harmony-based segmentation. Further research on cue weighting and cue integration, where multiple potential cues can be examined simultaneously, will reveal how much of the harmony information reported in this study can be beneficial in word segmentation.

REFERENCES

Abondolo, D. (1987). Hungarian. In B. Comrie (ed.), *The world's major languages*, 577–92. Oxford: Oxford University Press.

Aksu-Koç, A. & Ketrez, F. N. (2003). Early verbal morphology in Turkish: emergence of inflections. In D. Bittner, W. U. Dressler & M. Kilani-Schoch (eds.), *Mini-paradigms and the emergence of verb morphology*, 27–52. Berlin: Mouton de Gruyter.

Aslin, R. N., Saffran, J. R. & Newport, E. L. (1998). Computation of probability statistics by 8-month-old infants. *Psychological Science* **9**, 321–24.

Brent, M. R. & Cartwright, T. A. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition* **61**, 93–125.

Cairns, P., Shillcock, R., Chater, N. & Levy, J. (1997). Bootstrapping word boundaries: speech and statistical cues. A bottom-up corpus-based approach to speech segmentation. *Cognitive Psychology* **33**, 111–53.

Clements, G. N. & Sezer, E. (1982). Vowel and consonant disharmony in Turkish. In H. van der Hulst & N. Smith (eds.), *The structure of phonological representations* (*Part 2*), 213–55. Dordrecht: Foris Publications.

Cole, R. & Jakimik, J. (1980). A model of speech perception. In R. Cole (ed.), *Perception and production of fluent speech*, 136–63. Hillsdale, NJ: Erlbaum.

Curtin, S., Mintz, T. & Christiansen, M. (2005). Stress changes the representational landscape: evidence from word segmentation. *Cognition* **97**(3), 233–62.

Cutler, A. & Norris, D. G. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception & Performance* **14**, 113–21.

Demircan, Ö. (1996). *Türkçenin Sesdizimi*. Istanbul: Der Yayınevi.

Family, N. (2009). Lighten up: the acquisition of light verb constructions in Persian. *Boston University Conference on Language Development Proceedings* **33**, 139–50. Somerville, MA: Cascadilla Press.

Jassem, W. (2003). Polish. *Journal of the International Phonetic Association* **33**(1), 103–107.

Jusczyk, P. W. (1997). *The discovery of spoken language*. Cambridge, MA: MIT Press.

Jusczyk, P. W. (1999). How infants begin to extract words from fluent speech. *Trends in Cognitive Science* **3**, 323–28.

Jusczyk, P. W., Houston, D. M. & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology* **39**, 159–207.

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. *Journal of Laboratory Phonology* **1**, 207–224.

van Kampen, A., Parmaksiz, G., van de Vijver, R. & Höhle, B. (2008). Metrical and statistical cues for word segmentation: vowel harmony and word stress as cues to word boundaries by 6- and 9-month-old Turkish learners. In A. Gavarró & M. J. Freitas (eds.), *Language acquisition and development*, 313–24. Newcastle: Cambridge Scholars Publishing.

MacWhinney, B. (1974). How Hungarian children learn to speak. Unpublished doctoral dissertation, University of California, Berkeley.

Mattys, S. L., Jusczyk, P. W., Luce, P. A. & Morgan, J. L. (1999). Word segmentation in infants: how phonotactics and prosody combine. *Cognitive Psychology* **38**, 465–94.

Mintz, T. and Walker, R. (2006). Infant's sensitivity to vowel harmony and its role in word segmentation. Paper presented at the annual meeting of the Linguistic Society of America, Albuquerque, NM, 7 January 2006.

Morgan, J. L. (1996). A rhythmic bias in preverbal speech segmentation. *Journal of Memory and Language* **35**, 666–88.

Newport, L. N. and Aslin, R. N. (2004). Learning at a distance: statistical learning of non-adjacent dependencies. *Cognitive Psychology* **48**, 127–62.

Rodd, J. (1997). Recurrent neural-network learning of phonological regularities in Turkish. In T. M. Ellison (ed.), CoNLL97: *Computational Natural Language Learning*, 97–106. Somerset, NJ: Association of Computational Linguistics.

Saffran J, R., Newport, E. & Aslin, R. (1996). Word segmentation: the role of distributional cues. *Journal of Memory and Language* **35**, 606–621.

Slobin, D. (1982). Universal and particular in the acquisition of language. In E. Wanner & L. Gleitman (eds.), *Language acquisition: the state of the art*, 128–72. New York: Cambridge University Press.

Stone, G. (1987). Polish. In B. Comrie (ed.), *The world's major languages*, 348–66. Oxford: Oxford University Press.

Suomi, K., McQueen, J. M. & Cutler, A. (1997). Vowel harmony and speech segmentation in Finnish. *Journal of Memory and Language* **36**, 422–44.

Thiessen, E. D. & Saffran, J. R. (2003). When cues collide: use of stress and statistical cues to word boundaries by 7–9-month-old infants. *Developmental Psychology* **39**(4), 706–716.

Vroomen, J., Tuomainen, J. & Gelder, B. (1998). The roles of word stress and vowel harmony in speech segmentation. *Journal of Memory and Language* **38**, 133–49.

Weist, R., & Witkowska-Stadnik, K. (1986). Basic relations in child language and the word order myth. *International Journal of Psychology* **21**, 363–81.

Weist, R., Wysocka, H., Witkowska-Stadnik, K., Buczowska, E. & Konieczna, E. (1984). The defective tense hypothesis: on the emergence of tense and aspect in child Polish. *Journal of Child Language* **11**, 347–74.

Yang, C. D. (2004). Universal Grammar, statistics, or both. *Trends in Cognitive Sciences* **8**, 451–56.

APPENDIX: TOKEN FREQUENCIES OF HARMONIC AND NON-HARMONIC VOWEL PAIRS WITHIN WORDS (VV) AND ACROSS WORD BOUNDARIES (V#V)

TABLE A. *Harmonic and non-harmonic vowel pairs in Turkish child-directed speech*

Harmonic vowel pairs							Non-harmonic vowel pairs						
VV	*Tokens*	%	*V#V*	*Tokens*	%	*Total*	*VV*	*Tokens*	%	*V#V*	*Tokens*	%	*Total*
öü	335	100	ö#ü	1	0	336	iu	0	0	i#u	299	100	299
öe	706	99	ö#e	6	1	712	ıü	0	0	ı#ü	90	100	90
ou	2867	96	o#u	134	4	3001	oö	0	0	o#ö	28	100	28
aı	4698	94	a#ı	282	6	4980	üı	0	0	ü#ı	14	100	14
üü	709	87	ü#ü	106	13	815	uö	0	0	u#ö	78	100	78
öö	6	86	ö#ö	1	14	7	eo	5	0	e#o	1043	100	1048
ıı	922	85	ı#ı	165	15	1087	aü	3	2	a#ü	135	98	138
uu	1045	84	u#u	194	16	1239	eu	7	2	e#u	295	98	302
üe	494	83	ü#e	104	17	598	ıö	2	2	ı#ö	84	98	86
aa	5133	81	a#a	1185	19	6318	ıe	12	3	ı#e	457	97	469
oa	1594	81	o#a	385	19	1979	ui	12	4	u#i	295	96	307
ii	2792	80	i#i	699	20	3491	ue	27	5	u#e	557	95	584
ei	4174	79	e#i	1137	21	5311	eı	11	5	e#ı	191	95	202
ıo	1041	74	ı#o	359	26	1400	ıi	27	8	ı#i	327	92	354
ee	3048	72	e#e	1214	28	4262	ea	256	9	e#a	2663	91	2919
uo	592	69	u#o	263	31	855	oe	24	9	o#e	232	91	256
ua	1467	69	u#a	659	31	2126	iı	22	16	i#ı	113	84	135
ie	2107	68	i#e	1007	32	3114	aö	28	19	a#ö	122	81	150
ıa	957	55	ı#a	769	45	1726	oi	21	19	o#i	90	81	111
oo	100	47	o#o	111	53	211	ia	327	20	i#a	1348	80	1675
oı	10	32	o#ı	21	68	31	üa	32	23	ü#a	108	77	140
au	156	31	a#u	349	69	505	öo	1	25	ö#o	3	75	4
eü	29	16	e#ü	151	84	180	oü	30	32	o#ü	65	68	95
ao	100	14	a#o	598	86	698	üu	13	35	ü#u	24	65	37
üi	7	9	ü#i	67	91	74	öa	6	40	ö#a	9	60	15
iü	5	3	i#ü	152	97	157	ae	1003	44	a#e	1254	56	2257

Table A (*Cont.*)

Harmonic vowel pairs							Non-harmonic vowel pairs						
VV	*Tokens*	%	*V#V*	*Tokens*	%	*Total*	*VV*	*Tokens*	%	*V#V*	*Tokens*	%	*Total*
eö	2	1	e#ö	184	99	186	uü	52	49	u#ü	54	51	106
iö	1	1	i#ö	161	99	162	öu	1	50	ö#u	1	50	2
ıu	0	0	ı#u	248	100	248	ai	762	50	a#i	750	50	1512
uı	0	0	u#ı	0	0	0	io	1172	71	i#o	477	29	1649
öö	0	0	ö#ö	12	100	12	üo	284	88	ü#o	39	12	323
öi	0	0	ö#i	0	0	0	öı	0	0	ö#ı	0	0	0
Total	**35097**			**10724**		**45821**	**Total**	**4140**			**11245**		**15385**

TABLE B. *Harmonic and non-harmonic vowel pairs in Hungarian child-directed speech*

Harmonic vowel pairs							Non-harmonic vowel pairs						
VV	*Tokens*	%	*V#V*	*Tokens*	%	*Total*	*VV*	*Tokens*	%	*V#V*	*Tokens*	%	*Total*
éű	2	100	é#ű	0	0	2	aí	0	0	a#í	11	100	11
íő	2	100	í#ő	0	0	2	aű	0	0	a#ű	3	100	3
öű	20	100	ö#ű	0	0	20	aő	0	0	a#ő	14	100	14
öő	8	100	ö#ő	0	0	8	áí	0	0	á#í	15	100	15
űő	5	100	ű#ő	0	0	5	áő	0	0	á#ő	3	100	3
őö	19	100	ő#ö	0	0	19	éú	0	0	é#ú	16	100	16
öü	43	98	ö#ü	1	2	44	éó	0	0	é#ó	34	100	34
öö	358	96	ö#ö	13	4	371	íu	0	0	í#u	2	100	2
iű	17	94	i#ű	1	6	18	uí	0	0	u#í	1	100	1
üü	23	92	ü#ü	2	8	25	uü	0	0	u#ü	3	100	3
eü	442	91	e#ü	42	9	484	uö	0	0	u#ö	16	100	16
uo	369	90	u#o	39	10	408	úé	0	0	ú#é	11	100	11
áu	164	89	á#u	20	11	184	úü	0	0	ú#ü	2	100	2

üő	16	89	ü#ő	2	11	18	úö	0	0	ú#ö	6	100	6
eű	7	88	e#ű	1	13	8	oű	0	0	o#ű	2	100	2
uá	188	87	u#á	29	13	217	oő	0	0	o#ő	3	100	3
üö	69	86	ü#ö	11	14	80	óü	0	0	ó#ü	2	100	2
éü	85	86	é#ü	14	14	99	óő	0	0	ó#ő	1	100	1
óá	78	86	ó#á	13	14	91	üu	0	0	ü#u	6	100	6
áo	598	83	á#o	126	17	724	üú	0	0	ü#ú	2	100	2
úu	4	80	ú#u	1	20	5	űa	0	0	ű#a	55	100	55
óu	61	79	ó#u	16	21	77	űá	0	0	ű#á	6	100	6
áa	66	78	á#a	19	22	85	űu	0	0	ű#u	3	100	3
öe	388	77	ö#e	114	23	502	űo	0	0	ű#o	3	100	3
áá	190	77	á#á	56	23	246	őa	0	0	ő#a	51	100	51
uu	110	76	u#u	34	24	144	őá	0	0	ő#á	16	100	16
uó	66	73	u#ó	24	27	90	őu	0	0	ő#u	1	100	1
ua	374	73	u#a	138	27	512	őú	0	0	ő#ú	1	100	1
iő	37	73	i#ő	14	27	51	őó	0	0	ő#ó	10	100	10
őe	83	72	ő#e	33	28	116	óű	0	0	ó#ű	2	100	2
oá	420	70	o#á	178	30	598	üa	2	2	ü#a	110	98	112
íö	2	67	í#ö	1	33	3	oü	1	2	o#ü	40	98	41
űe	2	67	ű#e	1	33	3	ue	4	3	u#e	154	97	158
au	491	66	a#u	258	34	749	ué	1	3	u#é	36	97	37
ie	1630	64	i#e	934	36	2564	öa	6	3	ö#a	172	97	178
oo	569	63	o#o	329	37	898	üo	1	4	ü#o	24	96	25
üé	24	63	ü#é	14	37	38	oe	40	4	o#e	869	96	909
üi	95	63	ü#i	57	38	152	oé	11	5	o#é	226	95	237
őé	15	63	ő#é	9	38	24	oö	5	5	o#ö	102	95	107
ou	81	62	o#u	50	38	131	aü	5	5	a#ü	101	95	106
óo	46	61	ó#o	30	39	76	áe	19	5	á#e	376	95	395
iü	42	60	i#ü	28	40	70	óe	12	7	ó#e	159	93	171
öi	82	59	ö#i	56	41	138	ae	154	7	a#e	2022	93	2176
ee	3143	59	e#e	2161	41	5304	ea	190	7	e#a	2489	93	2679
eé	849	58	e#é	611	42	1460	öó	1	8	ö#ó	11	92	12
oa	1321	57	o#a	988	43	2309	üó	3	9	ü#ó	32	91	35
ao	1311	57	a#o	981	43	2292	aé	68	10	a#é	602	90	670
ée	965	57	é#e	732	43	1697	óö	2	11	ó#ö	16	89	18
úa	31	54	ú#a	26	46	57	eú	9	13	e#ú	62	87	71
oó	126	54	o#ó	109	46	235	eó	42	13	e#ó	277	87	319

Table B (*Cont.*)

Harmonic vowel pairs							Non-harmonic vowel pairs						
VV	*Tokens*	%	*V#V*	*Tokens*	%	*Total*	*VV*	*Tokens*	%	*V#V*	*Tokens*	%	*Total*
öé	82	53	ö#é	74	47	156	éo	38	14	é#o	233	86	271
óa	239	52	ó#a	217	48	456	iú	5	15	i#ú	29	85	34
óó	20	50	ó#ó	20	50	40	üá	4	17	ü#á	20	83	24
áó	31	47	á#ó	35	53	66	aö	46	18	a#ö	206	82	252
aa	1978	47	a#a	2235	53	4213	oí	5	19	o#í	21	81	26
üe	128	47	ü#e	145	53	273	öú	1	20	ö#ú	4	80	5
eö	186	44	e#ö	236	56	422	őo	3	20	ő#o	12	80	15
ei	989	44	e#i	1256	56	2245	éa	126	20	é#a	490	80	616
éi	345	43	é#i	453	57	798	eu	88	21	e#u	341	79	429
ii	354	42	i#i	485	58	839	eo	225	23	e#o	765	77	990
aá	521	42	a#á	715	58	1236	öo	9	24	ö#o	28	76	37
oú	12	40	o#ú	18	60	30	éá	29	25	é#á	87	75	116
úú	9	38	ú#ú	15	63	24	éu	16	25	é#u	48	75	64
úo	5	36	ú#o	9	64	14	öu	1	25	ö#u	3	75	4
aó	151	36	a#ó	273	64	424	eá	122	25	e#á	362	75	484
ié	152	34	i#é	290	66	442	óé	18	32	ó#é	39	68	57
úá	1	33	ú#á	2	67	3	ói	2	33	ó#i	4	67	6
őő	1	33	ő#ő	2	67	3	áö	19	35	á#ö	35	65	54
éö	22	31	é#ö	49	69	71	ia	719	35	i#a	1320	65	2039
éé	58	27	é#é	158	73	216	oi	329	39	o#i	511	61	840
úó	1	25	ú#ó	3	75	4	ía	10	42	í#a	14	58	24
ií	3	23	i#í	10	77	13	ai	827	43	a#i	1114	57	1941
iö	30	22	i#ö	104	78	134	áé	94	50	á#é	93	50	187
uú	2	22	u#ú	7	78	9	ió	128	51	i#ó	124	49	252
éő	1	17	é#ő	5	83	6	iu	109	52	i#u	99	48	208
aú	2	4	a#ú	44	96	46	úe	49	53	ú#e	44	47	93
eí	1	4	e#í	24	96	25	io	447	60	i#o	303	40	750
íe	1	3	í#e	33	97	34	áü	9	60	á#ü	6	40	15
áú	0	0	á#ú	11	100	11	úi	13	62	ú#i	8	38	21
eő	0	0	e#ő	0	0	0	öá	16	67	ö#á	8	33	24

éí	0	0	é#í	18	100	18
íé	0	0	í#é	2	100	2
íi	0	0	í#i	1	100	1
íí	0	0	í#í	9	100	9
íü	0	0	í#ü	0	0	0
íű	0	0	í#ű	0	0	0
üí	0	0	ü#í	0	0	0
üű	0	0	ü#ű	0	0	0
öí	0	0	ö#í	0	0	0
óú	0	0	ó#ú	6	100	6
űé	0	0	ű#é	3	100	3
űi	0	0	ű#i	0	0	0
űí	0	0	ű#í	0	0	0
űü	0	0	ű#ü	0	0	0
űö	3	0	ű#ö	0	0	0
űű	0	0	ű#ű	0	0	0
ői	0	0	ő#i	0	0	0
őí	0	0	ő#í	1	100	1
őü	0	0	ő#ü	0	0	0
őű	0	0	ő#ű	0	0	0
Total	**20465**			**15284**		**35746**

ío	11	79	í#o	3	21	14
ui	267	81	u#i	63	19	330
ái	671	81	á#i	156	19	827
iá	799	82	i#á	177	18	976
íó	5	83	í#ó	1	17	6
íá	44	94	í#á	3	6	47
áű	0	0	á#ű	0	0	0
íú	0	0	í#ú	0	0	0
uű	0	0	u#ű	0	0	0
uő	0	0	u#ő	0	0	0
úí	0	0	ú#í	0	0	0
úű	0	0	ú#ű	0	0	0
úő	0	0	ú#ő	0	0	0
óí	0	0	ó#í	0	0	0
űú	0	0	ű#ú	0	0	0
űó	0	0	ű#ó	0	0	0
Total	**5880**			**14955**		**20835**

TABLE C. *Harmonic and non-harmonic vowel pairs in Farsi child-directed speech*

Harmonic vowel pairs							Non-harmonic vowel pairs						
VV	*Tokens*	%	*V#V*	*Tokens*	%	*Total*	*VV*	*Tokens*	%	*V#V*	*Tokens*	%	*Total*
oa	1035	75	o#a	347	25	1382	eo	180	14	e#o	1139	86	1319
æe	2610	73	æ#e	977	27	3587	æo	499	37	æ#o	841	63	1340
ei	2212	68	e#i	1044	32	3256	io	703	42	i#o	981	58	1684
æi	1811	65	æ#i	964	35	2775	ea	1125	48	e#a	1235	52	2360
ie	1793	64	i#e	1014	36	2807	uæ	465	50	u#æ	457	50	922
aa	1379	63	a#a	799	37	2178	æu	504	57	æ#u	378	43	882
ææ	1924	63	æ#æ	1126	37	3050	aæ	1493	57	a#æ	1108	43	2601
ii	1510	61	i#i	973	39	2483	eu	608	59	e#u	427	41	1035
ou	174	55	o#u	144	45	318	oe	981	59	o#e	685	41	1666
oo	1128	54	o#o	948	46	2076	oi	701	59	o#i	480	41	1181
ee	1229	54	e#e	1034	46	2263	iu	553	60	i#u	366	40	919
au	342	52	a#u	315	48	657	oæ	1014	61	o#æ	636	39	1650
ua	304	51	u#a	289	49	593	æa	1386	67	æ#a	694	33	2080
iæ	983	48	i#æ	1072	52	2055	ae	2122	71	a#e	866	29	2988
uu	99	44	u#u	128	56	227	ui	556	71	u#i	224	29	780
eæ	1183	40	e#æ	1751	60	2934	ue	750	72	u#e	291	28	1041
ao	260	22	a#o	948	78	1208	ai	2728	75	a#i	907	25	3635
uo	34	9	u#o	359	91	393	ia	2940	77	i#a	882	23	3822
Total	**20010**			**14232**		**34242**	**Total**	**19308**			**12597**		**31905**

TABLE D. *Harmonic and non-harmonic vowel pairs in Polish child-directed speech. The character u represents both u and ó*

Harmonic vowel pairs							Non-harmonic vowel pairs						
VV	*Tokens*	%	*V#V*	*Tokens*	%	*Total*	*VV*	*Tokens*	%	*V#V*	*Tokens*	%	*Total*
uą	139	89	u#ą	18	11	157	ęo	30	7	ę#o	401	93	431
ua	1368	78	u#a	397	22	1765	io	167	13	i#o	1084	87	1251
uu	410	77	u#u	125	23	535	ęu	41	19	ę#u	175	81	216
oą	328	75	o#ą	109	25	437	ęa	152	20	ę#a	595	80	747
aą	324	72	a#ą	128	28	452	iu	116	29	i#u	280	71	396
ąu	82	71	ą#u	33	29	115	eu	344	31	e#u	783	69	1127
ię	187	68	i#ę	90	32	277	eo	984	32	e#o	2087	68	3071
ii	284	67	i#i	139	33	423	ea	1428	36	e#a	2529	64	3957
yi	303	66	y#i	155	34	458	yu	158	41	y#u	228	59	386
iy	313	62	i#y	195	38	508	ia	595	42	i#a	834	58	1429
ie	774	57	i#e	588	43	1362	aę	353	43	a#ę	468	57	821
ęe	397	57	ę#e	302	43	699	ąę	353	43	ą#ę	468	57	821
au	1082	56	a#u	848	44	1930	ią	37	45	i#ą	46	55	83
yy	257	55	y#y	214	45	471	ąi	43	45	ą#i	52	55	95
ei	610	54	e#i	510	46	1120	eą	140	47	e#ą	158	53	298
ou	800	53	o#u	698	47	1498	yo	554	47	y#o	624	53	1178
oa	2041	53	o#a	1781	47	3822	ąe	31	48	ą#e	34	52	65
ye	519	48	y#e	554	52	1073	ae	2663	49	a#e	2762	51	5425
aa	1902	48	a#a	2031	52	3933	oy	644	50	o#y	657	50	1301
ey	616	48	e#y	670	52	1286	oe	2198	52	o#e	1999	48	4197
ee	1489	47	e#e	1658	53	3147	ya	813	53	y#a	722	47	1535
ąą	10	45	ą#ą	12	55	22	ay	874	53	a#y	771	47	1645
uo	305	44	u#o	389	56	694	ąy	874	53	ą#y	771	47	1645
ęi	80	43	ę#i	106	57	186	ęą	29	54	ę#ą	25	46	54
ąa	160	40	ą#a	245	60	405	oę	441	55	o#ę	357	45	798
eę	196	39	e#ę	303	61	499	uę	161	59	u#ę	113	41	274
ao	1306	33	a#o	2636	67	3942	ue	1045	69	u#e	462	31	1507
oo	1063	33	o#o	2188	67	3251	yą	82	69	y#ą	36	31	118
ęy	50	32	ę#y	106	68	156	ai	1207	70	a#i	513	30	1720
yę	77	32	y#ę	166	68	243	uy	278	71	u#y	111	29	389
ęę	20	22	ę#ę	70	78	90	oi	1374	77	o#i	410	23	1784
ąo	28	13	ą#o	195	87	223	ui	724	89	u#i	90	11	814
Total	**17520**			**17659**		**35179**	**Total**	**18933**			**20645**		**39578**