Rhythmic contrast between Swedish and Albanian as an explanation for L2-speech?

Tronnier, Mechtild; Zetterholm, Elisabeth

Rhythmic contrast between Swedish and Albanian as an explanation for L2-speech?

Mechtild Tronnier[1], Elisabeth Zetterholm[2]
mechtild.tronnier@ling.lu.se, elisabeth.zetterholm@isd.su.se

[1]Centre for Languages and Literature, Lund University
[2]Department of Language Education, Stockholm University

Abstract. Based on observations of the rhythmic structure of L2-speech produced by L1-speakers of Albanian – which suggest the occurrence of transfer – a study is presented here that compares durational aspects between the two languages. In order to do this, speech read by Swedish and Albanian L1-speakers was recorded and investigated, and normalized durational factors were analysed. The results, however, do not support the assumption that there is variation in the rhythmic structure between the two languages. According to the results, transfer cannot explain previous observations.

Keywords: language rhythm, prominence, transfer, Albanian, Swedish

Introduction

The acquisition of rhythm and the contrast in prominence in a second language is challenging for learners. Not only does placement of stress on the appropriate syllable in the word has to be learned, but also features which are used to express such a contrast, and the extent to which they are used, have to be acquired, as well. Such features are determined by variation in sound intensity, segment and/or syllable length, the presence of tonal accents and the degree of articulatory precision. Furthermore, the level of prominence of a sequence of syllables in an utterance is not always binary, but may be primary, secondary or tertiary.

When studying the accented L2-speech of Swedish produced by L1-speakers of Albanian, it was not always clear which syllable in a word carried the highest level of stress (Tronnier & Zetterholm, 2013). Therefore, one of the foreign accent features, is not simply due to incorrect stress placement, but requires further explanations. Observations from auditory analysis were that the reduction of the vowel – which is usually required in unstressed syllables in Swedish – was not carried out sufficiently.
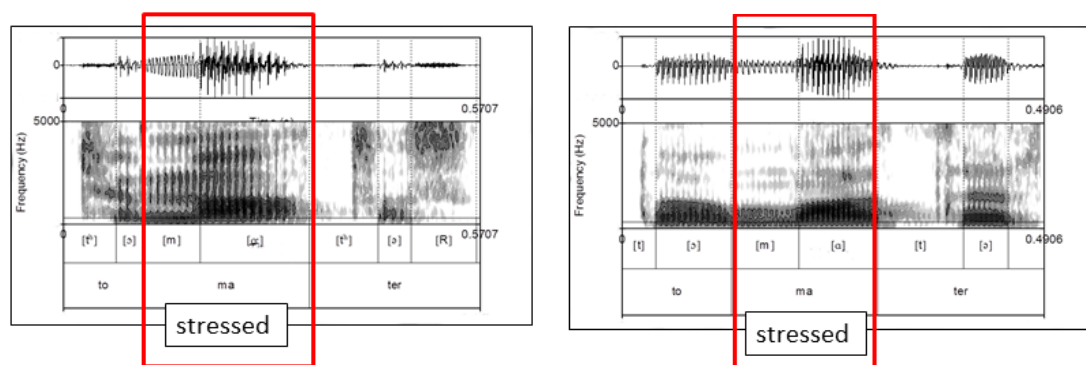
Figures 1 and 2. Examples of the word *tomater* [tʰɔˈmɑːtəʁ] "tomatoes", pronounced by a speaker with L1-Swedish (left) and an L1-speaker of Albanian producing L2-Swedish (right). The illustration shows differences in length of both the stressed syllable (in the red frame) and the unstressed syllables produced by the two speakers.

Visual inspection of the speech wave gave the impression that vowels seemed to be of a similar length and at almost equal distance from each other, whether or not these were part of an anticipated stressed

or unstressed syllable. An illustration of the difference between L1-speech and L2-speech of Swedish concerning that aspect is given in Figures 1 and 2.

To acquire insight into whether these observations are based on factors related to the rhythmic character of Albanian – i.e. the L2-learners' first language – and therefore a matter of transfer or on artefacts which emerged due to the experimental set up, a comparative study of the rhythmic structure of Albanian and Swedish, produced by L1-speakers of both languages, was carried out and the results are presented below.

The initial intention was to also analyse L2-Swedish produced by the same L1-speakers of Albanian. But due to poor performances during the recording sessions when reading the Swedish version of "The Northwind and the Sun" by most of the L2-speakers (i.e. extensive speech and reading errors, pauses, interruptions, hesitations and re-takes), that plan had to be abandoned.

Background

The nature of rhythm is based on the impression that a sequence of sounds, but also of other events recurs in a regular way. Such cyclic repetitions co-occur with the fact that subunits within the repeated parts are grouped together and that there is a clear division between the regularly repeated parts (cf. Bruce, 2012). With regard to sound perception, it has been reported from psychological experiments that listeners tend to impose a rhythmic structure onto clearly monotonous sound sequences (ibid.). Rhythm is therefore a basic human phenomenon.

Spoken language is also subject to rhythmic structuring and differences in rhythm are components which give the impression that different languages do not sound alike. The acquisition of the rhythmic structure in a foreign language is thus a matter that has to be taken into consideration in L2-instruction. It has been shown, however, that when presented with a non-lexical item - the CV-syllable "sa" - which preserved the original durational patterns of vowels and consonants from natural speech in synthesised speech samples, listeners could identify individual languages and distinguish between languages perceptually (Rasmus & Mehler, 1999).

In a historic description of language rhythm, languages were categorised as either stress-timed or syllable timed (Pike, 1945; Abercrombie, 1967). In that account, it was proposed that certain units appear in equal temporal intervals. Such isochrony was assumed to apply for the interval from one stressed syllable to the next in a stress-timed language and for each syllable interval for syllable-timed languages. The concept of isochrony has also been extended for the mora in a third type of languages, i.e. mora-timed languages. An example of a stress-timed language is English, of a syllable-timed language French and of a mora-timed language Japanese. However, the concept of isochrony has been questioned, and measurements of the intervals in focus have not given substantial support for its existence (e.g., Dauer, 1983). In addition, experiments with synthesized speech, in which isochrony has been strictly maintained, have shown that listeners experienced the rhythmic structure as being clearly unnatural (Bruce, 2012).

Instead, it has been proposed, that languages are more or less stress-based (Dauer 1983). In that way, different languages are found at different places along a scale of stress-basedness, depending on how central stress is in that particular language. Within that framework, rather than isochrony, three factors a) higher ratio of average syllable length between stressed and unstressed syllables, b) phonotactic complexity and c) vowel reduction in the unstressed syllable are considered to be representative for a language to fulfil the criteria to be placed at the higher end of the scale of stress-basedness. At that end of the scale, those languages have been placed, which traditionally were classified as stress-timed.

There is, however, evidence that some languages, like Polish, show mixed structures, e.g., phonotactically complex syllables co-occur with unstressed syllables that lack vowel reduction. Contrast in prominence can thus be manifested in a temporal domain, i.e. variation in duration, and in a quality domain, variation in articulatory excursion (c.f. Barry & Andreeva, 2001).

Interaction between different prosodic aspects can also have an influence on rhythmic structure, and languages with lexically flexible stress are mainly found among those which had previously been

classified as stress-timed. In addition, it should not be forgotten that phonological aspects of quantity can have an influence on the rhythmic character of a language. Along these lines, the interaction of prominence at different levels should be taken into account as well. The occurrence of lexical stress (primary, secondary), phrasal prominence, prominence based on the information structure within an utterance, voluntary focus and the degree and way in which adverse elements are neutralised contribute to the nature of rhythm of each language. Such interactions are relevant factors which one should bear in mind when comparing the rhythmic structures of two or more languages.

When comparing the rhythmic structure between languages or dialects, the question about which units are those affected and are subject to a perceived difference still remains. Thus, a difference can be related to variation in duration and/or articulatory excursion concerning syllables as a whole, codas, vowels only or consonants, and consonant clusters.

Rhythmic typology: Swedish

Swedish has been assumed to be a syllable-timed language, similar to other Germanic languages (c.f. Engstrand, 2004; Nishihara & van de Wejer, 2012). Experimental studies, however, have shown that isochrony cannot be confirmed either for interstress intervals or for syllables in general (Eriksson, 1991; Strangert, 1985). Other criteria proposed to better typify the rhythmic character of languages with an alternative classification, i.e. the notion of syllable-based languages (Dauer, 1983) are, however, fulfilled to a large extent. Along these lines, Swedish comprises stressed syllables which, on average, have longer durations in the flow of speech than unstressed syllables. Furthermore, in unstressed syllables, vowel quality is more neutralised than in stressed syllables. Reduction phenomena also occur in coda consonants in unstressed syllables. And finally: Swedish has a complex phonotactic structure, where the consonant clusters in mono-morphemic codas can contain three consonants (e.g., *hemsk* "terrible") and poly-morphemic codas can contain even more consonants (e.g., *skälmskt* "slyly"). It should, however, be pointed out, that in the latter case, consonant articulation is somewhat reduced in the flow of speech and not all consonants are pronounced completely.

On a lexical level, Swedish prosody incorporates aspects of quantity. The stressed syllable is always heavy and the unstressed syllable is light (Bruce, 2012). In addition, there is a structural diversity in the stressed syllable, which is also distinctive. This diversity manifests itself on the one hand by the occurrence of a long vowel followed by no consonant (*bo* [buː] "live, reside (infinitive form)") or a short and single consonant (*bok* [buːk] "book"), and on the other hand by the occurrence of a short vowel followed by a long consonant (*fall* [falː] "fall") or a consonant cluster (*falsk* [falsk] "false"). The contrast in syllable weight between stressed and unstressed syllables already provides a basis for Swedish to belong to those types of languages with a rhythmic structure which were classified as being strongly stress-based.

It should also be pointed out that Swedish has primary plus secondary stress on different elements in a compound. Syllables carrying secondary stress are subject to quantity rules just like syllables carrying primary stress, and they are therefore heavy. The difference in stress realisation lies in the absence of a tonal accent on the secondary stress, which is a salient feature of the primary stress.

In running speech, the lexical stress in function words becomes neutralised, which contributes to the rhythm of Swedish speech. Furthermore, phrasal accents, information structure and focus lead to the stronger prominence of certain syllables, which would otherwise not always be considered as being stressed.

Rhythmic typology: Albanian

Literature on the prosodic system of Albanian is not extensive. It is known that there is at least one stressed syllable in a word. However, it is not clear if this is distinctive and flexible, as in Swedish, or fixed. According to Garlén (1988), the example of the minimal pair 'bari "the grass" and ba'ri "(shep-)herds" indicates the occurrence of distinctive stress, whereas according to Lloshi (1999), stress is mainly fixed on the final syllable, which leads to a trochaic rhythm.

A distinction in vowel quantity is found in some Albanian dialects, like Gheg, which is spoken in Kosovo and northern Albania. This is pointed out by Granser and Moosmüller (2003) in their

investigation of vowel quality variation in stressed syllables. To what extent quantity distinction is restricted to stressed syllables is unclear, but it will be assumed that this is the case.

According to a survey on potential pronunciation problems for learners of English with Albanian as their L1 (Alimemaj, 2014), Albanian has length and weight on the last two syllables which co-occur with the stressed syllables. In addition, it is reported that Albanian is different from English in that every syllable is almost equal in length (ibid.). This remark is valuable for the current study, as this feature was observed to occur in L2-Swedish produced by L1-speakers of Albanian. However, no references to experimental studies are given in that survey (ibid.), which confirms the apparently impressionistic assessment.

Matter of contention

The focus of this contribution is a search for an explanation onto why L1 Albanian L2-learners of Swedish distribute prominence unlike L1-speakers of Swedish, rather than a search for a typological account for any of these two languages to be more or less stress-based. Details of rhythm-typology and stress-typology are presented above to demonstrate the complexity of the involved factors when approaching the current question.

The present study

Methodological approach

As a first approach, measurements of comparable units between the different languages were carried out in this study. As stress is very much connected to the vowel in a syllable and because it is easier to single out vowel onsets than syllable boundaries in connected speech, the procedure chosen was to measure the length of units from vowel onset to next vowel onset, thus obtaining the length of "quasi-syllables". For reasons of quantity features in the individual languages, this procedure was also preferred to measurements of e.g., vowel and/or consonant length. An example of segmentation can be seen in Figure 3.
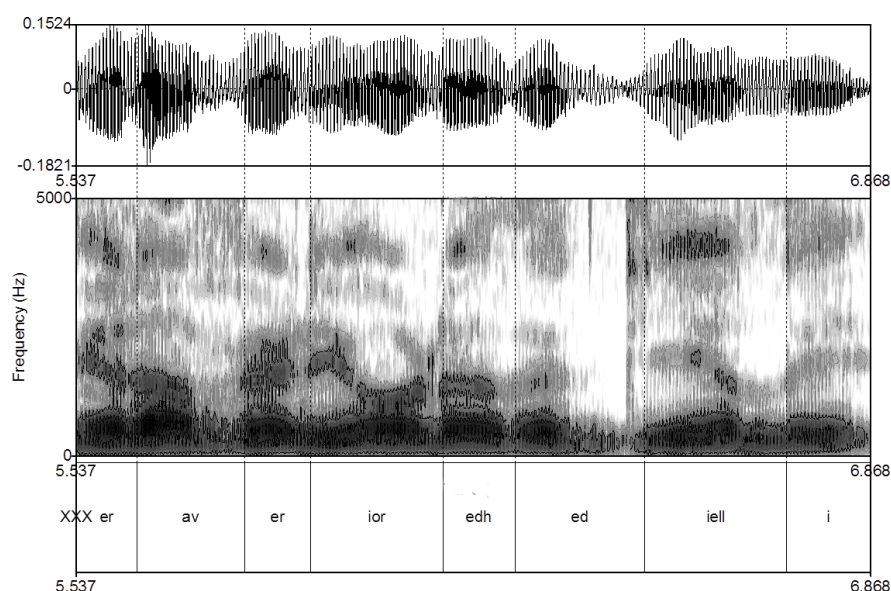

Figure 3. An example of the segmentation

Speakers, speech material and recordings

The speech material used for the present analysis was produced by seven L1-speakers of Albanian and seven L1-speakers of Swedish. All the Albanian speakers currently live in the South of Sweden and have been living there for a different period of time, between three months and 20 years. They all originate from Kosovo and they speak the Gheg variant of Albanian. The age of the Albanian speakers ranges from 25 to 54 years. The L1-speakers of Swedish also live in the South of Sweden, where they have been brought up. Therefore they also speak a variant of Southern Swedish, more precisely, they speak a variant of the Scanian dialect. Their age ranges from 23 and 49 years.

The material consisted of recordings of read speech of the story "The Northwind and the Sun", produced in Swedish by the Swedish L1-speakers and produced in Albanian by the L1-speakers of Albanian. The speakers were asked to read the story twice, and the second version was used for further analysis.

The recordings were made on various occasions and in varied settings and locations. The recordings of the Albanian speakers were all carried out in a quiet school classroom with a Roland Digital Audio Recorder R-05 and a directed lavalier microphone (Shure). The recordings with the Swedish speakers on the other hand, were carried out in studio-like booths, furnished with damped walls at Lund University, using the same recording equipment as in the classroom settings.

Data Analysis

The recordings were manually segmented in PRAAT by inserting marks/boundaries at vowel onsets. In that way, the duration of a segment from vowel onset to the successive vowel onset - a quasi-syllable - could be calculated.

For each speaker, the average ($\bar{x}$) and standard deviation (sd) of the duration of those quasi-syllables was calculated. The measure of the standard variation shows the average alteration of syllable length and is more important for the current analysis than the values of the average. The larger the standard deviation, the larger the contrast in syllable weight and vv. In addition, for each speaker, the standard deviation to average ratio was calculated ($\bar{x}$/sd) to normalise the data for differences in speech tempo. A low value of the ratio represents a larger variation in duration of the quasi-syllables.

As phrase-final lengthening is a known trait, its influence on durational variation was also tested. Therefore, statistic tests between the data from the two languages were carried out for both: a) the data set including the phrase final quasi-syllables, and b) the data set excluding the phrase final quasi-syllables. In that way, T-tests for independent samples were carried out for the values of the ratios between the two languages.

Results

The range of variation in syllable duration between the two languages is not significant (p >0.07 for the data including phrase final syllables and p >0.1 for the data excluding phrase final syllables). Thus, syllable length does not vary to a larger extent for the speakers of Swedish than for the speakers of Albanian. The results are depicted in Figures 4 and 5, where L1-speakers of Albanian are presented in the left cluster and L1-speakers of Swedish in the right cluster.

In addition, Figure 4 shows the distribution of the average ratios for the both groups of L1-speakers when values for phrase-final syllables are also included in the statistical analysis, thus ignoring the effect of phrase-final lengthening. Figure 5, on the other hand, shows the same distribution, but excludes the values for phrase-final syllables. For both languages, it can be confirmed that the data including the values of the phrase final syllables show no significant difference from the data excluding those values (p >0.9 for Albanian and p >0.9 for Swedish). From both figures, it can be seen that regardless of the inclusion or exclusion of the phrase-final syllable into the statistic calculation, that there is more conformity for the ratios obtained from the Swedish speakers than for the Albanian speakers. Hence, speaker variation is larger for the Albanian speakers.
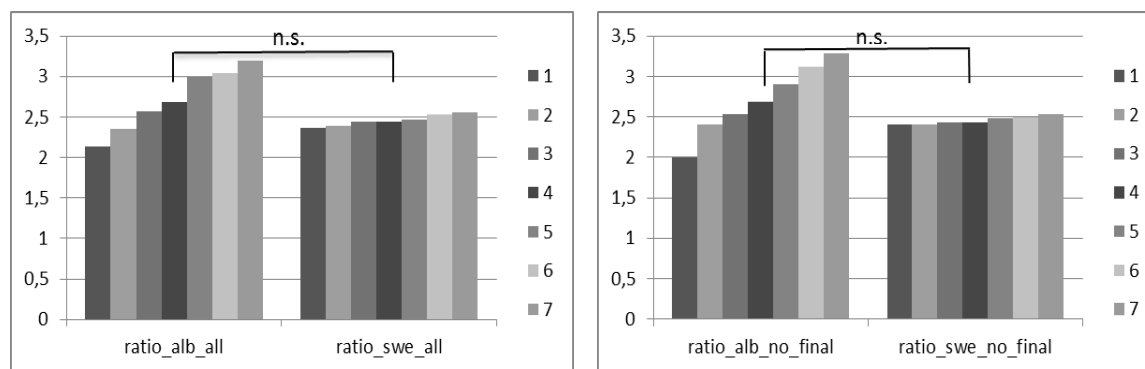
Figures 4 and 5. Average ratios for each speaker grouped by L1 (left block for Albanian and right block for Swedish in each figure). Figure 4 (left) includes phrase final quasi-syllables and Figure 5 (right) excludes phrase final quasi-syllables.

Discussion

The results of the present study do not support the assumption that neutralisation of syllable weight between stressed and unstressed in L2-Swedish produced by L1-speakers of Albanian, as found in an earlier study (Tronnier & Zetterholm, 2013), is based on transfer of rhythmic formation from L1. Based on the analysis of the range of length variation of quasi-syllables, the data and the analysis presented here do not show a significant difference in the rhythmic organisation between L1-Swedish and L1-Albanian speech. The results obtained here point to a similarity in length variation between the two languages, in that none of the languages presents us with larger variation in length of the quasi-syllables than the other. Larger variation could be interpreted to suggest that there is a clear difference in prominence between stressed and unstressed syllables. Lack of a clear length contrast as a feature in Albanian was pointed out by Alimemaj (2014) as a potential obstacle for L1-Albanian speakers learning English. Such dissimilarity does not seem to apply when comparing Albanian and Swedish, according to the analysis above.

One interesting aspect which the obtained data reveals (cf. Figures 4 and 5) is that the Swedish speakers show much more conformity in the ratio-values that represent the range of variation. The Albanian speakers show a more spread picture, where speaker 1, with a low value for the ratio, presents us with a large variation in duration, even larger than any Swedish speaker. The ratio obtained from speaker 7 in the Albanian group, however, tends to correspond to the expected outcome, based on previous observations. There is no explanation that can be given, other than external factors such as a different degree of comfort for the various speakers during the recording session.

In this study, however, only aspects of the length of rather large chunks of speech (the quasi-syllable) were analysed. This method had been chosen to overcome issues concerning quantity factors and questions of segmentation. Alternative duration measurements might represent a better way to find an explanation why a lack of rhythmic contrast in L2-Swedish produced by L1-speakers of Albanian was previously observed. In this sense, more detailed measurements of vowels and consonantal parts present in speech (*%V, ΔC, varcoC,* etc., cf. Dellwo, 2009) might give a better insight into the way in which Albanian differs from Swedish in its rhythmic structure, and if that could account for the rhythmic structure of L2-Swedish produced by L1-Albanian speakers.

Moreover, a closer investigation of qualitative factors (Barry & Andreeva, 2001) could give another insight into differences in rhythm between the two languages. Originally, however, alterations in durational factors in L2-speech were observed rather than differences in the use of e.g., articulatory reduction.

Conclusion

The results obtained in this study do not provide an explanation as to why L2-speakers of Swedish with Albanian as their L1 seem to vary the duration between stressed and unstressed syllables so little. As it was shown above, the durational variation between quasi-syllables in both languages is not as dissimilar as expected. On the basis of this study, transfer between L1-Albanian to L2-Swedish L2 as assumed in earlier observations cannot be accounted for. It must therefore be concluded that previous observations that syllable length varied less for Albanian L2-speakers of Swedish than for L1-Swedish speakers are based on behavioural grounds. For example, the production of L2-speech in a reading task might have led the L2-speakers to strongly focus on pronouncing the new text clearly and, therefore, produce fairly unnatural speech. Another explanation for the obtained results may be found perhaps in the unsatisfactory methodology used here, i.e. comparing the normalised duration of what was called "quasi-syllables". Other methods of analysis may thus be more suitable for this type of investigation, and will be considered in a follow-up study.

Acknowledgement

References

Abercrombie, D. (1967). *Elements of general phonetics.* Edinburgh, UK: Edinburgh University Press.

Alimemaj, Z. M. (2014) English phonological problems encountered by Albanian learners. *European Scientific Journal*, *10*, 1857-7431.

Barry, W. J., & Andreeva, B. (2001). Cross-language similarities and differences in spontaneous speech patterns. *Journal of the International Phonetic Association*, *31*(1), 51-66.

Bruce, G. (2012). *Allmän och svensk prosodi*. Lund: Studentlitteratur.

Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, *11,* 51-62.

Dellwo, V. (2009). Choosing the right rate normalization method for measurements of speech rhythm. In S. Schmid, M. Schwarzenbach, & D. Studer-Joho (eds), *La dimensione temporale del parlato* (pp. 13-32). Torriana: EDK Editore.

Engstrand, O. (2004). *Fonetikens Grunder.* Lund: Studentlitteratur.

Eriksson, A. (1991). *Aspects of Swedish Speech Rhythm.* Gothenburg Monographs in Linguistics 9. Gothenburg University: Department of Linguistics.

Garlén, C. (1988). *Svenskans fonologi*. Lund: Studentlitteratur.

Granser, T., & Moosmüller, S. (2001): The schwa in Albanian, In *Proceedings of the 7th International Conference on Speech Communication and Technology* (pp. 317-320). Aalborg.

Lloshi, X. (1999). Albanian. In U. Hinrichs (ed.), *Handbuch der Südost-Europa-Linguistik* (pp. 277-299). Wiesbaden: Harrassowitz Verlag.

Nishihara, T., & van de Weijer, J. M. (2012). On syllable-timed rhythm and stress-timed rhythm in world Englishes: Revisited. *Bulletin of Miyagi University of Education*, *46*, 155-163.

Pike, K. (1945). *The intonation of American English.* Ann Arbor, MI: The University of Michigan Press.

Rasmus, F., & Mehler, J. (1999). Language identification based on suprasegmental cues: A study based on resynthesis. *Journal of the Acoustical Society of America*, *105*(1)*,* 512-521.

Strangert, E. (1985). *Swedish speech rhythm in a cross-language perspective.* Almqvist & Wiksell International, Stockholm, Sweden.

Tronnier, M., & Zetterholm, E. (2013). Observed pronunciation features in Swedish L2 produced by L1-speakers of Albanian. *Studies in Language and Culture*, *21*, 85-88.