Marinis, Michail I.

What is missing to confirm a typology of rhythm? : theoretical observations and a preliminary application to two Greek varieties

Neograeca Bohemica. 2024, vol. 24, iss. [1], pp. 61-83

ISSN 1803-6414 (print); ISSN 2694-913x (online)

Stable URL (DOI): https://doi.org/10.5817/NGB2024-24-5
Stable URL (handle): https://hdl.handle.net/11222.digilib/digilib.81887

What is Missing to Confirm a Typology of Rhythm? Theoretical Observations and a Preliminary Application to Two Greek Varieties

Michail I. Marinis

Neograeca Bohemica | **24** | **2024** | **61–83**
https://doi.org/10.5817/NGB2024-24-5

Abstract

Since 1940, numerous efforts have been made to either verify or refute the hypothesis of a rhythm typology, yet no definitive conclusions have been reached. In this paper, I discuss the limitations of the reliability of data collection and processing methods, as well as the indices that dominate the attempts to measure the phenomenon, highlighting the obstacles to creating a rhythm typology. To highlight the issues under discussion, I conduct a test application of the frameworks from international literature on two varieties of the Greek language, the Amaliada variety and Cypriot Greek, based on the analysis of 192 intonational phrases, which were systematically and randomly selected from recordings of unscripted natural speech by two female speakers for each linguistic system. The analysis demonstrates such variability among speakers of the same dialect that it calls into question the validity of the rhythm measurement practices used to date. I propose the key pillars upon which rhythm research should be based in order to draw reliable conclusions and obtain cross-linguistically and inter-study comparable results, aiming to reach a definitive confirmation or refutation of a rhythm typology.

Keywords

phonological typology, rhythm classes, stress-timed languages, syllable-timed languages, Modern Greek dialects, linguistic variety

Acknowledgments

I extend my heartfelt thanks to the two anonymous reviewers for their careful reading of my manuscript and their constructive comments, which contributed significantly to improving this work. I am also grateful to Dimitris Papazachariou for our extensive and productive discussions on methodological issues, to Kateřina Bočková Loudová for her insightful observations on the flow of the text, and to Theodora Kolioni for our discussions on the equations and her assistance with the final typographical review of the manuscript. Any remaining errors are entirely my own responsibility.

1. The idea of a typology of rhythm[1]

Over the past eighty-four years, there has been an ongoing debate within the linguistic community about whether the world's languages can be perceptually distinguished into two or more broad rhythmic groups. While the theory continues to attract the interest of researchers even today (e.g., Gibbon 2023; Liu – Takeda 2021), it remains an open question, with passionate advocates as well as skeptics still engaged in the discussion.

The idea that groups of languages can be acoustically and perceptually distinguished from one another based on their rhythmic characteristics dates back to James (1940). He observed that the rhythm of certain languages, such as English, acoustically resembles the bursts of a machine gun, making it easy for listeners to differentiate from the rhythm of another group of languages, such as Spanish, which acoustically resembles Morse code. Shortly afterward, efforts began to linguistically interpret the factors that might underlie this perceptual distinction. For instance, Pike (1945) suggested that rhythmic differences may arise from certain languages, like English, tending to repeat the time intervals between stresses, while languages like Spanish tend to repeat the time intervals between syllables. Within this framework, languages that tend to repeat intervals between stresses can be referred to as stress-timed languages, while those that repeat intervals between syllables can be referred to as syllable-timed languages.

A little later, Abercrombie (1967) put forth the ambitious hypothesis that all spoken languages of the world could potentially be classified into one of the above rhythmic categories. He further introduced the concept of isochrony, which would dominate the field in the following decades. He observed that, on the one hand, stress-timed languages tend to exhibit equal time intervals either between stresses or rhythmic feet, while, on the other hand, syllable-timed languages tend to exhibit equal time intervals between syllables. As a result, within this framework, stress-timed languages tend to allow significant variation in the duration or size of syllables. In contrast, syllable-timed languages display considerable variability not in the temporal realization of syllables themselves but in the intervals between stresses.

Some researchers (e.g., Bloch 1942; Han 1962; Ladefoged 1975), attempting to account for the significant divergence of Japanese from both rhythmic groups

1 I use the term 'typology of rhythm' as it is used in Section 1 – and throughout this work – and in its sense as commonly employed in the literature, for example, by Ramus (2002) and Schmid (2012).

mentioned so far, extended the categorization by introducing a third rhythmic category: mora-timed languages. This category exhibits isochrony at a level smaller than the foot or the syllable. Since the observed isochrony occurs at the level of the mora, i.e., between the heads of syllables, it is readily understood that mora-timed languages are closer to syllable-timed languages than to stress-timed languages.

Experiments conducted primarily in English (Bolinger 1965; Shen – Peterson 1962; Uldall 1971; Lehiste 1977), as well as in a range of languages traditionally classified as stress-timed, provide clear evidence that refutes the theory of isochrony – at least at the level of the foot. These studies demonstrate that the duration of the foot is closely linked to the duration of the syllables comprising it. Extending this line of inquiry to languages classified as syllable-timed, as Arvaniti (2012) astutely notes, has not only failed to confirm the hypothesis of isochrony but has also led to its significant and, in many cases, complete discreditation.

The experimental refutation of the previously dominant idea of isochrony led to efforts to establish a typology of rhythm into a temporary deadlock. Building on Lehiste's (1977) assertion that the perception of languages as belonging to distinct rhythmic classes is not a matter of isochrony but rather a perceptual phenomenon, researchers turned their attention to syllable structure. They began to attribute rhythmic groupings of languages to the phonological properties of the permissible syllable structures in each language (e.g., Nespor 1990; Ramus et al. 1999; Ladefoged 1975). In a similar vein, Dauer (1983) proposed that all languages, to varying degrees, rely on the fundamental repetition of stressed syllables. Thus, what differentiates languages cannot be superficially sought at the level of isochrony but must be examined more deeply at the level of permissible syllable structures. Consequently, for the rhythmic categorization of a language, structural factors of the syllable must be considered, such as syllable structure, the presence or absence of vowel reduction phenomena, and word stress.

In this paper, I discuss the reasons why research conducted from the 1940s to the present has struggled to definitively establish a typology of rhythm. In Section 2, I present the factors that may perceptually influence the sense of different rhythmic classes. In Section 3, I critically review the measurement units used in international literature over the years for rhythm analysis. In Section 4, I explain the obstacles to developing a typology of rhythm. Following this discussion, I present a pilot study I conducted on the varieties of Amaliada and Cyprus, along with a discussion of its results. This study is based on a very small sample and does not aim to provide definitive or reliable conclusions about the rhythmic classification of these two varieties. Rather, its goal is to highlight, at a practical level, the problems with existing approaches to rhythm measurement and contribute to the design of a larger, more reliable study in the future. Thus, in Section 5, I detail the methodology of the present study. In Section 6, I present the study's results using commonly employed measurement approaches. In Section 7, I demonstrate how the findings from Section 6 can easily be challenged when viewed from different perspectives. The paper concludes (Section 8) with my proposals for the direction future rhythm research should take to definitively confirm or refute the existence of a typology of rhythm that could encompass the languages of the world.

2. What triggers the perception of different rhythmic groups?

Viewed through the lens described in Section 1, rhythm can be hypothesized as primarily the result of the alternation between consonantal and vocalic intervals. Within this framework, rhythm is influenced by the degree of complexity in the permissible syllable structures of a given language.

Languages whose permissible syllabic structures lack diversity perceptually create the sensation that earlier researchers described as syllable-level isochrony. Although true isochrony does not exist, the sensation arises from the property of a language having syllables whose durations do not vary significantly. This phenomenon is likely to occur in languages that do not permit complex consonant clusters. Such languages exhibit simple syllabic structures, often dominated by the least marked syllable type, consisting of an onset and a nucleus. In these languages, we do not expect significant vowel reduction. The combination of these features – (a) simple syllabic structures, (b) the absence of vowel reduction, and (c) the lack of complex consonant clusters – when present in a language, directly impacts the expected variability in syllable duration. In this paper, I propose that this is the true underlying cause of the perceptual sensation that earlier researchers referred to as isochrony.

Conversely, languages traditionally classified by researchers as stress-timed allow syllable structures that result in significant variation in syllable duration. This variation creates the perceptual impression that syllables do not share a common or similar realization time. Consequently, these languages are expected to permit: (a) complex syllable structures, (b) extensive vowel reduction, and (c) complex consonant clusters. This distinction underscores the direct relationship between phonotactic properties and the rhythmic profiles perceived across different languages.

3. But, how can we measure perception?

If the hypotheses outlined in Section 2 are correct, a critical question arises regarding the appropriate method for measuring the perception of different rhythmic groupings of languages. One issue in prior research, which will also be discussed in Section 4, is that different researchers in different studies have measured different parameters for rhythmic comparisons of the languages under investigation. From a typological perspective, this inconsistency complicates the task of establishing a unified typology of rhythm. In this study, the linguistic varieties under examination will be compared based on a range of metrics proposed by various researchers, each of which comes with its own advantages and limitations. We begin with the metrics first proposed by Ramus et al. (1999).

One of the most extensively studied metrics for the rhythmic classification of languages is the standard deviation of consonantal intervals (*delta consonantal ratio*, hereafter ΔC). This metric quantifies the variability in the duration of speech intervals consisting of consonants. In this study, ΔC will be calculated using formula (1), where n is the number of consonantal intervals in the speech sample, C_i represents the duration of an individual consonantal interval, and $\bar{C}$ is the mean duration of all consonantal intervals in the analyzed sample.

(1) $$\Delta C = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(C_i - \bar{C})^2}$$

Analogous to ΔC, which calculates the standard deviation of consonantal intervals, ΔV computes the standard deviation of vocalic intervals. It reflects the variability observed in the duration of vowels. In this study, ΔV will be determined using formula (2), where m is the number of vocalic intervals in the speech sample, V_i represents the duration of an individual vocalic interval, and $\bar{V}$ is the mean duration of all vocalic intervals in the analyzed sample.

(2) $$\Delta V = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(V_i - \bar{V})^2}$$

However, the standard deviation of vocalic intervals (ΔV) has been subject to significant criticism in the literature,[2] as it appears to be highly sensitive to speech rate. As an alternative, a preferable metric might be the percentage of the total duration of the analyzed sample that is occupied by vowels. For its calculation, I will use formula (3), where: $\sum_{i=1}^{m} V_i$ represents the sum of the durations of all vocalic intervals V_i in the speech sample. V_i is the duration of an individual vocalic interval, m is the number of vocalic intervals in the sample, and T represents the total duration of the speech sample.

(3) $$V\% = \left(\frac{\sum_{i=1}^{m} V_i}{T} \right) \times 100$$

Based on indicators (1)–(3), we expect that, compared to the so-called syllable--timed languages, the so-called stress-timed languages will exhibit higher values in the standard deviation of consonant durations (ΔC) and lower values in the percentage of the total sample covered by vowels (V%).

Since indicators (1)–(3) measure the overall variability of the intervals in the analyzed utterance corresponding to vowels or consonants, they do not account for sequential changes. For this reason, to measure the relative variability of the intervals in the analyzed sample corresponding to vowels or consonants – that is, how these intervals vary with respect to one another – the *Pair Variability Index* (PVI) has been proposed. The *raw Pair Variability Index* (rPVI, see Grabe – Low [2002]) is defined as the sum of the absolute differences between successive intervals, divided by the number of pairs in the speech sample (4), where m is the number of intervals (e.g., syllabic, vocalic, or consonantal) in the sample, d_k is the duration of the intervals, and d_{k+1} is the duration of the following interval. The Σ symbol represents the sum of all successive differences between intervals, from the first interval to the m – 1 interval.

2 For example, Ramus et al. (2003), aiming to test the reliability of metrics (1)–(3), decided to compare them by measuring these metrics in eight languages, each of which perceptually falls clearly into one of the traditional rhythmic categories: stress--timed (English, Dutch, Polish), syllable-timed (Spanish, Italian, French, Catalan), and mora-timed (Japanese). Although their data sets were carefully designed to ensure that speech rate would not influence the study, the ΔV metric remained unreliable. It failed to clearly classify the examined languages into rhythmic classes, as it demonstrated significant sensitivity to variations in speech rate.

(4) $$\text{rPVI} = \frac{1}{m-1}\sum_{k=1}^{m-1}|d_k - d_{k+1}|$$

To make equation (4) more useful for cross-linguistic comparisons, Grabe – Low (2002) proposed a normalized version, which adjusts the differences by taking into account the mean of two successive intervals. The normalized version of the rPVI (*normalized Pair Variability Index*, hereafter nPVI), as expressed in equation (5), is derived by dividing each absolute difference between successive intervals by their mean and then multiplying the result by 100.

(5) $$\text{nPVI} = \frac{100}{m-1}\sum_{k=1}^{m-1}\left|\frac{d_k - d_{k+1}}{\frac{d_k + d_{k+1}}{2}}\right|$$

This approach yields four metrics: the raw PVI for intervocalic intervals (*Intervocalic raw Pair Variability Index*, hereafter IrPVI), the raw PVI for vocalic intervals (*Vocalic raw Pair Variability Index*, hereafter VrPVI), the normalized PVI for intervocalic intervals (*Intervocalic normalized Pair Variability Index*, hereafter InPVI), and the normalized PVI for vowels (*Vocalic normalized Pair Variability Index*, hereafter VnPVI). Among these four metrics, derived from applying formulas (4) and (5) to measure vocalic and intervocalic intervals, not all exhibit the same capacity to distinguish between the languages under examination, and consequently, their degree of reliability differ. For instance, Ramus et al. (1999), as well as Ramus (2002, 2003), tested these metrics on established text corpora and concluded that the normalized version for vowels (VnPVI) demonstrates high reliability. Similarly, the raw version for calculating intervocalic interval variability (IrPVI) also shows strong reliability.

As with the case of the standard deviation of vocalic intervals (ΔV) discussed earlier, the failure of the VrPVI metric to distinguish between the examined languages, compared to the VnPVI metric, should be attributed to the significant influence of speech rate on the former. The normalized version of the metric, VnPVI, effectively eliminates the effect of speech rate as a factor influencing rhythm, thereby enhancing its reliability for distinguishing rhythmic patterns across languages.

On the other hand, for the calculation of consonants, the raw form of the pair variability index (IrPVI) proves to be more reliable. Its normalized version (InPVI) is problematic, likely because normalization results in the loss of critical information that reveals the presence or absence of consonant clusters. In other words, it obscures the degree of complexity in syllable structures. Additionally, speech rate has minimal impact on the articulation of consonants, further supporting the reliability of the raw metric.

A key issue in these measurements is the influence of speech rate on the perception of a language as either stress-timed or syllable-timed.[3] An attempt to address this issue is made through the use of the metric VacroΔC (6), which was proposed by Dellwo – Wagner (2003) and, in subsequent years, was adopted by several other researchers as an alternative to ΔC, whose formula was discussed in (1).

(6) $$\text{Vacro}\Delta C = \frac{\Delta C * 100}{\text{mean}(C)}$$

The metric VacroΔC constitutes a normalized version of ΔC, which facilitates the comparison of rhythmic variability across different speech samples by accounting for overall differences in the mean durations of consonantal intervals.

4. Obstacles in the attestation of a global rhythm typology

If the ultimate goal of rhythm research is to establish a typology of the world's languages, achieving this objective requires studies that produce results that are comparable across different studies. However, the fact that each study operates on distinct assumptions, measures different metrics (e.g., metrics 1–6), employs varying methodologies for collecting the audio material to be analyzed,

3 A series of experiments conducted over the past two decades have demonstrated that changes in speech rate can lead speakers of several languages to perceive rhythm differently. For example, Dellwo – Wagner (2003) asked native speakers to produce certain phrases at four different speech rates: very slow, slow, fast, and very fast. The results of their experiments showed that very slow utterances in languages traditionally considered stress-timed sometimes align with very fast utterances in languages traditionally regarded as syllable-timed. While this idea is intriguing, we find it methodologically problematic that the number of informants was not consistent, nor were their sociolinguistic profiles comparable across the studied language systems. At the same time, we recognize that speech rate is an extremely significant factor in distinguishing rhythmic classes, and it should neither be overlooked nor excluded from the analysis of the material under study.

and adopts different criteria for selecting the analyzed segments from the collected audio data makes cross-linguistic comparison impossible.

For example, when categorized using the metrics ΔC and V%, Thai appears syllable-timed. However, when categorized using the metrics VnPVI and IrPVI, Thai seems stress-timed (Grabe – Low 2002). It should be noted, however, that different metrics do not necessarily produce conflicting results (Ramus 2002), provided that these metrics are applied to carefully controlled material, from which the influence of speech rate has been meticulously removed or at least minimized to a statistically insignificant level. Therefore, to achieve comparable measurements, we need, among other things, appropriately selected material and a robust methodology.

Moreover, the overwhelming majority of previous studies on rhythm share a common denominator: the analysis of material that is carefully (pre)designed and produced in specialized recording booths. Native speakers of each language, often professional voice actors, are typically asked to read either a list of phrases or a short text, which is often translated into the other languages under study and performed in a similarly controlled and sterile manner. Rarely do these studies incorporate excerpts of spontaneous speech into their datasets. Even when they do, such speech is often elicited in a laboratory setting, albeit without direct guidance from the researcher. In these cases, participants are usually asked to engage in conversations with one another while remaining within the recording booths.

These techniques are problematic because, on the one hand, they completely disregard – and therefore make no attempt to mitigate – the observer's paradox. On the other hand, they imply the existence of a linguistically homogeneous system, entirely overlooking any form of variation. Most studies refer to "the rhythm of Greek," "the rhythm of English," etc., implicitly suggesting that a language is a uniform linguistic system spoken in exactly the same way – and therefore exhibiting the same rhythm – throughout its geographic range.

However, modern dialectology challenges this view of languages and highlights significant intralinguistic variation, both on a horizontal/geographic and a vertical/social level. Horizontally, linguistic realization differs from region to region within the same system, often leading to divergences so pronounced that speakers of one dialect may struggle to understand speakers of another dialect of the same language. Vertically, nearly all prior research has insufficiently accounted for the social use of language, which varies not just by social class or the criteria that define it (e.g., income, profession) but also within the speech of the same individual. Speakers adapt their language use to construct social identities and achieve their communicative goals in specific contexts.

5. Methodology[4]

Based on these considerations, in this study, I conduct a pilot application of the metrics used in the literature, which I summarize in Section 3. Unlike the common practice in rhythm research, the comparison is not made between different languages. Instead, I examine two varieties of Greek:[5] Cypriot Greek and the variety spoken in Amaliada. The varieties[6] were selected primarily due to their significant geographic distance and the historical particularities that influenced their development,[7] and additionally, they appear to exhibit fundamental phonetic differences that could potentially impact their rhythmic profile. For example, Cypriot Greek retains geminate consonants (see, among others, Arvaniti 1994; Arvaniti – Tserdanelis 2000), which have disappeared from the phonological system of the Amaliada variety and are now preserved in only a few Modern Greek dialects, such as Griko and Greko (see, for example, Marinis [forthcoming]).

However, I feel it is necessary to explicitly reiterate that the goal of this study is not to provide definitive or conclusive results regarding the rhythmic profiles of the two varieties. The dialects under investigation are used solely as case studies to highlight the limitations of rhythm measurement methodologies discussed in the preceding sections.

For all the reasons outlined in Section 4, I have consciously chosen to work with spontaneous, everyday conversational material recorded without any intervention or researcher presence. This approach aligns with the framework of ethnographically grounded data, incorporating the concept of the community of practice, as initially developed by Eckert – McConnell-Ginet (1992) and later expanded by Holmes – Meyerhoff (1999).

4 The measurements were conducted in 2013 as part of a research project I carried out – though I did not aim to publish it at the time – during the course of my postgraduate studies.

5 Over the past decades, considerable research on the rhythm of Greek has been conducted, primarily by Arvaniti and her colleagues. The present study references several of these works. However, a comprehensive and detailed review of the findings regarding Greek rhythm would exceed the scope of this study (see Section 1).

6 Moreover, cross-dialectal studies on Greek indicate that dialects differ not only at the phonological level but also exhibit significant divergences in other levels of linguistic analysis, such as morphology. For instance, Marinis (2020) and Marinis (2024) discuss cases where phonological changes that occurred in certain Greek dialects led to a restructuring of their inflectional paradigms.

7 For example, Amaliada is a region that became part of the Modern Greek state from its very inception in 1830 and, as a result, has been subject to the long-term influence of the Standard Modern Greek. In contrast, Cyprus has never officially been part of Greece and is today an independent state.

As outlined schematically in Table 1, this study analyzes 192 intonational phrases, systematically and randomly selected from approximately two hours of recordings for each linguistic system (Cypriot Greek and the variety of Amaliada). Intonational phrases with speaker overlap were excluded as unsuitable for the purposes of the research. Each of the two conversations involved two native speakers who were members of the same social network and shared similar sociolinguistic characteristics (age, social class, education, place of residence, etc.). For each of the four native speakers, 48 intonational phrases were analyzed: 16 short phrases consisting of only one phonological word, 16 medium-length phrases consisting of two to five phonological words, and 16 long phrases consisting of six or more phonological words.

		Examined intonational phrases		
Variety	Speaker	Short [1 phonological word]	Medium [2–5 phonological words]	Long [5+ phonological words]
Cyprus	K1	16	16	16
	K2	16	16	16
Amaliada	A1	16	16	16
	A2	16	16	16

Table 1: The intonational phrases selected, annotated, and measured for the extraction of the study's data

Phrases of varying lengths were selected, as the literature suggests that utterance length can influence the rhythmic profile of the linguistic system under study (Grabe – Low 2002; Nolan – Asu 2009; Dellwo 2010). Generally, longer utterances tend to exhibit stronger characteristics of stress-timed rhythm. Consequently, the length of the analyzed utterances can significantly impact results, particularly in studies relying on the *Pairwise Variability Index* (PVI).

The selected phrases were annotated using the phonetic analysis software PRAAT. For each speaker, and consequently for each group of phrases[8] (short, medium, long), a separate tier was used.

8 In hindsight, we realized that the approach of analyzing an equal number of intonational phrases from each length category was a top-down imposition on the data,

Each selected intonational phrase was segmented into vocalic and intervocalic intervals, with each interval type annotated on a separate tier. An intonational phrase was defined as any segment of speech occurring between two pauses, with a minimum pause duration set at 240 milliseconds.

During the segmentation of intonational phrases, I defined vocalic intervals as any part of the intonational phrase whose spectrogram and waveform exhibited the characteristics of a vowel. Specifically, these intervals showed five formant frequencies, higher intensity compared to the rest of the phrase, and periodicity in the waveform. Given the scope of the research, I did not differentiate or separately annotate vocalic intervals as monophthongs, diphthongs, or – on rare occasions – triphthongs. The boundaries of vocalic intervals were determined by the clear onset and offset of the second formant frequency (F2). Spectrogram sections containing vowels weakened to the point where their formant frequencies were not visible were not classified as vocalic intervals. Instead, they were annotated as intervocalic intervals. Finally, the semivowels [j] and [ç] were counted as vowels.

For the measurements, I used the CORRELATORE[9] software, into which I imported all annotated files in the form of annotation grids (TextGrid). Using CORRELATORE, I calculated all the metrics discussed in this study. Specifically, the calculations for the metrics ΔC, ΔV, V%, VnPVI, IrPVI, and VarcoC were based on the corresponding formulas presented and briefly explained in Section 3.

6. A first, typical presentation of results

Let us begin the presentation of the results by examining the dialects based on the metrics initially proposed by Ramus (2002). Of the three metrics (ΔC, ΔV, and V%), I exclude the standard deviation of vocalic intervals (ΔV) from my calculations, as nearly all prior studies concur that this metric is of questionable reliability (see Section 3). By examining the other two metrics, ΔC and V%, in correlation, we obtain the values displayed in Chart 1.

and thus manipulates the results. For example, we observed that short intonational phrases constituted a very small percentage of the total intonational phrases in the dataset.

9 The CORRELATORE software, currently at version 2.2, was designed, developed, and is freely distributed by Professor Paolo Mairano. It is available here: https://www.lfsag.unito.it/correlatore/index_en.html.

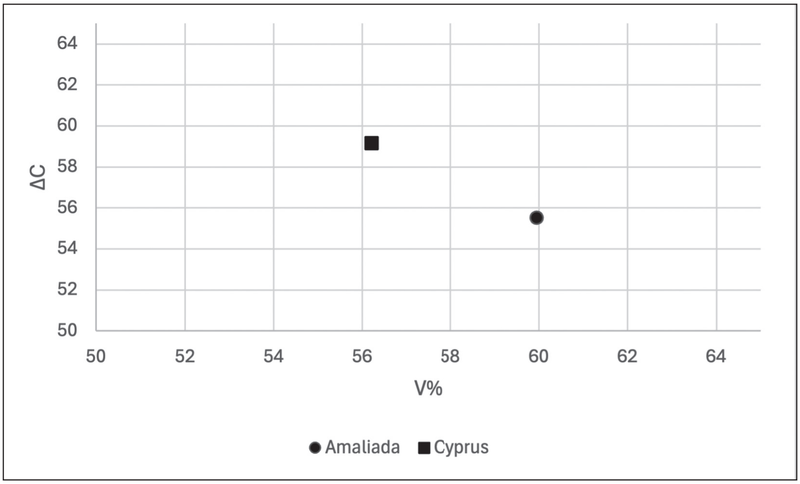

Chart 1: Results for the two examined varieties (Amaliada, Cyprus) presented through the correlation of ΔC values (vertical axis) and V% values (horizontal axis).

In Chart 1, for each of the two linguistic systems studied, I present the mean values of the standard deviation of intervocalic intervals (ΔC; vertical axis) and the percentage of intonational phrases covered by vowels (V%; horizontal axis) in correlation. In Cypriot Greek, the standard deviation of consonants (ΔC) exceeds that of the Amaliada variety. Specifically, while ΔC for Cypriot Greek is 59.14, the corresponding value for Amaliada drops to 55.51. Conversely, the V% value is higher for the Amaliada variety (59.96) than for Cypriot Greek (56.22). Based on the theoretical discussion and subsequent assumptions outlined in Sections 2 and 3, the findings depicted in Chart 1 support the classification of Cypriot Greek as closer to stress-timed languages compared to the Amaliada variety, which appears closer to syllable-timed languages. Therefore, at a preliminary level, the metrics proposed by Ramus (2002) seem to effectively categorize the two systems, providing a clear picture of their rhythmic differentiation.

Let us now attempt to reexamine the two linguistic systems based on the metrics proposed by Grabe – Low (2002), as presented in Equations (4) and (5). The results are shown in Chart 2, which correlates the normalized version of the pair variability index for vocalic intervals (VnPVI; vertical axis) with the raw form of the pair variability index for intervocalic intervals (IrPVI; horizontal axis). The first observation to note is that the distinction between the two varieties is not as clearly represented by these metrics as it is by the ΔC and V% indices shown in Chart 1.

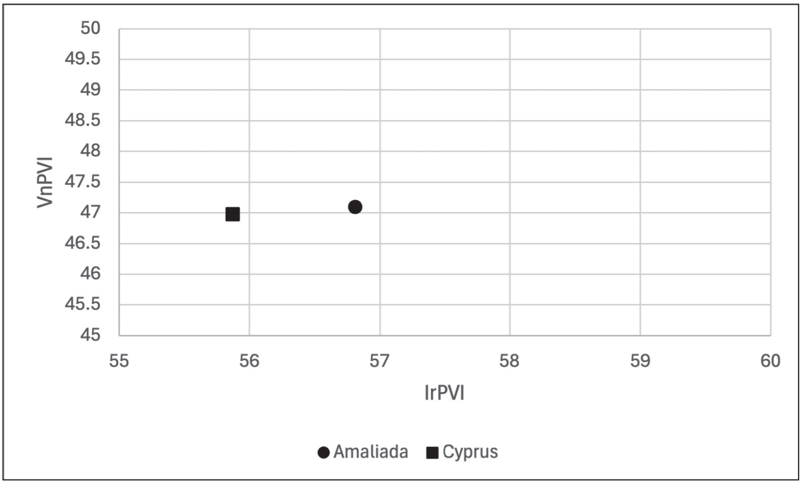

Chart 2: Results for the two examined varieties (Amaliada, Cyprus) presented through the correlation of VnPVI values (vertical axis) and IrPVI values (horizontal axis).

Regarding the VnPVI metric, Cypriot Greek shows a slightly lower value (46.97) compared to the Amaliada variety (47.09). However, the difference is extremely small and statistically insignificant. Similarly, for the IrPVI metric, Cypriot Greek scores marginally lower (55.87) than the Amaliada variety (56.81). Although a larger speech sample would be required to make a definitive claim, we must acknowledge that the results of Chart 1 contradict those of Chart 2.

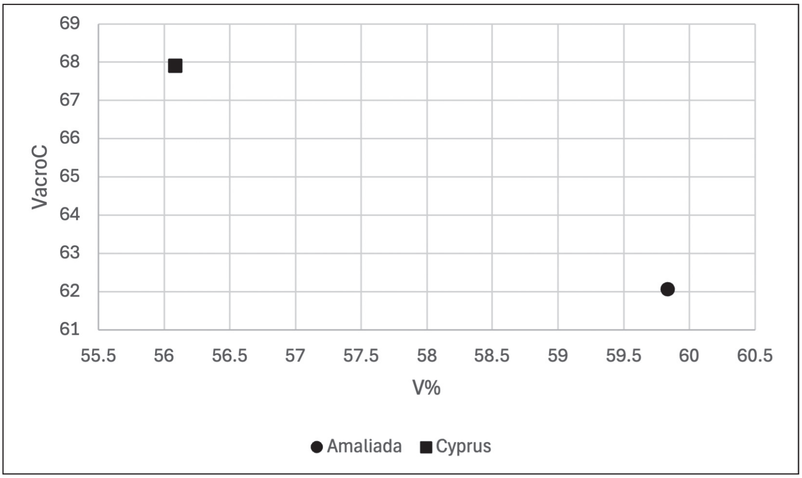

Chart 3: Results for the two examined varieties (Amaliada, Cyprus) presented through the correlation of VarcoC values (vertical axis) and V% values (horizontal axis).

Consequently, the proposals by Ramus (2002) and Grabe – Low (2002) cannot simultaneously be valid, as the results position the Cypriot and Amaliada varieties in opposite rankings along the conceptual continuum between stress-timed and syllable-timed languages.

Finally, let us attempt to examine the research data in Chart 3 by correlating the normalized version of the standard deviation of consonants (VarcoC; vertical axis), as proposed by Dellwo – Wagner (2003), with the percentage of the analyzed speech segments covered by vowels (V%; horizontal axis).

I briefly comment on the vertical axis, as we have already discussed the V% metric in Chart 1. The normalized version of the standard deviation of consonants (VarcoC) distinctly differentiates Cypriot Greek from the Amaliada variety. Cypriot Greek tends toward stress-timed languages, whereas, in comparison, the Amaliada variety leans toward syllable-timed languages. As somewhat expected, the values in Chart 3 produce a classification of the examined varieties that is quite similar to that derived from the metrics analyzed in Chart 1. Consequently, the results are in direct contrast to those produced by the metrics examined in Chart 2.

7. An alternative presentation of the data: variety, variety, variety

7.1. Variety in the methods of collecting and analyzing the data

Only very general observations can be made at this point, as any serious attempt to classify the examined varieties in terms of their rhythmic profile would require categorization based on comparisons not just between two systems but among many more – and, ideally, across all linguistic systems spoken worldwide. However, the results of this study cannot be compared with those of previous research, as different studies have employed (a) different methodologies, (b) different measurement metrics, (c) different corpora of material, and (d) even differing approaches to the definition of vocalic and intervocalic intervals.

As a result, the metrics become entirely incomparable. This is precisely why I limited the comparison to the two systems examined in this study (Cypriot Greek and the Amaliada variety) and initially attempted, using the widely employed metrics of rhythm theory, to position the studied varieties along a rhythmic continuum, as proposed in Scheme 1. At one end of the continuum, one might imagine stress-timed languages, while at the other end, syllable-timed languages could be placed.

<--->

stress-timed linguistic systems *syllable-timed linguistic systems*

Scheme 1: Conceptual continuum of linguistic systems ranging from stress-timed (left) to syllable-timed (right).

Of course, it must be emphasized that if one accepts the idea of a continuum, serious doubts arise regarding the very existence of distinct rhythmic groups. Could it be that, if we had achieved comparable results across the world's languages, the outcome would show that languages are not clustered at the extremes of stress-timed or syllable-timed languages? Might we instead find languages scattered across the entire spectrum of the continuum? To address this critical research question and provide answers to a phonological typology issue that has remained unresolved for 84 years, it is imperative to immediately create an open-access dataset, collected and annotated in a uniform manner. Such a resource would ensure that measurements across different studies are truly comparable.

7.2. Variation between speakers

The analysis so far has shown that not all metrics are equally reliable for categorizing linguistic systems and that they sometimes produce contradictory results. But what happens if we stop hiding behind means? Allow me to

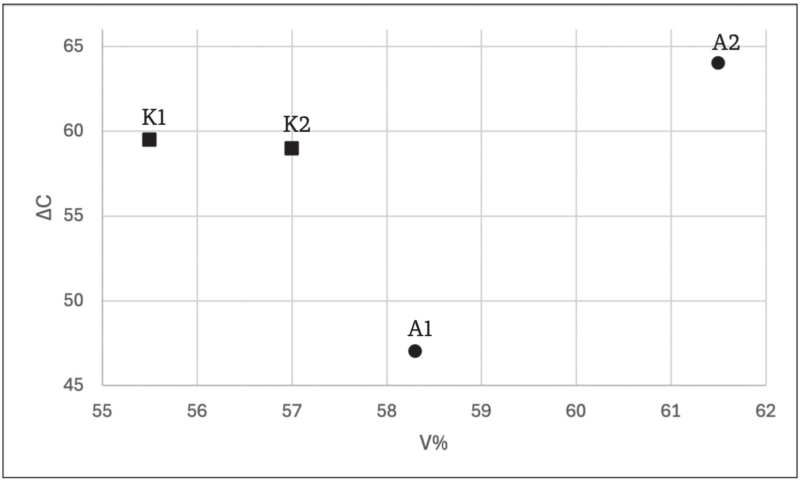

Chart 4: Results per speaker for each of the two examined varieties (circles for Amaliada speakers; squares for Cyprus speakers), plotted in terms of the correlation of ΔC values (vertical axis) and V% values (horizontal axis).

reintroduce the results of the study, not by language, but by individual speaker, following the order presented in Section 6 for each language to facilitate comparison. In Chart 4, I present the correlation between the mean values of the standard deviation of intervocalic intervals (ΔC; vertical axis) and the percentage of intonational phrases covered by vowels (V%; horizontal axis) for each of the two speakers analyzed in this study.

At first glance at the results presented in Chart 4, we are immediately confronted with the first major surprise. While we observe that on the horizontal axis (V%), the two speakers from each variety seem distinguishable in terms of their rhythmic profile, on the vertical axis (ΔC), the results are not only unclear but reveal something unexpected. Specifically, the second speaker from the Amaliada variety has a standard deviation of consonants that nearly aligns with the speakers from Cyprus – and, in fact, slightly exceeds them.

The situation becomes even more complicated in Chart 5, where the results of the VnPVI and IrPVI metrics are visually represented, not by linguistic variety, but by individual speaker.

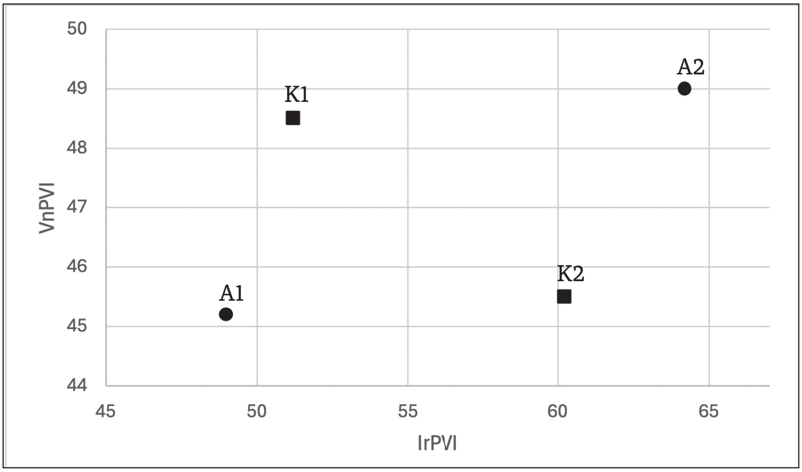

Chart 5: Results per speaker for each of the two examined varieties (circles for Amaliada speakers; squares for Cyprus speakers), plotted in terms of the correlation of VnPVI values (vertical axis) and IrPVI values (horizontal axis).

Here, absolute chaos prevails. If we compare one of the Cypriot speakers (K1) to one of the speakers from the Amaliada variety (A1) based on the VnPVI metric, we would classify Cypriot Greek as closer to stress-timed languages, while the Amaliada variety would fall closer to syllable-timed languages. However, if we compare the other two speakers (K2 for Cypriot Greek and A2 for the Amaliada variety), the results are reversed: Cypriot Greek appears closer to syllable-timed languages, while the Amaliada variety aligns more closely with stress-timed languages.

The same chaotic situation, with unclear and contradictory results, is evident on the horizontal axis (IrPVI). If we consider speakers K1 and A2 together, Cypriot Greek seems closer to syllable-timed languages than the Amaliada variety. Conversely, when we compare speakers K2 and A1 on the horizontal axis, the measurements suggest the exact opposite.

Finally, let us examine the performance of the VarcoC metric separately for each speaker, as depicted in Chart 6, where VarcoC is plotted against the V% metric.

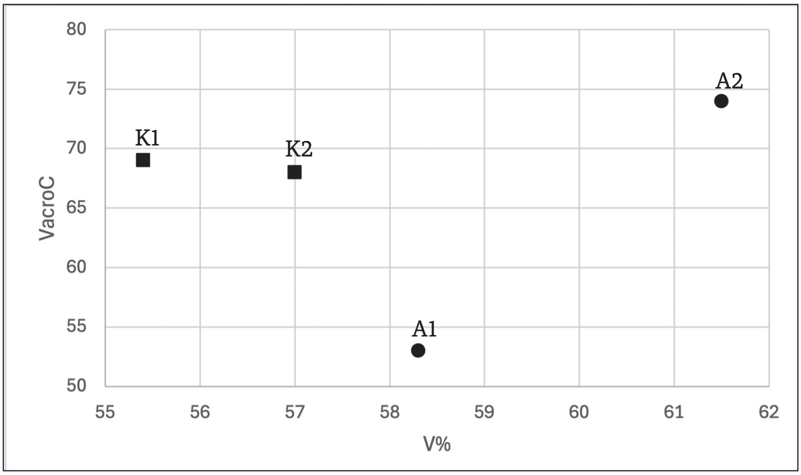

Chart 6: Results per speaker for each of the two examined varieties (circles for Amaliada speakers; squares for Cyprus speakers), plotted in terms of the correlation of VarcoC values (vertical axis) and V% values (horizontal axis).

In Chart 6, our observations regarding the classification provided by the VarcoC metric align with those we made earlier for ΔC. Specifically, the second speaker from Amaliada (A2) once again appears closer to the speakers from Cyprus (K1, K2).

The reexamination and presentation of the data undertaken in this section for each speaker individually highlights to a significant extent the inherent weakness of the core assumption underlying the vast majority of previous research on rhythm. These studies – almost universally – presume linguistic systems in which speakers' linguistic behavior is homogeneous. Consequently, they compare systems against each other and attempt to position them along the rhythmic continuum, closer to one group or the other.

This raises a critical methodological question: to what extent are we justified in making such generalizations while overlooking the factor of inter-speaker variation within the same variety? Specifically, in the context of this study's data, is it appropriate to calculate a mean value for the two speakers from the Amaliada variety?

7.3. Variation between utterances produced by the same speaker

Additionally, it is essential to investigate not only the variability between speakers but also the variability observed across different communicative contexts in the speech of the same speaker. Even if a study does not aim to explore rhythmic variation across different communicative situations, it remains imperative, in the design of such studies, to account for this critical parameter. Doing so is necessary if we are to achieve statistically significant, robust, and cross--linguistically comparable results. Furthermore, research designs should propose methods to address and mitigate this variability effectively.

8. Where do we go from here? Proposals for a global rhythm typology

Since the creation of a rhythmic typology capable of examining and potentially incorporating the languages of the world remains an open question in recent decades, it is necessary to reconsider the factors that hinder its realization and to reevaluate the prerequisites for its achievement.

To move toward the creation of cross-linguistically comparable measurements, the most critical prerequisite is the development of a comprehensive database of languages from around the world, adhering to common principles for data collection and annotation. For reasons extensively discussed in Sections 4 and 6, this database should exclude data based on preconstructed phrases produced on demand in laboratory recording conditions. Instead, the material should be derived from spontaneous, unscripted speech in natural conversational contexts between native speakers. The intonational phrases selected for annotation and analysis must be chosen through a systematically random process.

However, at this point, it is important to highlight the inherent limitation of this method. The random selection of phrases does not guarantee the statistical representation of the core phonological data present in the language. Ideally, the selection of intonational phrases should reflect the structure and phonological characteristics of the language under study. However, while this may seem ideal, it is practically extremely difficult to achieve. Even if we assume full knowledge of the structural and phonological phenomena of a linguistic system, we currently lack information on their proportional representation within the system. Furthermore, the large size of a database of comparable data does not, in itself, guarantee the success of the endeavor. These data must come from a wide range of participants, based on sociolinguistic criteria to ensure the representativeness of the sample. The composition of the informant sample must reflect the geographic distribution and societal structure of the language community being studied.

This proposal also offers a way to address the issue of the impact of speech rate on the reliability of rhythm measurements. By examining a very large dataset drawn from a diverse group of informants, carefully selected to represent the community, and collecting speech from various communicative contexts, we not only avoid isolating the factor of speech rate but instead embrace it as an integral element in our analysis.

We understand that this approach has not been followed to date for practical reasons. The workload involved in implementing this proposal is immense and, given that cross-linguistic comparison using existing data across studies is unfeasible for the reasons we have outlined in detail earlier, the method may seem utopian. How, then, can something that appears utopian become a reality?

Two decisive steps must be taken immediately to advance rhythm research:

(1) Establishing international standards for the rhythm research methodology, aimed at creating a global database that is accessible to all researchers at any time and capable of providing comparable data. Such standards would ensure consistency in the collection, annotation, and analysis of rhythm-related data across studies.

(2) Developing and universally adopting an algorithm for the segmentation of vocalic and consonantal intervals, eliminating the need for human intervention. This step is crucial as it would enable the processing of vast amounts of data derived from spontaneous speech. While this goal once seemed utopian, in the era of artificial intelligence, it has become entirely feasible.

Achieving these objectives would allow the theory of rhythmic distinction among languages to be definitively confirmed or refuted.

Bibliography

Abercrombie, D. 1967. *Elements of General Phonetics*. Chicago.

Arvaniti, A. 1994. Acoustic Features of Greek Rhythmic Structure. *Journal of Phonetics* 22/3, 239–268, doi.org/10.1016/S0095-4470(19)30203-7.

Arvaniti, A. 2012. The Usefulness of Metrics in the Quantification of Speech Rhythm. *Journal of Phonetics* 40/3, 351–373, doi.org/10.1016/j.wocn.2012.02.003.

Arvaniti, A. – Tserdanelis, G. 2000. On the Phonetics of Geminates: Evidence from Cypriot Greek. In *Proceedings of the 6th International Conference on Spoken Language*

Processing (Beijing, China, October 16–20, 2000), Vol. 2, 559–562, doi.org/10.21437/ICSLP.2000-330.
Bloch, B. 1942. Studies in Colloquial Japanese IV: Phonemics. *Language* 26/1, 86–125, doi.org/10.2307/410409.
Bolinger, D. 1965. *Forms of English: Accent, Morpheme, Order*. *Edited by I. Abe and T. Kanekiyo*. Cambridge, Mass.
Dauer, R. M. 1983. Stress-timing and Syllable-timing Reanalyzed. *Journal of Phonetics* 11/1, 51–62, doi.org/10.1016/S0095-4470(19)30776-4.
Dellwo, V. 2010. *Influences of Speech Rate on the Acoustic Correlates of Speech Rhythm: An Experimental Phonetic Study Based on Acoustic and Perceptual Evidence*. PhD Dissertation. Rheinische Friedrich-Wilhelms-Universität Bonn, hdl.handle.net/20.500.11811/4239.
Dellwo, V. – Wagner, P. 2003. Relations Between Language Rhythm and Speech Rate. In M. J. Solé – D. Recasens – J. Romero (eds.), *Proceedings of the XVth International Congress of Phonetic Sciences (Barcelona, Spain, August 3–9, 2003)*. Barcelona, 471–474 [online]. Available from: www.internationalphoneticassociation.org/icphs/icphs2003 [28 June 2024].
Eckert, P. – McConnell-Ginet, S. 1992. New Generalizations and Explanations in Language and Gender Research. *Language in Society* 28/2, 185–201, doi.org/10.1017/S0047404599002031.
Gibbon, D. 2023. The Rhythms of Rhythm. *Journal of the International Phonetic Association* 53/1, 233–265, doi.org/10.1017/S0025100321000086.
Grabe, E. – Low, E. 2002. Durational Variability in Speech and the Rhythm Class Hypothesis. In C. Gussenhoven – N. Warner (eds.), *Laboratory Phonology* 7, Berlin – New York, 515–546, doi.org/10.1515/9783110197105.2.515.
Han, M. S. 1962. The Feature of Duration in Japanese. *Onsei no Kenkyuu* (*Phonetics Research*) 10/1, 65–80.
Holmes, J. – Meyerhoff, M. 1999. The Community of Practice: Theories and Methodologies in Language and Gender Research. *Language in Society* 28/2, 173–183, doi.org/10.1017/S004740459900202X.
James, L. 1940. *Speech Signals in Telephony*. London.
Ladefoged, P. 1975. *A Course in Phonetics*. New York.
Lehiste, I. 1977. Isochrony Reconsidered. *Journal of Phonetics* 5/3, 253–263, doi.org/10.1016/S0095-4470(19)31139-8.
Liu, S. – Takeda, K. 2021. Mora-timed, Stress-timed, and Syllable-timed Rhythm Classes: Clues in English Speech Production by Bilingual Speakers. *Acta Linguistica Academica* 68/3, 350–369, doi.org/10.1556/2062.2021.00469.
Marinis, M. I. 2020. *Klisi kai organosi klitikon paradeigmaton*. PhD Dissertation. University of Patras. [Μαρίνης, Μ. Ι. 2020. *Κλίση και οργάνωση κλιτικών παραδειγμάτων*. Διδακτορική διατριβή. Πανεπιστήμιο Πατρών.] doi.org/10.12681/eadd/46842.
Marinis, M. I. 2024. Morphological Interpretations of Syncretism in the Panorama of Greek. *Folia Linguistica* 58/2, 541–562, doi.org/10.1515/flin-2024-2021.
Marinis, M. I. (forthcoming). Greek Enclaves in Southern Italy: An Introduction to Contemporary Language and its History. In B. D. Joseph (ed.), *Greek Around the World*. Publications of the Laboratory for the Study of Greek Language. Vol. 1.
Nespor, M. 1990. On the Rhythm Parameter in Phonology. In I. M. Roca (ed.), *Logical Issues in Language Acquisition*. Berlin – Boston, 157–176, doi.org/10.1515/9783110870374-009.
Nolan, F. – Asu, E. L. 2009. The Pairwise Variability Index and Coexisting Rhythms in Language. *Phonetica* 66/1–2, 64–77, doi.org/10.1159/000208931.

Pike, K. 1945. *The Intonation of American English*. Ann Arbor.
Ramus, F. 2002. Acoustic Correlates of Linguistic Rhythm: Perspectives. In *Proceedings of Speech Prosody 2002 (Aix-en-Provence, France, 11–13 April 2002)*, 115–120, doi.org/10.21437/SpeechProsody.2002-16.
Ramus, F. – Dupoux, E. – Mehler, J. 2003. The Psychological Reality of Rhythm Class: Perceptual Studies. In M. J. Solé – D. Recasens – J. Romero (eds.), *Proceedings of the XVth International Congress of Phonetic Sciences (Barcelona, Spain, August 3–9, 2003)*. Barcelona, 337–342 [online]. Available from: http://www.internationalphoneticassociation.org/icphs/icphs2003 [28 June 2024].
Ramus, F. – Nespor, M. – Mehler, J. 1999. Correlates of Linguistic Rhythm in the Speech Signal. *Cognition* 73/3, 265–292, doi.org/10.1016/S0010-0277(99)00058-X.
Schmid, S. 2012. Phonological Typology, Rhythm Types and the Phonetics-Phonology Interface. A Methodological Overview and Three Case Studies on Italo-Romance Dialects. In A. Ender – A. Leemann – B. Wälchli (eds.), *Methods in Contemporary Linguistics*. Berlin – Boston, 45–68, doi.org/10.1515/9783110275681.45.
Shen, Y. – Peterson, G. G. 1962. Isochronism in English. *Studies in Linguistics, Occasional Papers* 9. Buffalo, New York, 1–36.
Uldall, E. T. 1971. Isochronous Stresses in R.P. In L. L. Hammerich – R. Jakobson – E. Zwirner (eds.), *Form and Substance: Phonetic and Linguistic Papers Presented to Eli Fischer-Jørgensen*. København.