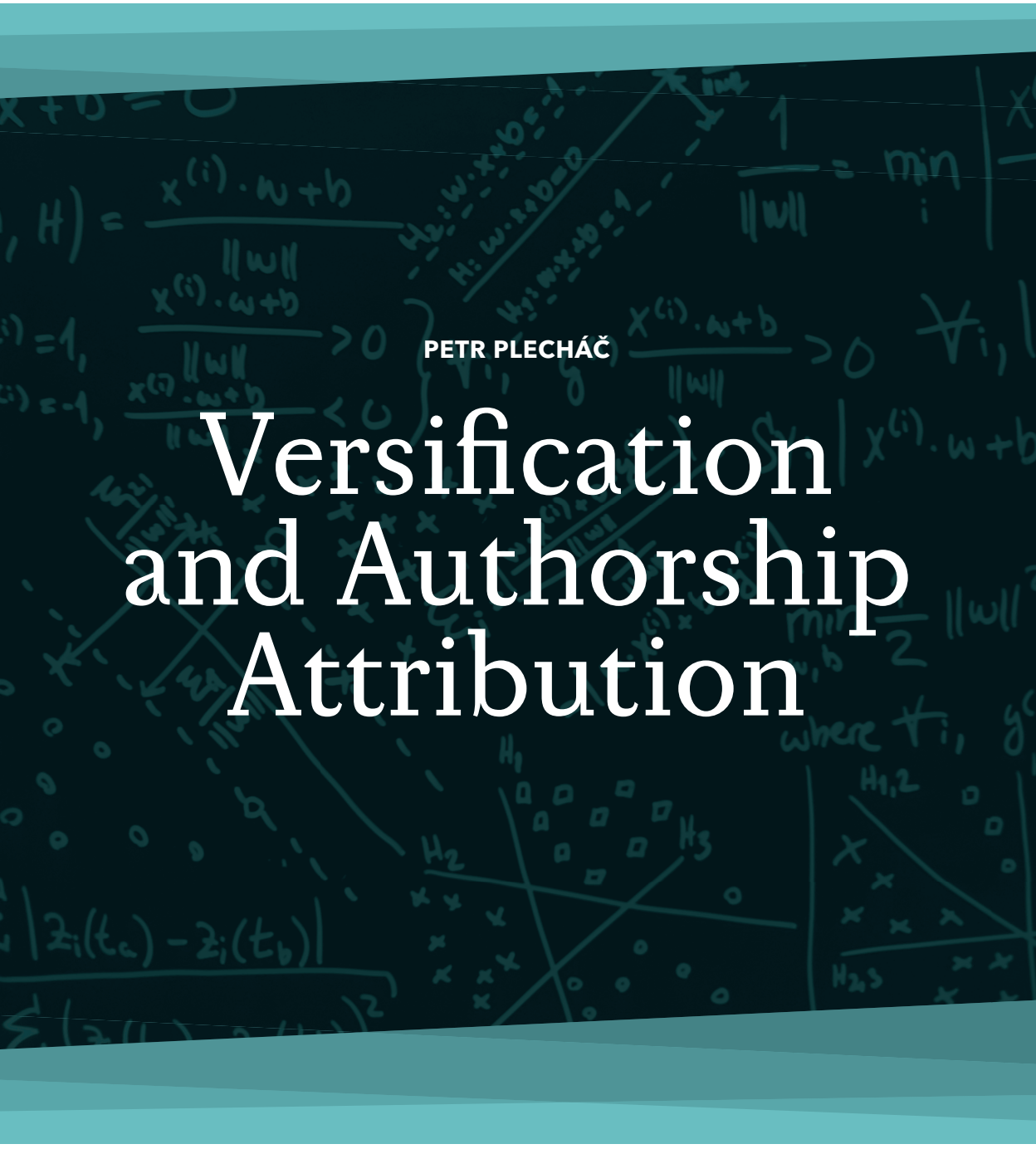
CZECH LITERATURE STUDIES



PETR PLECHÁČ

Versification and Authorship Attribution

INSTITUTE OF CZECH LITERATURE
KAROLINUM PRESS

2 Versification Features

Petr Plecháč
Institute of Czech Literature, Czech Academy of Sciences
e-mail: plechac@ucl.cas.cz

https://doi.org/10.14712/9788024648903.3

2 Versification Features

2.1 Rhythm

Since the time of Russian formalism, verse studies have distinguished between a poem's metre (i.e. the abstract pattern of each line) and its rhythm (i.e. the realisation of that metre through particular phonetic units). The relationship between strong (S) and weak (W) metrical positions and particular phonetic qualities is usually not predetermined but rather stochastic. Precisely the same metre may, thus, be achieved in very different ways in particular lines. As an example, we may consider the opening quatrain of the first canto of Karel Hynek Mácha's *Máj*, a well-known Czech narrative poem. All of the lines are written in accentual-syllabic iambic tetrameter with a strong ending (S) but the rhythmic realisation through stressed ("1") and unstressed ("0") syllables is different in each line:

	Byl pozdní večer — první máj —
rhythm:	0 1 0 1 0 1 0 1
metre:	W_1 S_2 W_3 S_4 W_5 S_6 W_7 S_8
	večerní máj — byl lásky čas.
rhythm:	1 0 0 1 0 1 0 1
metre:	W_1 S_2 W_3 S_4 W_5 S_6 W_7 S_8
	Hrdliččin zval ku lásce hlas,
rhythm:	1 0 0 1 1 0 0 1
metre:	W_1 S_2 W_3 S_4 W_5 S_6 W_7 S_8
	kde borový zaváněl háj.
rhythm:	0 1 0 0 1 0 0 1
metre:	W_1 S_2 W_3 S_4 W_5 S_6 W_7 S_8

https://doi.org/10.14712/9788024648903.3

It is widely accepted that the distribution of rhythmic patterns is not random in the works of a given author. Rather, it is an important part of their individual style. While the choice of metre is often based on general conventions with some metres reserved, for example, for a particular genre, the overall way that it is achieved (rhythmic style) may be applied as a mark of authorship.

There are two main methods of capturing rhythmic style in continental European verse studies, and both of them originate in the Russian tradition. They are (1) determining a *rhythmic profile* and (2) measuring the frequencies of *rhythmic types.*[8]

2.1.1 Rhythmic Profile

A rhythmic profile is a vector that tracks the frequency of stressed syllables in particular metrical positions. As an illustration, FIG. 2.1 presents the rhythmic profiles of all lines of iambic tetrameter with a strong ending in (1) *Máj*, (2) other works by Karel Hynek Mácha and (3)–(5) three books of poetry by a later author, Josef Václav Sládek.

FIG. 2.1 captures some important differences between the rhythmic styles of the two authors:[9]

(1) The initial W_1-position is stressed significantly more often in both Mácha samples than it is in Sládek's works.
(2) The S_2-position is stressed significantly less often in both Mácha samples than it is in Sládek's works.
(3) The line-ending S_8-position is stressed significantly more often in both Mácha samples than it is in Sládek's works.
(4) The W_3-position and W_5-position tend to be stressed slightly more often in both Mácha samples than they are in Sládek's works.

One disadvantage of the rhythmic profile method is that it completely disregards the context of particular syllables (cf. Dobritsyn 2016). FIG. 2.1 provides no information, for example, about what share of the approximately 13% of stressed syllables in the W_5-position appear in monosyllabic words:

8 These features are also known respectively as a *stress profile* and *rhythmic forms*.
9 For a thorough analysis of these differences, see, e.g. Červenka 1998; Červenka and Sgallová 1978; Jirát 1931–1932; Jakobson 1938/1995.

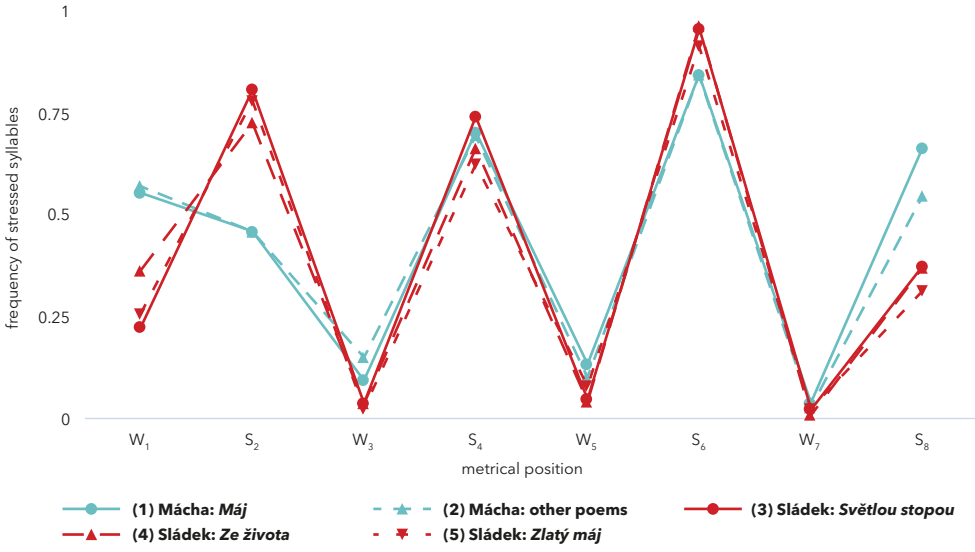


FIG. 2.1: Rhythmic profiles of all lines of iambic tetrameter with a strong ending in (1) *Máj*, (2) other works by Karel Hynek Mácha and (3)–(5) three books of poetry by a later author, Josef Václav Sládek.

„Kde Vilém můj?“ „Viz“, plavec k ní
rhythm: 0 1 0 1 1 1 0 0
metre: W_1 S_2W_3 S_4 W_5 S_6 W_7 S_8
(K. H. Mácha)

And, of course, we face the same question about the remaining share contained in polysyllabic words:

Kde borový zaváněl háj
rhythm: 0 1 0 0 1 0 0 1
metre: W_1 $S_2W_3S_4W_5S_6W_7$ S_8
(K. H. Mácha)

The most crucial problem, however, relates to so-called extrametrical syllables, i.e. cases where more than one syllable corresponds to a single metrical position. In Czech accentual-syllabic verse, these instances are rather rare:

Přistoupí strážce a lampy zář,
rhythm: 1 0 0 1 0 0 1 0 1
metre: W_0 S_1 W_1 S_2 $\llcorner W_2 \lrcorner$ S_3 W_3 S_4
(K. H. Mácha)

They are, however, very common in other syllabic accentual traditions. In English, for instance, we find:

Those trackless deeps where many a weary sail
rhythm: 0 1 0 1 0 1 0 0 1 0 1
metre: W_0 S_1 W_1 S_2 W_2 S_3 $\llcorner W_3 \lrcorner$ S_4 W_4 S_5
(P. B. Shelley)

The same holds true for metrical positions that are left blank (⌀), or what are sometimes called "headless lines":

Stay, the King hath thrown his warder down
rhythm: ⌀ 1 0 1 0 1 0 1 0 1
metre: W_0 S_1 W_1 S_2 W_2 S_3 W_3 S_4 W_4 S_5
(W. Shakespeare)

Since rhythmic profiling assumes metrical position to be a binary variable (achieved through either a stressed or unstressed syllable), it has no way to reflect these situations.

2.1.2 Rhythmic Type

The rhythmic type of a verse line describes the entire bit string that captures its rhythm. According to this approach, a poetic text can be represented based on the frequencies of its rhythmic types. TAB. 2.1 gives an example of one such representation. This is a 47-dimensional vector for the entire text of Mácha's *Máj*.

The rhythmic type method has no difficulty in resolving cases which the rhythmic profile approach cannot handle. Both extrametrical syllables (ranks 31–47, line "Přistoupí strážce…") and headless lines (ranks 31–47, line "Znovu v mdlobách…") are processed easily. Moreover, since this method does not focus on particular metrical positions but rather on entire lines, it can also be applied within systems where the

rank	rhythmic type	relative frequency	absolute frequency	example
1	10010101	0.2305	80	Večerní máj – byl lásky čas
2	01010101	0.1671	58	Byl pozdní večer – první máj
3	10010100	0.0922	32	Modré se mlhy houpají
4	01010100	0.0605	21	Já zatím hrob mu vyryji
5–6	01000100	0.0519	18	Vzdy zeleněji prosvítá
5–6	10100101	0.0519	18	Břeh je objímal kol a kol
7	01000101	0.0432	15	Tam při jezeru vížka ční
8	01001001	0.0288	10	Kde borový zaváněl háj
9–10	10011001	0.0230	8	Hrdliččin zval ku lásce hlas
9–10	10100100	0.0230	8	Dále zeleně zakvítá
			…	
31–47	100100101	0.0029	1	Přistoupí strážce a lampy zář
31–47	1010100	0.0029	1	Znovu v mdlobách umírá

TAB. 2.1: Rhythmic types of lines of iambic tetrameter with a strong ending in Karel Hynek Mácha's *Máj*.

number of positions varies (accentual verse) or where it makes no sense to distinguish them (free verse). On the other hand, the rhythmic type approach may produce rather sparse data. Some author-specific substrings may also end up being divided among a large number of less common types.

2.1.3 Rhythmic *N*-Grams

Given the limitations outlined in the previous sections, this book proposes using a method inspired by Forstall, Jacobson and Scheirer (2011) that charts a middle course. This method involves measuring the frequencies not of entire bit strings but their substrings. The latter are described here as ***rhythmic n-grams***.

From a verse line with k metrical positions, I extract all possible substrings that are of length n and start at the i-th position ($i \in \{1, 2, 3, \ldots, k - n + 1\}$). I then measure the frequencies of their rhythmic realisations. This can be illustrated by looking at the frequencies of rhythmic bigrams in Mácha's *Máj* (TAB. 2.2).

To capture the range of rhythmic variations as fully as possible, I represent the samples from my experiments with syllabic (Spanish) and accentual-syllabic (Czech) data through a combination of the frequencies of rhythmic 2-, 3- and 4-grams. In the case of the purely accentual (German) samples, I rely on the rhythmic type method for the reasons given in Section 2.1.2.

	Rhythmic realisations									
	00	**01**	**10**	**11**	**000**	**001**	**011**	**101**	**100**	**⌀1**
W_0S_1		0.4092	0.5533	0.0346						0.0029
S_1W_1	0.4611	0.0893	0.4467			0.0029				
W_1S_2	0.2017	0.7061	0.0836	0.0058			0.0029			
S_2W_2	0.2104	0.0749	0.6340	0.0490				0.0029	0.0029	
W_2S_3	0.0432	0.8012	0.1066	0.017		0.0029	0.0029			
S_3W_3	0.1239	0.0259	0.8242	0.0086				0.0029	0.0144	
W_3S_4	0.3083	0.6397	0.0345		0.0058	0.0086	0.0029			

TAB. 2.2: Rhythmic bigrams of lines of iambic tetrameter with a strong ending in Karel Hynek Mácha's *Máj*.

2.2 Rhyme

The peculiarities of rhyme are also generally recognised as author-specific. For my purposes, rhymes are represented as unordered pairs of the following features of both rhyming words:

(1) morphological features (for the Czech data, this refers to the first position of the Positional Tag[10] (=part of speech); for the German and Spanish data, this is the entire tag produced by the stochastic tagger *TreeTagger*[11]),
(2) word length measured by the number of syllables,
(3) number of syllables after the stressed syllable,
(4) final syllable coda,
(5) final syllable nucleus,
(6) onset of the final syllable + coda of the penultimate syllable (weak rhymes only) and
(7) nucleus of the penultimate syllable (weak rhymes only).

TAB. 2.3 breaks down the rhymes found in Johann Wolfgang Goethe's "Wandrers Nachtlied II" according to this schema:[12]

Über allen Gipfeln
Ist Ruh',
In allen Wipfeln

10 See Hajič 2004.
11 See <http://www.cis.uni-muenchen.de/~schmid/tools/ TreeTagger/>.
12 The International Phonetic Alphabet is used to represent sounds throughout this book.

	Gipfeln : Wipfeln	Ruh : du	Hauch : auch	Walde : balde
(1)	{NN, NN}	{NN, PPER}	{NN, ADV}	{NN, ADV}
(2)	{2, 2}	{1, 1}	{1, 1}	{2, 2}
(3)	{1, 1}	{0, 0}	{0, 0}	{1, 1}
(4)	{ln, ln}	{ø, ø}	{x, x}	{ø, ø}
(5)	{ə, ə}	{uː, uː}	{aʊ, aʊ}	{ə, ə}
(6)	{pf, pf}	–	–	{ld, ld}
(7)	{ɪ, ɪ}	–	–	{a, a}

TAB. 2.3: Rhymes contained in Johann Wolfgang Goethe's "Wandrers Nachtlied II".

Spürest du
Kaum einen Hauch;
Die Vögelein schweigen im Walde.
Warte nur, balde
Ruhest du auch.

The samples in my experiments are represented by the relative frequencies of these pairs within the relevant rhyme type (strong/weak ending).

2.3 Euphony

Finally, preferences for the accumulation of certain sounds or sound clusters within a short section of text (line, stanza) may also be understood as somehow author-specific. I describe this aspect of versification as *euphony*. Although there have been several attempts to capture phenomena of this kind through inferential statistics (e.g. Čech, Popescu and Altmann 2011), these approaches have always focused on only one subtype (e.g. the repetition of a sound within a single line). In my experiments in the next chapter, I use a very simple approximation of these phenomena: the samples are represented according to the frequencies of particular sounds.