Probing the “Temperature” Approach on Ukrainian Texts: Long-prose Fiction by Ivan Franko

Solomija Buk, Andrij Rovenchak (Lviv)

Abstract. This paper will analyze Ivan Franko’s long prose fiction, using the previously developed approach involving the “temperature” parameter defined from the frequency spectrum of texts. The scaling of this parameter with text size is studied and certain groupings of texts with respect to the scaling exponent and coefficient are observed. A new quantity is proposed to analyze relations between different sub-systems in texts.

Keywords: “temperature” of text; frequency spectrum; Ivan Franko; long prose fiction; author’s speech; direct speech.

1. Introduction

At the current state of scientific development there is an increasing interest in intercultural and interdisciplinary fields. Researchers are able to observe common laws working in exact and natural sciences as well as in social sciences and humanities. Techniques from the domain of statistical physics are applied in mathematics (Tran et al. 2004), biology (Jin et al. 2009), humanities and social sciences (Bohorquez et al. 2009; Palchykov et al. 2013; Colaiori et al. 2015; Mryglod et al. 2015). Linguistic studies are also well represented here (Eroglu 2014; Rodriguez et al. 2014), including approaches from the point of view of complex networks (Ferrer-i-Cancho et al. 2004; Čech et al. 2011; Mac Carrona, Kenna 2013).

The “temperature” concept with respect to languages was proposed in several ways (Mandelbrot 1953; Kosmidis et al. 2006), in particular, by analyzing the high-frequency words using the Boltzmann distribution (Miyazima, Yamamoto 2008; Rêgo et al. 2014). We introduced the “temperature” parameter in a different way, considering the low-frequency vocabulary (Rovenchak, Buk 2011a).

The idea of the present study is to extend the “temperature” approach to texts in Ukrainian written by Ivan Franko. The paper is organized as follows. Section 2 contains a brief description of Ivan Franko’s works of fiction. In Section 3, the approach applied for the quantitative text analysis is explained. Results and discussion are given in Section 4 followed by conclusions in Section 5.
2. Ivan Franko’s prose

Ivan Franko (1856–1916) was a famous Ukrainian writer, scholar, and public figure (Stech, Zhukovsky 2007). His long prose works are well described from the quantitative and statistical point of view (Buk, Rovenchak 2007; Buk 2010; 2011; 2012; 2013; Kelih et al. 2014). Note that there are also quantitative studies of Franko’s letters and poetry (Best, Zinenko 1998; 1999) and fables (Holovatch, Palchykov 2017). We have an ongoing project of compiling a tagged corpus of Ivan Franko’s prose (Buk 2007); this corpus could be a good base for further quantitative studies.

Ivan Franko authored over one hundred works of fiction in prose of different sizes, from short stories of some two pages to rather lengthy novels. There is no strict quantitative definition separating the types of narratives (namely, short stories, novellas, and novels) but both theorist and practitioners of literature generally agree on the upper limit of 20,000 words for a short story (cf. Thrall, Hibbard 1960: 457–458; Kotter 2008: 248–249; King 2010; Waas 2012: ix). However, this number is based on short stories written in English, and it should be somewhat lowered when dealing with a more synthetic language such as Ukrainian.

Within Franko’s œuvre, ten works are usually referred to as long prose fiction (cf. Pastukh 1996; Denysiuk 2008):

1. Boa constrictor (1st edition: 1878–84; 2nd edition: 1905–07);
2. Boryslav smijetsja (Boryslav Laughs) (1880–81);
3. Zakhar Berkut (1883);
4. Ne spytavšy brodu (Without Asking a Wade) (1885–86);
5. Dlja domašnjoho ohnyšča (For the Hearth) (1892);
6. Osnovy suspil’nosty (Pillars of Society) (1894–95);
7. Perekhresni stežky (The Cross-paths) (1900);
8. Velykyj šum (The Great Noise) (1907);
10. Lelum i Polelum [in Polish].
We will henceforth refer to the titles using the first letters of the Ukrainian transliteration, with this possibly followed by the edition number, i.e.: BC1, BC2, BSm, ZBe, NSB, DDO, OSu, PSt, VSh, and PD2. See Figure 1 for some title pages.
Interestingly enough, such a division of Franko’s prose is confirmed by quantitative data. In Figure 2 the size of each piece of prose is shown, measured in number of pages (Franko 1976–86). On average, a printed page corresponds to
340 words. So, the first edition of *Boa constrictor* (71 pages, ca. 25 thousand words) can be attributed as a novel whereas its nearest lower neighbor *Bez praci* (Without work, 66 pages, ca. 22 thousand words) can be attributed as a short story with some reservations if considering solely quantitative data. A qualitatively significant jump is observed only for the second edition of *Boa constrictor* (99 pages, about 34 thousand words).

### 3. The “temperature” approach

The approach used for the analysis in the present work was previously developed by the authors (Rovenchak, Buk 2011a; 2011b) and then applied with several modifications to obtain some hints on language evolution (Rovenchak 2014) as well as to some contrastive studies (Rovenchak 2015a; 2015b).

The idea of this approach is based on the analogy between the frequency spectrum of texts and the so-called Bose-distribution in statistical physics (Huang 1987: 183). The frequency spectrum $N_j$ is the number of tokens having absolute frequency exactly equal to $j$ (Tuldava 1996; Popescu et al. 2009). The value of $N_1$ thus corresponds to the number of *hapax legomena*. The set of $N_j$ is obtained for each text and subsequently fitted to the model of the Bose-distribution:

$$N_j = \frac{1}{z^{-1}e^{(j-1)\alpha/T-1}}. \quad (1)$$

Equation (1) contains three parameters: $z$, $\alpha$, and $T$. The first parameter is fixed by the number of hapaxes,

$$z = \frac{N_1}{N_1+1}. \quad (2)$$

The remaining parameters, $\alpha$ and $T$, are calculated by fitting the observed frequency spectrum to model (1) using the least-squares method. The relation $\tau = \ln T / \ln N_{\text{tot}}$, where $N_{\text{tot}}$ is the total number of words (tokens) in the given text, proved to be a parameter suitable for text classification alongside the exponent $\alpha$ (Rovenchak, Buk 2011a; 2011b; Rovenchak 2014).
Probing the “Temperature” Approach on Ukrainian Texts

Table 1
Fitting parameters of model (1) for Ivan Franko’s long prose fiction

<table>
<thead>
<tr>
<th>Title</th>
<th>Abbr.</th>
<th>(N_{tot})</th>
<th>(T)</th>
<th>(\alpha)</th>
<th>(\ln T / \ln N_{tot})</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Boa Constrictor</em> (1st ed.)</td>
<td>BC1</td>
<td>25427</td>
<td>1554</td>
<td>1.66</td>
<td>0.724</td>
</tr>
<tr>
<td><em>Boa Constrictor</em> (2nd ed.)</td>
<td>BC2</td>
<td>34215</td>
<td>1899</td>
<td>1.61</td>
<td>0.723</td>
</tr>
<tr>
<td><em>Boryslav smijetsja</em></td>
<td>BSm</td>
<td>77456</td>
<td>3192</td>
<td>1.54</td>
<td>0.717</td>
</tr>
<tr>
<td><em>Dlja domašnjoho ohnyšča</em></td>
<td>DDO</td>
<td>44841</td>
<td>2241</td>
<td>1.59</td>
<td>0.720</td>
</tr>
<tr>
<td><em>Ne spytavšy brodu</em></td>
<td>NSB</td>
<td>49170</td>
<td>2422</td>
<td>1.61</td>
<td>0.721</td>
</tr>
<tr>
<td><em>Osnovy suspiljnosty</em></td>
<td>OSu</td>
<td>67174</td>
<td>3038</td>
<td>1.58</td>
<td>0.721</td>
</tr>
<tr>
<td><em>Perekhresni stežky</em></td>
<td>PSt</td>
<td>93888</td>
<td>4099</td>
<td>1.56</td>
<td>0.727</td>
</tr>
<tr>
<td><em>Petriji i Dovbuščuky</em> (2nd ed.)</td>
<td>PD2</td>
<td>52751</td>
<td>2730</td>
<td>1.58</td>
<td>0.728</td>
</tr>
<tr>
<td><em>Velykyj šum</em></td>
<td>VSh</td>
<td>37005</td>
<td>1923</td>
<td>1.66</td>
<td>0.719</td>
</tr>
<tr>
<td><em>Zakhar Berkut</em></td>
<td>ZBe</td>
<td>50206</td>
<td>2521</td>
<td>1.54</td>
<td>0.724</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td></td>
<td></td>
<td><strong>1.59</strong></td>
<td></td>
<td><strong>0.722</strong></td>
</tr>
</tbody>
</table>

We further study the behavior of the “temperature” parameter \(T\) in the course of text production, i.e. how \(T\) changes for the first 1,000 of words in each text, for the first 2,000 of words, and so on. For simplicity, we fix the parameter \(\alpha\) at 1.6. It appears that the dependence of \(T\) on \(N\) is, to a good accuracy, described by a simple power law:

\[
T = tN^\beta. \quad (3)
\]

Note that such dependences occur in linguistics in various contexts (Naranan, Balasubrahmanayan 1998; Köhler 2002; Kaniadakis 2009).

Figure 3 demonstrates the results of fitting for full texts (with no division on the author’s and direct speech). Detailed results of calculations are given in the next section. The analysis was completed for texts as a whole as well as for the author’s and direct speech separately. These types of speech differ in the values of some text parameters, e.g., number of hapax legomena, dictionary and text richness, etc. (Buk 2011).

We can observe minor oscillations around the fitting curve of model (3). The nature of such behavior could itself become the subject of more detailed studies in the future (cf. Zörnig et al. 1990).
4. Results and their interpretation

The following figures demonstrate the values of the $t$ and $\beta$ parameters of Ivan Franko’s long prose fiction for full texts, author’s speech, and direct speech of all ten Ukrainian texts listed in Table 1. In Figure 4, the same scale is used for all the three panels while different scales (enlarged regions) are shown in Figure 5 for better visualization. Numerical values of the parameters and respective errors are given in Table 2.

In the direct speech data one can observe the highest concentration for PSt, PD2, and BC2; ZBe and VSh are located close together, as are BSm and OSu.

As expected, rather compact data are observed in author’s speech. For full texts, the highest concentration is found for NSB, PD2, and BC2. A separate location of PSt and OSu is also clearly seen. A somewhat dispersed group containing DDO, BC1, and VSh can be distinguished as well.

It would be interesting to find some common features of texts reflecting the abovementioned grouping. Below, we will analyze the correlation of the calculated parameters with some other numerical text data.

To extend the previously applied apparatus, one can also define the quantity...
\[ \mu = T \ln z \]  

known in physics as the chemical potential. For sub-systems in equilibrium, the chemical potentials are equal. We can estimate to what extent such a claim is applicable to our model by considering the direct and author’s speech as separate sub-systems within each text. To make the comparison, the following relative errors were calculated in each case:

\[ \delta \mu = 2 |\mu_a - \mu_d| / (\mu_a + \mu_d). \]  

The subscripts “a” and “d” correspond to the author’s and direct speech, respectively.

The results are presented in Table 2. While there is quite strong inverse correlation between \( \delta \mu \) and text size \( N \) (with the coefficient \( R = -0.61 \), see Table 3), we can still observe that the direct and author’s speech is best balanced in *Perekhresni stežky* and worst in *Petriji i Dovbuščuky* (2\textsuperscript{nd} ed.).

Curiously enough, such an estimation contradicts opinions of literary reviewers in the case of *Perekhresni stežky* (cf. Batsevych et al. 2007: 7-8).
Figure 4. Parameters of Ivan Franko’s long prose fiction for full texts, author’s speech, and direct speech
Figure 5. Parameters of Ivan Franko’s long prose fiction for full texts, author’s speech, and direct speech, enlarged regions compared to Figure 4.
Table 2
The values of parameters for Ivan Franko’s long prose. The extensions “-a” and “-d” correspond to the author’s and direct speech, respectively. The “size” column lists the total number of tokens in thousands (mathematical floor is applied).

<table>
<thead>
<tr>
<th>Text</th>
<th>$t$</th>
<th>$\Delta t$</th>
<th>$\beta$</th>
<th>$\Delta \beta$</th>
<th>$T$</th>
<th>$N_1$</th>
<th>$\mu$</th>
<th>$\delta \mu$</th>
<th>size</th>
</tr>
</thead>
<tbody>
<tr>
<td>BC1</td>
<td>0.126</td>
<td>0.011</td>
<td>0.922</td>
<td>0.009</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BC1-a</td>
<td>0.112</td>
<td>0.009</td>
<td>0.935</td>
<td>0.008</td>
<td>1270</td>
<td>5254</td>
<td>–0.242</td>
<td></td>
<td>13.4%</td>
</tr>
<tr>
<td>BC1-d</td>
<td>0.178</td>
<td>0.053</td>
<td>0.901</td>
<td>0.039</td>
<td>245</td>
<td>886</td>
<td>–0.276</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BC2</td>
<td>0.205</td>
<td>0.017</td>
<td>0.870</td>
<td>0.008</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BC2-a</td>
<td>0.149</td>
<td>0.015</td>
<td>0.908</td>
<td>0.010</td>
<td>1527</td>
<td>6522</td>
<td>–0.234</td>
<td></td>
<td>8.0%</td>
</tr>
<tr>
<td>BC2-d</td>
<td>0.378</td>
<td>0.041</td>
<td>0.794</td>
<td>0.013</td>
<td>443</td>
<td>1747</td>
<td>–0.254</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BSm</td>
<td>0.351</td>
<td>0.017</td>
<td>0.809</td>
<td>0.004</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BSm-a</td>
<td>0.204</td>
<td>0.008</td>
<td>0.868</td>
<td>0.004</td>
<td>2179</td>
<td>7930</td>
<td>–0.275</td>
<td></td>
<td>5.0%</td>
</tr>
<tr>
<td>BSm-d</td>
<td>0.657</td>
<td>0.023</td>
<td>0.725</td>
<td>0.004</td>
<td>1223</td>
<td>4231</td>
<td>–0.289</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DDO</td>
<td>0.122</td>
<td>0.006</td>
<td>0.913</td>
<td>0.005</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DDO-a</td>
<td>0.090</td>
<td>0.004</td>
<td>0.952</td>
<td>0.005</td>
<td>1393</td>
<td>5785</td>
<td>–0.241</td>
<td></td>
<td>8.7%</td>
</tr>
<tr>
<td>DDO-d</td>
<td>0.292</td>
<td>0.013</td>
<td>0.815</td>
<td>0.005</td>
<td>885</td>
<td>3368</td>
<td>–0.263</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NSB</td>
<td>0.190</td>
<td>0.011</td>
<td>0.872</td>
<td>0.005</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NSB-a</td>
<td>0.085</td>
<td>0.009</td>
<td>0.958</td>
<td>0.010</td>
<td>1479</td>
<td>6293</td>
<td>–0.235</td>
<td></td>
<td>4.9%</td>
</tr>
<tr>
<td>NSB-d</td>
<td>0.869</td>
<td>0.101</td>
<td>0.704</td>
<td>0.012</td>
<td>922</td>
<td>3738</td>
<td>–0.247</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OSu</td>
<td>0.264</td>
<td>0.014</td>
<td>0.836</td>
<td>0.005</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OSu-a</td>
<td>0.157</td>
<td>0.011</td>
<td>0.898</td>
<td>0.007</td>
<td>2051</td>
<td>7499</td>
<td>–0.274</td>
<td></td>
<td>8.8%</td>
</tr>
<tr>
<td>OSu-d</td>
<td>0.691</td>
<td>0.034</td>
<td>0.721</td>
<td>0.005</td>
<td>1121</td>
<td>4475</td>
<td>–0.251</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSt</td>
<td>0.246</td>
<td>0.014</td>
<td>0.847</td>
<td>0.005</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSt-a</td>
<td>0.176</td>
<td>0.005</td>
<td>0.888</td>
<td>0.003</td>
<td>2481</td>
<td>8640</td>
<td>–0.287</td>
<td></td>
<td>0.8%</td>
</tr>
<tr>
<td>PSt-d</td>
<td>0.367</td>
<td>0.013</td>
<td>0.793</td>
<td>0.003</td>
<td>1858</td>
<td>6524</td>
<td>–0.285</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PD2</td>
<td>0.195</td>
<td>0.014</td>
<td>0.874</td>
<td>0.007</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PD2-a</td>
<td>0.228</td>
<td>0.012</td>
<td>0.870</td>
<td>0.005</td>
<td>2035</td>
<td>6991</td>
<td>–0.291</td>
<td></td>
<td>18.3%</td>
</tr>
<tr>
<td>PD2-d</td>
<td>0.408</td>
<td>0.043</td>
<td>0.776</td>
<td>0.011</td>
<td>761</td>
<td>3142</td>
<td>–0.242</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VSh</td>
<td>0.144</td>
<td>0.011</td>
<td>0.894</td>
<td>0.007</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VSh-a</td>
<td>0.087</td>
<td>0.005</td>
<td>0.959</td>
<td>0.007</td>
<td>1176</td>
<td>5784</td>
<td>–0.203</td>
<td></td>
<td>13.2%</td>
</tr>
<tr>
<td>VSh-d</td>
<td>0.535</td>
<td>0.039</td>
<td>0.748</td>
<td>0.008</td>
<td>773</td>
<td>3332</td>
<td>–0.232</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ZBe</td>
<td>0.498</td>
<td>0.033</td>
<td>0.785</td>
<td>0.006</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ZBe-a</td>
<td>0.342</td>
<td>0.020</td>
<td>0.825</td>
<td>0.006</td>
<td>1769</td>
<td>6541</td>
<td>–0.270</td>
<td></td>
<td>6.6%</td>
</tr>
<tr>
<td>ZBe-d</td>
<td>0.503</td>
<td>0.037</td>
<td>0.760</td>
<td>0.008</td>
<td>829</td>
<td>2870</td>
<td>–0.289</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
As there were no immediately apparent text properties reflecting the behavior of \( t, \beta, \text{and} \mu \), we analyzed their correlation with text size \( (N_{\text{tot}}) \), year of publication, epithetization index (EI), as well as the so-called vocabulary richness (TTR). The epithetization index is the number of adjectives divided by the number of nouns in a given text (Ruszkowski 2004; Buk 2012). The vocabulary richness is equal to the ratio of the number of different lemmas and the total number of words in a given text (type-token ratio or TTR).

Table 3
Correlation coefficients between various text parameters

<table>
<thead>
<tr>
<th>( t )</th>
<th>( \beta )</th>
<th>( N_{\text{tot}} )</th>
<th>( \mu )</th>
<th>year</th>
<th>EI</th>
<th>TTR</th>
</tr>
</thead>
<tbody>
<tr>
<td>( t )</td>
<td>—</td>
<td>—0.96</td>
<td>0.42</td>
<td>—0.41</td>
<td>—0.42</td>
<td>—0.07</td>
</tr>
<tr>
<td>( \beta )</td>
<td>—0.96</td>
<td>—</td>
<td>—0.59</td>
<td>0.49</td>
<td>0.31</td>
<td>—0.13</td>
</tr>
<tr>
<td>( N_{\text{tot}} )</td>
<td>0.42</td>
<td>—0.59</td>
<td>—</td>
<td>—0.61</td>
<td>—0.08</td>
<td>+0.63</td>
</tr>
</tbody>
</table>

The highest correlation of \( \beta \) with TTR is in agreement with previously discovered relation between the “temperature”-related parameters and the level of language analyticity (Rovenchak, Buk 2011a; 2011b; Rovenchak 2014). Another interesting result is the absence of correlation between \( t, \beta \) and the epithetization index. The latter could signify, in particular, some complications in substantiating the argument that the groupings observed in Figures 4 and 5 were literary-based.

In view of a high correlation between \( t \) and \( \beta \), it is interesting to find the fitting function as a simple power law:

\[
\beta = bt^{-a} = 0.722t^{-0.114}.
\]

We thus obtain a very simple link between \( \tau = \ln T/\ln N_{\text{tot}} \) and \( \beta \):

\[
\tau = \beta + \frac{1}{a} \ln \frac{\ln h - \ln \beta}{\ln N_{\text{tot}}},
\]

where \( b = 0.722 \ldots \) and \( a = 0.114 \ldots \). Obviously, the second item vanishes as the text size becomes large (cf. Rovenchak 2015b).

5. Prospects and discussion

In this paper we analyzed Ivan Franko’s long prose fiction using a previously developed method inspired by a physical model. We were able to reconfirm previously revealed relations of the “temperature” parameter with the level of language analyticity. This time it is made indirectly, from the correlation between the “temperature” scaling exponent \( \beta \) and the vocabulary richness (type-token ratio with lemmas considered as types).
Moreover, some new results were also obtained. The observed small oscillations of the $T(N)$ dependence around the simple power-law fitting function might signify some supra-sentence peculiarities in the frequency structure of text (cf. Altmann 2014; Chen, Altmann 2015). In future studies it would be interesting to apply the analysis based on the introduced $\mu$ parameter for different sub-systems in texts, e.g. part-of-speech distribution. We also hope to obtain similar data for other languages and authors, to verify the approach developed in this study. It seems tempting to ascribe some unmeasurable properties to the analyzed texts on the basis of the calculated parameters. Whether it is possible – at least to a certain extent – remains so far an open issue. We expect that additional studies involving experts in the domain of literary criticism will clarify this problem.

Acknowledgements

A. R. acknowledges partial support by Project FF-30F (No. 0116U001539) from the Ministry of Educationand Science of Ukraine and by Grant F64/37-2016 (No. 0116U005055) from the State Fund For Fundamental Research of Ukraine.

References

Buk, S. (2012). The epithetization index in a work of fiction (on the basis of the text corpus of Ivan Franko’s long prose fiction). In: Obrębska, A. (ed.)
Probing the “Temperature” Approach on Ukrainian Texts


Issues in Quantitative Linguistics

4

edited by

Emmerich Kelih
Róisín Knight
Ján Mačutek
Andrew Wilson

Dedicated to Reinhard Köhler on the occasion of his 65th birthday

2016

RAM-Verlag
Studies in Quantitative Linguistics

Editors

Fengxiang Fan (fanfengxiang@yahoo.com)
Emmerich Kelih (emmerich.kelih@univie.ac.at)
Reinhard Köhler (koehler@uni-trier.de)
Ján Mačutek (jmacutek@yahoo.com)
Eric S. Wheeler (wheeler@ericwheeler.ca)


**ISBN: 978-3-942303-44-6**

© Copyright 2016 by RAM-Verlag, D-58515 Lüdenscheid

RAM-Verlag
Stüttinghauser Ringstr. 44
D-58515 Lüdenscheid
Germany
[RAM-Verlag@t-online.de](mailto:RAM-Verlag@t-online.de)
[http://ram-verlag.eu](http://ram-verlag.eu)
## Contents

**Gabriel Altmann**  
The first steps  

**Gabriel Altmann**  
On Köhlerian Motifs  

**Haitao Liu, Yu Fang**  
Quantitative Aspects of Hierarchical Motifs  

**Jiří Milička**  
Key Length Motifs in Czech and Arabic Texts  

**Germán Coloma**  
A Synergetic Regression Model of Language Complexity Trade-Offs  

**Lu Wang**  
Synergetic Studies on Chinese Lexical Structure  

**Haruko Sanada**  
A Measurement of Parts of Speech in Texts Using the Noun-Based Proportion  

**Hanna Gnatchuk**  
Testing Hypotheses on English Compounds  

**Maria Rukk**  
Context Specific Distribution of Word Meanings  

**Hans Goebl, Pavel Smečka**  
The Quantitative Nature of *Working Maps* (WM) and *Taxatorial Areas* (TA).  

**Sheila Embleton, Dorin Urîtescu, Eric S. Wheeler**  
Play with the Data!  

**Michele A. Cortelazzo, Arjuna Tuzzi**  
The First End-Of-Year Address by the New President of the Italian Republic Sergio Mattarella
Miroslav Kubát, Radek Čech
Thematic Concentration and Vocabulary Richness 150 - 159

Solomija Buk, Andrij Rovenchak
Probing the “Temperature” Approach on Ukrainian Texts: Long-prose Fiction by Ivan Franko 160 - 175

Sergey Andreev
Can Pronouns Change the Dynamic Visualization of the Poetic World? 176 - 182

Fan Fengxiang
A Study on Segmental TTR, Word Length and Sentence Length 183 - 195

Xiaxing Pan, Haitao Liu
Statistical Analysis of the Diachronic Development of Terminal Rhyme in Chinese Poetry 196 - 216

Gejza Wimmer, Ján Mačutek
Lexical Text Compactness with Link Length Taken into Account 217 - 227

Andrew Wilson
Continuous Modelling of Verse Lengths in Welsh and Gaelic Metrical Psalmody 228 - 236

Kamil Stachowski
German Loanwords in Polish and Remarks on the Piotrowski-Altmann Law 237 - 259

Emmerich Kelih, Ján Mačutek
Probleme der Modellierung von Lehnbeziehungen (am Beispiel von Serbokroatismen im Slowenischen) 260 - 272

Antoni Lozano, Bernardino Casas, Chris Bentz, Ramon Ferrer-i-Cancho
Fast Calculation of Entropy with Zhang’s Estimator 273 - 285

Gabriel Altmann
Der Emeritus 286 - 287