Attempting at parametrization of moderate-length poetic texts: *Moses*, a poem by Ivan Franko

Solomija Buk[1], Andrij Rovenchak[1*]

[1] Ivan Franko National University of Lviv, Ukraine
[*] Corresponding author's email: andrij.rovenchak@lnu.edu.ua; andrij.rovenchak@gmail.com

DOI: https://doi.org/10.53482/2022_53_399

ABSTRACT

The aim of this study is to find parameters that can be used for classification of not very long texts, for example, by author, genre, etc. We go through various known parameters and analyze to what extent they are useful for the intended purposes. We also suggest some improvements that need to be checked further. We calculate the values of parameters at various points of text comprising N tokens (running words) counted from the beginning of text. As parameters with prospects of author and/or language attribution we identify, in particular, the h-point scaling coefficient, Yule's K, relative repeat rate, and the fraction of dis legomena. These parameters demonstrate quite stable behavior in N. Another set includes scaling exponents of parameters with respect to N. Certain modifications are suggested for Lambda and entropy introducing logarithmic corrections being powers of ln N. The results are applicable for texts of thousands to tens of thousand words.

Keywords: texts classification, texts parametrization, Ivan Franko, poetry, Lambda, entropy.

1 Introduction

Classification of texts with respect to author's identity, author's sex, genre, time of writing, certain language features (e. g., analyticity/syntheticity), etc. remains a topical task despite various levels of success achieved upon application of efforts so far (Tuldava 2004; Kelih et al. 2005; Koppel et al. 2009; Nini and Grant 2013; Zörnig et al. 2016; Juola et al. 2018; Shen and Tao 2021). Potential results are promising both in purely theoretical and practical prospect, for example in machine learning, sociolinguistics, literary studies, forensic applications and so on.

We focus on approaches requiring minimal processing of texts, avoiding in particular lemmatization or part-of-speech tagging since such procedures are significantly language-dependent and they are difficult to make uniform across several languages. For the same reason, we do not apply syntactic parsing, morphological or phonological analysis and similar methods (cf. Golcher 2007; Berg 2014; Wilson and Harvey 2019).

Another focus of the present study is to analyze texts of moderate length (several thousand tokens). This, in turn, poses additional tasks of finding proper indicators being stable even for such text sizes. Our study is based on the experience of several researcher groups as described below. An analysis of a Twitter corpus and a corpus of essays was made by Juola et al. (2018). A stylometric analysis of British journalistic texts by Cortina-Borja and Chappas (2006) was based on the distribution of nouns only. So called constancy measures were analyzed by Tanaka-Ishii and Aihara (2015) for words and characters in several languages. A number of parameters were suggested by Popescu and Altmann (2006, 2007, 2008a, 2008b, but not limited to those) and by Popescu et al. (2009, 2011). Multivariate statistical analysis for Serbian texts of different genres was applied by Zörnig et al. (2016).

In the present paper, we analyze *Moses*, a poem written by Ivan Franko (1856–1916), a Ukrainian writer, polymath, and public figure. Although statistical studies of Franko's poetry date back to the writer himself (Franko 1912), they are not numerous (cf. Kovalyk et al. 1990; Best and Zinenko 1999; Holovatch and Palchykov 2007). Much more attention had been focused on the long prose so far (Buk and Rovenchak 2007, 2008, 2016; Buk et al. 2019; Rovenchak and Buk 2018; Buk 2021).

There are several reasons to work with texts by Ivan Franko. First of all, he originally wrote his works in various languages, namely, Ukrainian, Polish, German, and some others. A number of texts exist as self-translations allowing thus for studying language-dependent issues (excluding author's influence). Second, the same texts are available in different orthographies, so the influence of orthographic rules can be analyzed. Further advantages come from the availability of translations of Franko's works by other authors in various languages, often not being unique, making cross-language and cross-author comparisons possible. Other than that, it is important to mention that one of the dominant leading motives of his works (by the way, they cover over 100 volumes) is the theme of the independence of the state and the right to self-determination of the Ukrainian people, which remains topical even nowadays, more than a century after Franko.

The paper is organized as follows. In Section 2, a brief description of *Moses*, a poem by Ivan Franko, is given. Section 3 contains summary of the parameters applied in the analysis. Results of calculations are presented in Section 4 and discussion in Section 5 concludes the paper.

2 About the poem

Being written in 1905, this poem by Ivan Franko is getting more and more relevant today. Based on the reinterpretation of the biblical story about Moses, the author wrote a deep philosophical work about the future of his own fellow citizens and Ukrainian nation in general. In the world context it is about eternal problems of the relationships between the leader and his/her God (values), his/her people (mission), his/her temptations (doubts), as well as about the long way of searching for the place and mission by both the leader and the people. In 1904, Ivan Franko visited Italy and was impressed by Michelangelo's

masterpiece sculpture *Moses*. It was kind of a push to create his own poem that was written next year January through July (Mochulskyj 1956: 435–437; Yakymovych 2011).

The structure of the work consists of the prologue (an address to the people) and 20 chapters (songs, reflections). It starts at that moment of the biblical story when, after forty years of wandering in the desert, the Israelites, led by Moses, almost approached the promised land in Palestine. Tired and demotivated people start complaining and questioning God's commandments and promises together with Moses's right to lead them. Gradually losing his authority because of the main oppositionists Dathan and Abiram, the prophet starts to despair of his rightness; moreover, the voice of the devil Azazel increases his hesitation. Because of this, God punished the main protagonist with the fact that he will only be able to see the promised land, but will not be able to set foot on it, so he died. But with the death of Moses, his ideas and intentions did not die: Yehoshua continues the work of the predecessor and leads the Israelites to the promised land, he raised people who turned from nomads into conquerors. In this way, one can assume that Moses archived his aim because he strengthened the spirituality of the people and changed their way of thinking. The poem is nicely encrusted with the stories of the election of the king for the trees and the blind giant Orion.

When the poem of 20 chapters was prepared for publication by the printing house, a sudden mistake appeared: several pages at the beginning were empty, so it was recommended that Ivan Franko write a foreword to his work. Next day the author brought his famous Prologue created during one night (Mochulskyj 1956: 435–437; Yakymovych 2011). Now it is recognized as the most powerful and talented part of the *Moses*. In this introductory part, the author sees his nation on the cultural and geopolitical crossroad, confused about the direction of its development. At the same time, he expresses the hope that the mentality of Ukrainians will change from slaves to free people, and the country will relive (Zabuzhko 1993). This part is written in the rhyming verse form *terza rima*: the first line rhymes with the third, and the middle line rhymes with the first and the third lines of the following stanza, creating the following scheme: aba, bcb, cdc, … . In the remaining chapters, four-line stanzas are used with the rhyming scheme abcb.

The poem has been translated into several languages. Of those, we consider Polish, English, French, German, and Portuguese versions together with the Ukrainian text in the original orthography

- **UKR1905**: Ivan Franko. *Mojsej: poema* (Lviv: Z drukarni Naukovoho Tovarystva imeny Ševčenka, 1905).

and the modern one

- **UKR1976**: Ivan Franko. Mojsej. In: *Zibrannja tvoriv u 50-ty tomakh*. Tom 5 (Kyjiv: Naukova dumka, 1966), pp. 201–264.

Two Polish and two English translations are analyzed:

- **POL1914**: Iwan Franko. *Mojżesz: poemat*. Translated by Włodzimierz Kobryn (Lwów, 1914). The translation was authorized by Ivan Franko, see also discussion by Czetyrba-Piszczako (2016).
- **POL1993**: Iwan Franko. *Mojżesz*. Translated by Jerzy Litwiniuk (Warszawa, 1993).
- **ENG1938**: Ivan Franko. *Moses*. Translated by Waldimir Semenyna (New York, 1938).
- **ENG1973**: *Moses and other poems by Ivan Franko*. Translated by Vera Rich (New York, 1973). This version is also available from the Electronic Library of Ukrainian Literature at the following link: http://sites.utoronto.ca/elul/English/Franko/Franko-Moses.pdf.

Other language versions are as follows:

- **DEU** (German): Iwan Franko. *Moses*. Translated by Iryna Husar (Lwiw, 1996).
- **FRA** (French): Ivan Franko. *Moïse: poème*. Translated by André Swirko (Bruxelles, 1969).
- **POR** (Portuguese): Iván Frankó. *Moisés*. Translated by Wira Selansky and Helena Kolody (Rio de Janeiro, 1981).

A collage of some book covers is shown in Figure 1.

To make additional comparisons across texts in Ukrainian, we have included into consideration the following three poems of various genres:

- **UKR AK:** Абу Касимові капці / Abu Kasymovi kapci [Abu Kassem's slippers] by Ivan Franko;
- **UKR LM:** Лис Микита / Lys Mykyta [Mykyta the Fox] by Ivan Franko;
- **UKR Hajd:** Гайдамаки / Hajdamaky [Haidamaks] by Taras Shevchenko.

The scanned PDFs were optically recognized using the free online service OCR Convert (2010–2022). Texts then were thoroughly checked manually for errors and tagged to mark chapters (<c> tag) and some other data to avoid inclusion of non-relevant fragments in the analysis. In particular, Moses and Voice (Azazel) involved in a dialog in chapter XIII are marked with the <speaker> tag.

The raw data are available at Zenodo at https://doi.org/10.5281/zenodo.6951929. This set includes four texts of *Moses* (UKR1905, UKR1976, POL1914, ENG1973) plus three other used for comparisons (UKR AK, UKR LM, and UKR Hajd). Other analyzed translations of *Moses* are not deposited in order to avoid potential copyright violation.

Figure 1: Book covers of the original edition of *Moses* and some translations.

3 Parameters

For each text, we compiled the rank–frequency distribution. In order to avoid any ambiguities, tokens are strictly orthographic words, i. e., alphanumeric sequences between two spaces and/or punctuation signs. Obviously, quotation marks versus apostrophes and hyphens versus dashes were disambiguated before text processing. Hyphens and apostrophes are treated as letters. Such an approach may raise objections since apostrophes and hyphens are often used in contractions of two and more words, cf. English *it's* for *it is*, French *qu'est-ce* for *que est ce*, Portuguese *puxar-lhe* for *puxar lhe*, etc. In fact, though, there is no acceptable way to treat such situations separately from, e. g., article + preposition contractions in French (*du* < *de le*) or German (*zum* < *zu dem*) leaving us at the initial orthographic word position. Proper analysis is accessible only in lemmatized texts that is another issue to consider in future. Note that the English texts also contain several cases of specific apostrophe usage to mark contracted poetic forms (*'neath* for *beneath*, *'tis* for *it is*, *'twixt* for *betwixt*, *o'er* for *over*) and the German version is rather abundant in such poetic contractions (*günst'ge*, *lebend'ges*, *zukünft'ge*, *sei's* – these are found in the Prologue alone, with over 70 occurrences in other chapters).

Inclusion of two Ukrainian versions of *Moses* (published in 1905 and 1976) is meant to analyze the influence of certain orthographic changes over this period. For instance, in the original spelling of 1905,

the verbal reflexive particle *sja* '≈ -self' was written separately from the respective verb and was much more movable within a sentence comparing to the present-day rules. Some other examples include auxiliary words (*ne vže* versus *nevže* '≈ really?', *jak by* versus *jakby* '≈ if only', etc.) with separate spelling corresponding to the older text. Formally thus, the 1905 orthography would correspond to a more analytic language comparing to the one from 1976 text.

For all the available texts, we have calculated the following list of parameters:

- Number of tokens N.
- Number of types V.
- Type-token ratio $\mathrm{TTR} = V / N$.
- Number of hapax legomena N_1.
- Number of dis legomena N_2. Note that the fraction of dis legomena N_2/N was proposed as a rather stable parameter (weakly depending on text size) by Sichel (1986). Recently, it was suggested as a measure of text comprehensibility (Rovenchak and Rovenchak 2018).
- Fraction of hapax and dis legomena N_1 / N_2.
- Repeat rate (Zörnig et al. 2016)

(1) $$\mathrm{RR} = \frac{1}{N^2}\sum_{r=1}^{V} f_r^2 .$$

Here and below, f_r are absolute frequencies corresponding to rank r.

- Relative repeat rate (Zörnig et al. 2016)

(2) $$\mathrm{RR}_{\mathrm{rel}} = \frac{1-\mathrm{RR}}{1-1/V}.$$

- Yule's characteristic K (Yule 1944; Tanaka-Ishii and Aihara 2015; Juola et al. 2018)

(3) $$K = -\frac{1}{N} + \sum_{m=1}^{m_{max}} V(m,N)\left(\frac{m}{N}\right)^2 ,$$

where $V(m,N)$ is the frequency spectrum, i. e., the number of words appearing exactly m times in the text, and the summation runs up to the maximum frequency. Traditionally, the value of K is multiplied by a constant factor (say, 10^4) for convenience, to eliminate leading zeros in the calculated values. We will skip this procedure.

- Arc length L (Kubát et al. 2014; Juola et al. 2018):

$$L = \sum_{r=1}^{V-1} \sqrt{(f_r - f_{r+1})^2 + 1}\,. \tag{4}$$

This is just a sum of Euclidean distances between consecutive points of the rank–frequency curve.

- Lambda, aimed at balancing the dependence of the arc length on text size (Popescu et al. 2011; Čech 2015; Juola et al. 2018):

$$\Lambda = L\frac{\log_{10} N}{N}. \tag{5}$$

- h-point (Popescu and Altman 2008a)

$$h = \begin{cases} r & \text{if there exists a solution of } r = f_r \\ \dfrac{f_1 r_2 - f_2 r_1}{r_2 - r_1 + f_1 - f_2} & \text{otherwise; note that } f_1 > r_1 \text{ and } f_2 < r_2 \end{cases} \tag{6}$$

- h-point scaling a (Popescu et al. 2009)

$$a = \frac{N}{h^2}. \tag{7}$$

Such a parameter is based on observations that the values of h-point scale as $\sqrt{N}$ (cf. Popescu et al. 2009, p. 19).

- Writer's view φ (Popescu and Altmann 2007) being the angle between two vectors pointing from the h-point (h; h) to the first-ranked point (1; f_1) and to the last-ranked point (V; 1) as shown in Figure 2. Uniform scales along the horizontal and vertical axes are assumed.

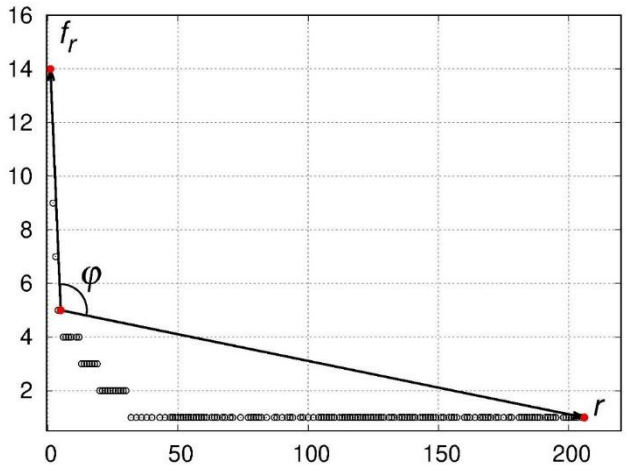

Figure 2: Illustration of the writer's view definition. The red point in the upper left corner is the first-ranked point with coordinates (1; f_1) and the one in the bottom right corner is the last-ranked point with coordinates (V; 1). The red vertex of the angle is the h-point (h; h).

The calculation is based on simple dot product definition yielding

$$\varphi = \arccos \frac{-[h(f_1 - h) + h(V - h)]}{\sqrt{[h^2 + (f_1 - h)^2][(V - h)^2 + h^2]}}. \tag{8}$$

The values of φ are measured in radians and lie within [π/2; π].

- Parameters C and s of Zipf's law

$$f_r = \frac{C}{r^s}. \tag{9}$$

In the present study, they are obtained by fitting the observed data using the above function starting at rank 10 in order to avoid the influence of the rank–frequency curve bending at low ranks.

- Parameter A

$$A = \frac{C}{(V - N_1/2)^s}, \tag{10}$$

where C and s are obtained by fitting the analyzed rank–frequency distribution using Zipf's law (9), V is the number of types and N_1 is the number of hapax legomena. This indicator was suggested by Popescu and Altman (2008b) as a measure of language analyticity/syntheticity. It was suggested that for highly synthetic languages, the Zipfian curve lies to the left side of the hapax legomena domain ($A \ll 1$) while for highly analytic languages this curve is located over hapaxes ($A \gg 1$). For a 'balanced' language, the Zipfian curve divides the domain of hapax legomena ranks into nearly equal parts ($A \simeq 1$).

- "Temperature" T and α are obtained by fitting the observed frequency $V(m, N)$ spectrum with

$$V(m, N) = \frac{1}{z^{-1} e^{(m-1)^\alpha / T} - 1}, \text{ where } z = \frac{N_1}{N_1 + 1}. \tag{11}$$

This approach was inspired by a physical model (Rovenchak and Buk 2011).

Special attention was paid to entropy S being defined as

$$S = -\sum_{r=1}^{V} p_r \ln p_r \ , \tag{12}$$

where $p_r = f_r / N$ are relative frequencies.

Various results regarding the (in)dependence of S on text length N are reported in the literature. A rough estimation can be obtained as follows. Assuming for simplicity the Zipfian model for frequencies,

$$f_r = \frac{C(N)}{r^s}, \tag{13}$$

we get

$$S = -\frac{1}{N}\sum_{r=1}^{V}\frac{C(N)}{r^s}\left[\ln\frac{C(N)}{r^s} - \ln N\right] = \ln N - \frac{C(N)\ln C(N)}{N}\sum_{r=1}^{V}\frac{1}{r^s} + \frac{sC(N)}{N}\sum_{r=1}^{V}\frac{\ln r}{r^s},$$

where the normalization

$$\sum_{r=1}^{V} f_r = \sum_{r=1}^{V}\frac{C(N)}{r^s} = N$$

is taken into account. Strictly speaking, the above summations reduce to harmonic numbers, the Riemann and the Hurwitz ζ-functions and their derivatives. But for the sake of our estimations, simple integrals might be used instead, yielding at large V in the leading order:

$$\int_1^V \frac{dr}{r^s} \sim \begin{cases} V/(1-s) & \text{for } s < 1, \\ \ln V & \text{for } s = 1, \end{cases}$$

$$\int_1^V \frac{\ln r\ dr}{r^s} \sim \begin{cases} (1-s)^{-1}\, V^{1-s}\ln V & \text{for } s < 1, \\ \dfrac{1}{2}\ln^2 V & \text{for } s = 1. \end{cases}$$

The next step is to estimate the total number of types V, which is known (Kornai 2002) to be

$$\ln N \lesssim V \sim N^{\rho},$$

yielding after simple transformations $S \sim \ln N$ (for $s < 1$) with a specific correction in the case of $s = 1$ given by $S \sim \ln N - \ln\ln N$. In practice, however, the relation $S / \ln N$ appeared too sharply declining with N, so a weaker dependence might be suggested as one of more suitable parameters to analyze, $S / \ln^{\eta} N$ with $\eta < 1$.

Moreover, it is worth mentioning that the expression ($\ln N - \ln\ln N$) represents the first two terms in the asymptotic expansion of the Lambert W function being the solution to the transcendental equation $We^W = z$. A possibility to scale the entropy with $W(z)$ should be noted as an option, additionally to logarithmic scalings discussed above.

4 Results

For certain parameters mentioned in the previous Section, their dependence on the number of tokens N was studied. First of all, we started with parameters well known to be increasing with N, namely: $V(N)$, $N_1(N)$, $N_2(N)$, $h(N)$, $T(N)$, and $L(N)$. To obtain detailed information on $\Lambda(N)$, we also analyzed the dependence $L \ln N\,(N)$. The fitting with a simple power law,

$$F(N) = BN^{\gamma}, \tag{14}$$

proved successful for all the above quantities. For each text, the parameter data were calculated at chapter boundaries. A selection of data is given in the Appendix while the complete set is available at Zenodo under https://doi.org/10.5281/zenodo.6951929. The respective scaling exponents are shown in Table 1.

The coefficient of determination is $R^2 > 0.99$ for all the seven quantities mentioned above in all the 12 analyzed texts.

For entropy, the dependences $S(N)$ yielded very small values of the fitting exponents ($\gamma < 0.1$) suggesting to analyze in more detail the relation $S(\ln N)$ as discussed in the previous Section. The fitting function for entropy is thus

$$G(N) = D \ln^{\eta} N \,. \tag{15}$$

The values of η are given in Table 1 together with the coefficient of determination for this fit in the last row.

The observed scaling of $L(N) \ln N \propto N^{\gamma}$ with $\gamma < 1$ (see Table 1) suggests that the Lambda parameter given by Eq. (5) in fact decreases as text size grows, confirming in particular observation by Čech (2015). On the other hand, the values of γ are indeed close to unity, so one can introduce a modified quantity, e.g.,

$$\Lambda^* = L \frac{\ln^{\theta} N}{N}, \tag{5*}$$

where $\theta > 1$. Proper value(s) of this exponent are yet to be defined. Such a parameter might be more stable in the range of text sizes describable as "moderate".

Table 1: The values of the scaling exponents γ (if functions depend on N) or η (in the case of entropy depending on $\ln N$) for fitting with Eqs. (14) or (15).

	UKR 1905	**UKR 1976**	**POL 1914**	**POL 1993**	**ENG 1938**	**ENG 1973**	**DEU**	**FRA**	**POR**	**UKR AK**	**UKR LM**	**UKR Hajd**
$V(N)$	0.782	0.781	0.785	0.812	0.646	0.652	0.727	0.729	0.719	0.835	0.790	0.747
$N_1(N)$	0.721	0.724	0.732	0.756	0.540	0.538	0.649	0.658	0.637	0.810	0.739	0.676
$N_2(N)$	0.996	0.976	0.939	1.015	0.733	0.758	0.861	0.841	0.821	0.895	0.867	0.832
$h(N)$	0.482	0.527	0.526	0.520	0.486	0.480	0.526	0.509	0.517	0.505	0.468	0.503
$T(N)$	1.050	1.028	0.971	1.044	0.797	0.835	0.906	0.893	0.875	0.900	0.896	0.862
$L(N)$	0.794	0.794	0.801	0.819	0.698	0.695	0.745	0.754	0.740	0.844	0.801	0.767
$L \ln N(N)$	0.913	0.914	0.921	0.939	0.810	0.812	0.864	0.868	0.857	0.965	0.912	0.881
$S(\ln N)$	0.648	0.637	0.613	0.705	0.460	0.473	0.586	0.592	0.603	0.669	0.560	0.598
R^2	0.989	0.991	0.994	0.995	0.972	0.967	0.981	0.986	0.982	0.998	0.999	0.992

With $V(N) = B_1 N^{\gamma_1}$, under condition $\frac{1}{2} < \gamma_1 < 1$ (cf. Table 1), one can estimate the large-N behavior of the writer's view φ given by Eq. (8) as follows. Let $h = \sqrt{N/a}$ and f_1 is defined as $f_1 = C(N)$ following Zipf's law (9) with $s = 1$. Then, $C(N) \sim N / \ln N$ in the leading order at large N. Keeping in Eq. (8) only the leading order quantities, we obtain:

$$\varphi \sim \arccos \frac{-hf_1}{\sqrt{f_1^2 V^2}} = \arccos \frac{-\sqrt{N/a}}{B_1 N^{\gamma_1}} = \frac{\pi}{2} - \frac{N^{1/2-\gamma_1}}{\sqrt{a}\, B_1} + \cdots .$$

The obtained asymptotic value of $\pi/2 = 1.570\ldots$ slightly differs from the golden section

$$\Phi = \frac{1+\sqrt{5}}{2} = 1.618\ldots$$

suggested by Popescu and Altmann (2008b). The difference is, however, not very significant and hardly can be caught with fitting procedures as discussed below.

We thus suggest the following function to fit $\varphi(N)$:

$$F_1(N) = X - YN^{\delta} \quad \text{with} \quad X = \frac{\pi}{2}. \tag{16}$$

The fitting parameters are given in Table 2 and respective illustrations are shown in Figure 3. Note that in two cases (POL1993 and UKR AK) we were not able to achieve a satisfactory fit as evidenced by the R^2 values.

Table 2: The values of the fitting parameters in Eq. (16) for the writer's view φ.

	UKR 1905	**UKR 1976**	**POL 1914**	**POL 1993**	**ENG 1938**	**ENG 1973**	**DEU**	**FRA**	**POR**	**UKR AK**	**UKR LM**	**UKR Hajd**
Y	4.28	2.34	4.25	1.64	9.79	9.26	2.00	1.85	4.82	2.60	11.2	26.2
δ	0.442	0.372	0.425	0.329	0.528	0.530	0.267	0.311	0.404	0.264	0.461	0.622
R^2	0.970	0.983	0.982	0.812	0.986	0.938	0.914	0.978	0.959	0.806	0.991	0.951

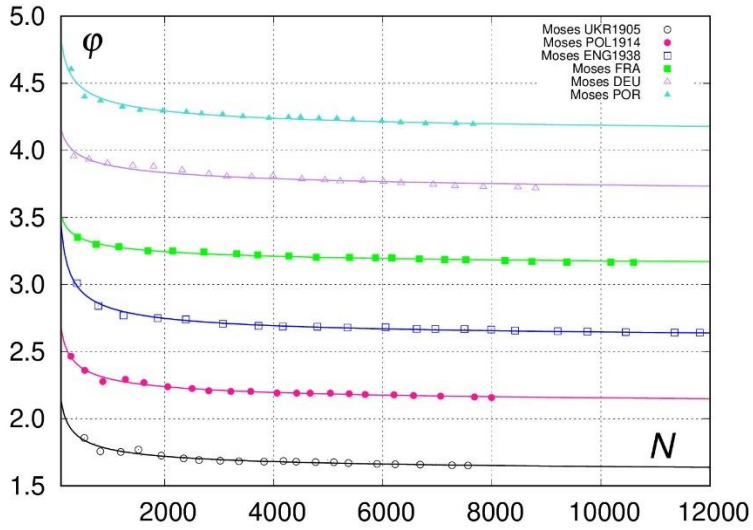

Figure 3: Fitting of the writer's view φ for several language versions of *Moses* using Eq. (16). Parameters are given in Table 2. For clarity, every subsequent set of data is shifted vertically by 0.5.

Curiously, letting the parameter X in Eq. (16) free, we were able to obtain different values for it, both around $\pi/2$ and the golden section Φ. For instance,

UKR1905: $X = 1.594 \pm 0.038$ (with $\delta = 0.523 \pm 0.158$)

UKR1976: $X = 1.564 \pm 0.030$ (with $\delta = 0.356 \pm 0.065$)

POL1914: $X = 1.582 \pm 0.026$ (with $\delta = 0.452 \pm 0.066$)

ENG1938: $X = 1.613 \pm 0.008$ (with $\delta = 0.700 \pm 0.047$)

Note that the first X value domain (for UKR1905) [1.556; 1.632] includes both $\pi/2 = 1.570\ldots$ and $\Phi = 1.618\ldots$. At present, we do not have sufficient data or theoretical justification to lean towards any fixed value for X except for $X = \pi/2$ derived above.

The most interesting are quantities exhibiting non-monotonous or insignificant variation with N. They can be discovered upon studying the correlation with text size. In Table 3, several parameters are shown having the Pearson correlation coefficient within $-0.5 < r < 0.5$ at least for one text. The dependences of these parameters on text size is demonstrated in Figures 4–10.

In Figure 4 (Dependence of the h-point scaling a on text size N) we can observe that both Polish translations are placed above $a = 10$ while all the remaining Franko's texts and translations demonstrate much more homogeneous behavior and are located in the lower part of the graph.

Table 3: The values of the Pearson correlation coefficient versus text size N.

	UKR 1905	UKR 1976	POL 1914	POL 1993	ENG 1938	ENG 1973	DEU	FRA	POR	UKR AK	UKR LM	UKR Hajd
a	−0.017	−0.552	−0.554	−0.007	+0.155	−0.052	−0.640	−0.490	−0.242	−0.438	0.667	0.148
RR_{rel}	0.102	0.027	−0.735	0.522	−0.489	0.019	0.591	0.825	0.634	−0.697	−0.642	−0.741
α	0.476	0.456	0.522	0.102	−0.876	−0.109	0.284	0.410	−0.885	0.577	−0.420	−0.242
$S/\sqrt{\ln N}$	0.659	0.672	0.693	0.781	−0.696	−0.521	0.426	0.503	0.484	0.854	0.892	0.733
N_2/N	0.255	−0.244	0.207	0.656	−0.920	−0.898	−0.709	−0.831	−0.816	−0.025	−0.903	−0.915
A	−0.898	−0.923	−0.803	−0.835	0.691	0.099	−0.377	−0.387	−0.624	−0.812	−0.591	−0.897
K	−0.670	−0.539	0.549	−0.748	−0.247	−0.696	−0.680	−0.804	−0.707	−0.777	−0.231	−0.708

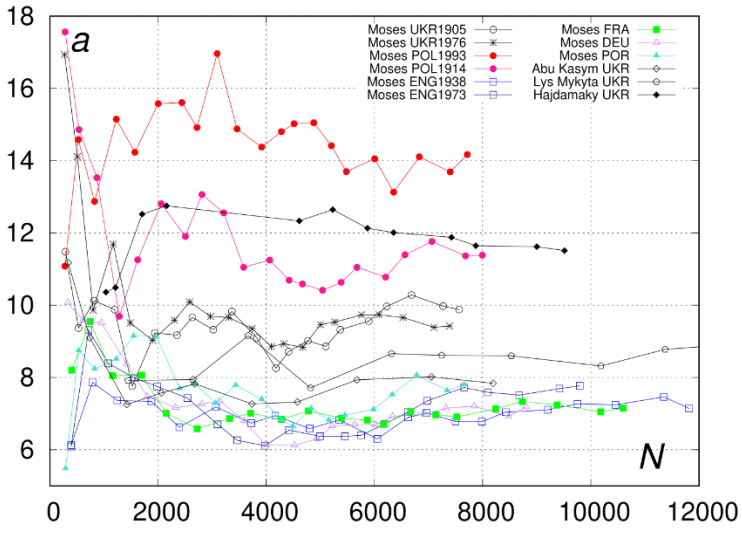

Figure 4: Dependence of the h-point scaling a on text size N.

Figure 5 (Dependence of the relative repeat rate RR_{rel} on text size N), as well as Figure 7 (Dependence of the logarithmically scaled entropy on text size N) and Figure 10 (Dependence of the Yule's K on text size N) show very similar picture: both English translations together with the French version visually create one group, while all the other texts create another group; just in Figures 5 and 7 this group is located in the upper part of the graph while in Figure 10 it goes in the bottom.

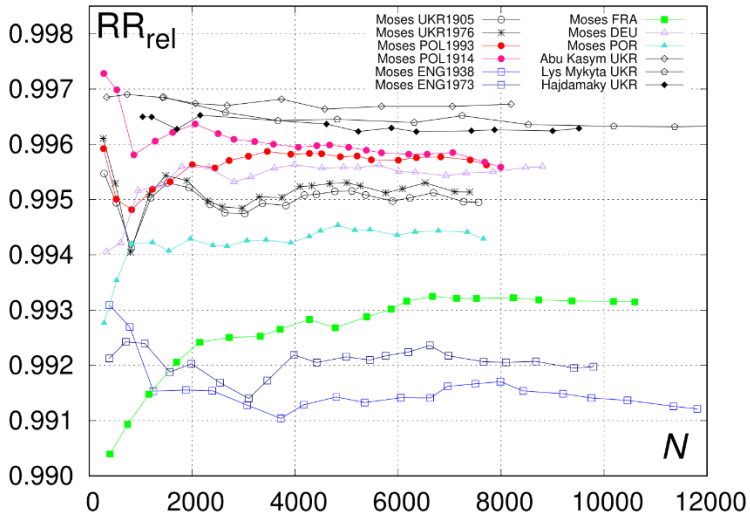

Figure 5: Dependence of the relative repeat rate RR_{rel} on text size N.

At the left side of Figure 6 (Dependence of the exponent α on text size N), anomalities in the values of α are observed for the Portuguese translation and for the Polish POL1993, namely, $\alpha > 2$. Certain texts demonstrate an opposite trend (α being significantly lower than the average of about 1.6–1.7) at low values of N. The explanation of such behavior of α is quite simple. As this parameter is obtained from fitting the frequency spectra, the collected data appear insufficient for small text sizes corresponding to $N < 2000$ and the respective frequency spectra do not demonstrate typical shapes that would ensure adequate application of the "temperature" approach.

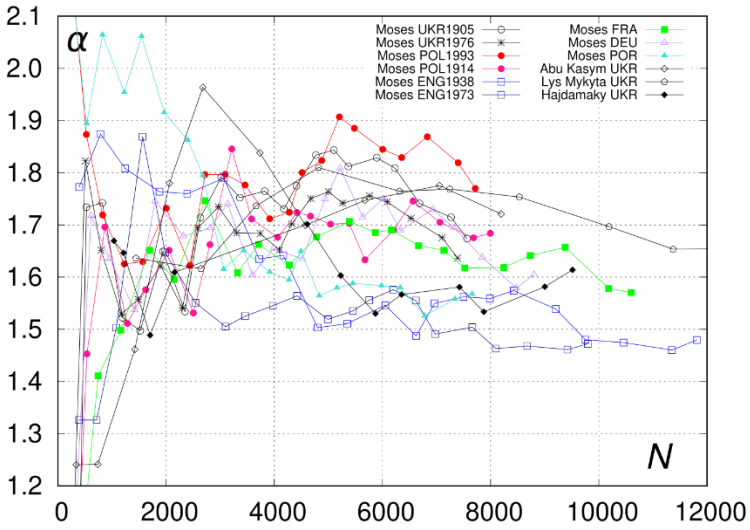

Figure 6: Dependence of the exponent α on text size N.

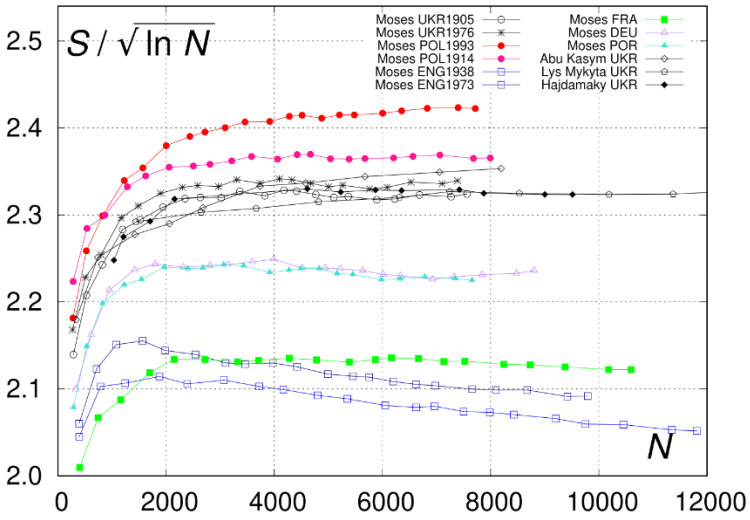

Figure 7: Dependence of the logarithmically scaled entropy on text size N.

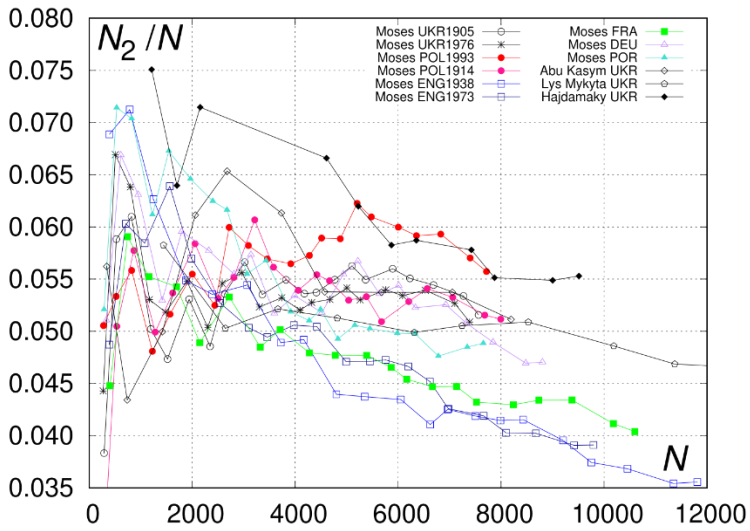

Figure 8: Dependence of the fraction of dis legomena N_2/N on text size N.

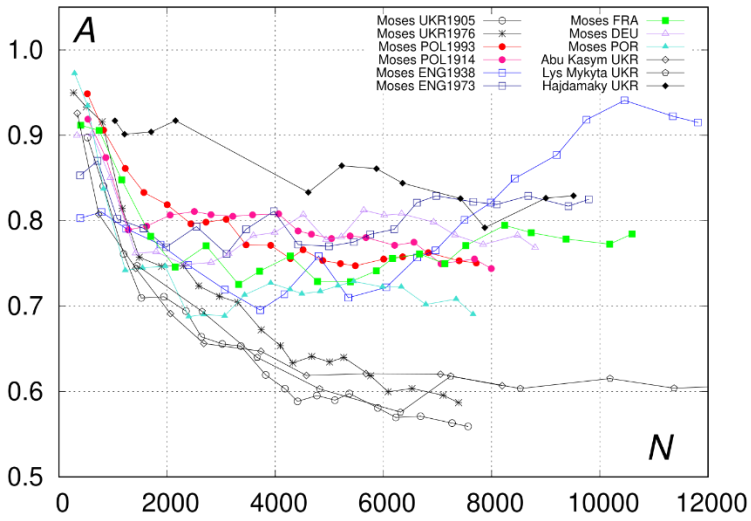

Figure 9: Dependence of the parameter A on text size N.

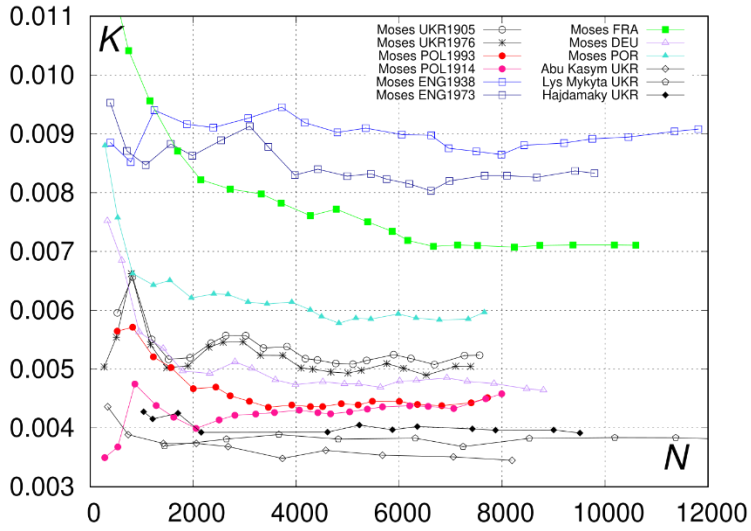

Figure 10: Dependence of the Yule's K on text size N.

Finally, we would like to mention observations regarding another combined parameter, N_1/N_2. While both quantities significantly depend on N (see, in particular, Figure 8), such a ratio achieves rather stable behavior already at text sizes of several thousand tokens, see Figure 11.

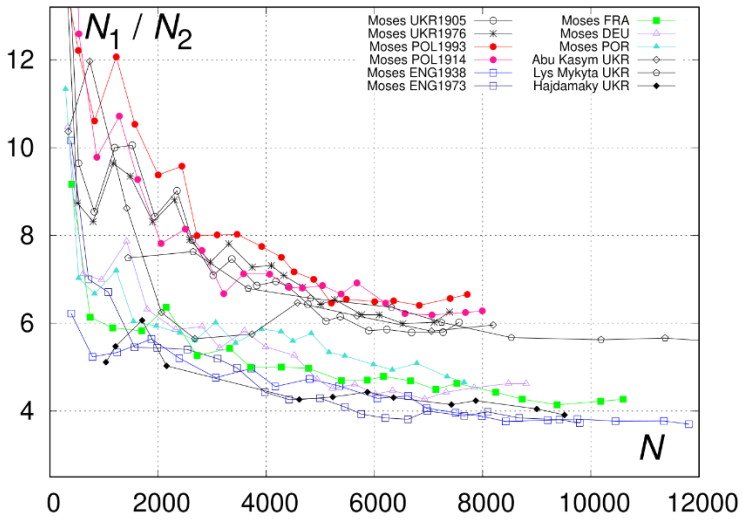

Figure 11: Dependence of the fraction of hapax and dis legomena N_1/N_2 on text size N.

5 Discussion

We have identified several parameters suitable for author or language attribution in relatively short texts. On the other hand, some parameters appear not relevant for studies of short or moderate length texts. One should keep in mind that we have analyzed poetic texts, and their peculiarities coming from limitations due to rhythm and rhyme requirements might influence the conclusions to a certain extent.

In particular, we have confirmed that h-point scales with text length N as $\sqrt{N}$. The behavior of the h-point scaling coefficient a with respect to text size suggests that this parameter might be author-specific: in Figure 4, Franko's texts are grouped separately from the Shevchenko's text UKR Hajd. Similar observations concern the parameter A, originally introduced as a measure of language analyticity/syntheticity and shown in Figure 9, where the differences between the two authors are even more pronounced. Further analysis with more texts of several authors is required to draw more substantiated conclusions.

From Figures 5 and 10 one can see that the relative repeat rate $\mathrm{RR}_{\mathrm{rel}}$ and Yule's K are quite stable in N. Since the values are different for the same author, these parameters might be genre-specific (and also language-specific). While such behavior could mean significant correlation between $\mathrm{RR}_{\mathrm{rel}}$ and K, in reality the Pearson correlation coefficient $r_{\mathrm{RRrel}-K}$ does not indicate this: it varies, both in sign and absolute value, from (−0.996) for FRA to (+0.547) for UKR AK.

The dependence of the fraction of dis legomena N_2/N as well as the fraction of hapax and dis legomena N_1/N_2 on text size (Figures 8 and 11) hints toward their author-specific nature. As with the parameters a and A mentioned above, the justification of such a claim remains an open issue so far.

Significant variations of the exponent α for a specific text mean that – at least for moderate text sizes considered here – this parameter is not suitable for text classification. On the other hand, the

"temperature" T scaling exponent γ_T, in particular, at a fixed value of α (e. g., an intermediate $\alpha = 1.5$) might serve a good discriminating parameter. Scaling exponents of other analyzed parameters, in particular, number of types $V(N)$, number of hapaxes $N_1(N)$, entropy $S(\ln N)$, are found to be text-specific with a tendency to lower values for languages with higher level of analyticity (especially English versus other studied languages), see Table 1.

Curiously, we have not discovered any significant variations of the calculated parameters with respect to the orthographies for the Ukrainian text. Only the relative repeat rate RR_{rel} (Figure 5) and Yule's K (Figure 10) demonstrate some systematic shifts. The RR_{rel} values for the 1905 original text UKR1905 appear to be shifted down and the values of K are shifted up, compared to the 1976 version UKR1976. Both directions clearly coincide with the domains corresponding to languages with higher analyticity level (in particular, English and French), as expected.

A modification of the Lambda parameter given by Eq. (5*) is suggested as a more stable quantity in text length, at least for texts of thousands to tens of thousand words. Logarithmic scaling (in powers of $\ln N$) can also ensure stable parameters based on entropy as shown in Figure 7. We also expect to apply nonadditive/nonextensive forms of entropy (Tsallis entropy, Rényi entropy, etc.; cf. Gell-Mann and Tsallis 2014; Tanaka-Ishii and Aihara 2015) and their properly scaled counterparts in future studies.

Acknowledgements

This research is supported by Project No. 0122U001613 from the Ministry of Education and Science of Ukraine. It would be impossible to conduct these studies without Armed Forces of Ukraine defending our homes and country against the full-scale invasion of the Russian Federation in Ukraine.

References

Berg, T. (2014). On the relationship between type and token frequency. *Journal of Quantitative Linguistics*, 21(3), pp. 199-222. https://doi.org/10.1080/09296174.2014.911505.

Best, K.-H., Zinenko, S. (1999). Wortlängen in Gedichten des ukrainischen Autors Ivan Franko. In: Genzor, J., Ondrejovič, S. (Eds.), *Pange lingua. Zbornik na počesť Viktora Krupu*, pp. 201-213. Bratislava: Veda.

Buk, S. (2021). *Velyka proza Ivana Franka: elektronnyj korpus, častotni slovnyky ta inši miždyscyplinarni konteksty*. Lviv University Press.

Buk, S., Krynytskyi, Yu., Rovenchak, A. (2019). Properties of autosemantic word networks in Ukrainian texts. *Advances in Complex Systems*, 22(6). article 1950016. https://doi.org/10.1142/S0219525919500164.

Buk, S., Rovenchak, A. (2007). Statistical parameters of Ivan Franko's novel Perekhresni stežky (The Cross-Paths). In: Grzybek, P., Köhler, R. (Eds.). *Exact Methods in the Study of Language and Text: Dedicated to Gabriel Altmann on the Occasion of his 75th Birthday*, pp. 39-48. Berlin-New York: Mouton de Gruyter. https://doi.org/10.1515/9783110894219.39.

Buk, S., Rovenchak, A. (2008). Menzerath-Altmann law for syntactic structures in Ukrainian. *Glottotheory*, 1(1), pp. 10-17. https://doi.org/10.1515/glot-2008-0002

Buk, S., Rovenchak, A. (2016). Probing the "temperature" approach on Ukrainian texts: Long-prose fiction by Ivan Franko. In: Kelih, E., Knight, R., Mačutek J., Wilson, A. (Eds.). *Studies in Quantitative linguistics 23: Issues in Quantitative Linguistics 4; dedicated to Reinhard Köhler on the occasion of his 65th birthday*, pp. 160-175. Lüdenscheid: RAM-Verlag.

Cortina-Borja, M., Chappas, C. (2006). A stylometric analysis of newspapers, periodicals and news scripts. *Journal of Quantitative Linguistics*, 13(2-3), pp. 285-312. https://doi.org/10.1080/09296170600892538.

Czetyrba-Piszczako, M. (2016). Twórczość literacka Iwana Franki w przekładach polskich. *Acta Neophilologica*, 18(2), pp. 235-247.

Čech, R. (2015). Text length and the lambda frequency structure of a text. In: Mikros, G. K., Mačutek, J. (Eds.). *Sequences in Language and Text*, pp. 71-87. Berlin-Boston: Mouton de Gruyter. https://doi.org/10.1515/9783110362879-006

Franko, I. (1912). *Kamenjari. Ukrajins'kyj tekst i pol's'kyj pereklad. Deščo pro štuku perekladannja*. Lviv.

Gell-Mann, M., Tsallis, C. (Eds.) (2014). *Nonextensive Entropy: Interdisciplinary Applications*. Oxford University Press.

Golcher, F. (2007). A stable statistical constant specific for human language texts. In: Angelova, G., Bontcheva, K., Mitkov, R., Nicolov, N., Nikolov, N. (Eds.). *International Conference "Recent Advances in Natural Language Processing" (RANLP-2007): Proceedings* (Borovets, Bulgaria, 27-29 September 2007), pp. 240-245. http://lml.bas.bg/ranlp2007/DOCS/RANLP2007.pdf.

Holovatch, Y., Palchykov, V. (2007). Fox Mykyta and networks of language. *Journal of Physical Studies*, 11(1), pp. 22-33. https://doi.org/10.30970/jps.11.022.

Juola, P., Mikros, G. K., Vinsick, S. (2018). Correlations and potential cross-linguistic indicators of writing style. *Journal of Quantitative Linguistics*, 26(2), pp. 146-171. https://doi.org/10.1080/09296174.2018.1458395.

Kelih, E., Antić, G., Grzybek, P., Stadlober, E. (2005). Classification of author and/or genre? The impact of word length. In: Weihs. C., Gaul, W. (Eds.). *Classification – The Ubiquitous Challenge*, pp. 498-505. Heidelberg: Springer. https://doi.org/10.1007/3-540-28084-7_58.

Koppel, M., Schler, J., Argamon, S. (2009). Computational methods in authorship attribution. *Journal of the American Society for Information Science and Technology*, 60(1), pp. 9-26. https://doi.org/10.1002/asi.20961

Kornai, A. (2002). How many words are there?. *Glottometrics*, 4, pp. 61-86.

Kovalyk, I. I., Oščypko, I. J., Poljuha, L. M. (1990). *Leksyka poetyčnykh tvoriv Ivana Franka.* Lviv.

Kubát, M., Matlach, V., Čech, R. (2014). *QUITA: Quantitative Index Text Analyzer*. Lüdenscheid: RAM-Verlag.

Locker, A. (2019). "Because the computer said so!" Can computational authorship analysis be trusted? *Journal of Language Works*, 4(1), pp. 23-37. https://tidsskrift.dk/lwo/article/view/115710.

Mochulskyj, M. (1956). Z ostannikh desjatylit žyttja Ivana Franka (1896-1916). In: Dej, O., Kornijenko, N. (Eds.). *Ivan Franko u spohadakh sučasnykiv. Knyha 1*, pp. 422-442. Lviv: Knyžkovo-žurnal′ne vydavnyctvo.

Nini, A., Grant, T. (2013). Bridging the gap between stylistic and cognitive approaches to authorship analysis using Systemic Functional Linguistics and multidimensional analysis. *International Journal of Speech, Language & the Law*, 20(2), pp. 173-202. https://doi.org/10.1558/ijsll.v20i2.173.

OCRConvert (2010-2022). *Online OCR Conversion*. Retrieved from https://www.ocrconvert.com (Accessed June 11, 2022).

Popescu, I.-I., Altmann, G. (2006). Some aspects of word frequencies. *Glottometrics*, 13, pp. 23-46.

Popescu, I.-I., Altmann, G. (2007). Writer's view of text generation. *Glottometrics*, 15, pp. 71-81.

Popescu, I.-I., Altmann, G. (2008a). On the regularity of diversification in language. *Glottometrics*, 17, pp. 94-108.

Popescu, I.-I., Altmann, G. (2008b). Hapax legomena and language typology. *Journal of Quantitative Linguistics*, 15(4), pp. 370-378. https://doi.org/10.1080/09296170802326699.

Popescu, I.-I., Altmann, G., Grzybek, P., Jayaram, B. D., Köhler, R., Krupa, V., Mačutek, J., Pustet, R., Uhlířová, L., Vidya, M. N. (2009). *Word Frequency Studies*. Berlin-New York: Mouton de Gruyter. https://doi.org/10.1515/9783110218534.

Popescu, I.-I., Čech, R., Altmann, G. (2011). *The Lambda-structure of Texts*. Lüdenscheid: RAM-Verlag.

Rovenchak, A., Buk, S. (2011). Application of a quantum ensemble model to linguistic analysis. *Physica A*, 390(7), pp. 1326-1331. https://doi.org/10.1016/j.physa.2010.12.009.

Rovenchak, A., Buk, S. (2018). Part-of-speech sequences in literary text: Evidence from Ukrainian. *Journal of Quantitative Linguistics*, 25(1). pp. 1-21. https://doi.org/10.1080/09296174.2017.1324601.

Rovenchak, A., Rovenchak, O. (2018). Quantifying comprehensibility of Christmas and Easter addresses from the Ukrainian Greek Catholic Church hierarchs. *Glottometrics*, 41, pp. 57-66.

Shen, Q., Tao, Y. (2021). Stance markers in English medical research articles and newspaper opinion columns: A comparative corpus-based study. *PLoS ONE* 16(3), article e0247981. https://doi.org/10.1371/journal.pone.0247981.

Sichel, H. S. (1986). Word frequency distributions and type-token characteristics. *Mathematical Scientist*, 11, pp. 45-72.

Tanaka-Ishii, K., Aihara, S. (2015). Computational constancy measures of texts—Yule's K and Rényi's entropy. *Computational Linguistics*, 41(3), pp. 481-502. https://doi.org/10.1162/COLI_a_00228.

Tuldava, J. (2004). The development of statistical stylistics (A survey). *Journal of Quantitative Linguistics*, 11(1-2), pp. 141-151. https://doi.org/10.1080/09296170512331383695.

Wilson, A., Harvey, R. (2020). Using rank-frequency and type-token statistics to compare morphological typology in the Celtic languages. *Journal of Quantitative Linguistics*, 27(2), pp. 159-186. https://doi.org/10.1080/09296174.2018.1560122.

Yakymovych, B. Z. (2011). Poema Ivana Franka "Mojsej": de, koly i jak narodžuvavsja odyn iz najvydatnišykh tvoriv svitovoji literatury XX st. *Visnyk of Lviv Commercial Academy. Series – The Humanities*, 10, pp. 149-158.

Yule, G. U. (1944). *A Statistical Study of Literary Vocabulary*. Cambridge: University Press.

Zabuzhko, O. (1993). *Filosofija ukrajins'koji ideji ta jevropejs'kyj kontekst*. Kyiv: Osnovy.

Zörnig, P., Kelih, E., Fuks, L. (2016). Classification of Serbian texts based on lexical characteristics and multivariate statistical analysis. *Glottotheory*, 7(1), pp. 41-66. https://doi.org/10.1515/glot-2016-0004.

Appendix

In the tables below, the first column marked by "Ch." corresponds to the poem chapters. The value "0" stands for the Prologue.

Table A1: The values of parameters for the Ukrainian original of the poem *Moses* UKR1905.

Ch.	V	N	N_1	N_2	N_2/N	S	$S/\sqrt{\ln N}$	a	Λ	φ	K
0	206	287	176	11	0.03833	5.0895	2.1394	11.48	1.83	1.932	0.00588
1	356	527	299	31	0.05882	5.5266	2.2076	9.37	1.94	1.856	0.00596
2	520	820	427	50	0.06098	5.8092	2.2427	10.12	1.99	1.757	0.00657
3	726	1195	600	60	0.05021	6.0777	2.2832	9.88	2.00	1.753	0.00550
4	882	1521	724	72	0.04734	6.2107	2.2944	7.76	1.97	1.770	0.00517
5	1078	1941	868	103	0.05307	6.3530	2.3089	9.23	1.96	1.725	0.00519
6	1273	2348	1028	114	0.04855	6.4598	2.3187	9.17	1.99	1.704	0.00544
7	1394	2631	1113	141	0.05359	6.5108	2.3201	9.66	1.97	1.692	0.00557
8	1544	3020	1213	171	0.05662	6.5664	2.3197	9.32	1.94	1.687	0.00557
9	1699	3363	1343	180	0.05352	6.6313	2.3270	9.83	1.94	1.683	0.00536
10	1853	3823	1440	210	0.05493	6.6689	2.3220	9.10	1.90	1.680	0.00538
11	2004	4179	1557	224	0.05360	6.7232	2.3284	8.25	1.89	1.685	0.00518
12	2087	4412	1619	237	0.05372	6.7421	2.3273	8.72	1.88	1.680	0.00516
13	2195	4769	1686	262	0.05494	6.7621	2.3235	9.02	1.85	1.677	0.00510
14	2290	5102	1735	287	0.05625	6.7790	2.3201	8.86	1.82	1.676	0.00508
15	2392	5369	1815	295	0.05495	6.8020	2.3210	9.32	1.82	1.669	0.00515
16	2563	5895	1924	330	0.05598	6.8282	2.3174	9.56	1.80	1.663	0.00525
17	2679	6229	2009	343	0.05507	6.8527	2.3184	9.97	1.79	1.661	0.00518
18	2836	6687	2107	364	0.05443	6.8940	2.3229	10.28	1.78	1.658	0.00508
19	3032	7269	2248	388	0.05338	6.9207	2.3210	9.97	1.77	1.653	0.00523
20	3151	7563	2348	390	0.05157	6.9456	2.3241	9.88	1.78	1.651	0.00524

Table A2: The values of parameters for the Ukrainian version of the poem *Moses* in the modern orthography UKR1976.

Ch.	V	N	N_1	N_2	N_2/N	S	$S/\sqrt{\ln N}$	a	Λ	φ	K
0	206	271	179	12	0.04428	5.1310	2.1678	16.94	1.93	1.864	0.00504
1	355	508	297	34	0.06693	5.5622	2.2284	14.11	1.99	1.801	0.00554
2	518	799	424	51	0.06383	5.8278	2.2543	9.86	2.04	1.754	0.00662
3	725	1169	598	62	0.05304	6.1041	2.2967	11.69	2.04	1.734	0.00542
4	881	1486	720	77	0.05182	6.2440	2.3104	9.51	2.01	1.747	0.00502
5	1076	1898	864	104	0.05479	6.3872	2.3248	9.03	2.00	1.724	0.00505
6	1270	2303	1022	116	0.05037	6.4878	2.3317	9.59	2.01	1.699	0.00537
7	1395	2583	1114	141	0.05459	6.5418	2.3339	10.09	2.01	1.688	0.00546
8	1546	2967	1220	165	0.05561	6.5957	2.3326	9.69	1.98	1.684	0.00546
9	1702	3306	1351	173	0.05233	6.6625	2.3405	9.66	1.98	1.683	0.00523
10	1852	3740	1448	199	0.05321	6.7010	2.3363	9.35	1.93	1.678	0.00523
11	1998	4094	1558	213	0.05203	6.7526	2.3414	8.86	1.92	1.681	0.00502
12	2079	4322	1616	228	0.05275	6.7717	2.3404	8.93	1.90	1.679	0.00500
13	2190	4674	1692	248	0.05306	6.7915	2.3364	8.84	1.88	1.678	0.00495
14	2284	5004	1742	271	0.05416	6.8071	2.3324	9.46	1.84	1.672	0.00493
15	2387	5264	1824	279	0.05300	6.8329	2.3342	9.53	1.85	1.668	0.00498
16	2548	5762	1926	311	0.05397	6.8560	2.3299	9.73	1.83	1.662	0.00509
17	2664	6085	2011	325	0.05341	6.8830	2.3317	9.74	1.82	1.661	0.00501
18	2823	6527	2109	352	0.05393	6.9284	2.3377	9.66	1.81	1.660	0.00490
19	3021	7101	2252	374	0.05267	6.9564	2.3360	9.39	1.81	1.655	0.00505
20	3140	7389	2352	376	0.05089	6.9821	2.3394	9.42	1.81	1.653	0.00504

Table A3: The values of parameters for the Polish translation of the poem *Moses* POL1914.

Ch.	V	N	N_1	N_2	N_2/N	S	$S/\sqrt{\ln N}$	a	Λ	φ	K
0	230	281	210	8	0.02847	5.2797	2.2235	17.56	2.05	1.966	0.00350
1	394	535	340	27	0.05047	5.7259	2.2845	14.86	2.08	1.860	0.00368
2	582	866	489	50	0.05774	5.9810	2.2997	13.53	2.07	1.778	0.00475
3	818	1282	686	64	0.04992	6.2393	2.3324	9.69	2.08	1.794	0.00438
4	984	1621	807	87	0.05367	6.3748	2.3449	11.26	2.04	1.769	0.00418
5	1177	2055	938	120	0.05839	6.5038	2.3548	12.81	2.00	1.738	0.00399
6	1368	2503	1083	133	0.05314	6.5910	2.3562	11.90	1.97	1.725	0.00414
7	1503	2810	1187	155	0.05516	6.6453	2.3582	13.06	1.96	1.709	0.00422
8	1673	3214	1301	195	0.06067	6.7123	2.3621	12.55	1.94	1.703	0.00423
9	1836	3580	1432	201	0.05615	6.7718	2.3672	11.05	1.94	1.704	0.00426
10	2010	4061	1559	219	0.05393	6.8149	2.3642	11.25	1.91	1.692	0.00430
11	2163	4421	1670	245	0.05542	6.8645	2.3693	10.69	1.91	1.690	0.00426
12	2258	4668	1741	256	0.05484	6.8879	2.3697	10.59	1.90	1.691	0.00424
13	2379	5040	1831	267	0.05298	6.9042	2.3646	10.41	1.87	1.690	0.00427
14	2501	5383	1913	287	0.05332	6.9297	2.3642	10.63	1.85	1.686	0.00432
15	2611	5676	1999	289	0.05092	6.9527	2.3648	11.05	1.85	1.680	0.00436
16	2796	6207	2118	328	0.05284	6.9908	2.3656	10.78	1.83	1.678	0.00438
17	2927	6564	2208	355	0.05408	7.0184	2.3673	11.40	1.83	1.673	0.00437
18	3099	7061	2326	376	0.05325	7.0518	2.3688	11.76	1.81	1.668	0.00433
19	3300	7683	2472	396	0.05154	7.0738	2.3650	11.37	1.80	1.662	0.00449
20	3422	7994	2569	409	0.05116	7.0915	2.3656	11.38	1.81	1.658	0.00458

Table A4: The values of parameters for the English translation of the poem *Moses* ENG1938.

Ch.	V	N	N_1	N_2	N_2/N	S	$S/\sqrt{\ln N}$	a	Λ	φ	K
0	221	392	168	27	0.06888	4.9968	2.0449	6.13	1.52	2.010	0.00885
1	399	786	293	56	0.07125	5.4290	2.1026	7.86	1.57	1.841	0.00852
2	571	1245	416	78	0.06265	5.6235	2.1065	7.37	1.55	1.770	0.00940
3	793	1877	581	103	0.05487	5.8045	2.1142	7.33	1.51	1.750	0.00916
4	932	2391	666	128	0.05353	5.8731	2.1057	6.62	1.45	1.740	0.00911
5	1137	3068	795	167	0.05443	5.9788	2.1100	7.18	1.44	1.708	0.00927
6	1303	3721	904	182	0.04891	6.0293	2.1027	6.74	1.42	1.692	0.00945
7	1386	4167	934	205	0.04920	6.0605	2.0992	6.94	1.37	1.687	0.00919
8	1503	4800	998	211	0.04396	6.0923	2.0925	6.58	1.32	1.686	0.00902
9	1626	5351	1069	234	0.04373	6.1194	2.0885	6.83	1.30	1.679	0.00910
10	1749	6052	1128	263	0.04346	6.1407	2.0809	6.30	1.25	1.681	0.00899
11	1849	6623	1181	272	0.04107	6.1658	2.0787	6.89	1.23	1.669	0.00898
12	1909	6963	1203	296	0.04251	6.1872	2.0800	7.02	1.22	1.669	0.00876
13	1990	7499	1243	314	0.04187	6.1950	2.0739	6.78	1.19	1.667	0.00870
14	2075	7986	1284	331	0.04145	6.2136	2.0729	6.77	1.18	1.665	0.00865
15	2146	8427	1316	350	0.04153	6.2246	2.0704	7.04	1.18	1.657	0.00881
16	2265	9205	1380	364	0.03954	6.2410	2.0658	7.10	1.16	1.653	0.00884
17	2318	9752	1393	365	0.03743	6.2418	2.0595	7.27	1.13	1.649	0.00892
18	2435	10456	1451	385	0.03682	6.2635	2.0589	7.24	1.12	1.646	0.00895
19	2560	11351	1517	402	0.03542	6.2727	2.0528	7.46	1.10	1.642	0.00904
20	2633	11813	1554	420	0.03555	6.2825	2.0516	7.14	1.10	1.641	0.00908

Table A5: The values of parameters for the French translation of the poem *Moses* FRA.

Ch.	V	N	N_1	N_2	N_2/N	S	$S/\sqrt{\ln N}$	a	Λ	φ	K
0	217	402	165	18	0.04478	4.9211	2.0096	8.20	1.52	1.852	0.01168
1	369	745	270	44	0.05906	5.3153	2.0669	9.55	1.54	1.800	0.01041
2	522	1159	377	64	0.05522	5.5442	2.0873	8.05	1.50	1.782	0.00956
3	733	1695	536	92	0.05428	5.7769	2.1186	8.06	1.52	1.752	0.00871
4	900	2147	668	105	0.04891	5.9105	2.1339	7.01	1.52	1.751	0.00822
5	1065	2721	763	145	0.05329	6.0007	2.1338	6.58	1.47	1.743	0.00806
6	1231	3321	874	161	0.04848	6.0679	2.1310	6.86	1.43	1.729	0.00798
7	1332	3708	929	186	0.05016	6.1140	2.1327	7.01	1.41	1.722	0.00782
8	1480	4277	1025	205	0.04793	6.1738	2.1351	6.84	1.39	1.713	0.00761
9	1627	4780	1134	228	0.04770	6.2092	2.1332	7.07	1.38	1.702	0.00772
10	1758	5389	1206	257	0.04769	6.2458	2.1308	6.87	1.35	1.701	0.00750
11	1877	5867	1284	273	0.04653	6.2843	2.1334	6.82	1.34	1.699	0.00734
12	1952	6168	1340	280	0.04540	6.3096	2.1358	6.70	1.33	1.698	0.00719
13	2057	6668	1397	298	0.04469	6.3351	2.1349	7.05	1.31	1.691	0.00709
14	2143	7134	1433	319	0.04472	6.3488	2.1314	6.97	1.29	1.687	0.00711
15	2239	7522	1505	325	0.04321	6.3690	2.1318	6.91	1.29	1.684	0.00710
16	2370	8240	1567	354	0.04296	6.3910	2.1283	7.13	1.26	1.677	0.00707
17	2471	8733	1618	379	0.04340	6.4092	2.1276	7.34	1.26	1.671	0.00710
18	2596	9377	1685	407	0.04340	6.4272	2.1252	7.24	1.24	1.668	0.00711
19	2744	10182	1769	419	0.04115	6.4466	2.1221	7.05	1.22	1.667	0.00711
20	2829	10598	1826	428	0.04038	6.4608	2.1222	7.15	1.22	1.664	0.00711

Table A6: The values of parameters for the German translation of the poem *Moses* DEU.

Ch.	V	N	N_1	N_2	N_2/N	S	$S/\sqrt{\ln N}$	a	Λ	φ	K
0	217	333	178	17	0.05105	5.0596	2.0994	10.07	1.71	1.954	0.00752
1	369	613	292	41	0.06688	5.4781	2.1623	9.58	1.73	1.933	0.00685
2	539	951	419	60	0.06309	5.7971	2.2137	9.51	1.74	1.903	0.00563
3	753	1416	590	75	0.05297	6.0251	2.2368	8.38	1.73	1.887	0.00535
4	892	1797	676	107	0.05954	6.1417	2.2436	7.48	1.68	1.880	0.00497
5	1064	2321	784	134	0.05773	6.2353	2.2398	7.16	1.61	1.853	0.00493
6	1249	2812	924	156	0.05548	6.3175	2.2418	7.27	1.60	1.825	0.00512
7	1348	3140	978	180	0.05732	6.3646	2.2429	7.35	1.57	1.806	0.00501
8	1494	3596	1084	186	0.05172	6.4264	2.2459	6.80	1.55	1.803	0.00482
9	1619	3987	1163	213	0.05342	6.4760	2.2491	6.13	1.54	1.809	0.00473
10	1742	4518	1232	234	0.05179	6.4947	2.2388	6.12	1.49	1.786	0.00478
11	1860	4939	1293	274	0.05548	6.5297	2.2390	6.30	1.47	1.778	0.00475
12	1934	5218	1334	296	0.05673	6.5481	2.2381	6.66	1.46	1.771	0.00475
13	2030	5636	1392	302	0.05358	6.5709	2.2359	6.70	1.43	1.775	0.00469
14	2116	6013	1433	327	0.05438	6.5827	2.2315	6.68	1.41	1.767	0.00480
15	2190	6335	1478	331	0.05225	6.5984	2.2302	6.92	1.39	1.758	0.00480
16	2327	6927	1555	364	0.05255	6.6195	2.2260	6.98	1.37	1.745	0.00485
17	2447	7325	1647	372	0.05078	6.6483	2.2286	7.15	1.37	1.735	0.00479
18	2587	7844	1740	384	0.04895	6.6817	2.2313	7.20	1.37	1.727	0.00475
19	2745	8478	1840	398	0.04695	6.7151	2.2328	6.92	1.36	1.725	0.00467
20	2846	8803	1914	414	0.04703	6.7388	2.2360	7.19	1.36	1.719	0.00464

Table A7: The values of parameters for the Portuguese translation of the poem *Moses* POR.

Ch.	V	N	N_1	N_2	N_2/N	S	$S/\sqrt{\ln N}$	a	Λ	φ	K
0	197	288	170	15	0.05208	4.9461	2.0785	5.48	1.74	2.105	0.00880
1	332	532	267	38	0.07143	5.3833	2.1488	8.74	1.78	1.900	0.00758
2	487	824	387	58	0.07039	5.6965	2.1984	8.24	1.80	1.869	0.00662
3	678	1226	540	75	0.06117	5.9193	2.2197	8.51	1.80	1.826	0.00643
4	812	1546	628	104	0.06727	6.0324	2.2261	9.15	1.76	1.800	0.00651
5	988	1966	754	127	0.06460	6.1687	2.2400	9.14	1.74	1.794	0.00621
6	1149	2401	867	150	0.06247	6.2447	2.2383	7.69	1.71	1.786	0.00628
7	1243	2677	925	165	0.06164	6.2904	2.2391	7.82	1.68	1.774	0.00627
8	1378	3064	1022	170	0.05548	6.3542	2.2427	7.29	1.66	1.769	0.00614
9	1487	3434	1082	195	0.05679	6.3959	2.2416	7.79	1.63	1.753	0.00611
10	1626	3914	1190	203	0.05187	6.4250	2.2339	7.40	1.60	1.741	0.00614
11	1739	4273	1266	218	0.05102	6.4667	2.2365	6.84	1.58	1.747	0.00601
12	1806	4492	1309	234	0.05209	6.4910	2.2383	6.64	1.57	1.747	0.00589
13	1900	4831	1372	238	0.04927	6.5190	2.2383	7.15	1.55	1.739	0.00578
14	1968	5159	1393	261	0.05059	6.5277	2.2326	6.82	1.52	1.738	0.00587
15	2044	5451	1440	274	0.05027	6.5459	2.2317	6.95	1.50	1.729	0.00585
16	2168	5982	1507	298	0.04982	6.5632	2.2256	7.11	1.48	1.717	0.00594
17	2257	6328	1555	315	0.04978	6.5873	2.2266	7.52	1.46	1.709	0.00587
18	2385	6783	1642	323	0.04762	6.6201	2.2288	8.07	1.46	1.700	0.00584
19	2512	7343	1702	356	0.04848	6.6443	2.2270	7.64	1.43	1.700	0.00585
20	2588	7656	1741	374	0.04885	6.6535	2.2249	7.80	1.43	1.695	0.00597