

# Measuring Rhythm Regularity in Verse: Entropy of Inter-Stress Intervals

Artjoms Šeļa<sup>1,2\*,†</sup>, Mikhail Gronas<sup>3,†</sup>

<sup>1</sup>*Institute of Polish Language (Polish Academy of Sciences), al. Mickiewicza 31, 31-120 Kraków*

<sup>2</sup>*University of Tartu, Ülikooli 18, 50090 Tartu, Estonia*

<sup>3</sup>*Dartmouth college, Hanover, NH 03755, USA*

## Abstract

Recognition of poetic meters is not a trivial task, since metrical labels are not a closed set of classes. Outside of classical meters, describing the metrical structure of a poem in a large corpus requires expertise and a shared scientific theory. In a situation when both components are lacking, alternative and continuous measures of regularity can be envisioned. This paper focuses on poetic rhythm to propose a simple entropy-based measure for poem regularity using counts of non-stressed intervals. The measure is validated using subsets of a well-annotated Russian poetic corpus, prose, and quasi-poems (prose chopped into lines). The regularity measure is able to detect a clear difference between various organizational principles of texts: average entropy rises when moving from accentual-syllabic meters to accentual variations to free verse and prose. Interval probabilities, when taken as a vector of features, also allow for classification at the level of individual poems. This paper argues that distinguishing between meter as a cultural idea and rhythm as an empirical sequence of sounds can lead to better understanding of form recognition and prosodic annotation problems.

## Keywords

rhythm, meter, poetry, regularity, entropy, diversity

## 1. Introduction

Identifying the meter of a poem is one of those tasks that is deceptively simple. The problem lies not with methods: a wide array of successful solutions exist, ranging from rule-based [3, 1, 18] to probabilistic [16] to deep learning [17, 8]. The problem, as often happens, is conceptual: metrical forms are treated as a closed set of classes, when in fact they are not; far from it. There is much more variation in organizational principles than common iambic or trochaic patterns. Not having a matching label for some unusual arrangement of stressed and unstressed syllables in a poem does not make that poem non-regular, or even non-metrical. And, indeed, quite often we do not have labels.

Here are some numbers from well-annotated, large poetic corpora that are indispensable in today's work. Only 2% of lines in the Russian corpus do not conform to any of the five classical

---

*CHR 2022: Computational Humanities Research Conference, December 12 – 14, 2022, Antwerp, Belgium*

\*Corresponding author.

†These authors contributed equally.

✉ artjoms.sela@ijp.pan.pl (A. Šeļa); mikhail.gronas@dartmouth.edu (M. Gronas)

🆔 0000-0002-2272-2077 (A. Šeļa)

 © 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

meters (iamb, trochee, dactyl, anapest, amphibrach); for Czech this number grows to 14%, in the Dutch song collection it is 15%, and finally the German corpus has an incredible rate of 68% ‘unrecognized’ lines (out of 170,000 in total).<sup>1</sup> This is not simply a failure of recognition systems—even if you suppose there is some noise. This is a problem of domain expertise and levels of variation in modern verse forms. Germanic versification systems, for example, stem from alliterative tonic verse and widely employ meters that are not based on stable, recurring units of rhythm (metrical feet) [6]. With enough domain expertise and a shared taxonomy (that stems from shared theory), all poems in a corpus, in principle, could be described with a meaningful metrical label.<sup>2</sup> That is how the Russian poetic corpus is described now [7].

Needless to say, ‘enough domain expertise’ and ‘shared theory’ are luxuries. More often than not, we will not have enough resources for the former nor enough scholarly consensus for the latter. Semi-annotated, theory-agnostic and unstructured data is the primary reality of computational work today. That is why parallel approaches—that are not based on labels—for describing metrical and rhythmic features of verse might be considered. This paper proposes a simple way to measure rhythm regularity as one continuous value based on measures from information theory and ecological diversity.

Conceptually, our approach focuses on rhythm alone and brackets the ‘meter’ (as an already known template) out. We try to ask how regularly the rhythmical features are organized in a text, to define a single scale that has highly regular iambs on one side, and irregular prose on the other—with many more intermediate and heterogeneous cases in between. Similar attempts to define one space for metrical forms have been made before [16], but suffered from not distinguishing between the metricality and regularity of verse. At the same time, scholars working with syllabic verse (mostly Romanic) often looked in the same direction, because similarly organized isosyllabic verse might have different rhythmic features across languages [4], or authors [13].

Our reasoning here is informed by the so-called ‘Russian school’ of metrics that have established a distinction between poetic meter and rhythm. In the classic formulation, meter is a theoretical abstracted scheme, while rhythm is an empirical realization of (and deviation from) that scheme in a given poem (with ‘wrong’ stresses, additional syllables, distribution of word boundaries, etc.; for a brief English overview see Starostin & Pilshchikov [22]). Czech, English, Russian and German literatures all had an idea of iambic meter by the 19th century, but the meter’s rhythm was different everywhere: English iambs allowed the most freedom and scheme transgressions (impossible in Czech and Russian); however, Slavic iambic meters still had enough rhythmic variation compared to German that was the strictest in following the iambic pattern (unstressed syllable followed by a stressed one) [23, 14]. Conversely, Spanish syllabic verse had strong iambic tendencies in its *rhythm* [4, 14], without having an iambic *meter* in its cultural repertoire per se.

This fundamental distinction was taken further by Maksim Shapir [20], who suggested that the relationship between meter and rhythm is dialectic, continuous, and non-hierarchical (in the sense that rhythm is not an outcome of a meter). Meter is just a rhythm that is repeated

---

<sup>1</sup>Numbers were provided by Petr Plecháč. For the detailed description of the mentioned corpora, see [19].

<sup>2</sup>Even if the label is just ‘free verse’, ‘tonic verse’, or something more exotic like: ‘3-ictus line with maximum inter-stress interval of 2 syllables and regular alternation of masculine and feminine clausulae’.

enough times and standardized in a tradition. For example, Russian folksongs allowed a recurring rhythmical sequence of five syllables with a strong preference for stress on position 3, ‘00100’ (often as a hemistich in 10-syllable line). This rhythm was a mere tendency, until it was recognized by poets and literati and parsed as metrical foot [2, 5]. It even received its early name—‘reduplicated amphibrach’—despite being open to equally plausible scansion within an already existing system: either as a trochaic ‘(1)0-10-0’ or anapestic ‘001-00’ meter.<sup>3</sup> This is how a ‘penton’ (*piatislozhnik*) was born in literary tradition for writing stylized folk-inspired poetry—when one of the rhythmical *possibilities* in folksongs became a rule.

Conversely, a meter could be ‘rhythmicized’ by breaking expectations and inertia in a given poem: actual rhythm can be so individual that it would not be possible to make a judgement about a governing scheme. From this perspective, meter is more of a cultural phenomenon, while rhythm is of prosodic nature. Meters, or measures, are crystallized technologies for the organization of speech: it is possible for them to be written down in poetics manuals, taught in schools, transmitted. They are (mis-)recognized by poets, debated, weaponized for ideological reasons or completely deconstructed [11]. Rhythm is a source of cultural conventions and a sub-product of linguistic affordances. Here, we concentrate on information from the sub-product alone, while tracing its relationship to culturally recognized meters using existing deep annotations in the corpus of Russian poetry.

## 2. Materials and methods

The underlying idea of a regularity measurement is simple. Any systematic outside force—like meter—that organizes speech prosody will leave its trace in the distribution of possible unstressed intervals. By measuring the shape, or (un)evenness of these distributions we acquire a proxy for rhythmical regularity. Ideal iambic meter allows only one type of unstressed interval—one syllable. In practice, it frequently allows one, often three; rarely five (pyrrhics happen instead of fully fledged iamb feet). In prose (at least in most of it) there is no reason to suspect systematic limitations on the allowed inter-stress interval. We can also think about it as an inequality problem—meters are tyrannical forces that allow only a fraction of linguistically possible intervals to dominate, while prosaic language has a more democratic (or just disinterested) outlook on the distribution of the possible intervals.

Our approach is summarized in Figure 1. Given a binary rhythmical annotation of an individual text, we extract all inter-stress intervals. We include in the notion of ‘interval’ an unstressed syllable that precedes the first instance of stress in a line (anacrusis), but we exclude unstressed syllables that follow the last stressed syllable (clausula).<sup>4</sup> The types (intervals of particular lengths) are then counted and transformed to probabilities. To measure how ‘uneven’ the resulting probability distribution is we use classic Shannon entropy. This measures the uncertainty of a probability distribution and has two features that are useful for us: it

---

<sup>3</sup>To be fair, both trochees and anapests were heavily employed as meters for imitation of folksongs—dactylic line endings played an important role here.

<sup>4</sup>Clausulae are dominantly regular in verse. However, the same pattern of line endings (that might depend on the prosody of rhyming words) can be shared across poetic forms of very different organization (syllabic, tonic, accentual-syllabic). Since the appearance of clausulae is an almost constant feature of traditional verse, counting them will not add much to our task of differentiating verse of different organization.



Corpus [7], prose from the collection of 19th c. narrative fiction [21].

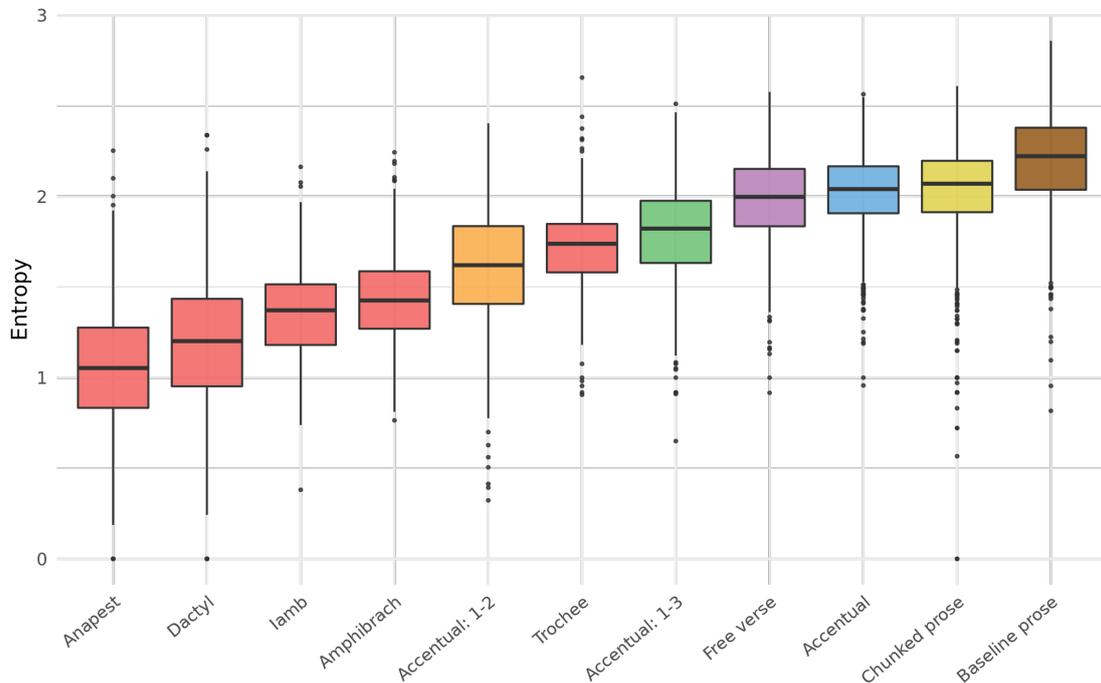
- **Classic meters** Sample of 700 poems for each of the classic accentual-syllabic meters (iamb, trochee, anapest, dactyl, amphibrach);
- **Baseline prose** Sample of 1000 paragraphs of Russian 19th century literary prose (2 paragraphs sampled from each of 500 texts);
- **Chunked prose** The calculated regularity of poetry, at least partly, emerges from division into lines that can artificially cut mid-sentence and mid-unstressed intervals. To emulate this behavior, we cut continuous prose into quasi-poems. For each chunk, its line length in syllables and overall length in lines is determined randomly by drawing values from the empirical distribution in the poetry corpus. In the end, our quasi-poems closely resemble the average dimensions of actual poetry. We sample 1000 quasi-poems (2 from each of 500 texts) ;
- **Free verse** Verse that is labeled ‘free’ in corpus annotations, and that should not be governed by any surface-level prosodic pattern (700 poems);
- **Accentual, 1-2 (A1-2)** Poems that take an intermediate position between accentual and accentual-syllabic verse by allowing variation of 1-2 syllables in inter-ictus intervals. A meter also known as *dolnik* (700 poems);
- **Accentual, 1-3 (A1-3)**. The same as A1-2, but allowing greater variation of 1-3 syllables. A meter known as *taktovik* (700 poems);
- **Accentual (A)**. Pure accentual: inter-ictus intervals are not under any regulation, the only measure is the tendency for a constant number of strong positions in a line (700 poems).

Intuitively, all these forms should be positioned at different parts of the scale of regularity. We can expect that the regularity of **Classic meters** > **A1-2** > **A1-3** > **A** > **Free verse (Chunked prose)** > **Baseline prose**. This is what we set out to test using the per-poem distributions of regularity measures.

### 3. Results

Figure 2 shows the distribution of entropy measures for individual texts in each category, arranged by median value. While the variance of observations is large, there is a clear increase in average entropy (decrease in regularity) from accentual-syllabic (median around 1) to baseline prose (median above 2). Hill numbers calculated for bootstrapped samples also show a clear difference in diversity profiles between prose, classical meters, and intermediate forms (see Appendix B). Overall, this supports our expectation of how things should be arranged on a regularity scale.

The interval distributions also reveal a shortcoming of this measure based only on inter-stress intervals: pure accentual, free verse and chunked prose all show similar median regularity. While for free verse and chunked prose this is not surprising, we know that accentual verse has a governing principle and is regular, it is just that this regularity is not reflected in the distribution of rhythmical patterns. Additionally, trochaic meters show unexpectedly large

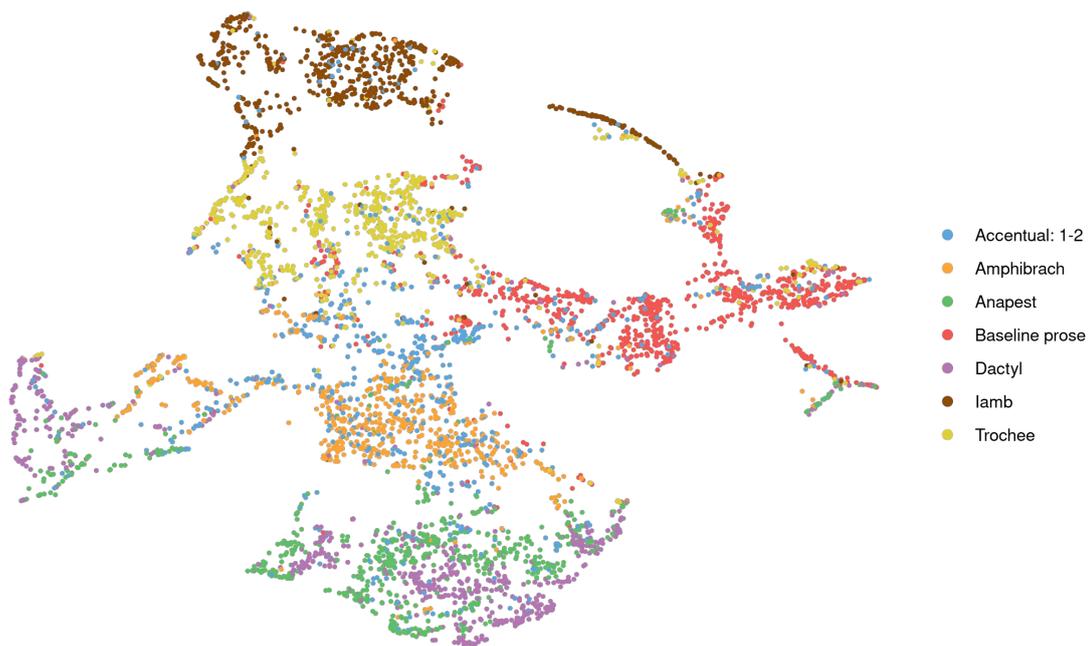


**Figure 2:** Distribution of entropy measures per different corpus subsets, arranged by median value. Red boxplots mark all classic accentual-syllabic meters, whiskers correspond to 95% confidence interval.

diversity of inter-stress intervals compared to all other accentual-syllabic meters. Partly this is a natural outcome of the measurement: trochee allows even-length intervals from anacrusis (unlike iamb), which can inflate entropy. However, it is important to note that Russian trochee can allow more metrical freedom than other meters, because of its historical source in folksongs. The perceived irregularity of the source material became a stylistic marker for folksong imitations that were often rendered in trochaic meters.

There is another potential problem with our observations—the length of the line. As one might expect, it is hard to cram a long unstressed interval into a short line. To account for this, we can formalize regularity differences across categories by building a linear model that estimates mean entropy for each group, conditioned on the average length of lines in a poem (for more details, see Appendix C). Posterior predictions made for the global average length across corpora are fully consistent with the observed trend. Based on this model, the difference that we are seeing does not come from difference in length of lines.

It is also possible to use raw probabilities of individual poem intervals to perform recognition of classic meters if every text is put in the same feature space. Figure 3 shows a UMAP [12] projection of a subset of our data, with each point corresponding to a single poem. Even with noisy, imperfect scansion, the potential for clustering is evident (*k*-means clustering with *k* set to the number of meters shows an Adjusted Rand Index of 0.6, suggesting a decent clustering force at the scale of individual poems).



**Figure 3:** UMAP projection of interval data using 5000 texts. Each point corresponds to a single poem. Prose and ‘Accentual,1-2’ were added for comparison. Adjusted Rand Index for  $k$ -means clustering for data on this plot is 0.41.

## 4. Discussion

Our entropy-based, continuous measure of rhythm regularity is able to adequately describe differences in verse organization at the level of individual poems. It is also independent from language and versification system, and could be used to compare poetic forms across languages and across meters—as long as the scansion is available. Of course the nature of this measurement makes it more suitable for accentual-syllabic foot-based verse, as that style has the greatest influence on the distribution of allowed inter-stress intervals. As we have seen in Figure 2, it is hard to distinguish pure accentual and free verse; purely syllabic poems will most likely show a similar range of entropy values. To account for this, regularity measurement would need to include both the regularity of stressed/strong positions (for tonic verse) and the regularity of line lengths (for syllabic verse)—but introducing additional dimensions would make comparisons on a single, continuous scale meaningless. There is little point in asking whether tonic verse is ‘more regular’ than syllabic.

We hope that this measure might be used for cross-linguistic comparisons, but its simplicity makes it especially useful in unseen and unstructured data scenarios, where expert annotators are unlikely to be available. This includes collections of self-published poetry, rap, and song lyrics. The regularity scale can be also used to answer long-standing questions about tendencies that are hard to see: e.g. whether all ‘free verse’ is completely free, or if some authors prefer

it with a bit of regularity (like micropolymetry)? We have seen that it is almost impossible to distinguish free verse from quasi-poems that were cut from prose; however, these quasi-poems have much more regulated line length per text, and, under certain conditions, visibly higher average entropy than free verse (see Appendix C).

One of the key features of this regularity measure is that it does not depend at all on the art of scansion. Imperfect, automated rhythmical annotation provides enough information both for entropy scores to make sense, and for meters to form distinct clusters. This reminds us that the rhythm of a poem as an empirical sequence of sounds is not equal to (and not always dependent on) meter as a cultural idea. By distinguishing the two concepts, we can focus our efforts on improving the rhythm annotation: deriving metrical labels is a parallel, and often much more costly, process.

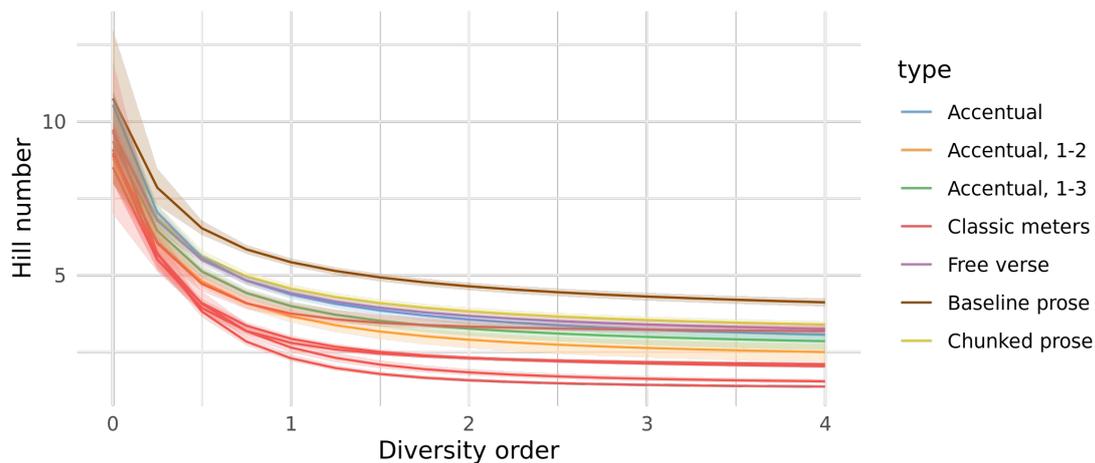
## Acknowledgments

AŠ was supported by the “Large-Scale Text Analysis and Methodological Foundations of Computational Stylistics” project (SONATA-BIS 2017/26/E/HS2/01019). We are deeply grateful to Benjamin Nagy and Petr Plecháč for their advice and discussion of the early drafts of the paper. We also would like to thank two anonymous reviewers for their helpful and thoughtful feedback.

## References

- [1] A. Anttila and R. Heuser. “Phonological and Metrical Variation across Genres”. In: *Proceedings of the Annual Meetings on Phonology 3.0* (2016). DOI: 10.3765/amp.v3i0.3679.
- [2] J. Bailey. “Literary Usage of a Russian Folk Song Meter”. In: *The Slavic and East European Journal* 14.4 (1970), pp. 436–452. DOI: 10.2307/305737.
- [3] K. Bobenhausen and B. Hammerich. “Métrique littéraire, métrique linguistique et métrique algorithmique de l’allemand mises en jeu dans le programme Metricalizer2”. In: *Langages* 199.3 (2015), pp. 67–88. DOI: 10.3917/lang.199.0067.
- [4] M. De Sisto. “The interaction between phonology and metre. Approaches to Romance and West Germanic Renaissance metre”. PhD Thesis. Radboud University, 2020.
- [5] M. L. Gasparov. “Russkii narodnii stikh i ego literaturnyie imitatsii”. In: *Izbrannyye trudy*. Vol. Iii. Moscow: Iazyki russkoi kul’tury, 1997, pp. 51–131.
- [6] M. Gasparov. *A History of European Versification*. Oxford, New York: Oxford University Press, 1996.
- [7] E. Grishina, K. Korchagin, V. Plungian, and D. Sichinava. “Poeticheskii korpus v ramkah NKRIA: obschaia struktura i perspektivy ispolzovania”. In: *Natsionalnii korpus russkogo iazyka: 2006-2008. Novye rezultaty i perspektivy*. St. Petersburg: Nestor-Istoria, 2009, pp. 71–113.

- [8] T. Haider. “Metrical Tagging in the Wild: Building and Annotating Poetry Corpora with Rhythmic Features”. In: *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*. Online: Association for Computational Linguistics, 2021, pp. 3715–3725. DOI: 10.18653/v1/2021.eacl-main.325.
- [9] M. O. Hill. “Diversity and Evenness: A Unifying Notation and Its Consequences”. In: *Ecology* 54.2 (1973), pp. 427–432. DOI: 10.2307/1934352.
- [10] M. Kestemont, F. Karsdorp, E. d. Bruijn, M. Driscoll, K. A. Kapitan, P. Ó. Macháin, D. Sawyer, R. Sleiderink, and A. Chao. “Forgotten books: The application of unseen species models to the survival of culture”. In: *Science* 375.6582 (2022), pp. 765–769. DOI: 10.1126/science.abl7655.
- [11] M. Martin. *The Rise and Fall of Meter: Poetry and English National Culture, 1860–1930*. Princeton University Press, 2012. DOI: 10.23943/princeton/9780691152738.001.0001.
- [12] L. McInnes, J. Healy, and J. Melville. “UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction”. In: *arXiv:1802.03426 [cs, stat]* (2020). URL: <http://arxiv.org/abs/1802.03426>.
- [13] A. Mittmann, P. H. Pergher, and L. d. S. Alckmar. “What Rhythmic Signature Says About Poetic Corpora”. In: *Quantitative Approaches to Versification*. 2019, p. 20.
- [14] P. Plecháč and D. Birnbaum. “Assessing the Reliability of Stress as a Feature of Authorship Attribution in Syllabic and Accentual Syllabic Verse”. In: *Quantitative Approaches to Versification*. 2019, pp. 201–210.
- [15] M. Ponomareva, K. Milintsevich, E. Chernyak, and A. Starostin. “Automated Word Stress Detection in Russian”. In: *Proceedings of the First Workshop on Subword and Character Level Models in NLP*. Copenhagen, Denmark: Association for Computational Linguistics, 2017, pp. 31–35. DOI: 10.18653/v1/W17-4104.
- [16] J. Porter. *The Space of Poetic Meter – Stanford Literary Lab*. 2018. URL: <https://litlab.stanford.edu/hooddistance/>.
- [17] J. de la Rosa, Á. Pérez, M. de Sisto, L. Hernández, A. Díaz, S. Ros, and E. González-Blanco. “Transformers analyzing poetry: multilingual metrical pattern prediction with transformer-based language models”. In: *Neural Computing and Applications* (2021). DOI: 10.1007/s00521-021-06692-2.
- [18] J. d. l. Rosa, Á. Pérez, L. Hernández, S. Ros, and E. González-Blanco. “Rantanplan, Fast and Accurate Syllabification and Scansion of Spanish Poetry”. In: *Procesamiento del Lenguaje Natural* 65.0 (2020), pp. 83–90. URL: <http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/6280>.
- [19] A. Šeļa, P. Plecháč, and A. Lassche. “Semantics of European poetry is shaped by conservative forces: The relationship between poetic meter and meaning in accentual-syllabic verse”. In: *Plos One* 17.4 (2022), e0266556. DOI: 10.1371/journal.pone.0266556.
- [20] M. Shapir and M. Shapir. “Metrum et rhythmus sub specie semioticae”. In: *Universum versus*. Vol. 1. Moscow: Iazyki russkoi kul'tury, 2000, pp. 91–128.



**Figure 4:** Hill numbers up to diversity order 4 for different corpus subsets.

- [21] O. Sobchuk and E. Lekarevitch. *Corpus of Russian narrative prose of the 19th century*. 2020. DOI: 10.31860/openlit-2020.10-C004.
- [22] A. Starostin and I. Pilshchikov. “Automated Analysis of Poetic Texts and the Problem of Verse Meter”. In: *Current Trends in Metrical Analysis*. Vol. 2. Littera. 2011, pp. 133–140.
- [23] K. Taranovskii. “Rannie russkie iamby i ikh nemetskie obraztsy”. In: *O poezii i poetike*. Moscow: Iazyki russkoi kul'tury, 2000, pp. 283–290.

## A. Code and data

The data pre-processing and analysis pipeline is openly available in a repository: [https://github.com/perechen/verse\\_regularity](https://github.com/perechen/verse_regularity)

## B. Rhythm diversity, Hill numbers

Figure 4 displays the decay of Hill numbers (diversity curves) for corpus subsets up to diversity order ( $q$ ) 4.  $q = 0$  corresponds to the simple number of types;  $q = 1$  to Shannon diversity (our regularity measure);  $q = 2$  to Simpson’s index. Higher positions on a curve signal higher relative diversity. We see that classical meters have the least diversity of unstressed intervals, followed by accentual variants and then followed by the highest, and distinct, curve for prose. The only exception is the behavior of trochee that displays a diversity of intervals (after  $q = 1$ ) approaching accentual verse and quasi poems. Lines show average Hill numbers for 1000 bootstrapped subsets (100 texts in each run), shaded areas correspond to 95% interval.

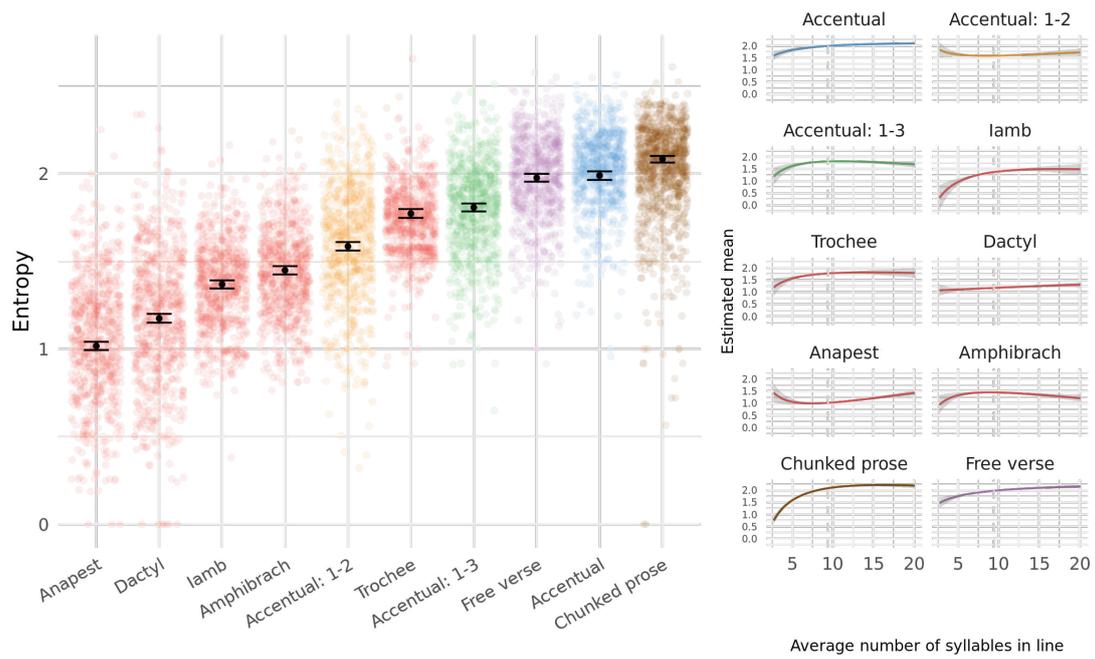
## C. Regularity and line length: posterior predictions

To formally model regularity difference between corpus subsets given the relationship entropy  $\sim$  line length, we build a Bayesian regression model using the ‘brms’ interface in R, where we estimate the average entropy ( $R$ ) for each group of texts ( $C$ ), conditioned on poem’s average line length in syllables ( $S$ ) within each group (interaction). We add a quadratic term for  $S$ , since the  $R \sim S$  relationship differs across meters and is not linear. We also log-transform  $S$  to have better predictions in the presence of outliers. In ‘brms’ formula notation (modeling  $R$  using normal distribution):

$$R \sim C * (S + I(S^2))$$

The left side of Figure 5 shows posterior predictions for fixed global average line length (dotted lines on the right side plot). Since there is an interaction between each subset and length, each prediction for each group uses its own scale. Estimates are consistent with the empirically observed differences in Figure 2, with one exception: quasi-poems at the average corpus line length show greater average entropy than accentual and free verse. Note that while some forms show a slight increase in entropy as line lengths increases, the relationship is not clear in many cases, except for free verse and chunked prose that are unconstrained by metrical regularities.

Additionally, we omit prose samples from modeling, because the division to lines was absent from them, which means that ‘line length’ was just equal to the length of a paragraph (in syllables). This would introduce unreasonable predictions for unobserved lengths, like iambic poems with an average line length of, say 400 syllables (200-foot iamb! iambic diakosiameter!). The right side of the Figure 5 already shows wide confidence intervals for extremely short and extremely long lengths for which we don’t have many (if any) observations.



**Figure 5:** Left: posterior estimates (black) for corpus subsets superimposed on empirical data. Predictions are made for global average line length (9.4 syllables). Right: posterior relationships between entropy and line length for each group. Vertical line marks the point that was used for predictions on the left.