Computational Linguistics and Intellectual Technologies:
Proceedings of the International Conference “Dialogue 2020”

Moscow, June 17–20, 2020

# QUANTILE-BASED APPROACH TO ESTIMATING COGNITIVE TEXT COMPLEXITY

**Eremeev M. A.** (maks5507@yandex.ru)
Lomonosov Moscow State University (MSU), Moscow, Russia

**Vorontsov K. V.** (vokov@forecsys.ru)
Moscow Institute of Physics and Technology (MIPT),
Moscow, Russia

This paper introduces an approach to measuring the cognitive complexity of texts on various language levels. While standard readability indices are based on the linear combination of primary statistics, our general approach allows us to estimate complexity on morphological, lexical, syntactic, and discursive levels. Each model is defined by the tokens for the specific language level and the complexity function of a single token. We then use the reference collection of moderately complex texts and the quantile-based approach to spot the abnormally rare tokens. The proposed supervised ensemble, based on the ElasticNet model, incorporates models from all language levels. Having collected a labeled dataset through crowdsourcing, consisting of pairs of articles from the Russian Wikipedia, we consider several models and ensembles and compare them to common baselines. Suggested models are flexible due to the freedom in choosing the reference collection. The described experiments confirm the competitiveness of the proposed approach, as the ensembles demonstrate the best target metric value.

**Key words:** cognitive complexity, language levels, ElasticNet, supervised learning, exploratory search

**DOI:** NN.NNNNN/ANNNNNNNANNNNNNN-N

# КВАНТИЛЬНЫЙ ПОДХОД К ОЦЕНИВАНИЮ КОГНИТИВНОЙ СЛОЖНОСТИ ТЕКСТА

**Еремеев М. А.** (maks5507@yandex.ru)
Московский Государственный Университет
им. М. В. Ломоносова (МГУ), Москва, Россия

**Воронцов К. В.** (vokov@forecsys.ru)
Московский Физико-Технический
Институт (МФТИ), Москва, Россия

В данной статье описан подход к оцениванию когнитивной сложности текста на разных уровнях языка. В отличие от индексов удобочитаемости, которые основаны на линейной комбинации текстовых статистик, мы предлагаем обобщенный подход, поззволяющий оценивать сложность на морфологическом, лексическом, синтаксическом и дискурсивном уровнях языка. Мы используем референтный корпус текстов и квантильный подход для определения токенов с аномальной частотой. Собрав выборку размеченных пар документов русской Википедии, мы также обучаем и исследуем линейную комбинацию моделей со всех уровней языка. Приведенные в статье результаты экспериментов показывают конкурентоспособность предложенного подхода.

**Ключевые слова:** когнитивная сложность, уровни языка, обучение с учителем, разведочный поиск

## 1. Introduction

Automated text complexity measurement tools have been proposed in order to help teachers to select textbooks that correspond to the students' comprehension level and publishers to explore whether their articles are readable. Thus, plenty of readability indexes (RIs) were developed. Readability indexes focus on estimating complexity by evaluating aggregated syntactic and lexical features of the whole texts. There are many well-known RIs, such as *Automated Readability Index* [13], *Flesch-Kincaid readability tests* [7], *Gunning fog* [8] and fairly modern ones like *Linsear Write Formula* [11]. They all use statistics like the total number of words, mean number of words per sentence, or the number of syllables to evaluate how complex given text is. By combining these statistics, RIs assign the given document a *complexity score*. For instance, an Automated Readability Index (ARI) has the following form for the document $d$:

$$ARI(d) = 4.71 \times \frac{c}{w} + 0.5 \times \frac{w}{s} - 21.43, \tag{1}$$

where $c$ refers to the total number of letters in document $d$, $w$ is the total number of words, and $s$ denotes the total number of sentences in $d$.

RIs are interpretable and easy to implement. However, due to the significant amount of constants, they are language-dependent and, most of the time, tailored to the US grade level system. That restrains the number of possible applications a lot.

As for research on complexity estimation of the Russian text, it is worth highlighting works of I. Oborneva [12], where she derives new version of *Flesch Readability Ease* (FRE) [10], customized for the Russian language.

$$FRE(d) = 206.836 - (1.52 \times ASL) - (65.14 - ASW), \tag{2}$$

where *ASL* stands for the mean number of words per sentence, *ASW*—for mean syllables per word. In 2018 V. Solovyev [14] obtains new readability formula created explicitly for Russian documents. Text complexities are valuable in different areas, e.g., [4] describes complexity formulas for legal documents in Russian.

In 2007 [1] introduced psychophysiological (cognitive) methods of measuring text complexity, highlighting the following assumptions:

1. Any text can be considered as a sequence of tokens (codes)—parts of the finite alphabet—letters, syllables, sentences, words, etc.
2. When reading the text, our nervous system decodes the tokens, progressively on the following language levels: morphological, lexical, syntactic, discursive, and semantic.
3. Decoding processes occur in different nervous system zones (e.g., part of the cortex). Each zone is responsible for the specific token on a specific language level. When the zone finishes the decoding process, it moves into the state of refractoriness and needs time to recover. During the recovery, the zone cannot execute decoding and forces another zone to take the load. Such a redistribution of nervous system resources diminishes effectiveness of the nervous system as a whole, and the person starts perceiving the document with more effort.
4. Thus, if the token's distance to the previous occurrence exceeds some threshold, the nervous system must allocate additional resources to decode it. Such terms are considered complex. Hence, the complexity of the document is a combination of abnormally frequent (complex) tokens.

In [1], authors propose to count the mentioned threshold as a quantile of the empirical distribution, calculated over the large set of simple texts (*reference collection*). They explore the morphological level, considering letters as a token. [2] introduces a lexical level model, assuming the word complexity is determined only by its length. [19] features the model on the discursive level, counting the number of connector words and phrases in each sentence.

Based on the assumptions above, in this paper we elaborate our research presented in [6], offering models on the morphological, lexical, and syntactic levels, and then training the linear model, obtaining the all-levels complexity model. Experiments were performed on two datasets. We compare models with readability indexes and cognitive models proposed in [1], [2], [19].

## 2. General Model

Let $d$ be the arbitrary document, consisting of tokens $x_1, \ldots, x_n$ from a fixed token alphabet $A_h$. Here, $h$ refers to the language level, i.e., morphological, lexical, syntactic, or discursive. So, the tokens may be letters, syllables, sentences, words, etc. We denote $c_i$ to be the *cognitive complexity score* of token $x_i$, and $w_i$—its weight. The *document complexity score* then is a sum of weights over tokens having abnormal complexities.

To measure the token complexity, we use a reference collection—a set of moderately complex texts—to calculate empirical distributions of complexity scores for each token. Thus, the token's complexity is abnormal when it is greater than a $\gamma$-quantile of the counted distribution (figure 1) (assumption 4). In our experiments we use *Russian Wikipedia* and *Noosphere* (noosphere.ru) open corpora as reference collections. Former comprises more domain-specific documents (1.5M in total), while the latter incorporates various types of texts, including fiction and poems (200K in total).

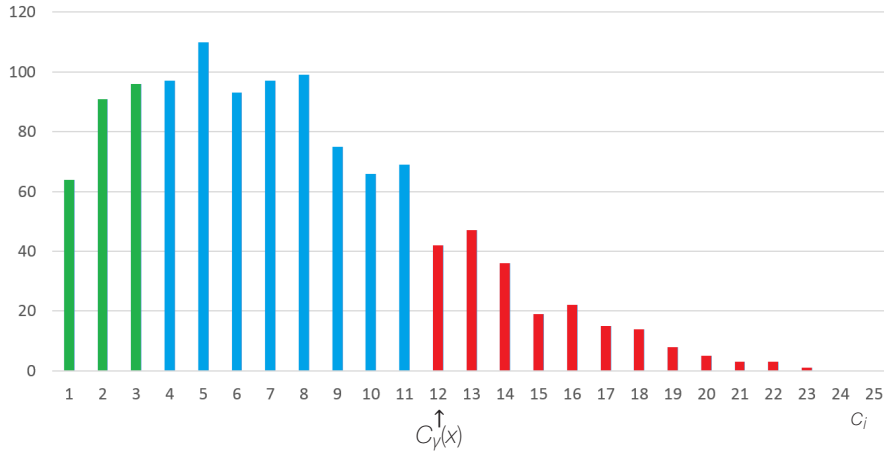


**Figure 1:** Sample distribution of complexity scores and its $\gamma$-quantile

Finally, document $d$ complexity $W(d)$ is calculated by aggregating complexity scores $c_i$ of complex tokens in $d$.

$$W(d) = \sum_{i=1}^{n} w_i \left[c_i > C_\gamma(x_i)\right], \tag{3}$$

where [ ] refers to the Iverson notation (i.e. $[true] = 1$, $[false] = 0$), $n$ is the number of tokens from $A_h$ in document $d$. Some examples of interpretable weights $w_i$ are presented in table 1.

Raising the weight to the $p$ power, we obtain a nonlinear sum of weights:

$$W(d) = \sum_{i=1}^{n} w_i^p \left[c_i > C_\gamma(x_i)\right], \tag{4}$$

where $p > 0$ is an integer power.

If the token $x_i$ does not appear in the reference collection, we set $C_\gamma(x_i)$ equal to $-\infty$, therefore always counting it as abnormally complex.

Thus, to set up the model, we need to specify the reference collection $D$, the alphabet $A_h$, token complexity function $c$, weights $w$, and power $p$.

**Table 1:** Weights $w_i$ examples

| $w_i$ | Meaning of $w_i$ |
|---|---|
| 1 | number of complex tokens |
| $1/n \times 100\%$ | complex tokens percentage |
| $c_i$ | total complexity |
| $c_i/n$ | mean complexity |
| $c_i - C_\gamma(x_i)$ | excessive complexity |
| $(c_i - C_\gamma(x_i))/n$ | mean excessive complexity |

## 3. Token complexity functions

Firstly, we indicate two approaches to estimating the complexity of a single token.

### 3.1. Distance-based complexity function

According to assumptions 3–5, let $r_i$ be a distance from previous token occurrence $x_i$ to its current occurrence in the text:

$$\dots \quad \boxed{x_{i-r_i} = a} \quad \underbrace{x_{i-r_i+1} \quad x_{i-r_i+2} \quad \dots \quad x_{i-2} \quad x_{i-1}}_{r_i} \quad \boxed{x_i = a} \quad \dots$$

Equally,

$$r_i = \min_{1 \le j < i} \{i - j \mid x_i = x_j\}. \tag{5}$$

If $i$ is the first occurrence of term $t_i$ in document $d$, there is no previous occurrence, so $r_i$ is undefined. To solve this issue. we redefine $r_i$ so that sum of $r_i$ over all tokens $x_i = a$ is equal to $n$.

For example, if $A_h$ consists of the letters:

**Table 2:** $r_i$ and redefined $r_i$ examples for letter-based model

| token | t | h | e | g | r | e | a | t | g | a | t | s | b | y |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $r_i$ | — | — | — | — | — | 3 | — | 7 | 5 | 3 | 3 | — | — | — |
| redefined $r_i$ | 4 | 15 | 11 | 9 | 14 | 3 | 11 | 7 | 5 | 3 | 3 | 14 | 14 | 14 |

Then, we define token complexity function as some decreasing function $f$ of $r_i$:

$$c_i = f(r_i) \tag{6}$$

The $f$ should be decreasing according to the assumption 4, as only the most frequent terms put pressure on the nervous system. Example of $f$:

$$c_i = -r_i, \tag{7}$$

Hence, we build an empirical distribution of complexities $\{f(r_i)|x_i{=}a\}$ for all tokens $a \in A_h$, count corresponding quantiles $C_\gamma(x_i)$ and, finally, calculate the complexity score, according to formula (4).

### 3.2. Counter-based complexity functions

In the counter-based approach, we assume every term has fixed complexity score (not depending on position in the text), so alphabet $A_h$ includes the only token: $A_h = \{a\}$. In other words, the token's complexity is defined only by its linguistic properties (e.g., length of the word or sentence).

Taking that into account, we construct single empirical distribution over all tokens. Therefore, the quantile is one for all tokens $C_\gamma(x_i) = C_\gamma$ and model (4) takes the following form:

$$W(d) = \sum_{i=1}^{n} w_i^p \, [c_i > C_\gamma] \tag{8}$$

## 4. Considered models

Trying different combinations of tokens and complexity functions, we want to share models on four language levels.

### 4.1. Morphological complexity models

At the morphological level, tokens are letters, morphemes, syllables, or, in general case, $n$-grams. Also, we can sort the letters in $n$-gram, therefore lessening the vocabulary size to acquire more reliable distributions. Indeed, our brain easily handles local letter permutations, so they do not affect the complexity much.

In our experiments, we use a distance-based model with complexity function (7) for both *letters, sorted and unsorted syllables*.

The examples of empirical distributions for letter-based models over the Russian Wikipedia and Noosphere reference collections are introduced in **figure 2**. Comparison of the distributions for syllables-based and sorted-syllables-based models are presented in **figure 3**.

### 4.2. Lexical complexity models

Here we use separate words as tokens. However, in such a case, the vocabulary turns out to be vast and makes the distributions less precise. To shrink it, we eliminate all short words (less than a length of 3) and too rare words (that appears only once on the whole reference collection).

#### 4.2.1. Distance-based model

The distance-based complexity model uses complexity function (7) as it calculates the distributions of the score for every word (*lexical distance model*). The example of the distribution is shown in **figure 4**.

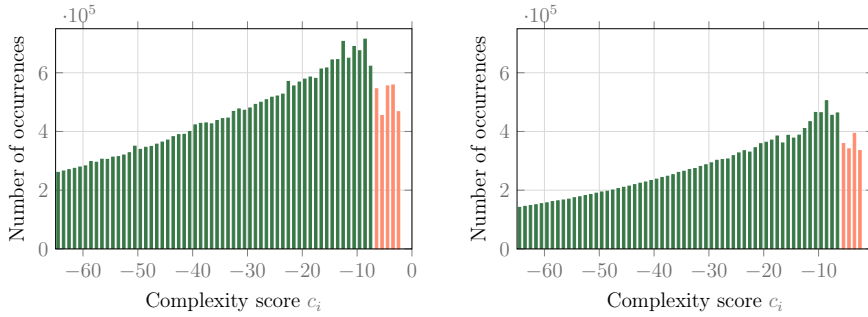


**Figure 2:** Distribution of $c_i$ for the letter «У», calculated over the Russian Wikipedia and Noosphere collections. The orange part of the distribution correspond to $c_i > C_\gamma(x)$, $\gamma = 0.95$

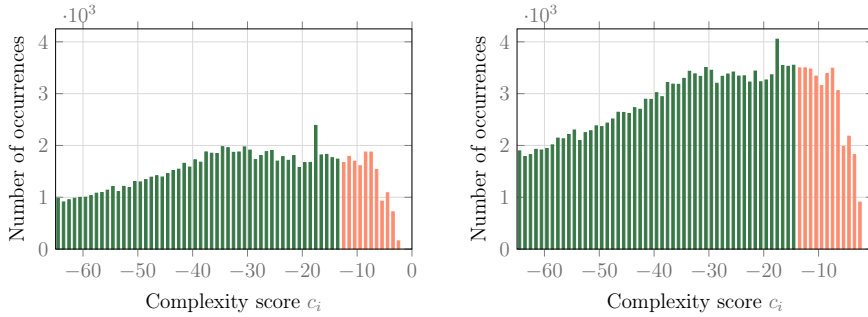


**Figure 3:** Distribution of $c_i$ for the syllable «ЛОК», calculated over the Russian Wikipedia collection for models with and without sorting. The orange part of the distribution corresponds to $c_i > C_\gamma(x)$, $\gamma = 0.95$

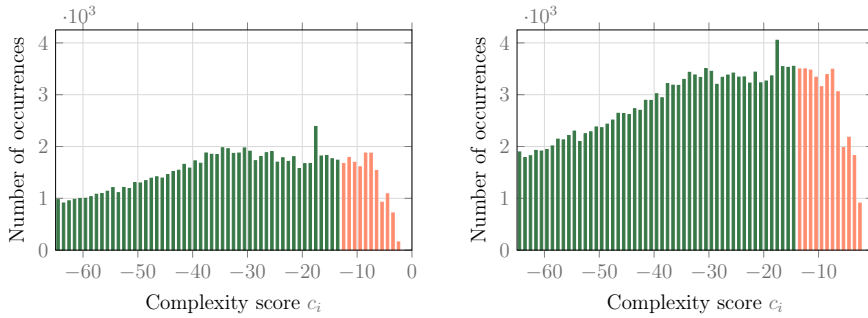


**Figure 4:** Distribution of $c_i$ for the word «МАТЕМАТИКА», calculated over the Russian Wikipedia collection. The orange part of the distribution corresponds to $c_i > C_\gamma(x)$, $\gamma = 0.95$

#### 4.2.2. Counter-based models

We explore two functions here. Firstly, [2] defines the complexity of the word as its length (*lexical length model*). Therefore, the model builds empirical distribution over all words' lengths and counts the word as complex if it is long enough.

Advancing this approach, we consider not the word length, but its counter value $count(x_i)$, which is the number of times word $x_i$ appears in reference collection (*lexical counter model*). The complexity function should be a decreasing function of $count(x_i)$. For example:

$$c_i = -count(x_i) \tag{9}$$

### 4.3. Syntactic complexity models

To estimate syntactic complexity, we use UDPipe [15] to extract syntactic dependencies, part of speeches (noun, verb, adjective, etc.) and sentence parts (subject, object, attribute, etc.). Using derived information, we propose two models.

#### 4.3.1. Distance-based model

Let $A_h$ be a product of $PoS$—set of all parts of speech may occur, and $SP$—set of all sentence parts. Therefore each $a \in A_h$ is a pair $(p, s)$, where $p \in PoS$ and $s \in SP$ are part of speech and sentence part respectively. We call such pairs *syntgams*.

We apply the distance complexity function (7) to such tokens to receive a distance-based syntactic model (*syntactic syntgam model*).

#### 4.3.2. Counter-based model

Using the syntactic dependencies returned by the parser, we define the complexity function as a length of the dependency (alike using word length [2]) and acquire the counter-based syntactic model (*syntactic length model*). The examples of distributions are shown in **figure 5**.

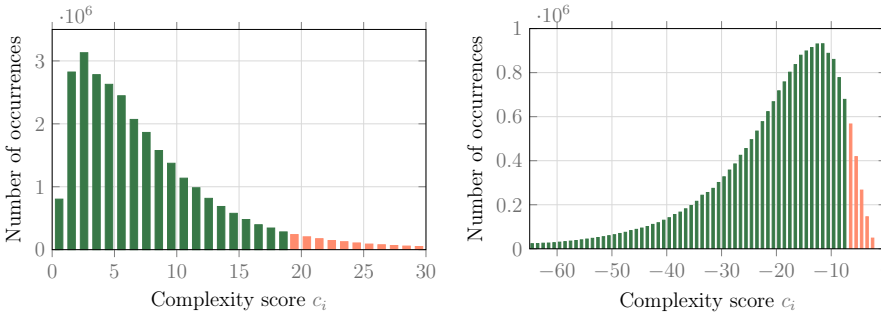


**Figure 5:** Distribution of syntactic dependencies' lengths and distance-based $c_i$ for syntgam (verb, root), calculated over the Russian Wikipedia dataset. The orange part of the distribution corresponds to $c_i > C_\gamma(x)$, $\gamma = 0.95$

### 4.4. Discursive complexity models

The last but not least language level we consider is the discursive level, initially proposed in [19]. On this level, model evaluates the meaningfulness of text, its coherence, and consistency.

To evaluate the complexity the vocabulary of common connector-words for the Russian language (i.e., «который», «из-за того что», «с тех пор как», etc.) is used. Thus, the more such connectors appear in the document, the more complex it is.

Therefore, we define a counter-based model with sentences as tokens, and complexity function equal to the number of connectors in the sentence *(discursive connectors model)*.

## 5. Dataset

We used a crowdfunding platform Yandex.Toloka to gather a labeled dataset of pairs of Russian Wikipedia pages.

Assessors were asked to label 10K pairs of Russian Wikipedia articles. We suggested them to read both pages carefully and choose which is more challenging to comprehend. The interface consisted of two links to evaluated articles and four options to choose from: "LEFT" or "RIGHT" when an assessor assumes the left or the right document is more complex, "EQUAL" in case the assessor cannot determine which document is more challenging to comprehend and "INVALID" option if the documents in given pair lie in different domains. The interface is shown in **figure 6**.

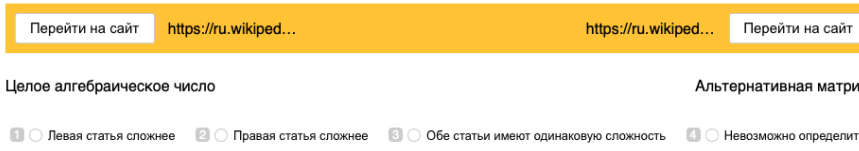


**Figure 6:** Interface for assessors at Yandex.Toloka

We chose documents from math, physics, medicine, and programming areas. The topic modeling approach [9], namely the Additive Regularization of Topic Models (ARTM) theory [16], was used to cluster the documents by fields. ARTM features an effective way to build structured multimodal topic models [17], [3]. We included the modalities of words and word collocations, obtained with TopMine algorithm [5]. Then, documents from a single domain and with almost identical lengths formed the pairs. Examples of document pairs to assess are introduced in **table 3**.

Each pair was labeled by two assessors to avoid human factor mistakes. We assume that the pair was correctly labeled if labels were not controversial, i.e., one assessor labeled the first document as more complex while others chose the second document. If the pair was labeled as 'INVALID' at least by a single person, we also eliminated it from the final dataset.

**Table 3:** Examples of labeled document pairs

| Left Document | Right Document | Which document is more complex |
|---|---|---|
| Matrix | Tensor | RIGHT |
| Rational number | Fraction (mathematics) | LEFT |
| Proton | Neutron | EQUAL |
| Mac OS X | Convex Hull | INVALID |

So, 8K pairs out of 10K were correctly labeled and formed the dataset

$$D = \{(d, d') \mid d' \text{ is more complex than } d)\}.$$

To shorten the calculations and formulas, let's denote $(d, d') \in D$ as $d \prec d'$.

## 6. Ensembling models

Having the dataset, we can train a supervised model to piece together all the proposed models. Such an ensemble combines estimations from all language levels.

We chose a linear combination to be the resulted model:

$$W(d, \alpha) = \sum_{k=1}^{K} \alpha_k W_k(d), \quad \alpha_k \geq 0, \tag{10}$$

where vector $\alpha$ is the solution to the optimization problem:

$$\sum_{d \prec d'} \mathscr{L}\Big(\underbrace{W(d', \alpha) - W(d, \alpha)}_{\text{pair-wise margin}}\Big) \to \min_{\alpha}, \tag{11}$$

where $\mathcal{L}(M)$ is a smooth, non-increasing function of margin $M$.

To avoid overfitting, we use ElasticNet [18] method of combining L1 and L2 regularizes:

$$\frac{1}{2|D|} \sum_{d \prec d'} \mathscr{L}\big(W(d', \alpha) - W(d, \alpha)\big) + \lambda \left( (1 - \beta) \sum_{k=1}^{K} \alpha_k^2 + \beta \sum_{k=1}^{K} |\alpha_k| \right) \to \min_{\alpha}, \tag{12}$$

where $\beta$ is a mixing parameter between ridge ($\beta = 0$) and lasso ($\beta = 1$), $\lambda$ controls the regularization impact.

For $\mathcal{L}$ function we consider three options:

- **Negative SE**: $\mathcal{L}(M) = -M^2$
- **Negative sigmoid**: $\mathcal{L}(M) = -\sigma(M)$, where $\sigma(x) = 1/(1 + \exp x)$—sigmoid function
- **Negative AE**: $\mathcal{L}(M) = -|M|$

The results of testing all models above and the ensemble are described in the Experiments section.

## 7. Experiments

We tested every model and the ensemble trained on the dataset mentioned above. For all experiments, we used Wikipedia as a reference collection. The accuracy score was selected as a quality metric.

$$\text{accuracy(c)} = \frac{\sum\limits_{d \prec d'} [c(d) < c(d')]}{|D|} \tag{13}$$

To validate the ensembles, we preliminarily split the dataset into train $D_{train}$ and test $D_{test}$ parts, so having 6K training objects and 2K testing.

### 7.1. Single models

We compare all aforementioned quantile-based models to various readability indexes and baselines proposed in [1], [2] and [19]. As for hyperparameters, we used $w_i = c_i / n$ (for text length not to affect the scores), $p = 1$, and $\gamma = 0.95$ for all models proposed. The results are exposed in table 4.

**Table 4:** Comparison of readability indexes performance to proposed models

| Model Class | Model | Accuracy |
|---|---|---|
| Readability Indexes | Automated Readability Index | 50.5% |
| | Flesch-Kincaid Grade | 44.7% |
| | Gunning FOG | 44.4% |
| | Flesch Reading Ease | 50.7% |
| | Dale–Chall | 37.0% |
| | Linsear Write | 45.2% |
| | Coleman-Liau | **52.1%** |
| Morphological | Letter [1] | 63.7% |
| | Syllables | 70.9% |
| | Sorted Syllables | 73.1% |
| Lexical | Length [2] | 42.4% |
| | Distance | **75.0%** |
| | Counter | 71.2% |
| Syntactic | Length | 62.0% |
| | Syntgam | 64.2% |
| Discursive | Connectors [19] | 62.5% |

The lexical distance model demonstrates the best performance in terms of accuracy among all the described models. Moreover, all quantile-based models, except for lexical distance one, outperform readability indexes. The sorted-syllables model performs better than unsorted, which proves the assumption about the sustainability of distributions in the sorted-syllables model.

### 7.2. Ensembles

**Table 5:** Comparison of ensembles with different margin functions to the best models on different language levels

| Model | Margin Function | Accuracy |
|---|---|---|
| Coleman-Liau | — | 52.1% |
| Morphological Sorted Syllables | — | 73.1% |
| Lexical Distance | — | 75.0% |
| Syntactic Syntgams | — | 64.2% |
| Connectors | — | 62.5% |
| Ensemble | Negative SE | **88.1%** |
| Ensemble | Negative sigmoid | 84.6% |
| Ensemble | Negative AE | 85.1% |

To validate ensembles trained on $D_{train}$, we first evaluate all models on $D_{test}$ part of the dataset to get comparable results. In table 5, we compare the best models from all language levels with ensembles with various margin functions. We set the hyperparameters equal $\beta = 0.5$ and $\lambda = 10$ for all models.

As can be seen, Negative SE works best for fitting an ensemble, while all ensembles demonstrate quality growth compared to other models.

### 7.3. Noosphere Reference Collection

Here we explore the impact of the reference collection on the models' performance. We fitted the models with Noosphere corpora as a reference collection. This collection is less scientific and formal, featuring diverse literary works. We still evaluate the models on the labeled dataset, introduced in Section 5. The results are exposed in table 6.

**Table 6:** Comparison of models fitted on Noosphere reference collection

| Model Class | Model | Accuracy |
|---|---|---|
| Morphological | Letter [1] | 60.3% |
| | Syllables | 69.2% |
| | Sorted Syllables | 70.5% |
| Lexical | Length [2] | 39.8% |
| | Distance | 72.1% |
| | Counter | 66.9% |
| Syntactic | Length | 63.1% |
| | Syntgam | 66.4% |
| Discursive | Connectors [19] | 60.2% |
| Ensembles | Negative MSE | **83.1%** |

All scores are lower, except for the syntactic models. There are understandable reasons for that. Firstly, the Noosphere collection is smaller than Wikipedia, resulting

in less accurate empirical distribution estimations. Secondly, the collection consists of the non-scientific documents and does not contain specialized terms. Nevertheless, syntactic models improve their performance, mainly because of the absence of formulas in the reference collection.

Overall, the ensemble's accuracy is still higher than 80%, which outperforms both the readability indices and cognitive model baselines.

## 8. Conclusion

In conclusion, we presented new quantile-based models to measure cognitive text complexity. All models are based on psychophysiological assumptions. We explored models dealing with tokens from morphological, lexical, syntactic, and discursive language levels. All complexity scores are calculated with respect to the reference collection—a set of adequately simple documents used to obtain the empirical distributions of the token complexities. The reference collection should be chosen carefully and be large enough, but it gives high flexibility to the discussed approach. By varying the reference collection, we can obtain complexity scores concerning a particular domain. We introduced the way to measure the quality of the cognitive complexity models, based on crowdsourcing. By ensembling models from various language levels, we attain an accuracy score of more than 88% and 83% using Russian Wikipedia and Noosphere reference collections, respectively. Suggested models outperform the readability indices and previously proposed cognitive complexity models.

## 9. Acknowledgements

This work is supported by the Russian Foundation for Basic Research, grant 20-07-00936.

## References

1. *A. A. Birkin:* Speech Codes. Hippocrat, Saint-Peterburg, 2007.
2. *A. A. Birkin:* Nature of Speech. Likbez, Moscow, 2009.
3. *N. A. Chirkova.:* Additive regularization for hierarchical multimodal topic modeling. Machine Learning and Data Analysis, 2:187–200, 01 2016.
4. *Aryna Dzmitryieva:* The art of legal writing: A quantitative analysis of russian constitutional court rulings. Sravnitel'noe konstitucionnoe obozrenie, 3:125–133, 01 2017.
5. *Ahmed El-Kishky, Yanglei Song, Chi Wang, Clare Voss, and Jiawei Han:* Scalable topical phrase mining from text corpora. Proceedings of the VLDB Endowment, 8, 06, 2014.
6. *M. A. Eremeev:* and Konstantin Vorontsov. Lexical quantile-based text complexity measure. In RANLP, 2019.
7. *R. Flesh:* How to test readability. New York, Harper and Brothers, 1951.
8. *Robert Gunning:* The technique of clear writing. McGraw-Hill, New York, 1952.

9. *Thomas Hofmann:* Probabilistic latent semantic indexing. In Proceedings of the 22Nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '99, pages 50–57, New York, NY, USA, 1999. ACM.
10. *J. Peter Kincaid, Robert P. Fishburne, Richard Lawrence Rogers, and Brad S. Chissom.:* Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel. 1975.
11. *William Lidwell, Kritina Holden, and Jill Butler:* Universal principles of design, revised and updated: 125 ways to enhance usability, influence perception, increase appeal, make better design decisions, and teach through design. 2010.
12. *Irina Oborneva:* Avtomatizirovannaya otsenka slozhnosti uchebnykh tekstov na osnove statisticheskikh parametrov [semiautomatic evaluation of the complexity of academic texts on the base of statistic parameters]. 2006.
13. *R. J. Senter and E. A. Smith:* Automated readability index. AMRL-TR, 66(22), 1967.
14. *Valery Solovyev, Vladimir Ivanov, and Marina Solnyshkina:* Assessment of reading difficulty levels in russian academic texts: Approaches and metrics. Journal of Intelligent and Fuzzy Systems, 34:1–10, 04 2018.
15. *Milan Straka and Jana Strakova:* Tokenizing, pos tagging, lemmatizing and parsing ud 2.0 with udpipe. pages 88–99, 01 2017.
16. *K. V. Vorontsov and A. A. Potapenko:* Additive regularization of topic models. Machine Learning, Special Issue on Data Analysis and Intelligent Optimization with Applications, 101(1):303–323, 2015.
17. *Konstantin Vorontsov, Oleksandr Frei, Murat Apishev, Peter Romov, Marina Suvorova, and Anastasia Yanina:* Non-bayesian additive regularization for multimodal topic modeling of large collections. 10 2015.
18. *Hui Zou and Trevor Hastie:* Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society, Series B, 67:301–320, 2005.
19. *В. М. Тютюнник, А. А. Биркин и Ю. Г. Гущин:* Основы лингвистической психофизиологии. изд-во МИНЦ «Нобелистика» Тамбов; М.; СПб.; Баку; Вена; Гамбург, 2016.