Научная статья

УДК 519.766.48
DOI 10.25205/1818-7900-2025-23-4-62-73

Оценка качества перевода художественного текста с амхарского на английский язык с использованием методов сжатия данных

Йешевас Гетачеу Лулу

Новосибирский государственный университет
Новосибирск, Россия

j.lulu@g.nsu.ru; https://orcid.org/0009-0006-8054-9846

Аннотация

Оценка качества перевода является важной задачей в области компьютерной лингвистики. В данном исследовании рассматривается использование методов сжатия данных для оценки точности перевода путем выявления характерных языковых закономерностей. Традиционные методы оценки перевода основаны на анализе стилистических показателей и машинном обучении, однако на эти подходы часто влияют длина текста и предопределенные лингвистические особенности. Чтобы устранить эти ограничения, мы используем теоретико-информационный метод, основанный на сжатии данных.

Наша методология использует алгоритмы сжатия для анализа перевода с целью оценки качества. Мы оцениваем неосознанный стилистический вклад переводчиков, сравнивая несколько переводов одних и тех же литературных произведений. Кроме того, мы применяем классификацию на основе сжатия, чтобы различать оригинальные тексты на амхарском языке, тексты, переведенные человеком с амхарского на английский, и тексты, переведенные компьютером. В наших экспериментах мы использовали шесть оригинальных романов на амхарском языке для анализа авторских стилей, а для оценки качества перевода – известные произведения, переведенные как переводчиками-людьми, так и компьютерными переводчиками. Среди различных алгоритмов сжатия данных без потерь были протестированы следующие: Prediction by Partial Matching (PPM), кодирование Хаффмана, преобразование Барроуза – Уилера (BWT) и алгоритм Лемпеля – Зива – Маркова (LZMA) с целью оценки их эффективности. Согласно коэффициенту V Крамера, рассчитанному по результатам различных экспериментов, алгоритм Prediction by Partial Matching (PPM) показал наивысшую стабильность и поэтому был выбран для всех последующих анализов.

Результаты показывают, что алгоритм PPM достигает наивысшей точности классификации: коэффициент Крамера (V) составил 0,89 для авторских текстов на амхарском языке, 0,762 и 1 для текстов, переведенных человеком с английского на амхарский, 0,91 для текстов, переведенных компьютером с амхарского на английский, и 0,53 для задач компьютерного перевода с английского на амхарский.

Исследование демонстрирует, что методы сжатия данных обеспечивают жизнеспособный, не зависящий от языка подход к оценке качества перевода, особенно для языков с ограниченными ресурсами, таких как амхарский. Эти результаты подчеркивают потенциал теоретико-информационных методов в лингвистическом анализе и компьютерных исследованиях перевода.

Ключевые слова

сжатие данных, перевод текста на амхарский язык, лингвистический анализ, оценка качества перевода, коэффициент Крамера

Благодарность

Автор выражает искреннюю благодарность научному руководителю, профессору Борису Рябко за ценное время, опыт и содержательные отзывы об этой исследовательской работе. Его конструктивные комментарии и продуманные предложения сыграли ключевую роль в повышении качества и ясности статьи.

Для цитирования

Лулу Й. Г. Оценка качества художественного перевода с амхарского на английский язык с использованием методов сжатия данных // Вестник НГУ. Серия: Информационные технологии. 2025. Т. 23, № 4. С. 62–73. DOI 10.25205/1818-7900-2025-23-4-62-73

Assessment of Amharic-English Literary Translation Quality Through Data Compression Techniques

Yeshewas Getachew Lulu

Novosibirsk State University
Novosibirsk, Russian Federation

j.lulu@g.nsu.ru; https://orcid.org/0009-0006-8054-9846

Abstract

Translation quality assessment are crucial challenges in computational linguistics. This study explores the use of data compression techniques to evaluate translation accuracy by identifying distinct linguistic patterns. Traditional methods for translation evaluation rely on style metric analysis and machine learning; however, these approaches are often influenced by text length and predefined linguistic features. To address these limitations, we employ an information-theoretic method based on data compression.

Our methodology utilizes compression algorithms to analyze translation of quality assessment. We assess the unconscious stylistic contribution of translators by comparing multiple translations of the same literary works. Additionally, we apply compression-based classification to distinguish between original Amharic texts, human-translated Amharic-to-English texts, and computer-translated texts. In our Experiments were conducted using six original Amharic novels for authorship styles and for translation quality assessment we utilize well-known translated works by human translator and computer translators. Among various lossless data compression algorithms, the following were tested: Prediction by Partial Matching (PPM), Huffman coding, Barrows-Wheeler Transform (BWT) and Lempel-Ziv-Markov Algorithm (LZMA), in order to evaluate their performance. According to the Cramer's V coefficient calculated from different experiments, the Prediction by Partial Matching (PPM) algorithm showed the highest stability and was therefore selected for all subsequent analyses.

Results indicate that PPM achieves the highest classification accuracy, with a Cramer coefficient (V) of 0.89 for Amharic authorship works, 0.762 and 1 for human-translated English-to-Amharic texts,0.91 for computer based translated Amharic-to-English texts and 0.53 for English Amharic computer translated tasks.

The study demonstrates that data compression techniques provide a viable, language-independent approach for translation quality assessment, particularly for low-resource languages like Amharic. These findings highlight the potential of information-theoretic methods in linguistic analysis and computational translation studies.

Keywords

Data compression, Amharic text translation, Linguistic analysis, Translation quality assessment, Cramer coefficient

Acknowledgements

We would like to express our sincere gratitude to the supervisor Professor Boris Ryabko for their valuable time, expertise, and insightful feedback on this research work. Their constructive comments and thoughtful suggestions have played a pivotal role in enhancing the quality and clarity of the manuscript.

For citation

Yeshewas Getachew Lulu. Assessment of Amharic-English Literary Translation Quality Through Data Compression Techniques. *Vestnik NSU. Series: Information Technologies*, 2025, vol. 23, no. 4, pp. 62–73 (in Russ.) DOI 10.25205/1818-7900-2025-23-4-62-73

1. Introduction

In an increasingly interconnected global landscape, translation functions as a critical conduit for cross-linguistic communication, enabling the dissemination of information across domains such as literature, journalism, film, and digital social platforms. The escalating demand for high-quality

translations has catalyzed substantial developments in both human and machine translation methodologies, thereby stimulating scholarly inquiry into their relative efficacy and fidelity [1].

Traditionally, literary scholars have evaluated translations using established analytical methods, accumulating extensive insights into both individual works and broader translation principles. In recent years, mathematical and computational approaches have emerged, offering new ways to assess translation quality [2; 3]. However, defining what constitutes a high-quality translation remains a challenge due to the subjective nature of style and interpretation.

From a computational perspective, translation quality is formulated as a classification problem that relies on distinctive linguistic features to capture an author's writing style [1; 3]. Style metric analysis, which includes vocabulary richness, word frequency distributions, and lexical repetition, is widely used for this purpose. However, as noted by Madigan et al [4], many of these metrics are highly dependent on text length, making them difficult to apply reliably. Researchers have explored alternative stylistic markers, such as word class frequencies, syntactic structures, word collocations, grammatical errors, and document structure (e.g., sentence and paragraph length) [5–8]. Additionally, machine learning techniques, including Support Vector Machines [9], Neural Networks [10], and Decision Trees [11], have been employed for authorship classification and translation quality evaluations.

To avoid the need for predefined linguistic features, some researchers have proposed using Data compression models for authorship attribution and translation quality assessment [1–3], as well as related tasks such as text categorization [12], language identification [13], genome classification, and clustering [14].

Amharic is a Semitic language spoken in North Central Ethiopia by the Amhara. It is the second most spoken Semitic language after Arabic, and the official language of Ethiopia. Amharic is also the official or working language of several of the states, including Amhara Region and the multi-ethnic Southern Nations, Nationalities, and People's Region [15]. Amharic has been spoken in Ethiopia since the late 12th century in various industries including the legal system, commerce, communications, the military and religion [16].

an additional challenge arises in translation between low-resource languages, such as Amharic to English. Unlike widely spoken languages, these translations suffer from limited linguistic data, inadequate machine translation models, and a lack of comprehensive studies on quality assessment. The scarcity of research in this area highlights the need for more systematic approaches to improve translation accuracy and preserve the nuances of both languages.

2. Methodology

In this section, we will briefly describe Ryabko et al.'s [2; 3] approach to literary text attribution and the Information-Theoretic Method for assessing translation quality. We will also explain how we use the data compression method for evaluating the quality of translations between Amharic and English in both directions, applying our analytical approach. Here below in figure 1 outline the steps to conduct this research.

2.1 Data collection and parameter selection

We began by selecting six well known Amharic novels, "*Oromai*", "Fiker Eskemekaber", "YeHelina Dewel", "Shotelay", "Keadmase Bashager", and "Emego" written by Dr. Hadis Alemayehu, Bealu Girma, Mamo Wodneh, and Alemneh Wassie respectively for the first experiment.

For our experiment, we prepared our dataset by selecting a 64kb training sample from each book. Each sample was then divided into 32 slices, each with a size of 2 KB.

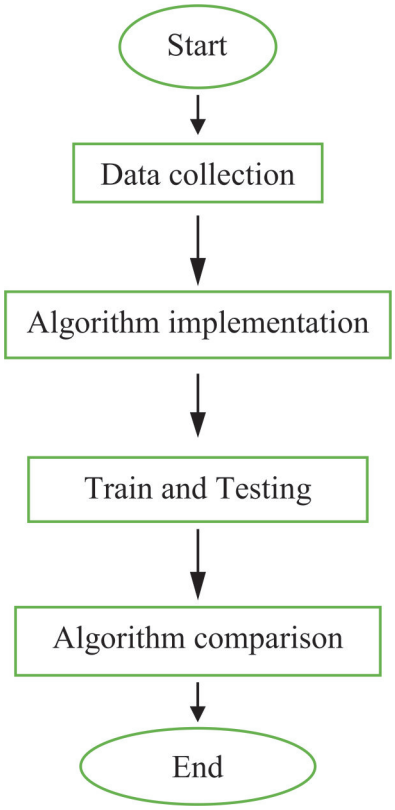

Fig. 1. Flow chart of the research stage

In the second experiment, we selected two well-known English books that had been translated into Amharic by two different translators. Specifically: the first book entitled "*Les Miserable*" by Victor Hugo, which was translated into Amharic as "*Menduban*" by Yohannes G. Meskel and "*Mekergunchu*" by Sehale Selassie Berhan and the second book "*Crime and Punishment*" by Fyodor Dostoevsky, which was translated into Amharic as "*Wenjel Ena Ketat*" by Muluberhan and "*Wenjel Ena Ferde*" by Kassa

For the third experiment, due to unavailability of one books translated from Amharic to English by several human translators we utilized three computer translator that support Amharic language two of them the common language translator google translator [17], Yandex Translator [18] and one of them is translator that support Artificial intelligence large language model and neural machines translation model that is lingvanex translator [19]. we select famous novel in Amharic entitled "*Fiker Esk mekaber*" (Love up to death) written by Dr. Hadis alemayhu for this work. For the last experiment we select an English book entitled "Born a crime" by Trevor Noah and translated to Amharic by applying the above mentioned computer translators.

2.2 Algorithm implementation and selection

In our investigation of we utilized the following modern lossless data compression algorithm: GNU ZIP (Gzip) [20], BZIP2[21], Lempel–Ziv–Markov chain algorithm (LZMA) [22], Brotli[23], Zstandard (Zstd) [24], and Prediction by partial matching (PPM)[25].

We cheek all mentioned potential modern lossless Data compressor and calculating their chi-square and crammer coefficient V in order to select the efficient algorithm for our experiment.

The Cramer coefficient V varies from 0 (corresponding to no association between the variables) to +1 (complete association). It is based on Pearson's chi-squared statistic. Here's below the steps followed to calculate Cramer's V for each contingency table for Data compressors used in the test:

1. Construct the Contingency Table:
   - Create a contingency table that shows the frequency distribution of the variables.
2. Calculate the Chi-Squared Statistic (χ^2):
   - Use the formula for Pearson's chi-squared test:

$$x^2 = \frac{\sum (Oij - Eij) 2}{Eij}.$$

where O_{ij} is the observed frequency and E_{ij} is the expected frequency under the assumption of independence.

3. Determine the Total Sample Size (N):
   - Sum all the observed frequencies in the table to get N.
4. Find the Minimum Dimension (k):
   - Determine the smaller value between the number of rows (r) and the number of columns (c) in the table:

$$k = \min(\mathrm{r} - 1, \mathrm{c} - 1)$$

5. Compute Cramér's V:
   - Use the formula

$$V = \sqrt{\frac{x^2}{N * K}}.$$

2.3. Training and Test the model

In this section we use the proposed method is based on the approach developed by [2; 3] for the comparative analysis of two characteristics of translation quality, based on a comparative quantitative analysis of the unconscious style of translation texts. First, we quantify the contribution of the unconscious author's style by comparing different Amharic original work of different Authors. Secondly, we indirectly compare the contribution of the author of the work and the translator by analyzing translations of the same work by different translators.

For our investigation we split our Data set in two exclusive sets which is the training 64kb and the test set from each book. Then we split the test set into 32 slices with size 2kB to Test the model with the given training set.

Modern digital systems employ data compression techniques mentioned [19–25], known as archivers, which rely on principles from information theory and formal grammars. These tools reduce text size by detecting patterns in character frequency, enabling efficient data storage and retrieval [16]. This study based on an information-theoretic method for determining translation quality assessment from Amharic to English and vice versa.

3.0. experimental setup and Results

We prototype a python program for our experiment because Python is easy to import data compression libraries and mathematical manipulation using numpy library and data visualization mod-

ules. You can find the research implementation and Data in GitHub repository, https://github.com/yeshewas/Information-Theoretic-Method-for-Assessing-the-Quality-of-Translations.

Here below in the subsection 3.1 we investigate the authorship style of Amharic works, in subsection 3.2 we show results of English to Amharic translation of two books each with two different human translators and four different English books by a single Amharic translator, in section 3.3 we show the results of one Amharic work to English by three computer translators and finally in section 3.4 we investigate one English work to Amharic by computer translator's.

3.1 Authorship style of Amharic original texts

For authorship identifications experiment, we used original Amharic works by Dr. Hadis Alemayehu, Bealu Girma, Mamo Wodneh, and Alemneh Wassie, with N = 6 and m = 32. From these authors, we created six training samples, X1, X2, ..., X6, each with a size of 64kb. Additionally, we extracted 32 test samples from each author's work—Y1j (j = 1, ..., 32) from Hadis Alemayhu, Y2j from Bealu Girma, and so on, up to Y6j from alemneh Wassie. Each test sample was 2 KB in size.

Table 1: Results of the experiments. The data obtained for the PPM archiver, training set 64kb, and 32 Test slice each with 2 kB size. For All Potential Data compressor results calculated there Cramer coefficient in the same fashion and the result attached in Appendix 1.

PPM Cramer coefficient V=0.89

	Hadis-alemayhu	Bealu-Girma	Mamo-wedneh	Tsegaye-G/medhin	Michel-kebede	Alemayhu-waasie
Hadis-alemayhu	32	0	0	0	0	0
Bealu-Girma	0	30	0	0	0	2
Mamo-wedneh	0	0	31	1	0	0
Tsegaye-G/medhin	0	0	0	32	0	0
Michel-kebede	0	0	0	0	32	0
Alemayhu-waasie	0	2	0	0	0	30

From Table 1 presents a confusion matrix showing the attribution of texts to six different authors based on stylistic features. The results demonstrate a high degree of accuracy in distinguishing between the authors' styles.

Most notably, *Hadis Alemayhu*, T*segaye G/Medhin*, and *Michel Kebede* each achieved perfect classification, with all 32 samples correctly attributed to them. *Mamo Wedneh* also shows very strong consistency, with 31 texts correctly identified and only one misclassified as *Tsegaye G/Medhin*.

Slight confusion appears between *Bealu Girma* and *Alemayhu Waasie*, where each had 30 texts correctly attributed to them, but two samples were misclassified as each other. This suggests some stylistic overlap or similarity between these two authors.

Overall, the matrix indicates that the stylistic signatures of each author are distinct and reliably identifiable, with only minor exceptions.

3.2. English to Amharic translated books by two different real translators

In this experiment we utilize two English books translated to Amharic by real Amharic writers, firstly, a Book from vector ego known as "lie measurable" translated two Amharic novels "menduban" and "mekergonchu" written by Yohans G/mesekel and salhalselasi berhan respectively. Secondly a book from crime and punishment by devestoky to Amharic novels the books entitled "wenjel ena keta " and "wenjel ena ferde" by two translator muluberhan and kassa respectively, Each books with Training size of 64kb and Test slice of 32 with Slice size 2kB.

Table 2 presents a contingency matrix showing the attribution of Amharic translations of the English work "*Lie Measurable*" to two novels—"*Menduban*" by Yohans and "*Mekerognochu*" by Sehaleselasi—using a Partial Prediction Matching (PPM) compression-based method. The Cramer's V value of **0.762** indicates a **strong association** between the predicted and actual sources, suggesting the method is effective at distinguishing between the stylistic features of the two translations.

PPM V=0.762

	Menduban	**Mekerognochu**
Menduban by Yohans	31	1
Mekrognchu by sehaleselasi	7	25

From table 2 The high Cramer's V value (0.762) reflects a substantial level of distinguishability between the two translated texts. While *"Menduban"* displays a highly distinctive and consistent style, *"Mekerognochu"* appears to share certain stylistic traits with *"Menduban"*, leading to some misclassifications. This overlap may point to either subtle stylistic similarities between the authors or influences in translation that blur the boundaries between their styles.

Overall, the PPM compression method shows strong performance in identifying authorial style, particularly for Yohans's *"Menduban"*, though some refinement may be needed to improve accuracy for Sehaleselasi's *"Mekerognochu"*.

Table 3: contingency table of translation of English work "crime and punishment" to Amharic Novels "Wenjel ena ketat" and "Wenjel ena ferde" by Muluberhan and Kassa respectively Cramer V=1 for predication partial matching(PPM) compressor.

PPM v=1

	Wenjl ena ferde	**Wenjel ena ketat**
Wenjel ena ferde	32	0
Wenjel ena ketat	0	32

From Table 3 The results demonstrate that the PPM method can perfectly distinguish between the translation styles of Muluberhan and Kassa. This suggests that each translator imposed a highly distinct stylistic signature on their respective Amharic versions of *Crime and Punishment*. The absence of any misclassification indicates no stylistic overlap, making this an ideal case of translation style differentiation.

Table 4: For this experiment, I analyzed three literary works translated from English into Amharic by the renowned Ethiopian translator and author Sahle Sellassie Berhane Mariam, along with one of his original novels. The translated works included:

- Victor Hugo's *Les Misérables* (Amharic title: *Mekergochu*)
- Charles Dickens' *A Tale of Two Cities* (Amharic title: *Ye Hulet Ketemawech Weg*)
- Pearl Buck's *The Mother* (Amharic title: *Emye*)
- Sahle Sellassie's original work *Basha-Ketaw*

was included in the analysis. The experiment utilized a training corpus of 64 kB, with 32 test slices of 2 kB each for textual analysis.

Writers	**Sahle Sellassie**	**Victor Hugo**	**Charles Dickens**	**Pearl Buck**
Sahle Sellassie	32	0	0	0
Victor Hugo	0	32	0	0
Charles Dickens	0	0	29	3
Pearl Buck	2	8	1	21

From the table 4, we observe that the original style of Sahle Sellassie's novel was perfectly preserved in all 32 instances, as his translations consistently matched his own stylistic traits. Similarly, Victor Hugo's translations were distinctly different from those of other writers, with all 32 samples closely aligning with his unique style. However, Charles Dickens' style was less accurately preserved—three of his translated excerpts bore a closer resemblance to Pearl Buck's works.

Pearl Buck's style proved the most challenging to retain, with eleven of her translated slices leaning more toward other authors' styles. Overall, Sahle Sellassie's translations exhibited the worse preservation of the original author's style. This observation is supported by the Cramer's V coefficient of 0.77, indicating a strong but imperfect association between authors and their translated styles.

Notably, certain writers present exceptional difficulties for translators. Pearl Buck stands out as one such author—her distinctive style remains, among the most difficult to translate faithfully.

3.3. Amharic to English translated work by computer translators

In third experiment We used three computer translator that support Amharic language two of them the common language translator google translator, Yandex and one of them is translator that support Artificial intelligence large language model and neural machines translation model that is lingvanex translator [25].for this task our Data source is a famous novel in Amharic entitled "Fiker esk mekaber" (Love up to death) written by Dr. Hadis alemayhu with training size 64 kB and 32 Test slices each with 2 kB.

Based on the above experimental setup draw the contingency table and calculate the Cramer coefficient for all potential compressor (Gzip, LZMA, BZIP2, PPM, brotli, zstad) to select the efficient one. So, the result shows that almost all compressor scores approximate Cramer coefficient value which is 0.78 except the PPM V=0.83.

Table 5: contingency table of translation of Amharic work "Feker Esk mekaber" to English by computer translators. Cramer V=0.91 for predication partial matching (PPM) compressor.

PPM V=0.91

	Google-translators	**Yandex**	**lingvanex**
Google-translators	29	3	0
Yandex	4	28	0
Lingvanex	0	0	32

From Table 5 demonstrates that the translations produced by Google, Yandex, and Lingvanex differ significantly in style. This indicates that each translator retains elements of their own unconscious stylistic patterns. The Cramer's V coefficient of 0.91 further confirms that the distinctive styles of the translators are strongly expressed—while the original style of the classic novel is not as clearly preserved.

3.4. English to Amharic translated work by computer translators

Lastly, we used three computer translator that support Amharic language two of them the common language translator google translator, Yandex and one of them is translator that support Artificial intelligence large language model and neural machines translation model that is lingvanex translator [25]. for this task our Data source is a famous novel in English novel entitled "Born a crime" written by Trevor Noah with training size 64 kB and 32 Test slices each with 2 kB.

Table 6: contingency table of translation of English work "Born a crime" to Amharic by computer translators. Cramer V=53 for predication partial matching (PPM) compressor.

PPM V=0.53

	Google	Yandex	lingvanex
Google-translators	16	4	12
Yandex	1	27	4
Lingvanex-AI	0	3	29

From Table 6, we observe that the style of Lingvanex translations stands out significantly from the styles of other translators. Specifically, 29 out of 32 slices are most similar to Lingvanex's own translations. In comparison, Yandex's translation style is less distinct—four of its translation slices resemble Lingvanex's style more closely, and one aligns with Google's. Finally, Google's translation style is the least consistently conveyed, with 16 of its slices being more similar to those of other translators.

To assess high-quality translation, one essential condition must be met: the translator's influence on the translation style should be minimal—or ideally, absent altogether. The data in this table suggest a near-ideal scenario, as indicated by the Cramer's V coefficient value of 0.53. This supports the conclusion that the translator's impact on the translated text is relatively small.

4.0 Conclusion

This study demonstrates the potential of data compression techniques in assessing translation quality and authorship style analysis. By leveraging information-theoretic methods, we successfully analyzed linguistic patterns and stylistic similarities across original and translated texts. The results confirm that lossless data compressors, particularly PPM, effectively capture differences in authorship styles and translation styles of Amharic to English literary texts.

Our experiments revealed that translations by different human translators exhibit distinguishable linguistic patterns, whereas machine-translated texts show varying degrees of similarity depending on the algorithm used. The application of Cramer's V metric allowed us to quantify these differences, reinforcing the validity of data compression as a tool for evaluating translation fidelity and authorial influence.

The findings also highlight the challenges associated with translations between low-resource languages like Amharic and English. The variability observed among machine-generated translations underscores the need for improved natural language processing models that preserve linguistic degrees more effectively.

Future research should explore expanding the dataset to include more language pairs and integrating hybrid models that combine traditional linguistic features with data compression techniques. Additionally, refining machine translation algorithms based on compression-based quality metrics could lead to significant advancements in translation accuracy and authorship identification.

Appendix 1:

Calculating Cramer coefficient of All compressor:

LZMA V=0.86

	Hadis-alemayhu	Bealu-Girma	Mamo-wedneh	Tsegaye-G/medhin	Michel-kebede	Alemayhu-waasie
Hadis-alemayhu	32	0	0	0	0	0
Bealu-Girma	0	32	0	0	0	0

	Hadis-alemayhu	Bealu-Girma	Mamo-wedneh	Tsegaye-G/medhin	Michel-kebede	Alemayhu-waasie
Mamo-wedneh	0	0	32	3	0	0
Tsegaye-G/medhin	0	0	0	32	0	0
Michel-kebede	0	0	0	5	27	0
Alemayhu-waasie	0	4	0	0	0	28

BZIP2 V=0.75

	Hadis-alemayhu	Bealu-Girma	Mamo-wedneh	Tsegaye-G/medhin	Michel-kebede	Alemayhu-waasie
Hadis-alemayhu	19	0	0	13	0	0
Bealu-Girma	0	23	0	0	8	1
Mamo-wedneh	0	0	29	3	0	0
Tsegaye-G/medhin	0	0	0	32	0	0
Michel-kebede	0	0	0	8	24	0
Alemayhu-waasie	0	1	0	3	0	28

Zstand V= 0.87

	Hadis-alemayhu	Bealu-Girma	Mamo-wedneh	Tsegaye-G/medhin	Michel-kebede	Alemayhu-waasie
Hadis-alemayhu	32	0	0	0	0	0
Bealu-Girma	0	30	0	0	0	2
Mamo-wedneh	0	0	32	0	0	0
Tsegaye-G/medhin	0	0	0	32	0	0
Michel-kebede	0	0	0	3	29	0
Alemayhu-waasie	0	2	0	3	0	30

Brotli V=0.88

	Hadis-alemayhu	**Bealu-Girma**	**Mamo-wedneh**	**Tsegaye-G/medhin**	**Michel-kebede**	**Alemayhu-waasie**
Hadis-alemayhu	32	0	0	0	0	0
Bealu-Girma	0	28	0	0	0	4
Mamo-wedneh	0	0	32	0	0	0
Tsegaye-G/medhin	0	0	0	32	2	0
Michel-kebede	0	0	0	1	31	0
Alemayhu-waasie	0	0	0	0	0	32

References

1. **Malyutov M.B.** Authorship Attribution of Texts: A Review General Theory of Information Transfer and Combinatory. Lecture Notes in Computer Science, vol. 4123. Springer, Berlin, Heidelberg. DOI.org/10.1007/11889342_20.
2. **Ryabko B.Y., Savina N.** Information-Theoretic Method for Assessing the Quality of Translations. Entropy 2022, 24, 1739. DOI.org/10.3390/e24121739
3. **Ryabko B.Y., Savina N.** Using Data Compression to Build a Method for Statistically Verified Attribution of Literary Texts. Entropy 2021, 23, 1302. DOI.org/10.3390/e23101302
4. **Madigan D; Genkin A; Lewis D. D; Argamon, S; Fradkin, D; Ye L.** Author identification on the large scale. 2005, June. In *Proceedings of the 2005 Meeting of the Classification Society of North America (CSNA*).
5. **Argamon S; Šarić M; Stein S.S.** Style mining of electronic messages for multiple authorship discrimination: first results. 2003, August. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 475–480.
6. **Forsyth R.S; David I.H.** Feature-finding for text classification. Literary and Linguistic Computing, 1996, vol. 11, no. 4. pp. 163–174.
7. **Juola P.** Authorship attribution. Foundations and Trends® in Information Retrieval. 2008 Mar 6, vol. 1, no. 3. pp. 233–334.
8. **Katirai Hooman; Waterloo Ontario; Dale Schuurmans.** Filtering junk e-mail. 1999. Department of Electrical & Computer Engineering, University of Waterloo.
9. **Cover T. M.** Elements of information theory. 1999. John Wiley & Sons.
10. **Williams C. B.** Mendenhall's studies of word-length distribution in the works of Shakespeare and Bacon.1975. Biometrika, pp. 207–212.
11. **Thompson J. W; Padover S. K.** Secret diplomacy: espionage and cryptography. 1963, pp. 1500–1815.
12. **Mendenhall T. C.** The characteristic curves of composition. Science. 11 Mar 1887, no. 9, issue 214, pp. 237–246. DOI: 10.1126/science. ns-9.214S.237.
13. **Bahtin M.** Problemi poetike Dostojevskoga. Zadar: Sveučilište u Zadru; 2020.
14. **Brinegar C. S.** Mark Twain and the Quintus Curtius Snodgrass letters: A statistical test of authorship. 1963. Journal of the American Statistical Association. vol. 58, no. 301, pp. 85–96.

15. **Bende M. L.** The origin of Amharic. *Ethiopian Journal of Languages and Literature*. 1983, vol. 1, pp. 41–52.
16. **Adugna, Gabe.** Research: Language Learning - Amharic: Home. library.bu.edu. Retrieved 2025-05-10.
17. https://translate.google.ru/?hl=en&sl=it&tl=ru&op=translate
18. https://translate.yandex.com/en/
19. https://lingvanex.com/translate/
20. **Abdelfattah M.S; Hagiescu A; Singh D.** Gzip on a chip: High performance lossless data compression on fpgas using opencl. In Proceedings of the international workshop on openCL. 2013 & 2014, pp. 1–9.
21. **Seward J.** bzip2 and libbzip2. avaliable at http://www. bzip. Org. 1996, pp. 8–18.
22. **Onuma Y., Terashima Y., Kiyohara R.** Compression method for ECU software updates. IEEE Tenth International Conference on Mobile Computing and Ubiquitous Network (ICMU). 2017 Oct 3, pp. 1–6.
23. **Alakuijala J., Farruggia A., Ferragina P., Kliuchnikov E., Obryk R., Szabadka Z., Vandevenne L.** Brotli: A general-purpose data compressor. ACM Transactions on Information Systems (TOIS).2018, vol. 37, pp. 1–30.
24. **Collet Y., Kucherawy M.** Z-standard Compression and the application/zstd Media Type. 2018, no. rfc8478.
25. **Shkarin D.** PPM: one step to practicality. In: Proceedings of the Data Compression Conference. vol. DDC ’02, pp. 202. IEEE Computer Society (2002)

Information about the Author

Yeshewas Getachew Lulu, PhD Student in specialty of Mathematical and software of computer systems, complexes and computer network in Faculty of information Technology, Novosibirsk state university, Russian Federation.

Сведения об авторе

Йешевас Гетачеу Лулу, аспирант факультета информационных технологий Новосибирского государственного университета

Статья поступила в редакцию 09.06.2025;
одобрена после рецензирования 31.07.2025; принята к публикации 31.07.2025

The article was submitted 09.06.2025;
approved after reviewing 31.07.2025; accepted for publication 31.07.2025