



KARŞILAŞTIRMALI ÇEVİRİ ANALİZİNDE KH CODER'IN KULLANIMI: LE PETIT PRİNCE ÇEVİRİLERİ THE USE OF KH CODER IN COMPARATIVE TRANSLATION ANALYSIS: LE PETIT PRINCE TRANSLATIONS¹

İLYAS BOĞUSOĞLU

Sayı 4, ss., 8-25

Makale Türü: Araştırma Makalesi

Geliş Tarihi: 01.10.2024

Kabul Tarihi: 31.10.2024

Değerlendirme: Çift Taraflı Kör Hakemlik

Etik Bildirim: lisaniyatstudies@gmail.com

İntihal Denetimi: Yapıldı. (İntihalnet)

Etik Beyan: Bu çalışma, etik kurul izni gerektirmeyen nitelikte olup kullanılan veriler literatür taraması/yayınlanmış kaynaklar üzerinden elde edilmiştir. Çalışmanın hazırlanma sürecinde bilimsel ve etik ilkelere uyulduğu ve yararlanılan tüm çalışmaların kaynakçada belirtildiği beyan olunur.

Lisans Bilgisi: Dergimizde yayımlanan çalışmalar, CC-BY-NC-ND lisansı ile açık erişim olarak yayımlanmaktadır.

Year 3, Issue 4, pp., 8-25

Article Type: Research Article

Submitted: 01.10.2024

Accepted: 31.10.2024

Evaluation: Double-Blind Review

Ethics Declaration: lisaniyatstudies@gmail.com

Plagiarism Check: Done. (İntihalnet)

Ethics Statement: This study does not require ethics committee approval, and the data used was obtained through a literature review/published sources. It is hereby declared that scientific and ethical principles were adhered to throughout the preparation, and all references are cited in the bibliography.

License Information: The studies published in our journal are published as open access with the CC-BY-NC-ND license.

Atıf Bilgisi / Reference Information:

BOĞUSOĞLU, İ. (2024). The Use Of KH Coder In Comparative Translation Analysis: Le Petit Prince Translations, *Lisaniyat Studies Uluslararası Filoloji Araştırmaları Dergisi*, 4, 8-25

İlyas Boğusoğlu

Bağımsız Araştırmacı

bogusoglu@hotmail.com

ORCID: 0000-0001-7817-1115

stract

This study aims to evaluate the usability and effectiveness of KH Coder, a text mining tool, for translation comparison research, focusing on the concepts of 'equivalence' and 'matricial norms.' Text mining involves analyzing natural language texts to derive structured data. KH Coder is adept at processing and analyzing data for text mining, computational linguistics, and quantitative content analysis. 'Matricial norms,' as proposed by Gideon Toury, guide how the formal values of the source text should be represented and modified in the target text, encompassing substitutions, additions, and deletions. 'Equivalence' is assessed through two lenses: 'adequacy,' where a translation closely aligns with the source, and "acceptability," where it resonates with the target culture.

In this study, Antoine de Saint-Exupéry's book 'Le Petit Prince' and its translations of 'The Little Prince' in English, '星の王子さま' (Hoshi no Ōjisama) in Japanese and 'Маленький принц' (Malen'kij Princ) in Russian are considered as the objects of study. The analysis commenced with organizing these texts into an Excel format before inputting them into KH Coder for analysis. Then findings have been compared within the theoretical framework. Ultimately, the study concluded that KH Coder is both usable and effective for translation comparison studies framed by matricial norms and equivalence.

Keywords: Translation Studies, Text Mining, Matricial Norms, Target Oriented Theory, Le Petit Prince

Özet

Bu çalışmanın amacı, bir metin madenciliği inceleme aracı olan KH Coder'ın, 'eşdeğerlik' ve 'matriks normlar' ışığında yapılan karşılaştırmalı çeviri çalışmalarındaki kullanılabilirliğini ve yeterliliğini test etmektir. Metin madenciliği, doğal dil metnini veri kaynağı olarak kabul eden ve metin üzerinden yapısalştırılmış veri elde etmeyi amaçlayan bir yöntemdir. KH Coder, metin madenciliğinin yanı sıra hesaplamalı dilbilim ve nicel içerik analizi çalışmalarında kullanılabilen, verilerini doğrudan işleyip analiz edebilen bir yazılımdır. Gideon Toury'nin öne sürdüğü çeviri normlarından 'matriks normlar' ise kaynak metnin biçimsel değerlerinin erek metinde nasıl yer alacağına ve ne kadar değiştirileceğine yön veren, metindeki her türlü yer değiştirme, ekleme ve çıkarma gibi biçimsel değişiklikleri kapsayan normlardır. Diğer yandan 'eşdeğerlik' iki farklı kavramla belirlenebilmektedir: 'yeterlilik' ve 'kabul edilebilirlik.' Çeviri, kaynak dizgeye yakın olması durumunda 'yeterli' çeviri, erek dizgeye yakın olması durumunda 'kabul edilebilir' çeviri olarak kabul edilir.

Bu çalışmada inceleme nesnelere olarak Antoine de Saint-Exupéry'in 'Le Petit Prince' adlı eseri ile İngilizce 'The Little Prince', Japonca '星の王子さま' (Hoshi no Ōjisama) ve Rusça 'Маленький принц' (Malen'kij Princ) çevirileri ele alınmıştır. İlk olarak inceleme nesnelere Excel dosyası formatında düzenlenmiş, ardından KH Coder'a tanıtılmış ve analiz bulguları çıkarılmıştır. Elde edilen bulgular kuramsal bağlamda karşılaştırılmıştır. Çalışmanın

¹This article is derived from the master's thesis titled A Comparative Study With Text Mining: Le Petit Prince Translations (2024) under the supervision of Zeynep Gençer Baloğlu at the Department of Language of Science and Cultural Studies, Institute of Social Sciences, Pamukkale University.



sonunda KH Coder'ın matris normlar ve eşdeğerlik bağlamında yapılan karşılaştırmalı çeviri çalışmalarında kullanılabilir ve yeterli olduğu gözlemlenmiştir.

Anahtar Kelimeler: Çeviribilim, Metin Madenciliği, Matris Normlar, Erek Odaklı Kuram, Le Petit Prince

Structured Abstract

The aim of this study is to observe the functioning of the text mining method, which has been introduced as a new method in translation comparison studies. Accordingly, the study evaluates the applicability and effectiveness of KH Coder, a text mining tool, in comparative translation analysis by focusing on the theoretical concepts of 'equivalence' and "matrix norms". Text mining involves extracting structured data from natural language texts. KH Coder is a versatile tool capable of processing and analyzing data for text mining, computational linguistics, and quantitative content analysis. As proposed by Gideon Toury, "matricial norms" guide the representation and modification of formal values from the source text in the target text, encompassing substitutions, additions, and deletions. "Equivalence" is assessed through two lenses: "adequacy," where a translation closely aligns with the source, and "acceptability," where it resonates with the target culture. The study focuses on Antoine de Saint-Exupéry's *The Little Prince* and its English, Japanese, and Russian translations.

In the initial phase of the study, the texts to be analysed have been prepared in a format suitable for KH Coder. English, Japanese and Russian translations have been converted to .txt format and loaded into the software. Since each language has its own morphological and structural features, these features have been taken into account in the comparisons between languages. In particular, the differences between word types have showed how important the structural features of languages are in the analysis process. Since Japanese has more word types than Russian, some limitations have been imposed on the comparison made in this study. For this reason, comparisons have been made on the basis of four basic word types: noun, adjective, adverb and verb. Word types in each language have been analysed using language analysis tools such as Standard POS Tagger for English, FreeLing for French and Russian, and ChaSen for Japanese. Comparisons have been made on the word frequency lists and other analysis data provided by the software, and these data have been used as important tools in understanding the additions, deletions or changes in place in the translations.

The concepts of translation adequacy and acceptability are other important elements that form the theoretical background of this study. While adequacy evaluates how faithful the translation is to the source text, acceptability examines how well the translation adapts to the target language and culture. As a result of the study, it has been determined that the English translation is closer to the source text, while the Japanese and Russian translations are more adaptable to the target culture. In addition, Hepburn Style has been used in the transliteration of Japanese text, and ISO9 standard has been used in the transliteration of Russian text. These transliteration methods have played an important role in the analysis of languages with different alphabet systems. KH Coder's ability to analyse texts in these languages increases its usability in multilingual translation studies. Analyses conducted in languages with different alphabet systems, especially Japanese and Russian, show how KH Coder processes linguistic diversity.

This study makes significant contributions to the body of literature on translation studies and provides new perspectives on how text mining methods can be used in translation studies. The quantitative data provided by the software facilitates the structural analysis of translations and provides an important tool for understanding differences between languages. While the data provided by KH Coder facilitates the structural analysis of translations, it also provides great convenience in understanding how additions, deletions and changes in translation processes are made. However, it has been emphasized that the data provided by tools such as KH Coder should be analyzed carefully, considering the different structural features of languages. KH Coder stands out as a robust and systematic analysis tool in translation comparisons. The findings obtained



throughout the study have showed that the software can be used in translation studies and provides important data in understanding the formal changes made in translation processes.

As a result of this study, text mining in translation studies, which has emerged as an increasingly important field, has shown itself as an innovative and effective tool with the methods used in this study. The text mining

method was found to be very useful and sufficient especially in comparative translation studies. Nevertheless, it is believed that more translation studies using this method and various text mining tools should be conducted.

Introduction

Language is a complex whole that presents various aspects which can only be approached separately through abstraction (Vardar, 1998, p. 11). Understanding and analyzing this complexity is one of the fundamental motivations of linguistics. This discipline aims to understand the essence of language by examining its structure, usage, and development. With technological advancements, linguistics has also found a place in computational environments. Computational linguistics analyzes language using various techniques and algorithms to understand its cognitive aspects. Areas of interest in this field include language parsing, syntactic analysis, semantic analysis, machine translation, pragmatics, and language teaching/learning (Türker, 1991, p. 139). Another significant area of interest is text mining, which utilizes text as a data source and aims to extract structured data from it (Seker, 2015, p. 30).

Anthony Pym (2010) emphasizes the evolving role of technology in translation, noting that computational tools enable deeper insights into translation strategies and decision-making processes. Based on this, the lack of adequate use of text mining in comparative translation studies is one of the major motivations of this study. This study observes the functioning of the text mining method, which has been introduced as a new method in translation comparison studies and tests the applicability of the selected examination tool in translation studies. In this context, KH Coder, a text mining software, has been selected as the examination tool for this study. KH Coder can effectively process, analyze, and visualize complex and large text data files. Considering the function of text mining, it has been deemed appropriate to utilize Gideon Toury's (2012) Target Oriented Theory and translation norms as the theoretical framework for this comparative translation study. In the Target Oriented Theory, which is a descriptive approach, it can be determined whether translations are 'adequate' or 'acceptable,' while translation norms can provide insights into the translation process through various actions that guide decisions during the translation process. This study employs 'matricial norms' among the translation norms. Since matricial norms enable the quantitative analysis of translation processes, they are considered the most suitable translation norms for KH Coder.

As the objects of analysis, *Le Petit Prince* and its translations into English '*The Little Prince*,' Japanese '*星の王子さま*' (Hoshi no Ōjisama), and Russian '*Маленький принц*' (Malen'kij Princ) are selected, considering the theoretical framework of the study and the text processing capacity of the tool used in the methodology. In this context, the fact that the selected objects have been translated into multiple languages and show various differences in their translations contributes to the theoretical framework of the study. Additionally, the lengths of the texts are considered suitable for the tool used in the analysis. Using KH Coder as the examination tool, the data obtained from the analysis have been examined concerning one of Toury's (2021, p. 201) proposed translation norms, 'matricial norms.' After the examination, the findings have been compared to determine the acceptability and adequacy of the translations. The analysis focuses solely on structural values, without addressing the semantic dimension of the texts.

In the literature review, Nalan Bozan, who addresses both text mining and translation norms in her article titled "A Text Mining Approach to 'Operational norms'" (2021) and in her master's thesis titled *Extending the Boundries in Descriptive Translation Analysis: A Text Mining Approach* (2022), has tested her analysis in the context of "operational norms" using text mining using a specially developed software as a tool for her study.



As a result of her studies, she stated that the tool she used could be improved and tested in more studies, and that various information obtained using this method could contribute to the field. Unlike the studies mentioned, this study tests the use of the text mining method in comparative translation studies using a different analysis tool. In addition, the usability of the analysis tool used in comparative translation studies is also observed.

The primary aim of this study is to provide a model for researchers and field studies that wish to use text mining method or for future studies. The secondary aim is to test the usability and adequacy of KH Coder in comparative translation studies conducted within the framework of target-oriented translation theory and "matricial norms." To achieve these aims, the study seeks answers to the following questions:

- Does KH Coder analysis yield sufficient results in evaluating matricial norms?
- Is KH Coder usable and adequate in a comparative study within the context of translation norms?
- Can KH Coder be a tool used in comparative translation studies?

Following the literature review, it has been observed that text mining is frequently used as a research method in many fields, but there is little occurrence of this method in translation-related studies. This indicates that the method's place in such studies is relatively new. In this context, it is believed that the study will contribute to the diversity of translation comparisons and that it will be an important work for better understanding text mining as a review method in studies within this field.

Text Mining

In today's digital age, where information is abundant, diverse, and easily accessible, the ability to extract meaningful insights from large amounts of textual data has become crucial. This has led to the need for automated methods to analyse and extract valuable information from this wealth of data (Atan, 2020, p. 222). Text mining analysis has been developed as a response to this need, offering computational techniques to examine and interpret large volumes of textual information. This method treats text as a data source, aiming to extract structured data from it (Seker, 2015, p. 30). Furthermore, it uses methods that can handle numerous words and structures in natural language while also addressing ambiguities, thereby revealing hidden information (Hotho et al., 2005). The evolution of text mining can be attributed to factors such as technological advancements, interdisciplinary collaboration, and an increasing demand for automated data analysis methods.

One of the most significant factors in the development of text mining analysis is the progress made in the field of Natural Language Processing (NLP), a branch of artificial intelligence (Jurafsky & Martin, 2008, p. 37). This emerging scientific field, which has gained momentum with the advancements in information technology, has gradually evolved into the domain of computational linguistics (Adalı, 2012). NLP focuses on the design and development of computer systems capable of performing actions such as the analysis, interpretation, and generation of texts in natural language (Aslan & Kuşçu, 2015). Advances in NLP, including text parsing, part-of-speech tagging, and semantic analysis, have laid the foundation for effectively processing and analyzing text data using text mining techniques. These NLP techniques have formed the backbone of many text mining algorithms, enabling computers to extract meaning from unstructured texts. Additionally, text mining analysis has greatly benefited from interdisciplinary collaboration, drawing extensively from fields and methodologies such as computer science, linguistics, statistics, and information science. Additionally, the rise of statistical and machine learning methods has played a significant role in shaping the evolution of text mining analysis. These methods provide powerful tools for automatically identifying patterns, trends, and relationships in textual data. Techniques such as clustering, classification, and topic modeling have become central to text mining analysis, enabling researchers to uncover implicit meanings and extract valuable insights from textual datasets. Therefore, it can be said that text-mining studies can yield both structural and semantic data about the texts being examined.



Text mining, a computational technique for extracting patterns from large textual datasets, has become increasingly important in corpus-based and comparative translation studies. By analyzing large volumes of texts, text mining enables scholars to uncover translation norms, identify linguistic trends, and explore translation strategies across different languages and cultures. This method provides a quantitative complement to the more traditional, qualitative approaches in the field.

The relevance of text mining is particularly evident in relation to the work of scholars. Maeve Olohan (2004) highlights the benefits of text mining for identifying stylistic shifts, terminological consistency, and the subtle differences between source texts and target texts. Susan Bassnett (2002, p. 2) notes that the relationship between translation and advancing technology has become increasingly prominent. On the other hand, Mona Baker (1993, p. 248) highlights the urgent need to explore the potential of large computer corpora in translation studies, emphasizing the various components available for this purpose and noting the necessity of appropriate research methodologies and robust tools to develop a descriptive branch of translation studies. The integration of text mining aligns well with the principles of Descriptive Translation Studies (DTS), as pioneered by Gideon Toury (2012). Toury emphasizes the importance of empirical data in identifying translation norms, which are the regularities observed in translators' decisions. Although he doesn't directly address computational tools, his focus on systematic analysis laid the foundation for incorporating text mining into translation research. By processing large corpora of translations, text mining supports Toury's objective of discovering patterns in translation behavior that may otherwise have gone unnoticed. Ultimately, text mining enhances the methodological scope of translation studies, building on Toury's DTS framework by providing new ways to empirically explore the norms and patterns that guide translation practices.

In this context, it is evident that the text mining method can provide various tools and benefits in the field of translation studies. By analyzing parallel texts and translation memory systems, text mining helps translators identify patterns, assess translation quality, optimize translation processes, extract domain-specific terminology, and particularly enhance machine translation systems. Utilizing text mining techniques in translation studies can provide valuable insights and tools to improve translation efficiency, consistency, and accuracy in multilingual communication.

Target-Oriented Theory

Considering the function of text mining, it has been deemed appropriate to utilize Gideon Toury's (2012) Target Oriented Theory and translation norms as the theoretical framework in this study. According to Toury (2021, p. 197), translation is a 'cultural' activity, and thus the translator must perform their function in a manner deemed appropriate by a particular community. In this context, it is natural for the translator's behaviors during the translation process to be guided or restricted by rules and norms established by the target culture.

One of the significant concepts emphasized in Toury's Target-Oriented Theory is 'equivalence,' which serves as a tool for determining the prevailing understanding of translation in the target culture by re-examining actions in the translation process (Yazıcı, 2005, p. 135). To establish equivalence, Toury proposes two distinct concepts: 'adequacy' and 'acceptability.' If the translation closely aligns with the norms of the source culture, it is deemed an 'adequate' translation. Conversely, if the translation is aligned with the norms of the target culture, it is considered an 'acceptable' translation. Furthermore, Toury (2012, p. 70) emphasizes that no translation can be entirely source-oriented or target-oriented; every source-oriented translation contains some degree of target orientation, and every target-oriented translation contains some source-oriented shifts.

In his article titled 'The Nature and Role of Norms in Translation' (2021), Toury discusses norms in a socio-cultural context and emphasizes their impact on translations. He states that translation reflects the traditions



of at least two languages and two cultures, containing two different norm systems at both the linguistic and cultural levels (Toury, 2021, p. 199). Toury categorizes the various norms that guide translation into three main headings: initial norms, preliminary norms, and operational norms.

Initial Norm

Initial norm is the decisions made by the translator before commencing the translation. The translator must decide whether their translation will be source-oriented or target-oriented. In other words, the translator must determine whether to consider the norms of the source text for an adequate translation or to adhere to the norms of the target culture for an acceptable translation. Although the concept of initial norm is intended to serve as an explanatory tool, it is still possible and useful to clarify micro-level decisions with these terms, even if no clear macro-level tendency toward adequacy or acceptability is detected (Toury, 2021, p. 200).

Preliminary Norms

Preliminary norms are those that help to understand certain choices related to translation before the translation takes place. Toury addresses these as two interconnected categories (2021, p. 201). In Translation Policy, Toury points out that the selection of the text to be translated is also significant in this process. Factors such as why the translator chose a particular text and what influenced their choice can provide insights into the translation. And Directness of Translation encompasses the translator's decision on whether to translate directly from the source language or through an intermediary language. Understanding the decisions made at this point is facilitated by this norm.

Operational Norms

Operational norms are a group of norms that guide decisions made during the act of translation. These norms influence the text at both the formal and linguistic levels, thus directing the relationship between the source and target texts, determining what changes and what remains the same (2021, p. 201-202). Like preliminary norms, operational norms can also be divided into two categories. Matricial Norms directly affect the formal structure of the text, guiding how the formal/matricial values of the source text are represented in the target text and how much they will be altered. Matricial norms encompass all formal changes, such as additions, omissions, and alterations. Textual-Linguistic Norms guide the selection of linguistic values to be used in the target text, influencing micro-level values and grammatical structures during the translation process.

In this study, the focus is on matricial norms. The formal values of the source text are obtained and compared with the formal values in the translations. Since the main aim of the matricial norms is structural analysis, the semantic dimension of objects is not addressed in this study. As a result of comparing the obtained matricial values, the adequacy and acceptability of the translations in the context of equivalence are determined.

Objects of Study

Le Petit Prince has been the subject of research in various fields since its first publication due to factors such as its subject matter, its intended messages, its visuals, and its readership across different age groups. One of these factors is the book's translation into many languages, which has led to numerous translation studies. These studies often focus on challenges in the translation process, wordplay, cultural references, cultural adaptations, interdisciplinary themes, and language education. Such examinations have been conducted in both source text and translation text contexts.

Le Petit Prince



Published in 1943 by French author Antoine de Saint-Exupéry, the book is notably meaningful for both children and adults, holding an important place in the literary canon. It also reflects the author's life experiences as a pilot. The narrative follows the interaction between the pilot and the Little Prince, whom he

meets after his plane crashes in the desert, exploring the Prince's experiences and adventures on different planets. The work consists of twenty-seven chapters, enriched with the author's illustrations, and presents deep thoughts on themes like friendship, love, humanity, growth, loneliness, and the meaning of life. While the text is written in a simple language, it has profound meanings behind straightforward sentences.

A phenomenon worldwide, the book has been translated and published in over 500 languages and dialects, selling more than 200 million copies since its release. The languages with the most translations include Chinese (Mandarin), English, Italian, Japanese, Spanish, Portuguese, Persian, Serbian, Greek, and Turkish. Additionally, it maintains its classic status in the literary canon as the second most translated work after the Bible. This study examines a 2017 edition of the book, which serves as the source text (ST).

The Little Prince

The first English translation of the work was done in 1943. Under the title *The Little Prince*, it gained significant popularity in English-speaking countries. While there have been many translations by various individuals at different times, the most famous and widely used translation is by Katherine Woods.

Woods (1886-1968) was known as a writer and translator and gained fame through her translation of *The Little Prince*. This translation, being the first, remained the only English translation available until 1995 (Roldán-Riejos, 2021). In her translation, Woods effectively reflects the original text faithfully and fluently, which has been appreciated by both readers and critics. This study examines the 2017 edition of Woods's translation, which serves as the first target text (TT1).

星の王子さま (Hoshi no Ōjisama)

The first Japanese translation of the work was published in 1953 by Arō Naitō under the title of *星の王子さま* (Hoshi no Ōjisama), meaning 'The Prince of the Stars.' Since then, there have been many editions published by different publishing houses. These translations often include adaptations to fit Japanese culture, encountering modifications of some cultural and linguistic features from the original text.

While there are many translators of the work, this study focuses on Yū Ōkubo's translation, which serves as the second target text (TT2). Born in 1982, Ōkubo began translating at a young age and has continued to present various translations in literature and art. He currently continues his studies in translation and works actively as a freelance translator and writer.

Маленький принц (Malen'kij Princ)

The first known Russian translation of *Le Petit Prince* was published in 1960 under the title of 'Маленький принц' (Malen'kij Princ) during the Soviet Union era. This translation by Nora Gal is one of the significant translations into Russian. This study examines a 2012 edition of Nora Gal's translation, which serves as the third target text (TT3).

Nora Gal, born Eleonora Yakovlevna Galperina (1912-1991), was a Soviet-Russian translator, writer, and literary critic. Known for her translations from French and English literature into Russian, Gal played a crucial role in bringing many Western literary works to Russian readers. Her translation of *The Little Prince* is particularly renowned for its lyrical quality and fidelity to the spirit of the original text. In addition to her



translation work, Gal was a respected literary critic, contributing significantly to the field with essays and articles on translation theory and the art of literary translation.

METHOD

KH Coder, developed by Japanese sociologist Koichi Higuchi, was first released in 2001. This software can be used for text mining, computational linguistics, and quantitative content analysis. It processes and analyzes data directly without requiring additional data processing tools. KH Coder is available for download on its official website and supports Windows, Linux, and Macintosh operating systems. Initially, it only supported the analysis of Japanese texts, but its latest update has expanded the language options to thirteen, including Japanese, English, French, German, Italian, Catalan, Dutch, Russian, Portuguese, Spanish, Slovene, Simplified Chinese, and Korean. Its multilingual support allows for analysis of multilingual datasets. The tool itself features six different interface languages: Chinese, English, French, Japanese, Korean, and Spanish.

Thanks to its user-friendly interface, KH Coder enables easy execution of complex analyses. Some key features of KH Coder used for analyzing large text datasets include Content Analysis, Statistical Analysis, Integrated Data Mining Tools, Language Processing Capabilities, Various Analysis Techniques, Graphical Visualization, Social Network Analysis, Machine Learning Integration and Coding Analysis.

The first literary work analysis conducted with this tool was performed by its developer, who published two consecutive articles illustrating how KH Coder can be applied to literary analysis by examining *Anne of Green Gables* (Higuchi, 2016; Higuchi, 2017). An example of similar studies in the context of literary work analysis is Gençer Baloğlu's (2022) work.

Higuchi conducted the aforementioned studies using a two-step quantitative content analysis approach he previously proposed, testing the accuracy of the findings with KH Coder. These steps are as follows (Higuchi, 2016, p. 77):

- Firstly, automatically extract a word list from the existing data and analyze them statistically to obtain the bigger picture and eliminate biases.
- Secondly, establish coding rules, such as "If a specific expression exists, we consider it as an appearance of concept A," to derive concept categories from the data. Then, statistically analyze the concepts to deepen the analysis.

According to him, quantitative analysis has two main advantages: it facilitates the exploration of data, leading to a better understanding of the results, and it increases the reliability of the analysis (Higuchi, 2017).

Tool Features Used in the Study

For this study, the consideration of statistical data collection has been deemed more beneficial for evaluation in a theoretical context, thus focusing on Higuchi's "Step 1." Three different features have been selected in the context of assessing matricial norms. Below are details about these features.

Count of Unit Words – Sentences – Paragraphs

After preprocessing, database statistics are displayed in KH Coder's main window. The relevant statistics include number of tokens, word types, sentences and paragraphs.

In addition to the shown paragraph count, the distribution of paragraph counts in the sections of the text can also be determined. The numbers in parentheses indicate the count identified by KH Coder.

Word Frequency List



This feature, located under the 'Words' menu, arranges the words used in the text according to their frequency, from the most to the least common. It displays the first 100 words at a time.

Word Types Frequency List

The data ranks the types and counts of words according to their frequency of usage. All word types are displayed in columns within a single file, showing both the words and their usage counts. However, the types of words may vary depending on the word extraction software used according to the language of the file.

Data Preparation and Collection

Since KH Coder can only process textual data, the analysis is conducted solely on the main texts found within the chapters. Sections containing book covers, visuals, author/translator/editor introductions, prefaces, afterwords, and notes are not included because these sections, unlike the main texts, vary from edition to edition. Initially, the necessary materials for the analysis have been gathered. The printed versions of the texts in French, English, and Russian have been obtained, while the Japanese text has been accessed from publicly available online sources. All texts have been formatted appropriately for upload into the software. KH Coder version 3.Beta.07f (2023) has been utilized for the analysis.

First, all texts to be analysed have been converted to plain text format (.txt) and transferred into separate files within an Excel (.xlsx) document. The first row of the Excel file has been labeled with columns A-B-C as 'Text-Chapter-Part,' with each paragraph organized into its own row under the 'text' column. Each chapter title from the total of 27 chapters has been extracted into the 'Chapter' column, while all chapter titles have been compiled under the 'Part' column. The prepared files have been then individually uploaded into the software. Different software programs have been used for word extraction from each language: 'FreeLing' for French, 'Standard POS Tagger' for English, 'ChaSen' for Japanese, and 'FreeLing' for Russian. Considering that each language's software extracts varying numbers of word types, Russian, which has the lowest count, has been selected for a balanced comparison. The analysis has been limited to four types: noun, adjective, adverb, and verb. Before uploading the projects to the archive, unnecessary word lists associated with the examined languages have been removed to facilitate a more balanced analysis, and these words have been included in the analysis. After finalizing the settings, the files have been undergone preprocessing one by one which is essential for converting unstructured text into a form that can be analyzed by algorithms. Upon completion of these procedures, the files have been prepared to utilize the desired features.

Following preprocessing, the initial data has been generated through the database results in the opening window of KH Coder. The features used in the analysis have been selected from the 'words' section. First, word frequency lists have been obtained. Subsequently, these lists have been compiled into Excel files to create lists of word type frequencies. Finally, all numerical data collected for comparison in the analysis has been manually tabulated. Additionally, Hepburn style (Heebon Shiki) has been utilized for the transliteration of Japanese words, while the ISO9 standard has been applied for the transliteration of Russian words.²

Analysis

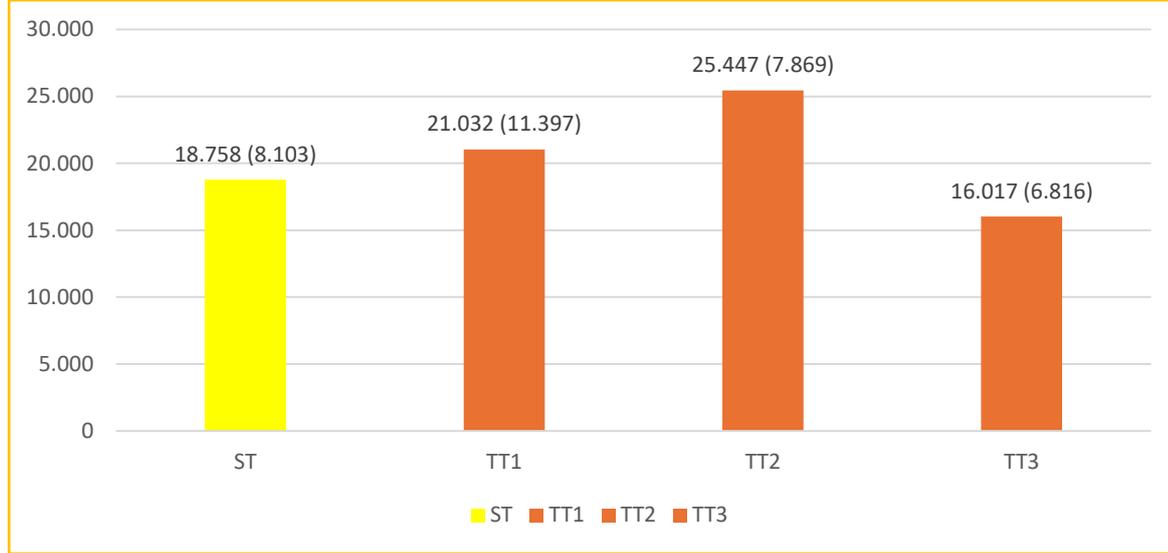
Comparison Of Token Counts

When comparing the token counts of the texts, it is observed that the text closest to the ST is the TT1. TT1 has 2,274 more tokens than the ST, making it the closest translation. Conversely, the TT2 has 6,689 additional tokens, indicating it is the farthest from the ST. On the other hand, the defined token count closest to the ST is TT2, while TT1 is the farthest.

International Organization for Standardization. (1995). Information and documentation - Transliteration of Cyrillic characters into Latin characters - Slavic and non-Slavic languages (ISO Standard No.9:1995). <https://www.iso.org/standard/3589.html> (20.05.2024)



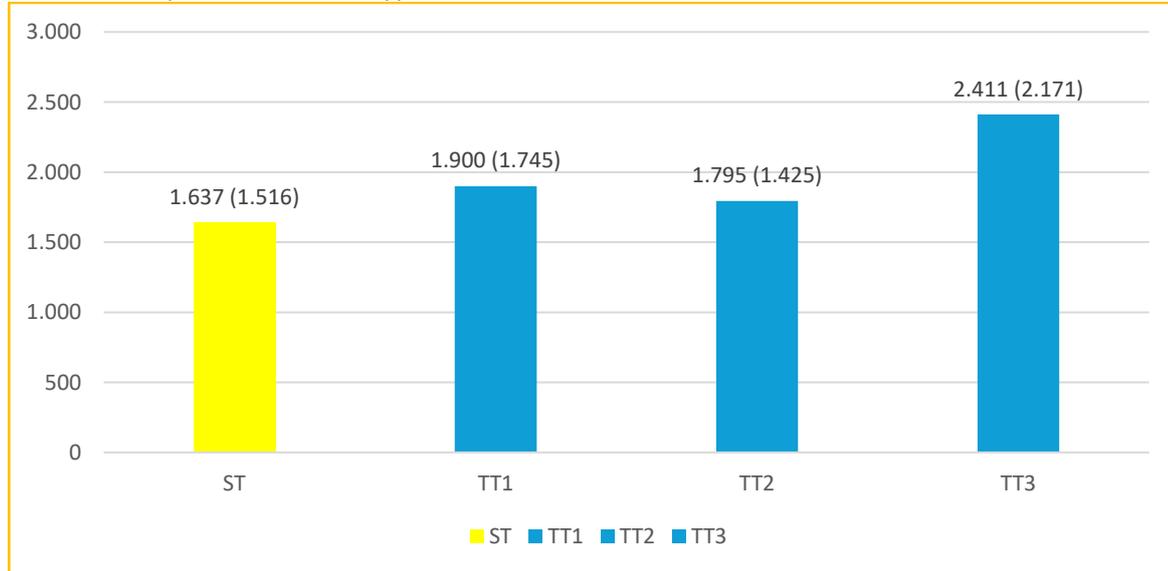
Table 1. Comparison of Token Counts



Comparison Of Word Types Counts

When comparing the number of word types, it is found that the text closest to the ST is TT2, with 158 more types of word. In contrast, TT3 is the farthest from the ST, with a difference of 774 words. This situation also holds true for the defined word types counts.

Table 2. Comparison of Word Types Counts



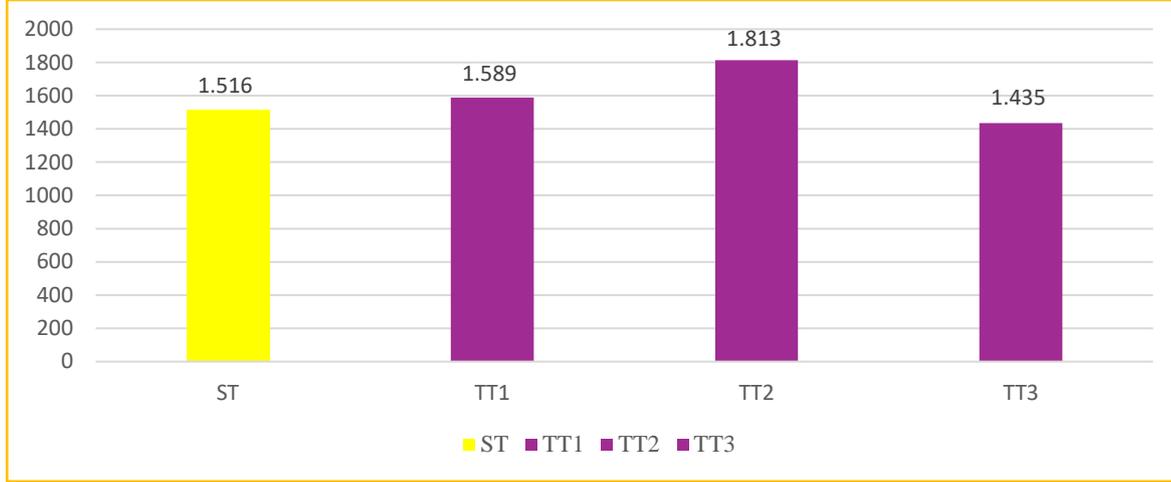
Comparison Of Sentence Counts

When examining the sentence counts, it is observed that TT1 has the closest value to the ST, with an additional 73 sentences. This may suggest that the translator divided existing sentences in the ST or added sentences. On the other hand, TT2 is the farthest from the ST, with a difference of 297 more sentences. Similar to TT1, this implies that the translator may have also divided or added sentences in the ST. In contrast,



TT3 has 91 fewer sentences than the ST, indicating that the translator may have combined sentences or deleted some from the ST.

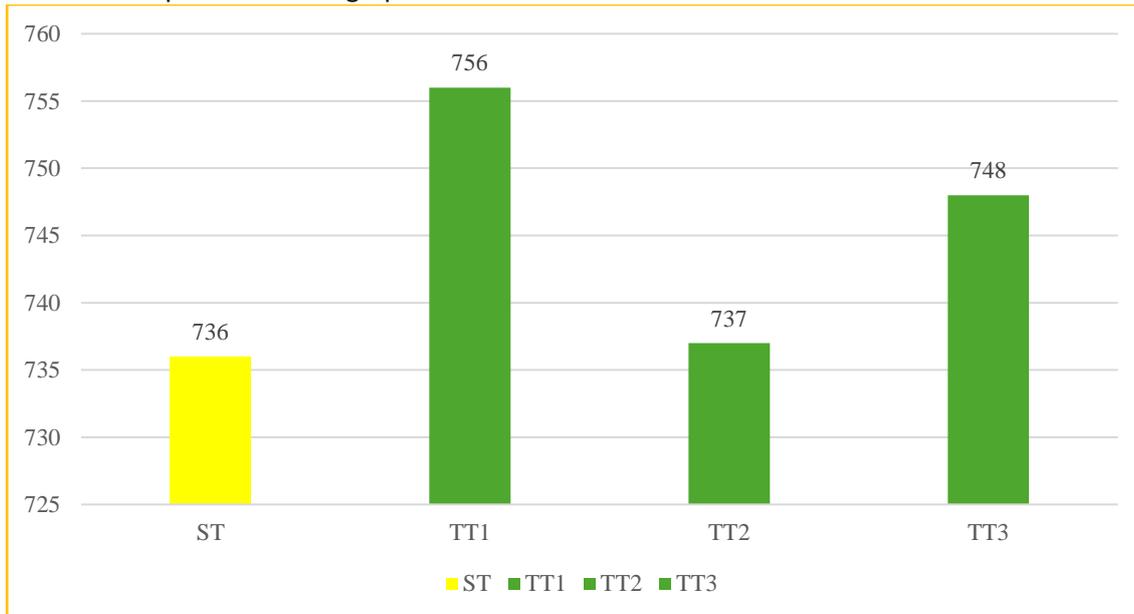
Table 3. Comparison of Sentence Counts



Comparison Of Paragraph Counts

When comparing paragraph counts, it is observed that the TT2 has the closest number of paragraphs to the ST. The ST has a total of 736 paragraphs, while TT2 has 737, indicating that the translator made almost no changes at the paragraph level. TT1 has 756 paragraphs, exceeding the ST by 20. When considering the increase in sentence count alongside the increase in paragraph count, this difference may not be very noticeable. However, the situation is slightly different for TT3. TT3 has 12 more paragraphs than the ST. The decrease in sentence count in TT3 correlates inversely with the increase in paragraph count, suggesting that the translator may have divided the paragraphs in the ST into smaller ones.

Table 4. Comparison of Paragraph Counts





Comparison Of Paragraph Counts In Chapters

The distribution of paragraph counts across the objects is presented in the table below. When comparing the objects, the following conclusions can be drawn:

- In terms of paragraph distribution, TT2 aligns most closely with the ST, as 18 out of 27 chapters have an equal number of paragraphs.
- TT1 has 10 chapters where the paragraph count matches that of the ST.
- The object with the least equality in chapters compared to the ST is TT3, with only 8 matching chapters.
- While there are generally 1 or 2 variations in paragraph counts, significant discrepancies in a few chapters can help identify where the translators made additions or omissions, such as the 4-paragraph increase in TT1 in chapters 4 and 5 or the 5-paragraph decrease in TT3 in chapter 21

Table 5. Comparison of Paragraph Counts in Chapters

Object Chapter	ST	TT1	TT2	TT3
1	10	13	10	10
2	33	31	35	32
3	29	30	29	29
4	12	16	11	11
5	18	22	17	20
6	12	13	13	13
7	31	34	31	31
8	25	27	25	27
9	16	17	16	17
10	58	57	58	58
11	20	19	20	20
12	11	11	11	12
13	53	53	53	53
14	38	38	38	37
15	48	50	48	49
16	5	5	5	5
17	30	31	30	32
18	8	8	8	8
19	9	9	8	8
20	11	11	11	12
21	62	63	63	57
22	17	17	17	18
23	7	8	8	8
24	29	30	29	32
25	48	48	49	49
26	86	85	86	87
27	10	10	9	13
Total	736	756	737	748



Overall, the results indicate that TT2 is the closest to the ST. Although TT1 and TT3 also have matching chapters, they are further from the ST than TT2.

Comparison Of Word Frequency Lists

In the word frequency list comparison, as shown in Table 6, the first 10 most frequently used words in the objects are examined. The comparison is conducted based on the words in the ST, one by one.

Être: The most frequently used word in the ST is the most commonly used word in TT1 and TT3, as it ranks as the fourth most frequently used word in TT2. However, it has been demonstrated in various studies that Japanese is a "suru (doing) language" (Ikegami,1981; Kuribayashi, 2018). In other words, Japanese here is a "language of doing" as opposed to French and English, which are among the "languages of being".³

Avoir: The word "Avoir" means "to have," and it is observed that it is used in the same meaning only in TT1's "have." It is believed that this situation arises from the similarity of language structures.

Ne: This negative adverb appears to be used in each language, although not in the same order. The words are as follows: "not," "nai," "net." Furthermore, the more frequent use of this word compared to its synonyms in the translation texts may imply that the negative usages in the translation texts are less than in the ST. On the other hand, the word "net" in TT3 is a negative term used as a noun meaning "no." In TT2, "nai" is used as an adjective since it negates adjectives more often in the text.

Petit: This adjective is represented as "little" in TT1 and "маленький" (malen'kij) in TT3. No corresponding word for this adjective is found in TT2. The reason for the absence of its equivalent in TT2 is that the title of TT2 means "The Prince of the Stars." The word "hoshi," meaning "star," is the sixth most frequently used word in TT2. Additionally, it is noted that the word "little" appears in two different forms in Russian, where the adjective "little" used to describe the word "prince" is identified as a noun.

Dire: This word which means "to say" has been frequently used in all translations, similar to "being." The words that convey this meaning are as follows: "say," "iu," and "skazat'."

Pas: This word is a negative adverb like "ne," which ranks third in the ST. The word can be evaluated in the same way as "ne" since it is used in the structure "ne ... pas."

Prince: This word which means "prince" has been used in all translations. Analyzing the frequency of the words "Prince," "ōji," and "princ" reveals that all are used more frequently than in the ST. While this word ranks seventh in the ST, it appears earlier in the translations.

Faire: This word meaning "to do" corresponds to "do" in TT1 and the most frequently used word "suru" in TT2.

Bien: No words corresponding to the word meaning "good" have been found in the translations examined.

Fleur: This word means "flower" and is used as "flower" in TT1 and "hana" in TT2.

³ Kuribayashi (2018) analyzed whether Turkish, like Japanese, tends to use the verb "yap-" instead of "ol-" based on "être/ol-" in Japanese and Turkish translations of Le Petit Prince, and whether Turkish, like Japanese, should be included in the category of "do (yap) language". The researcher states that the main feature of "ol (be)" expressions in Japanese is that they report becoming, not doing, by making the whole sentence intransitive, in other words, by hiding the subject. Another feature is that it is used to achieve polite discourse by blurring the meaning and functions as modality.



When examining the first ten most frequently used words in the ST, it is determined that 8 out of the 10 words from the ST has matched 7 words in the most frequently used list of TT1. Furthermore, 6 words in the ST has matched 5 words in TT2. Finally, it is found that 6 words in the ST has matched 6 words in TT3.

Table 6. Comparison of Word Frequency Lists

Object Freq. No	ST	TT1	TT2	TT3
1.	être (v) (537)	be (v) (870)	する (v) (350)	byt' / быть (v) (197)
2.	avoir (v) (344)	have (v) (271)	いう (v) (229)	принс / принц (n) (176)
3.	ne (adv) (323)	little (n) (253)	王子 (n) (175)	skazat' / сказать (v) (131)
4.	petit (adj) (219)	say (v) (224)	なる (v) (174)	malen'kij / маленький (n) (117)
5.	dire (v) (196)	not (adv) (196)	子 (n) (155)	оцен' / очень (adv) (69)
6.	pas (adv) (179)	prince (n) (185)	星 (n) (121)	malen'kij / маленький (adj) (67)
7.	Prince (n) (172)	do (v) (153)	ある (v) (104)	planeta / планета (n) (65)
8.	faire (v) (111)	very (adv) (81)	ひと (n) (104)	мощ' / мочь (v) (49)
9.	bien (adv) (100)	make (v) (80)	花 (n) (99)	sprašivat' / спрашивать (v) (46)
10.	fleur (n) (66)	then (adv) (74)	いる (v) (78)	нет / нет (n) (44)

Comparison Of Frequency Counts By Word Types

Looking at the table, the numerical values of each word type in the objects are shown. The comparison by word types is as follows:

Noun: The object closest to the ST is TT1. In TT2, the usage of nouns drops to 340, making it the furthest from the ST. TT3 shows an increase in noun usage.

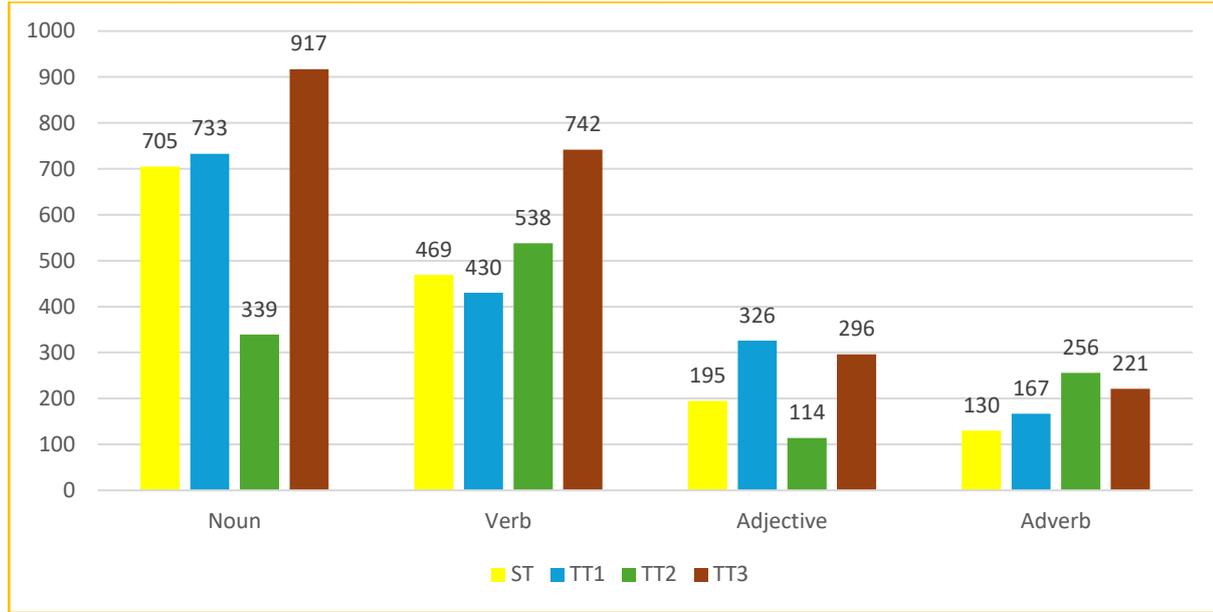
Verb: The usage of verbs in the objects shows less fluctuation compared to noun usage. Although the number of verbs in TT1 decreased compared to the ST, it remains the closest value. TT2 has not deviated much from the ST in verb usage. Finally, TT3 is again the furthest from the ST, as it is with nouns.



Adjective: TT1 has disrupted the closeness found in nouns and verbs, becoming the object furthest from the ST in terms of adjectives. Although the adjective usage in TT2 appears quite low, it has the closest value to the ST. TT3 falls between the values of TT1 and TT2.

Adverb: In this word type, the ST is observed to be the object that uses adverbs the least. The closest value is found in TT1, while TT2 has the furthest value. TT3 is positioned between TT1 and TT2, similar to adjective usage.

Table 7. Comparison of Frequency Counts by Word Types



Overall, it is understood that the object closest to the ST in all values is TT1. TT2 and TT3 are almost identical in their distance from the ST.

Evaluation

The analyzed findings have been compared under subheadings in the context of matricial norms. The equivalence evaluation of the objects is made based on the subheadings involved in the comparison, and secondly, the overall positions of the translation texts are determined as a result of these comparisons. Accordingly, the equivalence evaluations of the translations at the micro level are as follows:

- At the word count level, TT1 is determined to be an "adequate" translation, while TT2 and TT3 are classified as "acceptable."
- In the evaluation of different word counts, TT2 is identified as an "adequate" translation, and TT3 is considered "acceptable." TT1 is closer to "adequate" compared to the other translations.
- At the sentence count level, TT1 is an "adequate" translation, whereas TT2 and TT3 are both "acceptable."
- In the paragraph count evaluation, TT2 is deemed an "adequate" translation, while TT1 and TT3 are "acceptable."
- For the number of paragraphs in sections, TT2 is again an "adequate" translation, with TT1 and TT3 classified as "acceptable."
- In the word frequency evaluation, TT1 is an "adequate" translation, while TT2 and TT3 are "acceptable."
- Finally, in the evaluation of word type frequency, TT1 is "adequate," whereas TT2 and TT3 are "acceptable."



As a result of these findings, the equivalence evaluation of the translations in the context of matricial norms is as follows:

- TT1 is found to be in the "adequate" category in most comparisons. Thus, it is concluded that TT1 is an "adequate" translation due to its closeness to the source text in the context of matricial norms.
- TT2 is observed to be nearly evenly split between "adequate" and "acceptable." However, in the overall evaluation, it is concluded to be an "acceptable" translation due to its closer relation to the target text rather than the source text in the context of matricial norms.
- TT3 is seen as "acceptable" in all comparisons. It is concluded that TT3 is an "acceptable" translation due to its proximity to the target text in the context of matricial norms.

Conclusion

The importance of technological developments in translation studies and the need for new methods and powerful tools to be used in the field have been emphasized commonly in Translation Studies. However, there are not many studies in which text mining has been used in this context. In this study, both text mining as a method has been observed and KH Coder as a text mining tool has been tested. The usability and adequacy of KH Coder in comparative translation studies has been examined in the light of matricial norms, one of the translation norms. The analysis has focused on *Le Petit Prince* and its English, Japanese, and Russian translations. Initially, findings obtained with KH Coder have been analysed, followed by their comparison.

The evaluation has been concluded with the result that KH Coder offers valuable insights within the context of matricial norms, confirming the equivalence of the compared findings. The English translation has been found 'adequate,' while the Japanese and Russian translations have been classified as 'acceptable.'

Based on the findings and evaluations, the research has come up with the following answers:

Does KH Coder analysis yield sufficient results in evaluating matricial norms?

KH Coder effectively presents interpretable data in the context of matricial norms through various feature extractions. All findings have been interpretable within this framework, demonstrating the software's adequacy in evaluating matricial norms. Consequently, it has been concluded that KH Coder is suitable for such evaluations in relevant studies.

Is KH Coder usable and adequate in comparative study within the context of translation norms?

During the comparisons, imbalances in KH Coder data from different structural languages have been observed. Therefore, certain restrictions have been implemented for a more accurate analysis. Ultimately, KH Coder has been found to be usable in the context of translation norms, though it may not always be adequate for comparisons involving structurally different languages. However, it is believed that analyses of structurally similar languages yield more reliable results, making the software both usable and adequate.

Can KH Coder be a tool used in comparative translation studies?

Considering the features mentioned in the methodology section, the findings from the analysis, and the final evaluations, it has been concluded that KH Coder is a suitable tool for comparative translation studies.

During the data collection phase of the study, restrictions have been imposed because the research objects were from languages with different typologies. The study has been concluded more soundly with the restrictions imposed. As in the example of French and English, if all the objects had been from the same typology, perhaps there would have not been any need for restrictions. However, since the objects of examination have been selected from languages we are fluent in, restrictions have had to be made. In future studies to be conducted with KH Coder in comparative translation studies, it is recommended that languages with typological similarities be addressed or restrictions be imposed.



Another suggestion is that the tool should be developed to process multiple files simultaneously for comparative studies. Thus, it is thought that a possible comparison feature will be helpful to researchers not only in translation studies but also in studies conducted in other areas such as social media and survey studies.

In light of the obtained results, it is believed that this study will contribute to the field of translation studies regarding the diversity of translation comparisons, and that text mining will be better understood as a review method in this area. Additionally, it is anticipated that this work will serve as a valuable guide for future studies utilizing KH Coder.



Bibliography

- Adalı, E. (2012). Doğal Dil İşleme. Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi, 5(2).
- Aslan, E., & Kuşçu, E. (2015). Çeviride Yapay Zeka Uygulamaları. Atatürk Üniversitesi Kazım Karabekir Eğitim Fakültesi Dergisi, 63-77.
- Atan, S. (2020). Metin Madenciliği: İmkânlar, Yöntemler Ve Kısıtlar. *Mehmet Akif Ersoy Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*(31), 220-239.
- Baker, M. (1993). Corpus Linguistics and Translation Studies - Implications and Applications. In M. Baker, G. Francis, & E. Tognini-Bonelli, *Text and Technology: In Honor of John Sinclair* (pp. 233-252). John Benjamins Publishing Company.
- Bassnett, S. (2002). Translation Studies. London: Routledge.
- Bozan, N. (2021). A Text Mining Approach to 'Operational Norms'. *transLogos Translation Studies Journal*, 57-80.
- Bozan, N. (2022). Extending the boundaries in descriptive translation analysis: A text mining approach. İstanbul: İstanbul 29 Mayıs Üniversitesi, Sosyal Bilimler Enstitüsü, Çeviribilim Ana Bilim Dalı, Yayımlanmamış Yüksek Lisans Tezi.
- Gençer Baloğlu, Z. (2022). Yapay Zekâ Söseki'yi Nasıl Okur?-Eyleyenler Modeli Oluşturma ve Duygu Analizi Denemesi. *Humanitas-Uluslararası Sosyal Bilimler Dergisi*, 10(20), 160-182.
- Higuchi, K. (2016). A Two-Step Approach to Quantitative Content Analysis : KH Coder Tutorial using Anne of Green Gables (Part I). *Ritsumeikan Social Sciences Review*, 52(3), 77-91.
- Higuchi, K. (2017). A Two-Step Approach to Quantitative Content Analysis : KH Coder Tutorial Using Anne of Green Gables (Part II). *Ritsumeikan Social Sciences Review*, 53(1), 137-147.
- Hotho A., Nürnberger A. & Paaß G. (2005). A Brief Survey of Text Mining. *Journal for Language Technology and Computational Linguistics*, 20(1), 19-62.
- İkegami, Y. (1981). *Suru to Naru no Gengogaku: Gengo to Bunka no Taiporoji e no Shiron*. Tokyo: Taishūkanshoten.
- Jurafsky, D., & Martin, J. H. (2008). 8. Speech Synthesis. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. içinde Prentice Hall.
- KH Coder 3.Beta.07f. (2023). Retrieved from <https://kncoder.net/en/>
- Kuribayashi, Y. (2018). Karşılaştırmalı Dilbilim Bakış Açısıyla Türkçede ol- Fiili. *Yeni Türkiye*(99), pp. 433-437.
- Ökubo, Y. (2014). *星の王子さま*. Retrieved from https://www.aozora.gr.jp/cards/001265/files/46817_24670.html
- Olohan, M. (2004). *Introducing Corpora in Translation Studies*. Routledge.
- Pym, A. (2010). *Exploring Translation Theories*. Routledge.
- Roldán-Riejos, A. (2021). Study of The Little Prince.
- Saint-Exupéry, A. d. (2012). *Маленький принц*. (N. Gal, Trans.) Moskova: EKSMO.
- Saint-Exupéry, A. d. (2017). *Le Petit Prince* (1 ed.). Genç Destek Yayınları.
- Saint-Exupéry, A. d. (2017). *The Little Prince*. (K. Woods, Trans.) Londra: Egmont.
- Seker, S. E. (2015). Metin Madenciliği (Text Mining). *YBS Ansiklopedi*, pp. 30-32.
- Toury, G. (2012). *Descriptive Translation Studies - and beyond* (revised ed.). Amsterdam: John Benjamins Publishing Company.
- Toury, G. (2021). The Nature and Role of Norms in Translation. In L. Venuti, *The Translation Studies Reader* (4. ed., pp. 197-210). Routledge.
- Türker, F. (1991). Dilbilimin çalışma alanlarında bilgisayar kullanımı. *Dilbilim Araştırmaları Dergisi*, 2, 137-143.
- Vardar, B. (1998). *Dilbilimin Temel Kavram ve İlkeleri*. İstanbul: MULTILINGUAL.
- Yazıcı, M. (2005). *Çeviribilim Temel Kavram ve Kuramları*. İstanbul: MULTILINGUAL.