T.C.
İSTANBUL 29 MAYIS ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
ÇEVİRİBİLİM ANABİLİM DALI

EXTENDING THE BOUNDRIES IN DESCRIPTIVE TRANSLATION ANALYSIS: A TEXT MINING APPROACH

(YÜKSEK LİSANS TEZİ)

Nalan BOZAN

Danışman:
Prof. Dr. Işın ÖNER

İSTANBUL
2022

T. C.
İSTANBUL 29 MAYIS ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
ÇEVİRİBİLİM ANABİLİM DALI

EXTENDING THE BOUNDRIES IN DESCRIPTIVE TRANSLATION ANALYSIS: A TEXT MINING APPROACH

(YÜKSEK LİSANS TEZİ)

Nalan BOZAN

Danışman:
Prof. Dr. Işın ÖNER

İSTANBUL

2022

T. C.

İSTANBUL 29 MAYIS ÜNİVERSİTESİ

SOSYAL BİLİMLER ENSTİTÜSÜ MÜDÜRLÜĞÜNE

Mütercim Tercümanlık Anabilim Dalı, Çeviribilim (İngilizce) Tezli Yüksek Lisans Programı'nda 010518YL09 numaralı Nalan BOZAN'ın hazırladığı "*Extending the Boundaries in Descriptive Translation Analysis: A Text Mining Approach*" konulu yüksek lisans tezi ile ilgili tez savunma sınavı, 01/02/2022 günü 13:00 – 14:00 saatleri arasında yapılmış, sorulan sorulara alınan cevaplar sonunda adayın tezinin başarılı olduğuna oy birliği ile karar verilmiştir.

Prof. Dr. Işın ÖNER
İstanbul 29 Mayıs Üniversitesi
(Tez Danışmanı ve Sınav Komisyonu Başkanı)

Prof. Dr. Ayşe Banu KARADAĞ
Yıldız Teknik Üniversitesi

Dr. Öğr. Üyesi Nilüfer ALİMEN
İstanbul 29 Mayıs Üniversitesi

BEYAN

Bu tezin yazılmasında bilimsel ahlak kurallarını uyulduğunu, başkalarının eserlerinden yararlanılması durumunda bilimsel normlara uygun olarak atıfta bulunulduğunu, kullanılan verilerde herhangi bir tahrifat yapılmadığını, tezin herhangi bir kısmının bu üniversite veya başka bir üniversitede başka bir tez çalışması olarak sunulmadığını beyan ederim.

Nalan BOZAN

01/02/2022

ÖZ

Gideon Toury 1970'lerin sonunda Betimleyici Çeviri Araştırmaları (BÇA) kapsamında 'çeviri normları' kavramını geliştirmiştir. Buna göre; amaç, çeviri davranışına ilişkin genellemeler yapabilmek için çeviri olgusunu betimlemek ve açıklamaktır. Toury, insan çabasının kısıtlılığı nedeniyle çeviri çözümlemesinde her şeyin ele alınmasını "savunulamaz" olarak değerlendirir. Bu tezde, metin madenciliğinin 'çeviri süreci normları' bakımından yapılacak çözümlemeyi kolaylaştıracağını savunuyorum. Metin madenciliği, yapısal olmayan metin verilerinden yeni bilgiler keşfetmek için metnin bir yazılım aracılığıyla otomatik olarak çözümlenmesidir. Çalışmanın amacına özel olarak geliştirilen yazılımın insan çabası karşısındaki etkililiğini değerlendirmek amacıyla, başka bir yüksek lisans tezinin bütüncesi karşılaştırma için referans noktası olarak alındı. Bütüncenin hazırlanması ve işlenmesinin ardından, salt insan çabasıyla elde edilemeyecek ve araştırmada incelenen çevirilere dair araştırmacılara yeni bilgiler sunan sonuçlar elde edildi. Bu bakımdan çalışmanın BÇA için dikkate değer bir katkı sunma potansiyeli olduğunu düşünüyorum. Tezde kullanılan yazılımın farklı metin türleri üzerinde uygulanabilirliğini doğrulamak için daha fazla çalışma gerekmektedir.

Anahtar Kelimeler

Çeviri Normları, Çeviri Süreci Normları, Betimleyici Çeviri Araştırmaları, BÇA, Metin Madenciliği, Kapsamlı Çözümleme

ABSTRACT

The notion of ‘translational norms’ was introduced by Gideon Toury in the late 1970s within Descriptive Translation Studies (DTS). The aim was to describe and explain translations to make generalizations regarding translational behavior. Toury deems examination of everything “untenable” due to human limitation. In this thesis, I argue that text mining can facilitate exhaustive analysis on ‘operational norms.’ Text mining is the automated analysis of text data to extract and discover new information. For the purpose of this research, a customized tool was designed by a software developer. In order to test the effectiveness of the software versus human effort, the corpus of a master’s thesis was taken as a reference point to compare the results. The corpus underwent a preparation stage for achieving optimal results. After processing the corpus, operational norms were extracted to be interpreted. Consequently, the study provides results that cannot be obtained by sole human effort and enables researchers to gain further insights about the analyzed translations. This research therefore has a potential for making notable contribution to DTS. Further efforts are needed to improve the tool and verify its viability for different text types to ensure such contribution.

Keywords

Translational Norms, Operational Norms, Descriptive Translation Studies, DTS, Text Mining, Exhaustive Analysis

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to my advisor, Prof. Dr. Işın ÖNER. It was her profound guidance and valuable support that made this thesis possible. I am honored to work with such a distinguished and kind scholar who dedicates her life to her students, her profession, and research in Translation Studies.

I am also indebted to Prof. Dr. Ayşe Banu KARADAĞ who taught the course where I first studied the subject of this thesis. Her invaluable insights and continuous encouragement helped me to build the foundation for this thesis.

I would like to offer my thanks to Asst. Prof. Nilüfer ALİMEN for her kindness and all the assistance she has provided so far. I am fortunate to be the student of such valuable academicians. Their boundless enthusiasm for research, learning and teaching has been a source of inspiration.

A special thanks to my husband, Kerim BOZAN for always being there to support me on this journey. It wouldn't have been possible to carry out this research without his efforts and contributions as a software developer. Also, I want to thank my sister, Dilara Nihal, for keeping me motivated throughout this process.

Finally, I would like to extend my thanks to TUBİTAK for awarding me 2210-A General Domestic Graduate Scholarship and providing me with the financial means to complete this thesis.

TABLE OF CONTENTS

TEZ ONAY SAYFASI ii
BEYAN iii
ÖZ iv
ABSTRACT v
ACKNOWLEDGEMENTS vi
TABLE OF CONTENTS vii
LIST OF ABBREVIATIONS viii
LIST OF TABLES ix
LIST OF FIGURES x

INTRODUCTION 1
1. THEORETICAL BACKGROUND 3
1.1 Key Concepts 3
1.1.1 Operational Norms 3
1.1.2 Text Mining 5
1.2 Research Problem 7
1.3 Chronological Background of the Thesis 8

2. LITERATURE REVIEW 11
2.1 Group 1: Translation as the Main Focus 11
2.2 Group 2: Translation as a Research Object 14
2.3 A Previous Case-Study 15

3. METHODOLOGY 17
3.1 Corpus 17
3.2 Tool: O.N.E. 18
3.2.1 Alignment Feature 20
3.2.2 Search Feature 21
3.2.3 Textual-Linguistic Analysis Feature 22
3.2.4 Matricial Analysis Feature 25
3.3 Research Procedure 26

4. RESULTS 30
4.1 Alignment 31

4.2 Matricial Analysis 36
4.2.1 Word Count 37
4.2.2 Sentence Count 38
4.2.3 Paragraph Count 39
4.2.4 Footnote/Endnote Count 40
4.2.5 Image Count 40
4.2.6 Average Word Length 41
4.2.7 Average Sentence Length 42
4.2.8 Average Paragraph Length 43
4.2.9 Content 44
4.3 Textual-Linguistic Analysis 45
4.3.1 Most Frequent Words 45
4.3.2 Personal Titles 47
4.3.3 Proper Nouns 47
4.3.4 Exclamations 49
4.3.5 Non-defined Words 49
4.3.6 Measurement Units 51
4.3.7 Idioms 52
4.3.8 Slang 54

5. DISCUSSION 57

CONCLUSION 59
REFERENCES 60
APPENDICES 63
ÖZGEÇMİŞ 76

LIST OF ABBREVIATIONS

abbr.	abbreviation
DTS	Descriptive Translation Studies
cf.	*confer*, compare
ç.n.	*çevirmenin notu* [translator's note]
DM	data mining
docx	file extension for Microsoft Word documents
e.g.	*exempli gratia*, for example
etc.	*et cetera*, and so forth
fig.	figure
IE	information extraction
i.e.	*id est*, that is
KDD	knowledge discovery in database
n.d.	no date
no.	number
NER	named entity recognition
OCR	optical character recognition
p.	page
PDF	portable document format
rev. ed.	revised edition
segm.	segment
ST	source text
s.v.	*sub verbo*, under the word
TS	Translation Studies
TT-1	Target Text 1
TT-2	Target Text 2
TTs	Target Texts

LIST OF TABLES

Table 1. Corpus of Duymaz's (2020) Research 18
Table 2. Corpus of the Current Research 18
Table 3. Analyzed Elements 30
Table 4. Exact Matches of TT-1 (From Segment 589 Onwards) 34
Table 5. Alignment Results for TT-2 (by Segment Range) 35
Table 6. Content of ST and TTs 45
Table 7. Most Frequent Words (First 20 Entries) 46
Table 8. List of Slang 54

LIST OF FIGURES

Figure 1. Upload page of O.N.E. 19
Figure 2. Search panel of O.N.E. 21
Figure 3. Identification process for proper nouns. 24
Figure 4. OCR on ABBYYFineReader. 27
Figure 5. OCR of TT-1 on ABBYYFineReader. 28
Figure 6. Interface of O.N.E. 28
Figure 7. First aligned segment of TT-1. 32
Figure 8. First aligned segment of TT-2. 32
Figure 9. Beginning of misalignment for TT-1 on segment 217. 33
Figure 10. Partial alignment for TT-1 between segments 217-224. 33
Figure 11. End of alignment for TT-1 and TT-2. 36
Figure 12. Results for word count. 37
Figure 13. Results for sentence count. 38
Figure 14. Results for paragraph count. 39
Figure 15. Results for footnote/endnote count. 40
Figure 16. Results for image count. 41
Figure 17. Results for average word length. 42
Figure 18. Results for average sentence length. 43
Figure 19. Results for average paragraph length. 43
Figure 20. Results for length (side-by-side). 44
Figure 21. Results for personal titles. 47
Figure 22. Results for proper nouns (first 8 entries). 48
Figure 23. Results for exclamations. 49
Figure 24. Results for non-defined words (first 11 entries). 50
Figure 25. Results for measurement units (first 7 entries). 51
Figure 26. Results screen for idioms. 53
Figure 27. Results screen for slang. 56

INTRODUCTION

When Translation Studies (TS) is examined historically, one sees several approaches. These studies had been guided by prescriptive researches for many years. However, this dominant prescriptive approach left its place to a descriptive one. Gideon Toury (2012, 20) highlights the importance of descriptive branch in a discipline and describes his endeavors within target-oriented paradigm as "geared primarily towards the descriptive-explanatory goal of supplying exhaustive accounts of whatever has been presented/regarded as translational within a target culture." Nevertheless, he considers analysis of everything "untenable" referring to the difficulties in the establishment of sampling rules (71). Even though no strict statistical methods are pointed out in his work, Toury does not leave the advances in computer sciences out of the picture indicating the possibility to analyze much bigger samples (92). Previous research has shown that using computerized corpora and text mining can make a substantial contribution by supplying more data to analyze. However, such research does not place particular emphasis on 'operational norms.' Therefore, the purpose of this thesis is to facilitate extraction of operational norms in descriptive translation analysis through text mining in order to make exhaustive analysis possible, which can offer larger amount of data hence further insights on the norms.

In accordance with the aim of this thesis, a tool was developed by a software developer. This study seeks to explore if the tool -or text mining- allows for a full-scale analysis, and if so, how it achieves this. Additionally, capabilities and limits of the tool are tested. The software is designed for analysis exclusively on operational norms. The justification behind this choice will also be explained in the thesis. Finally, the study aims to evaluate the contribution of this thesis to descriptive translation analysis.

The thesis is organized in five chapters. Chapter 1 outlines the theoretical framework of the thesis, introducing the key concepts and theoretical background that this research is built around. Additionally, some background information is given on the research itself: the milestones it has passed so far and how its methodology have evolved throughout the process. Chapter 2 is devoted to the overview of the key literature that is relevant to this research. Previous studies are discussed under two groups according to the disciplines in which they are carried out. A special focus will be given on a specific

study with a view to providing the foundation for the thesis. Chapter 3 describes the methodology employed in this research, including the procedures, the corpus, and the software used in the analysis. In Chapter 4, results are presented based on the tasks that the software performs on source and target texts. Chapter 5 focuses on the interpretation of the results in general. Finally, conclusion chapter brings together the main areas covered in the thesis and includes the final comments on the research.

1. THEORETICAL BACKGROUND

The theoretical background of this study is formed by two concepts that walk in parallel to each other. One of them is the 'operational norms' as introduced by Toury (1978, 1995, 2012); and the other is text mining. First concept constitutes the theoretical framework, while the latter is a method that I think is correspondent with this framework. In this chapter, these concepts will be explicated along with the research problem and purpose.

1.1 Key Concepts

As an interdisciplinary research, this thesis is based on two concepts from two disciplines: (i) 'operational norms' as a term of TS, and (ii) "text mining", which is a term of Computer Sciences. In order to state the research problems, these two terms and the theoretical framework that 'operational norms' fall within should be explained first.

1.1.1 Operational Norms

According to Toury (2012), main goal of TS is to describe, explain and predict translation phenomena in order to become an autonomous discipline, and he proposes a research methodology under the umbrella of 'Descriptive Translation Studies (DTS)'. As a key concept in DTS, 'translational norms' were introduced by Toury in 1978 for regularities of translational behavior. Describing translation as "a culturally-determined kind of activity" (2012, 61), Toury points out the cultural significance of translation activities and the social role that translator plays. This socio-cultural dimension of translation includes norms.

With reference to concepts of sociology, anthropology and sociopsychology, Toury defines norms as social values shared by a community concerning appropriate behavior and refers to translation as a norm-governed activity. Nevertheless, the notion falls into a descriptive category rather than a prescriptive one. In other words, the focus is not on evaluating translations, but investigating the regularities of translational behavior. Toury postulates that norms determine "the type and extent of equivalence actually exhibited by a translation vis-à-vis its source" (2012, 85). Investigation of norms therefore occupies an important place in understanding how this postulate of equivalence

is realized whether in a single translation, or in the work of a single translator or translators, in a particular historical period or in any other subgroup.

Toury (2012, 81-85) introduces three types of norms: "initial norms," "preliminary norms," and "operational norms." The initial norm refers to the basic choice that the translator – whether consciously or unconsciously – makes between the norms valid in the source culture or those that are active in the target culture. Overall, there are two extreme orientations to choose from: Leaning on the source norms determines the "adequacy" of translation whereas adherence to the norms of target culture determines the "acceptability." Having said that, the term "initial" does not infer that such choice is made only once during the act of translation or it is always the first one to be made. Rather, it can be repeated throughout the act.

Preliminary norms, on the other hand, deal with two interrelated considerations: (i) "translation policy", which refers to the selection of text types, individual texts, human agents and groups etc. and (ii) "directness of translation" which refers to the threshold of tolerance for indirect translation. According to Toury, it can be said that translation policy exists in the degree that the choices are found to be non-random. As far as directness of translation is concerned, Toury elucidates tolerance for indirect translation through the following questions:

> Is indirect translation permitted at all? In translating from what languages / text-types / periods (etc.) is it permitted / preferred / prohibited / tolerated? What are the permitted / prohibited / tolerated / preferred mediating languages and why are they given that status? (Toury 2012, 82)

The third type is operational norms**.** They are divided into two further types: "matricial norms" and "textual-linguistic norms." These two types of norms entail the decisions that the translator makes during the act of translation process. Operational norms affect the text's "textual make-up and verbal formulation" and its "matrix -i.e., the way linguistic material is distributed in it" (2012, 82).

Matricial norms are related to the distribution of linguistic material in the text, completeness of translation, and segmentation of target-text material. Toury (2012, 83) suggests that norms may also determine the extent to which additions, omissions, segmentation operations and location changes are applied in the translation. The scholar

underlines that what we want is not a true explanation anyway, but explanatory hypotheses.

Textual-linguistic norms, on the other hand, affect how linguistic material is selected, target text is formulated or, the original material is replaced. According to Toury (2012, 83), although some of these norms may be similar to the norms governing the production of non-translational texts in the target culture, such similarity cannot be used as a framework for actual research.

1.1.2 Text Mining

One of the key concepts that this research is built around is 'text mining'. Peter Jackson and Isabelle Moulinier (2002, 2-3) defines this concept as "function of software or hardware components in a computer system which analyze or synthesize spoken or written language." Main objective in text mining is to "turn text into data so that it is suitable for analysis" (Zanini and Dhawan 2015).

According to Chunyu Kit and Jian-Yun Nie (2015, 510), "text mining is considered a variation or extension of data mining (DM), which is also known as knowledge discovery in database (KDD)" and the researchers point to a difference between text mining and data mining: "Instead of working on databases, [text mining] works on unstructured texts." A pre-processing stage is required before algorithms can work on unstructured data. In another words, unstructured data must turn into a structured form. Then text mining tasks, including but not limited to 'information extraction,' 'text categorization,' 'text clustering,' 'text summarization,' and 'sentiment analysis' can be performed. These tasks can be described briefly as follows:

- **Information extraction (IE):** "Extracting prespecified types of facts from written texts or speech transcripts, and converting them into structured representations" (Ji 2009). Typical subtasks of IE are named entity recognition (NER), relation recognition, event recognition, and terminology extraction.
- **Text categorization:** Automatically assigning textual documents into some predefined categories based their content. It is also known as text classification (Shen, Text Categorization 2009).

- **Text clustering:** Automatically grouping textual documents into clusters based on similarity in their content (H. Li 2009).
- **Text summarization:** Extracting the most important information from a text in order to create an abridged version for a certain job and user (Shen, Text Summarization 2009).
- **Sentiment analysis:** "Research work of mining user opinion data. It is also known as opinion mining" (Cai and Sun 2009).

Text mining is used in a variety of applications in different areas. For instance, it is employed in social media platforms in order to understand target audience's reactions to a specific event/idea/product. Another example of applied areas is bioinformatics where text mining is used to identify entities such as drugs, receptors, enzymes, toxins, genes and their features. We can also see text mining applied in marketing to analyze customer behavior and identify trends. There are many more areas where text mining is utilized like customer care services (e.g., automated analysis of customer feedback), insurance (e.g., fraud detection), academia (e.g., plagiarism detection), and many more purposes that it can be used for such as knowledge management, risk management, cybercrime prevention, business intelligence, spam filtering and so on.

A closely related concept is 'natural language processing' or NLP. Basically, NLP helps computers to understand and handle human languages. There are "two major goals" of NLP according to Olivia Kwong Oi Yee (2015, 563): (i) "to enable human-computer interaction with human languages as the medium," and (ii) "to build language application systems which require considerable human language abilities and linguistic knowledge." Text mining systems use several NLP techniques like 'tokenization,' 'lemmatization,' 'parsing,' 'stemming,' and 'stop word removal' in order to build inputs for machine learning models. These techniques are defined as follows:

- **Tokenization:** Process of splitting text into tokens (word units)
- **Lemmatization:** Process of mapping inflectional variants of a word to its base form or lemma (Kit and Jian-Yun 2015, 497)
- **Parsing:** Process of analyzing a sentence in terms of its syntactic structure (Oi Yee 2015, 567)

- **Stemming:** "Process by which word endings or other affixes are removed or modified in order that word forms which differ in non-relevant ways may be merged and treated as equivalent" (Paice 2009, 2790)
- **Stop word removal:** Process of removing semantically irrelevant words in the text, such as frequently occurring words like articles ('a', 'an', and 'the') or first-person pronoun 'I'.

1.2 Research Problem

Bringing text mining and extracting operational norms arises from a point in Toury's theory. Toury (2012, 71) mentions difficulties in the establishment of sampling rules for translational behavior or the results of such behavior. He argues that an absolute consideration of everything is "untenable". Nevertheless, he does not rule out the possibility of advances in computing world which may enable more and much bigger samples and remarks that "much energy should still be directed toward the crystallization of systematic research methods, including quantitative ones" (92).

Considering the fact that today's technology and resources allow translation researchers investigate operational norms to a larger extent, such possibility can be seen as a door left open for researchers to go through stimulating search for new research methods or tools. To put it in another way, the motivation here is not to fill a gap in Toury's theory but to develop his hypothesis on 'crystallization of systematic research methods.'

Another theoretical motivation for this research can be found in Mona Baker's study (1993). Baker discusses the impact of availability of corpus on the study of translation in relation to some of the applications of corpus techniques in the applied branch of TS. According to Baker, "the potential for using large computerised corpora in translation studies" needs to be explored for the development of descriptive branch of TS. Even though the corpus-driven part of her study is not directly relevant to this thesis, Baker's approach is noteworthy because it calls for new techniques and a set of tools "to develop a descriptive branch of the discipline with well-defined objectives and an explicit program" (1993, 248).

1.3 Chronological Background of the Thesis

The previous sections explain the reason for initiating this study and its theoretical starting point. From that point to this current thesis, the work has passed through a number of stages, or to put it more precisely, milestones. This section aims to briefly talk about these stages and show how the study evolved throughout the whole research process.

To divide the process in very broad terms, the research has passed four milestones until it reached its current state. The study first began with a research conducted as part of a master's course in which an interdisciplinary perspective was the focus. The identification of the abovementioned theoretical base was originally the result of that research. The aim was to test if text mining enables exhaustive analysis on translational norms, which Toury deems "untenable." The choice of the corpus to be examined was one of the key parts of the research as it required texts where sole human effort would be insufficient in descriptive analysis or at least provide incomplete results. Therefore, a large volume of text and two translations of that text were determined as source and target text(s) respectively. The source text was Henry Fielding's novel *The History of Tom Jones, a Foundling* (1749). The target texts are Mina Urgan's translation into Turkish (Tom Jones: Sokakta Bulunmuş Bir Çocuğun Hikayesi 1945) and its retranslation (Tom Jones 1990) published at a later date. The reason behind the choice of target text is actually hidden in Urgan's own words. The first explanation is found in the translator's foreword of the retranslation:

> Bu önsözü bitirirken, çevirmen olarak bir açıklama yapmam gerekiyor: *Tom Jones*'u eskiden de çevirmiştim. Milli Eğitim Bakanlığı Dünya Klasikleri arasında kırk yıl önce çıkan çevirim, yanlış olmamakla birlikte, Türkçe açısından kötü, hatta çok kötüydü. Nurullah Ataç, bana kalırsa yüzde yüz haklı olarak, kullandığım Osmanlıca sözcüklerden ötürü beni uyarmıştı o sırada. Eski çevirimden nerdeyse bir tek tümce bile almayan bu yeni çeviriyi yaparken, vaktiyle işlediğim bir günahın bağışlanmasını da istedim. Eğer çevirileri birine sunmak geleneği olsaydı, yeni çevirimi Nurullah Ataç'ın anısına sunardım. (Fielding 1990, 66)

My translation:

> Concluding this foreword, I must make a statement as a translator: I have translated *Tom Jones* before. Although it was not inaccurate, my

> translation that was published forty years ago among the World Classics of the Ministry of National Education was bad, even terrible in terms of Turkish. At that time, Nurullah Ataç gave me a one hundred percent rightful warning about the Ottoman words I used. Making this new translation, which did not include almost a single sentence from my old translation, I also asked for forgiveness for a sin I once committed. If there was a tradition of dedicating translations to someone, I would dedicate my new translation in memory of Nurullah Ataç.

Describing the Turkish that she used in her first translation as "bad", Urgan further discusses the reason for retranslating *Tom Jones* in her book titled *Bir Dinozorun Anıları* ("Memoirs of a Dinosaur"):

> Gençliğimde o güzel Türkçeyi bilmediğimi açıkça itiraf etmeliyim şimdi. Dilbilgisi kurallarına uygun bir Türkçe kullanmakla yetinirdim sadece. Sözcük dağarcığım perişan bir durumdaydı. Fransızca ya da İngilizce bir sözcüğün, ancak bir ya da iki Türkçe karşılığı aklıma gelirdi. Örneğin, "grue" ya da "whore" sözcüklerine karşılık, ancak "orospu" ya da "fahişe" sözcüklerini düşünebilirdim. Oysa sürtük vardı, yosma vardı, sokak kadını vardı, aşifte vardı, fettan vardı, kaltak vardı; bunların arasında da ince farklar vardı. Ama dilimin zenginliklerini ve inceliklerini bilmediğim için, gençliğimde çevirdiğim kitapları - bu arada çok uzun bir kitap olan *Tom Jones*'u- başından sonuna kadar yeniden çevirmek zorunda kaldım. (Urgan 2013, 70)

My translation:

> I must openly confess now that I didn't know that beautiful Turkish when I was young. I would just content myself with using a grammatically correct Turkish. My vocabulary was in a miserable state. I could only think of one or two Turkish equivalents of a French or English word. For example, as equivalent of "grue" or "whore," I could only think of *"orospu"* or *"fahişe"*. However, there was a *yosma, sokak kadını, aşifte, fettan, kaltak*; and there were subtle differences between them. But because I didn't know the treasures and intricacies of my language, I had to retranslate the books I had translated when I was young from cover to cover – among them is *Tom Jones,* which is a very long book by the way.

In my eyes, these statements turned into an open invitation for an analysis in terms of operational norms. Examining the source text of 81,184 words and target texts of 59,410 and 59,327 words only by human effort and analyzing them in terms of translational norms would have taken too much time and wouldn't have been feasible, or

at least a thorough analysis might not be achieved. Precisely for this reason, my hypothesis was that text mining would facilitate analysis on operational norms by providing more data and allowing a more detailed examination.

The second milestone was my research carried out as part of a master's course on scientific research methods. At this stage, the study did not include an analysis done by a human to test the hypothesis outlined in the previous section. Therefore, it was not possible at that stage to compare the effectiveness of pure human effort with the performance of text mining tools. For that reason, the corpus of the research was changed to John Ronald Reuel Tolkien's novel *The Hobbit* and its Turkish translations, namely the corpus of this master's thesis research. This corpus had been subjected to descriptive analysis before by a human in another study. Thus, it would allow for comparisons with the results of that study.

The third stage, which can be described as another milestone, was the publication of the work done in the earlier stage (see Bozan 2021). The publication process served as a close-up of the work in which scientific presentation of the research was a matter of utmost concern to the author. The experience was illuminating regarding issues such as organization of the text, the clear and comprehensible transfer of sentences or words to the reader, compliance with the current scientific writing guidelines, and accurate citation. In other words, the review process led to small yet significant changes to eliminate ambiguity in some expressions. It even helped the author realize the need for an update in the software used in this research to provide user with clearer data. The publication phase was therefore a valuable process that clarified the author's view on the research in general.

Finally, the present thesis can be considered as the last phase of this research process. In the aforementioned second stage, there were some parts regarding textual-linguistic norms section of the analysis that required further development. For this reason, this research has carried on without changing its corpus for the sake of integrity of the findings and to have a baseline for comparison. Ultimately, it would maximize the benefit to gain from text mining.

2. LITERATURE REVIEW

There have been a number of studies which employ similar processing methods used in this research. Although these studies are of interdisciplinary nature, they can be divided into two according to how they deal with translation. First group researches take translation as the main focus of analysis to develop the relevant field of TS while the latter group uses translation as a research object to test methods or applications in the relevant branch of Computer Sciences. Yet, to this author's knowledge, no research in the context of text mining has been explicitly driven by the 'operational norms' as defined within Toury's study (1978, 1995, 2012).

Only one research (Çetin 2009) can be considered indirectly relevant to the present thesis since the author approaches 'translational norms' as recommendations to develop algorithms of computerized translation systems working with artificial intelligence. Naming those recommendations as "artificial translation norms", Özge Çetin (2009) proposes a model comprising of various steps based on Toury's translational norms. Yet, the model seems to serve merely as a questionnaire for human translators rather than algorithmic recommendations considering its inapplicability to machine translation.

In this chapter these two groups of studies will be reviewed. Additionally, a special focus will be given on a previous case-study. The study is significant for this thesis since it serves as a theoretical incentive to study on 'operational norms.'

2.1 Group 1: Translation as the Main Focus

Mona Baker (2000) argues that little or no attention has been paid thus far to describe the style of translator(s) and therefore outlines a methodological framework to investigate a literary translator's style while introducing a translational corpus. The corpus includes several translations by individual experienced literary translators and also several works by the same author translated by different translators. All the translations in the corpus are made in English from a variety of source languages, both European and non-European. Baker also uses a software tool that was developed to process corpora semi-automatically. For the purpose of investigating the style of individual literary translator, Baker identifies some patterns related to "type/token ratio; average sentence length; variation across texts; frequency and patterning of SAY (the most frequent reporting verb

in English)" (2000, 249). In her analysis, she aims to determine if individual literary translator has a distinctive style other than the source author's style by comparing two different translators' translations into English. One major drawback of such comparison is that the source texts subject to analysis in Baker's study are not only different texts, but they are also dissimilar as their overall number of words and even source languages are not identical. In that case, identifying what is attributable to the translator and what is a direct reflection of the source text would be questionable. In order to address this problem, one should compare the translations of the same source text to eliminate source text-distinctive features. Even Baker directs the same question at the end of her study.

> How can we best distinguish stylistic elements which are attributable only to the translator from those which simply reflect the source author's style, general source language preferences, or the poetics and preferences of a particular subset of translators? Should we try to? [...] Should we perhaps be comparing different translations of the same source text into the same target language, by different translators, thus keeping the variables of author and source language constant? (Baker 2000, 261)

However, reaching such conclusion can be explained with her study's exploratory nature as the study itself is an overall attempt to raise this type of questions. All in all, it gives an idea on how researchers can benefit from conducting a research using sub corpora of a large corpus and computational linguistics.

Another corpus-assisted translator style study was carried out by Defeng Li (2017) incorporating meta and paratexts into the analysis. Li first sets out the discussion on the issue of style in translation and mentions the increasing interest in translator's thumbprints in translation. Then he elaborates on translator style studies reviewing the existing corpus-assisted translator style studies and discussing their methodological issues. Li argues for "an integration of quantitative and qualitative analysis" (103) and stresses the importance of sense-making of the statistical data to examine translator's style in a larger socio-cultural and political context (109-111). Finally, he performs a case study comparing English translations of the classical Chinese novel *Hongloumeng* to illustrate his argument. For the quantitative analysis, the author sets two factors to examine: type-token ratio and sentence length; and makes use of three different software tools: (i) Wordsmith 4.0, (ii) a self-compiling software and (iii) Microsoft Access Database (2017,

118). Li's study is noteworthy for its emphasis on why and how findings of statistical analysis should be interpreted within their socio-cultural, ideological and political context.

In their interdisciplinary research collaboration, Tom Cheesman et al. (2017, 739) carry out an experimental project and develop "a web-based system which enables users to create parallel, segment-aligned multi-version corpora" revealing variation in multiple translations through visual interfaces. The intuition behind this project is to understand the variation among multiple translations and their reasons. The researchers' curiosity leads to the development of a toolset 'Translation Array Prototype' and an analysis involving a corpus of 38 digital texts of German translations of *Othello*, Act 1, Scene 3. The project offers an important opportunity to understand how digital tools "could be developed to explore patterns in variation among (re)translations" (743). Significance of Cheesman et al.'s project for the present thesis is the possibilities that the researchers projected, i.e. more sophisticated analysis, greater scope, mapping texts to Wordnets, dictionaries and thesauri, inclusion of variant translations evaluated by human and so on (755-756).

Johanna Monti et al. (2012, 1) adopt a different approach to text mining and suggest it as a way to "improve the documentary competence of translators" with a special focus on technical translation. The researchers' hypothesis is that text mining can speed up the preliminary phases of translation workflow, i.e. looking for information required for the translation task and terminological search for technical translation. In order to test their hypothesis, the researchers use a text mining tool, CATALOGA, in combination with an information retrieval application and/or a machine translation/translation memory system to perform analysis on Italian texts and their English translations (2). It can be said that the paper rather describes the capabilities and possibilities that the tool can offer to technical translators with limitation on multiword units, e.g. compounds. One criticism of this research can be made regarding the confirmation of the hypothesis, i.e. whether text mining can speed up the translation process, since there is no experiment on the efficiency of the tool against other ways of initial translation process.

2.2 Group 2: Translation as a Research Object

Kemal Altıntaş, Fazlı Can and Jon M. Patton (2007) use time-separated parallel translations as corpus for quantification of language change. Focusing on the Turkish language during the second half of the twentieth century, the researchers determine a set of features to quantify diachronic language changes e.g. vocabulary richness, word lengths, word stems, and suffixes. Due to agglutinative nature of Turkish language, Altıntaş et al. develop stemming tools to avoid inaccuracy in the measurement of *types* (distinct words). According to the concluding observations, it is seen that length of Turkish words have become significantly longer while the opposite is true for their stems, and also a significant increase in suffix lengths and shrinkage in vocabulary richness have been noted. From the perspective of this thesis, this work exemplifies a case where translation serves as an object to apply text mining methods on for any purpose other than studying translation itself ultimately for the development of TS. It also suggests some methods that can be adopted in the present research.

Fazlı Can, Ethem F. Can and Ceyhun Karbeyaz (2010, 117) "introduce a method that aims to quantify translation relationship with no information other than the texts themselves." According to the method, texts -i.e. source text and its translation- are divided into blocks and then source and target texts are clustered separately to examine whether there is similarity between them, which will be considered as an indication of translation relationship by the authors. Because it offers a language-independent method, this work can pave the way for further researches on less investigated language pairs, especially those involve languages having different grammar rules such as English-Turkish.

Jon M. Patton and Fazlı Can (2012) perform a comparative stylometric analysis on James Joyce's *Dubliners* and its Turkish translation by Murat Belge using five style markers: (i) sentence length in terms of words, (ii) the most frequent words, (iii) token length, (iv) type length, and (v) type-token ratio. In other words, they use statistical methods to analyze and compare writing styles of the author and the translator. The researchers explain their aim as "to identify style-related features of the original work which are retained in translation" (210).

Another study involving translation as data on which text mining methods are employed is Sevilay Çalışkan's master's thesis "Text Mining Analysis of Translation, Social Communication and Literary Writing for Turkish." Çalışkan (2020, iii) "analyze[s] different types of Turkish text from different points of views, having an overall review on text mining in Turkish at the end". The research consists of four chapters, but only the first chapter is relevant to this thesis as it involves translation. The other chapters include either monolingual analysis of different text types or multilingual analysis of different texts in the same domain. In the first chapter, Çalışkan evaluates the "loyalty of translations" for the Turkish novel *Benim Adım Kırmızı* (*My Name is Red*) into English, French, and Spanish. Since the novel contains different fictional characters hence different styles, the evaluation is based on the hypothesis that "the styles of characters will be preserved to some point in the translations if the style has not [been] changed consciously by translator" (3). Six different stylistic features are determined in Çalışkan's analysis: (i) number of tokens, (ii) number of types, (iii) average word length for both types and tokens, (iv) average sentence length, (v) average vowel count for both token and types, and (vi) frequency of the most frequently used words.

2.3 A Previous Case-Study

There is one study that does not fit in any of the groups above yet requires special focus for this thesis. It is a research conducted by Işın Bengi-Öner (1999) involving an investigation on translational norms in three translations of the novel *La Dame aux Camélia* by Alexandre Dumas Fils. Bengi-Öner first mentions the importance of descriptive translation studies in the "tripartite network of relations" between the descriptive, applied and theoretical fields of the discipline. Accordingly, descriptive studies will bring a new perspective to the history of translation by revealing the facts related to translation and ensuring translations take their rightful place in diachronic and synchronic axes of the history. In addition, the findings of such researches will broaden the boundaries of translation theory and their data will ultimately contribute to the applied field (45).

Referring to the shift from translation to Translation Studies, Bengi-Öner remarks that such change is a reflection of the dominant approaches in the second half of the

twentieth century to dissolve and reconstruct institutionalized structures through questioning. She considers criticism as one of the effective parameters in the formation of institutionalized structures in the world of translation. Bengi-Öner argues that translation criticism is confined to the applied field, ignoring the findings of descriptive studies. Thus, she examines translations of *La Dame aux Camélias* made by Ahmed Midhat (1880), Mustafa Nihad Özön (1937) and Mithat Cemal (1937) (cf. Bengi-Öner 1999). The aim of her research is to determine the translational norms of these translations and compare the findings with the views of literary critic İsmail Habib Sevük on these translations. The results show that there is a great difference between the two.

One of the findings of Bengi-Öner's research is particularly striking from the point of view of this thesis. The results show a big difference in length between Ahmed Midhat's translation and the other translations. Sevük interprets this difference by suggesting that Ahmed Midhat did not translate *La Dame Aux Camélias* word-by-word but paraphrased it. However, Bengi-Öner's analysis reveals that the length of translated text is actually equivalent to the length of the source text even though the number of pages and paragraphs are less compared to the source text. Bengi-Öner justifies the difference explaining that Ahmet Midhat wrote in Ottoman script, in which vowels are often omitted, and noting that Midhat merged the paragraphs in one part. In this respect, her research has an influence on the history of literature in the sense that its findings changed the hitherto determinations of the era's literary historians. On the other hand, the research has particular significance for this thesis as Bengi-Öner has limited the scope of her analysis to the sections 1 and 26 of the corpus to determine the matricial norms stating that "it is not possible to use a full 299-page book" (1999, 47; my translation). From this point of view, Bengi-Öner's work can be considered as a case confirming the need for facilitating descriptive analysis through text mining.

3. METHODOLOGY

The method adopted in a research is closely related to the nature of the research. As the title of this thesis suggests, this research aims to make exhaustive analysis possible through text mining within the framework of DTS. This aim thus requires a corpus suitable for testing this hypothesis, and a tool employing text mining techniques. This chapter gives information on the corpus and the software used in this thesis. The chapter also provides an overview of the research procedures that were followed to carry out this study.

3.1 Corpus

Three criteria were considered when determining the corpus of the research: (i) It must comprise of a voluminous source text, and hence voluminous target texts which would make a thorough analysis infeasible or at least difficult due to human limitation. (ii) The translations must have been analyzed before by a human within the scope of translational norms. (iii) There must be at least one re-translation of the source text so that results could be compared for two different translations.

There is a text fulfilling all the three criteria: The fantastic novel *The Hobbit*, which was written by John Ronald Reuel Tolkien in 1937, and its translations from English into Turkish. Translations of *The Hobbit* into Turkish have been analyzed before by another TS researcher within the framework of DTS (see Duymaz 2020). This established a common ground or a reference point for comparison. Namely, this corpus would enable the researcher to compare the results obtained through text mining to those that have already been obtained by sole human effort. It should be noted that the intention of such comparison is not to evaluate the quality of findings, but rather to explore the capacity or limits of text mining. Consisting of 96,707 words (according to the software developed for his research), *The Hobbit* met the criterion for text length. Also, the novel has been translated into Turkish more than once.

Table 1 provides the details regarding the corpus of Duymaz's (2020) research, while Table 2 presents the corpus analyzed in this research. It can be seen that Duymaz analyzes four translations of *The Hobbit* into Turkish. In the present research, e-pub edition of the source text (Tolkien, The Hobbit 2009) is processed, and it will be referred

to as ‘ST’ hereinafter. For translations of the novel, we examine the scanned version of Emel İzmirli’s translation (Tolkien, Hobbit: Oradaydık ve Şimdi Buradayız 1996) and e-book edition of Gamze Sarı’s translation (Tolkien, Hobbit 2015). They will be referred to as ‘TT-1’ and ‘TT-2’ respectively. Also, the abbreviation ‘TTs’ will be used to refer to the two target texts in general.

Table 1

Corpus of Duymaz's (2020) Research

	Title	Year	Publisher	Translator
Source Text	The Hobbit	1990	HarperCollins Publishers	-
Target Texts	Hobbit, Oradaydık ve Şimdi Buradayız	1996	Altıkırkbeş Yayın & Mitos Yayıncılık	Emel İzmirli
	Hobbit, Oradaydık ve Şimdi Buradayız	1997	Altıkırkbeş Yayın	Esra Uzun
	Hobbit	2007	İthaki Yayınları	Gamze Sarı
	Hobbit	2018	İthaki Yayınları	Gamze Sarı

Table 2

Corpus of the Current Research

	Abbr.	Title	Year	Publisher	Translator	Edition
Source Text	ST	The Hobbit	2009	HarperCollins Publishers	-	E-book
Target Texts	TT-1	Hobbit, Oradaydık ve Şimdi Buradayız	1996	Altıkırkbeş Yayın & Mitos Yayıncılık	Emel İzmirli	Paperback
	TT-2	Hobbit	2015	İthaki Yayınları	Gamze Sarı	E-book

3.2 Tool: O.N.E

The corpus will be analyzed by a tool that was developed by a software developer[1] exclusively for the purpose of this research. The software is named *O.N.E* (*Operational Norm Extractor*), which will be referred to as ‘O.N.E’ or ‘the software’ from this point forward. O.N.E is a web-based software employing text mining techniques. The software makes it possible to upload as many texts as users wish to examine depending on the

[1] Kerim BOZAN. E-mail: kerim@bozan.biz

capacity of their servers. In other words, the more texts are uploaded, the more time it will take the software to process them.

First, users upload the source text and target text(s) to O.N.E as shown in figure 1 in Microsoft Word format (docx). Once the files are uploaded, the software extracts the contents of them and parses the extracted xml documents immediately to save them into corresponding tables in the database as sentences and paragraphs based on their tags. This process eliminates the necessity to upload the same documents again later as long as their file names are not changed, and it expedites the further operations. Following the upload, users can perform certain tasks under 4 main categories:

1. Alignment
2. Search
3. Textual-Linguistic Analysis
4. Matricial Analysis

It should be noted that each task is standalone. In other words, they run independently form the other tasks e.g. tasks under textual-linguistic analysis do not use the data obtained from the alignment task.

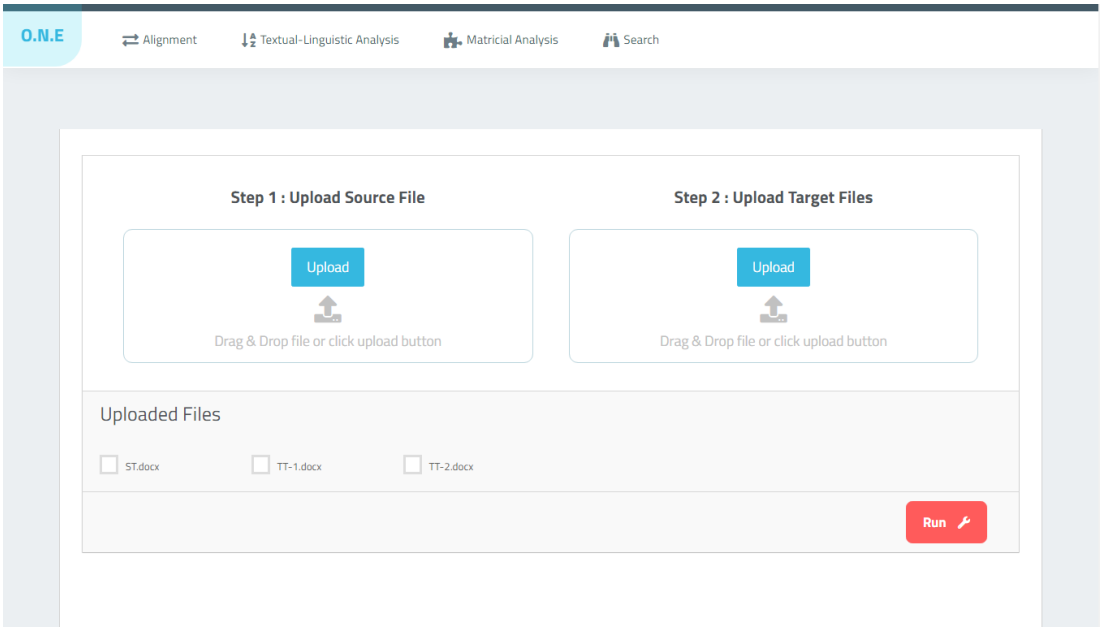


Figure 1. Upload page of O.N.E.

3.2.1 Alignment Feature

In simple terms, alignment feature compares a source text with its translation(s) to find the part of the target text(s) that corresponds to the relevant part in the source text. We can divide this process mainly into two stages: (i) detection of the first aligned segment[2], and (ii) alignment of the rest of the texts. For the first stage, O.N.E follows the procedure summarized below.

First, the software translates the segment in the source text using a bilingual database. Then, it calculates the match rate between the translation by O.N.E and the target segments. Here, it is necessary to specify the minimum number of characters that the target segment must contain to be encompassed in alignment. The purpose of such limitation is to eliminate the irrelevant data, e.g. contents in table of contents, cover page, title page, copyright page etc., and to detect the actual starting segment that the translation takes place. In our research, this number is determined in a way to maintain the optimum level of accuracy in results.

Next, O.N.E checks the match rates of the target segments and excludes the segments which do not pass a certain similarity threshold. In order to detect the first target segment containing translation in the most efficient way, this operation is limited to the first 100 sentences of the source text and the first 500 sentences of the target text(s). At this point, the software looks for the correspondent of the relevant source segment that achieves the highest match rate. For a segment to be set as the first segment of translation, meeting the highest match rate condition alone is not enough. Its segment number must also have the lowest value among the other candidates. It should be noted that this process does not take place at one time for the whole text but rather repeats itself for each sentence of the source text.

In the case of multiple target texts as in our research, this process takes place for the other target text(s) as well. Ultimately, O.N.E determines a first aligned segment for all the target texts; however, the target text that follows the source text i.e. the one which is uploaded first, after source text is uploaded- is taken as basis to identify the first segment of translation.

[2] A segment is a unit that O.N.E split from the source/target text. It can be a single word or a group of words like a phrase or a sentence.

Once the first segment is determined, it is aligned with the relevant source segment. This is where the second stage come in. The software examines the content of the aligned segment in target text 1 and compares it with the content of segments in target text 2. The segment that meets a certain match rate is accepted as the corresponding segment. When a match is found in target text 2 corresponding the segment of target text 1, this process ends.

The second stage involves the alignment of the rest of the texts. O.N.E handles this process using a certain algorithm. Whenever necessary, the software divides or merges segments in order to establish an accurate alignment.

3.2.2 Search Feature

O.N.E enables users to find a specific word unit in the source text and/or target text(s) through the search feature (see fig. 2). This feature shows the total number of instances where the searched word unit is used in the text. Also, it provides the user with the context in which the searched word(s) is used highlighting the searched entry in the sentence where it is used.

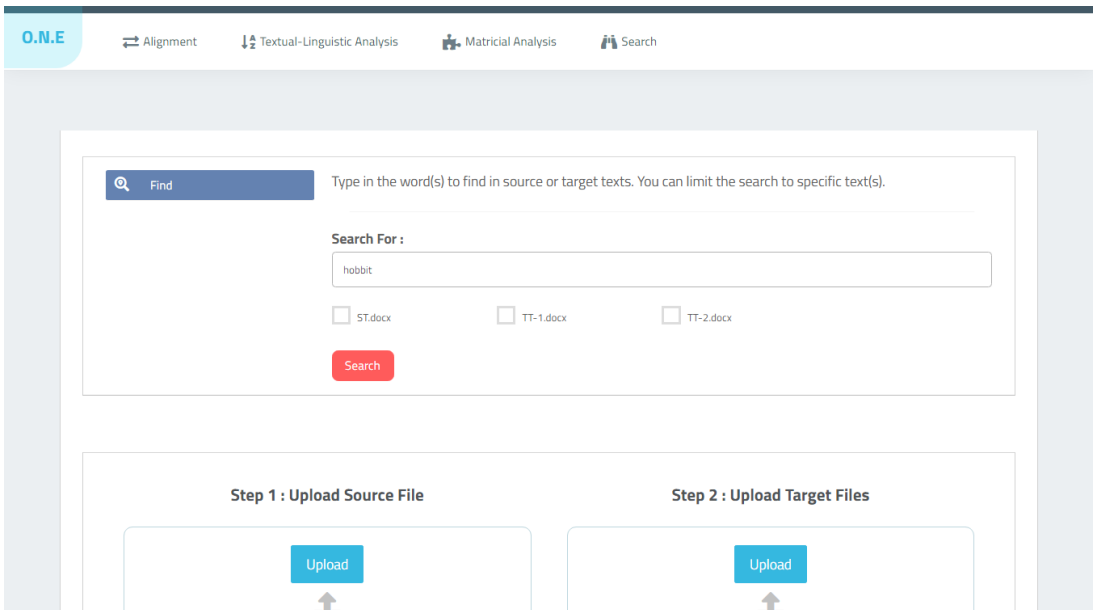


Figure 2. Search panel of O.N.E.

3.2.3 Textual-Linguistic Analysis Feature

Users can perform a number of tasks under the feature of Textual-Linguistic Analysis. Those tasks are described briefly as follows:

1. **Most Frequent Words:** This task searches for the most frequent words in source and target texts and displays them along with their frequency rate. The software first removes stop words and personal titles in the texts to provide the user with only important information.
2. **Personal Titles:** This task searches for the personal titles (e.g. Mr., Ms., Sir etc.) used in source and target texts. If a personal title is not included in the current library that the software uses for search, users can add more personal titles as entries.
3. **Proper Nouns:** This task searches for the proper nouns used in source and target texts. It finds matches according to a process involving certain criteria which is demonstrated in figure 3.
4. **Idioms:** This task searches for the idioms used in source and target texts based on a dataset compiled from *Cambridge International Dictionary of Idioms* and Turkish idioms category of Wiktionary[3]. The results are displayed along with their context.
5. **Slang:** This task searches for the slang used in source and target texts based on a dataset compiled from two websites for English slang and *Türkçenin Büyük Argo Sözlüğü* for Turkish slang[4]. Users can exclude a word unit which was identified as slang from the search results. (Duckworth n.d.) (Rader 1996) (Aktunç 1998)
6. **Non-defined Words:** This task searches for the words in source and target texts that does not exist in any dictionary or dataset used by the software.
7. **Exclamations:** This task searches for the exclamations used in source and target texts (e.g. hey, ouch, wow etc.). The software first searches for exclamation marks

[3] *Cambridge International Dictionary of Idioms* (Cambridge, UK: Cambridge University Press, 1998); "Kategori:Türkçe deyimler," Wiktionary: The Free Dictionary, Wikimedia Foundation, last modified September 10, 2020. https://tr.wiktionary.org/wiki/Kategori:Türkçe_deyimler.

[4] "A Dictionary of English Slang & Colloquialisms," Ted Duckworth, accessed December 23, 2021, http://www.peevish.co.uk/slang/index.htm; "The Online Slang Dictionary (American, English, and Urban slang," Walter Rader, accessed December 23, 2021, http://onlineslangdictionary.com; Hulki Aktunç, *Türkçenin Büyük Argo Sözlüğü (Tanıklarıyla)* (İstanbul: Yapı Kredi Yayınları, 1998).

in the text and displays the total number of strings followed by an exclamation mark. In order to eliminate the instances where exclamation mark is used to emphasize the whole sentence as in exclamatory sentences, only the sentences consisting of maximum three words are listed as entries. However, it should be noted that exclamatory sentences containing three words may still appear in results. The software shows both the total number of exclamations (based on the use of exclamation mark) and the list of exclamations confirmed based on the maximum-three-words rule.

8. **Measurement Units:** This task searches for measurement units such as units indicating length (e.g. centimeter), volume (e.g. liter), area (e.g. acre), mass (e.g. gram), and temperature units (e.g. Fahrenheit) etc. The results are displayed along with the word preceding and following the measurement unit. If a measurement unit is not included in the current library that the software uses for search, users can add more measurement units.

Apart from datasets indicated above, O.N.E uses two datasets as the main dictionaries: WordNet 3.0 database (Princeton University 2010) and Zargan Dictionary (Bilgin 2016) are used for English and Turkish respectively.

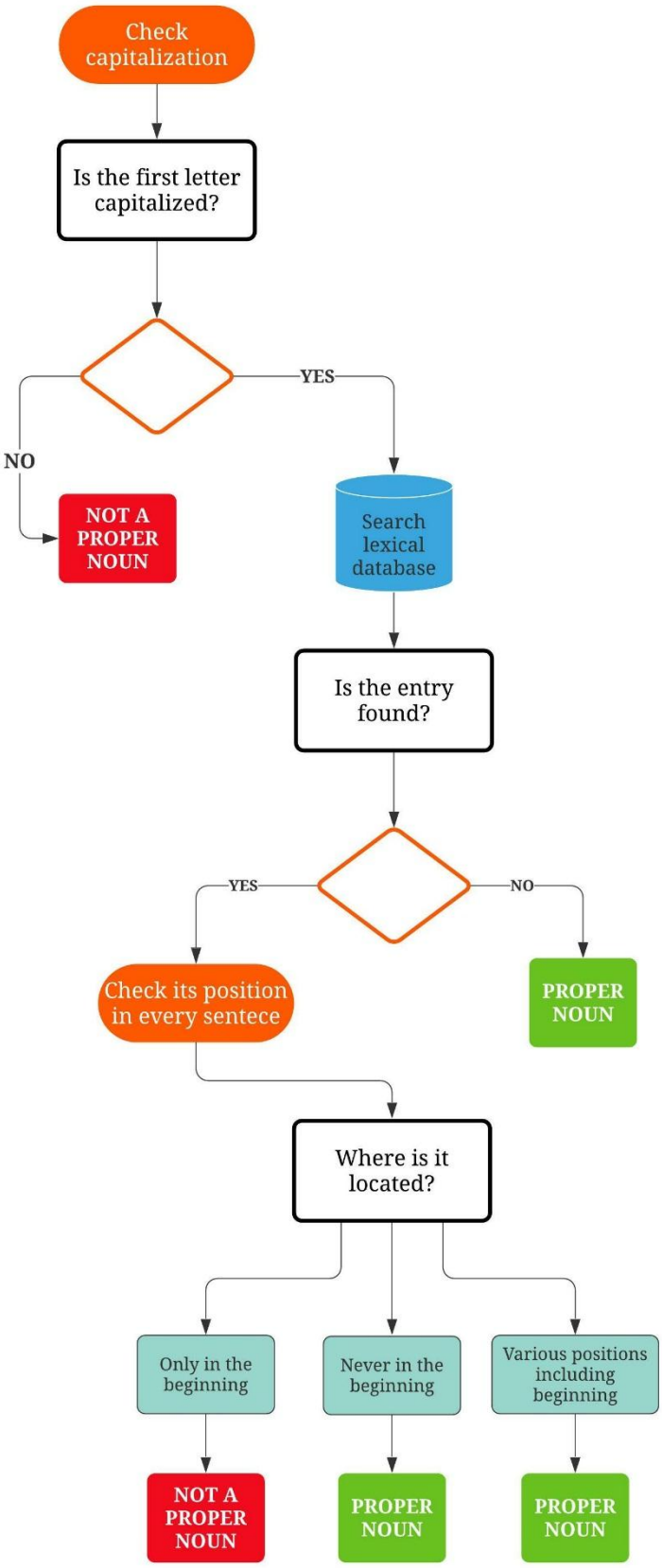


Figure 3. Identification process for proper nouns.

3.2.4 Matricial Analysis Feature

Matricial Analysis feature enables researchers/users to perform following tasks under three categories:

1. **Count:** The tasks performed under this category count the words/sentences/paragraphs/footnotes/images in the source and target texts.
    a. Word Count
    b. Sentence Count
    c. Paragraph Count
    d. Footnote/Endnote Count
    e. Image Count
2. **Length:** The tasks under this category calculates the average length of words/sentences/paragraphs in the source and target texts.
    a. Average Word Length
    b. Average Sentence Length
    c. Average Paragraph Length
3. **Content:** This task looks for if certain content elements exist in source and target texts. The software performs the task based on the structural tags provided by the word processor. It searches for the below-given content and identifies the parts that exist in the texts. Researcher/user can select the part(s) that they wish to search for specifically or they can exclude the part(s) from search.
    a. Book Cover
    b. Title Page
    c. Copyright Page
    d. Dedication
    e. Foreword
    f. Preface
    g. Prologue
    h. Acknowledgements
    i. Author’s Notes
    j. Author’s Bio
    k. Translator’s Notes

l. Table of Contents
m. Introduction
n. Epilogue
o. Appendix
p. Bibliography
q. Glossary
r. Index

3.3 Research Procedure

This section outlines the framework for how the research was conducted. The stages included in this research can be listed as follows:

1. Developing the software
    a. Identifying the tasks according to ‘operational norms’
    b. Developing the algorithms
    c. Creating the database
    d. Collecting the datasets
    e. Testing the software
    f. Fixing bugs
2. Preparing the corpus
3. Processing the corpus
4. Extracting and analyzing operational norms

The first stage entails the software's design and development, which requires interdisciplinary collaboration. O.N.E was developed by a software developer outside of the field of TS. Therefore, the required tasks, functions or features were first reported based on the aspects investigated within the scope of ‘operational norms.’ Then, the algorithms were developed accordingly. After the software developer created the database, which stores the data to be analyzed and other data that is required for analysis, datasets were collected. Finally, the software was tested and, where necessary, bugs were fixed.

The second stage involves the preparation of the corpus to convert it into a computer-readable digital format. First, e-pub version of ST was converted into PDF to perform optical character recognition (OCR) on it. For OCR, ABBYY FineReader was used and the resulted text was edited within the same application as demonstrated in figure 4. The editing process included the following operations: removing the (hyper)links, recognizing image areas, adjusting heading levels, replacing the dagger symbols (†) on page 10 of ST with the phrase 'See p. 28 and 55' to establish cross-references.

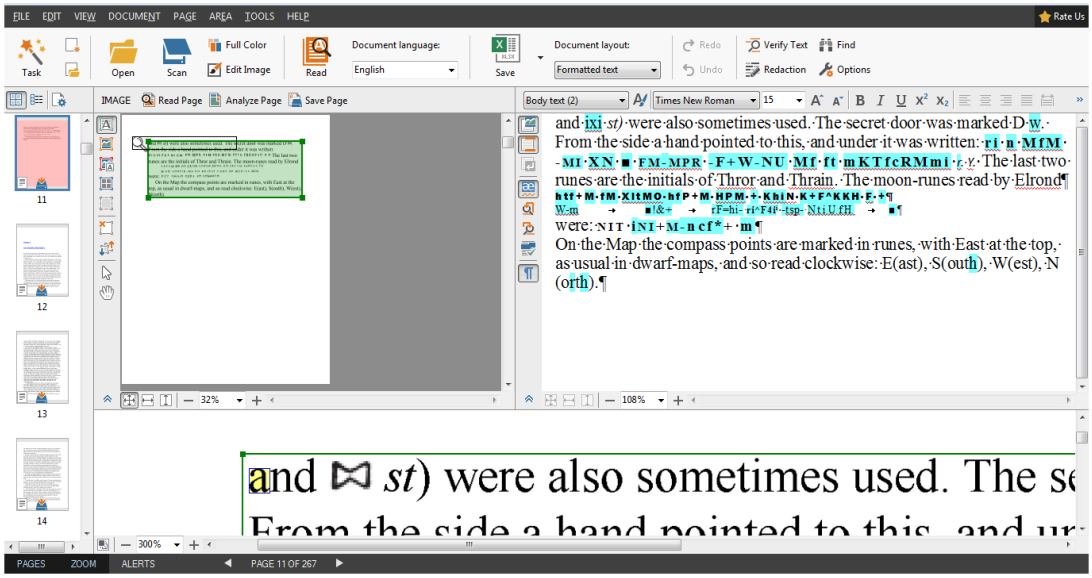


Figure 4. OCR on ABBYYFineReader.

Secondly, TT-1 and TT-2 were undergone OCR. OCR process for TT-1 (see fig. 5) required more effort and time compared to other texts in the corpus because the document was a scanned version of the paperback edition. The resulted text was edited, and the following tasks were performed as part of editing: correcting typos, removing redundant page/line breaks, recognizing image areas, adjusting heading levels, removing incorrectly detected headers and footers, adding line breaks where necessary. The key point of this stage was to provide the most flawless version of a document. In other words, the better optically recognized a scanned document is, the more accurate results that O.N.E produces.

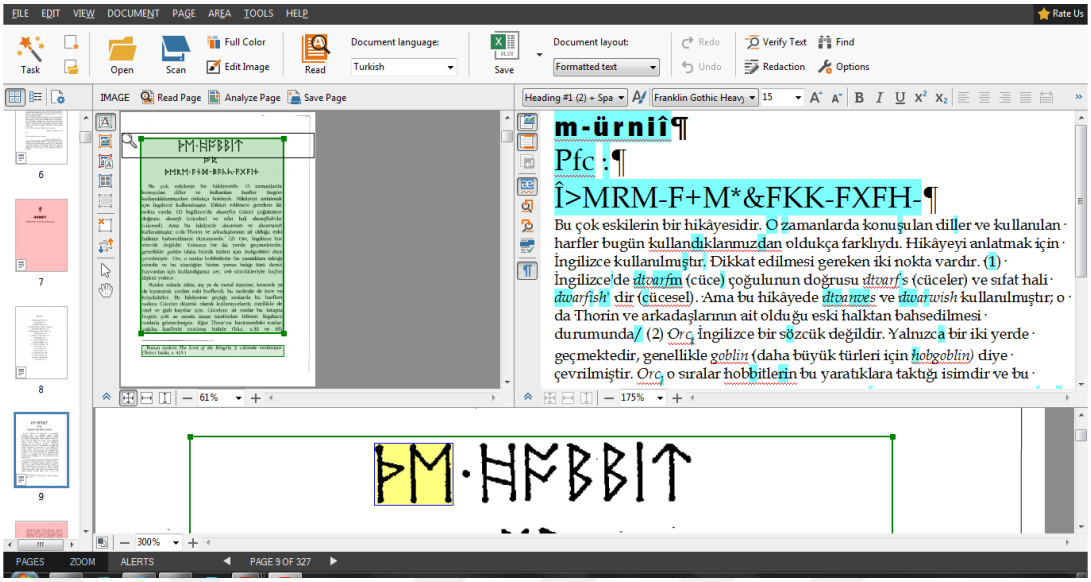


Figure 5. OCR of TT-1 on ABBYYFineReader.

In the third stage, the corpus was processed using the software developed in the first stage. In order to perform any task, the source and target text(s) must be uploaded into the system. In our case, one source text and two target texts were uploaded to the relevant area on the landing page. After that, the user/researcher can run the tasks through a user-friendly interface (see fig. 6).

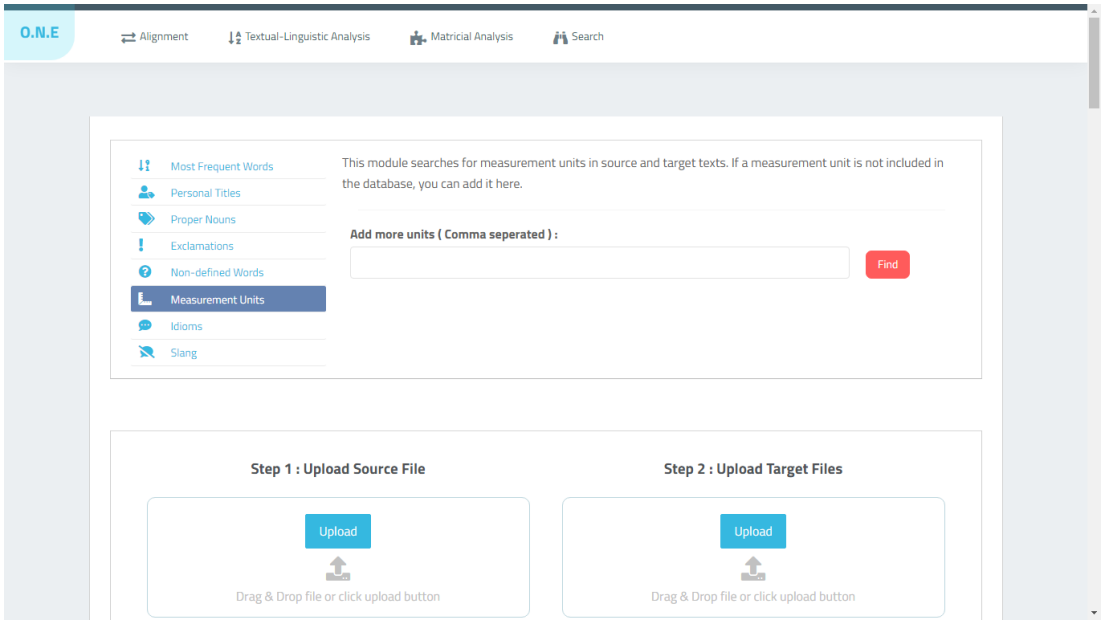


Figure 6. Interface of O.N.E.

Final stage provided the results of processing according to the tasks identified within the framework of ‘operational norms.’ O.N.E presented the results in a neat and easy-to-follow way via visualization through graphs and/or tables, which ultimately helped the researcher to analyze ‘operational norms.’

4. RESULTS

In this chapter, the results obtained from the analysis made by O.N.E on the corpus. In order to be able to compare the software's results to those acquired through human effort, this research involves the corpus and analysis of another study (see Duymaz 2020) as a reference point for comparison. For this reason, a closer look is required into the relevant elements examined in Duymaz's study. Table 3 shows the elements covered in her analysis vis-à-vis the elements of analysis of the present research.

Table 3

Analyzed Elements

Duymaz's (2020) Research	**Current Research**
Matricial Norms*	
-	Word count
	Sentence count
	Paragraph count
	Average word length
	Average sentence length
	Average paragraph length
	Footnote/Endnote count
	Image count
	Content
Textual-Linguistic Norms	
Translation of proper nouns	Proper nouns
Translation of neologisms	Non-defined words
Translation of pun	-
Translation of idioms	Idioms
Translation of colloquialisms and dialects	-
Translation of exclamations	Exclamations
Translation of food and beverage names	-
Translation of games	-
Translation of measurement units	Measurement units
-	Slang
-	Personal titles
-	Most frequent words

*****Duymaz (2020) does not specify any subcategories for matricial norms. Yet, details such as title of the texts, content of the covers, content parts (images, forewords etc.), number of pages, length of the texts, number of chapters and how they are presented are explained in paragraphs and then compared.

As shown in table 3, while some of the elements of Duymaz's analysis are covered in this thesis under the same category, there are differences in some parts. For example, Duymaz examines the matricial norms by describing them in paragraphs rather than presenting them a structured way. In this research, on the other hand, quantitative and statistical data are compartmentalized into levels such as character, word, sentence, and paragraph. One can justify this difference by the difference in nature and goals of these two studies.

Another difference applies for the elements examined under the category of 'Textual-Linguistic Norms.' While proper nouns, non-defined words, idioms, exclamations, and measurement units correspond to an element in Duymaz's research, translation of pun, dialects, food and beverage names, and games could not be included in this research because these elements are found unique to the analyzed texts and thus no datasets could be complied for them. 'Slang,' 'most frequent words,' and 'personal titles' are the elements that were not covered in Duymaz's analysis. Although the elements were identified based on the relevant part of Duymaz's research, the results will be presented according to the tasks performed by O.N.E as outlined in chapter 3.2. Screenshots from the software will also be added wherever it is necessary.

4.1 Alignment

The researcher can refer to the results of alignment task to confirm their prediction regarding the results of other tasks. For instance, they can see the source text content and its translation(s) side-by-side to interpret the results obtained from tasks under count, length and content categories.

As explained in chapter 3.2.1, TT-1 was uploaded after ST and taken as basis to determine the first aligned segment instead of TT-2. It is because some segments in ST and TT-1 are missing in TT-2, which causes a lower segment to be designated as the first segment while there is an alignable segment available in upper segments of TT-1. In other words, although TT-1 contains segments that can be aligned with the segments in ST, those segments would be excluded from alignment when TT-2 was uploaded first. A concrete example can be given over the target texts analyzed in this research:

As figure 7 shows, the first aligned segment in TT-1 corresponds to the 84th segment of ST. The first aligned segment in TT-2, however, corresponds to the 102nd segment of ST (see fig. 8). If ST was followed by TT-2 instead of TT-1, its corresponding segment would become the first aligned sentence. The alignment would consequently start from the 102nd segment of ST, hence exclude the segments between 84-102 of TT-1 from the results.

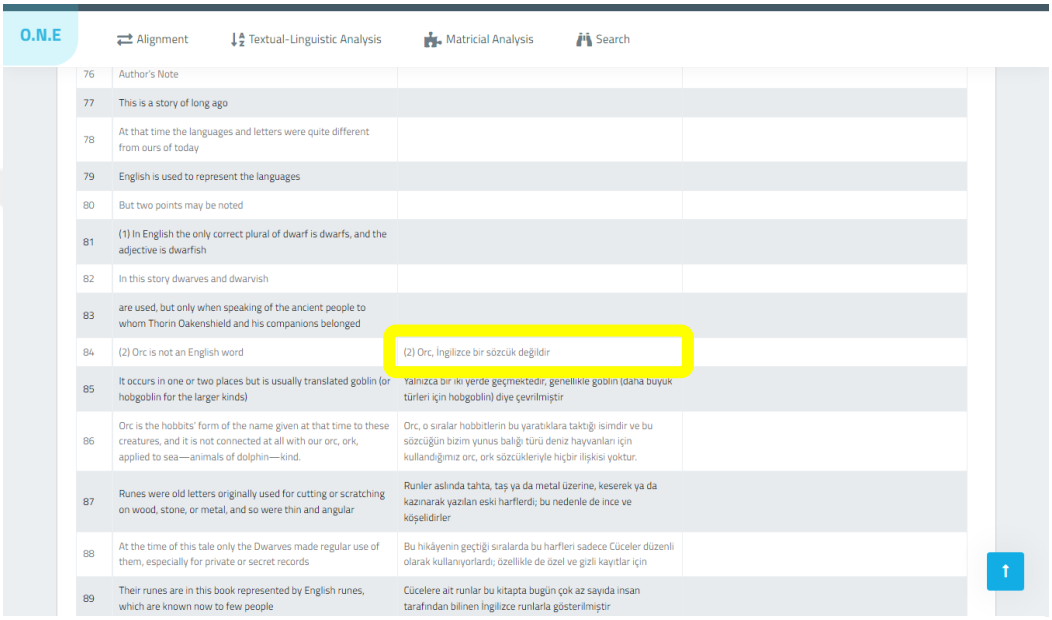


Figure 7. First aligned segment of TT-1.

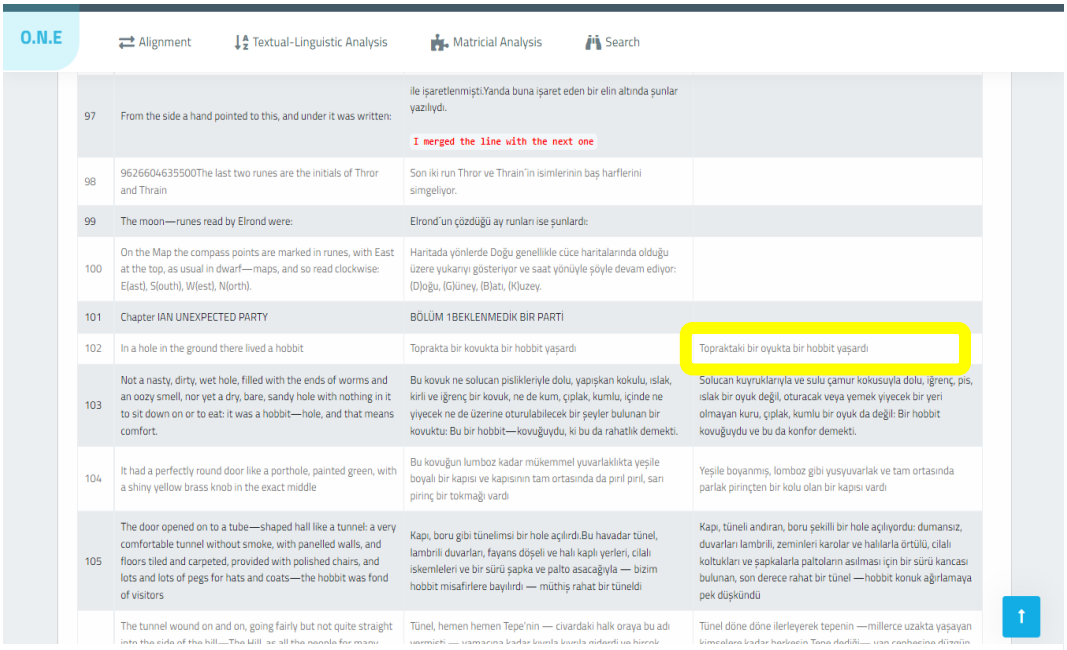
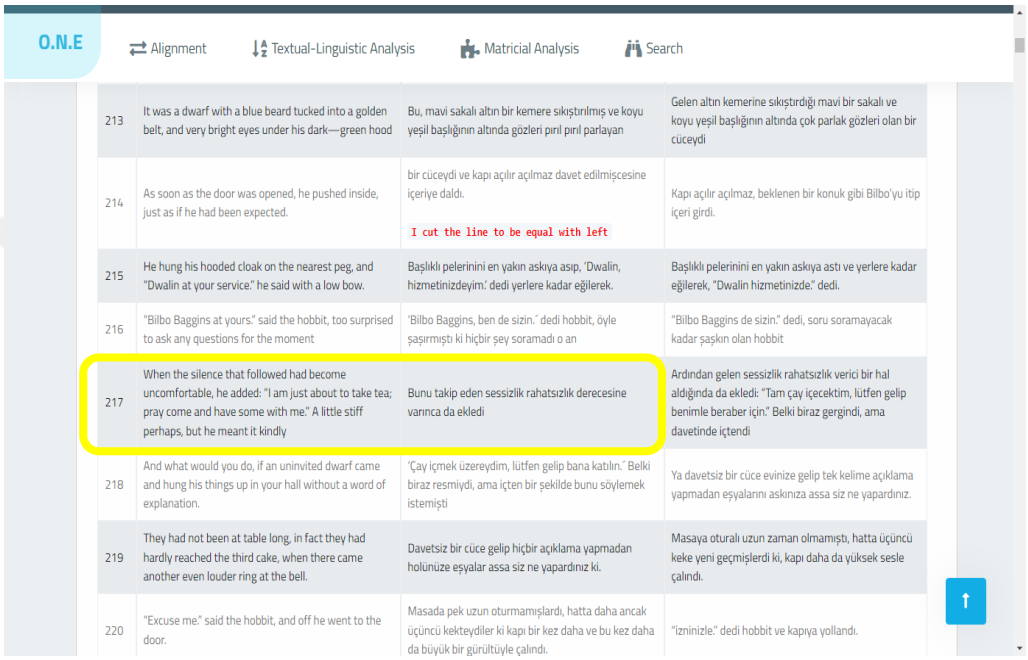


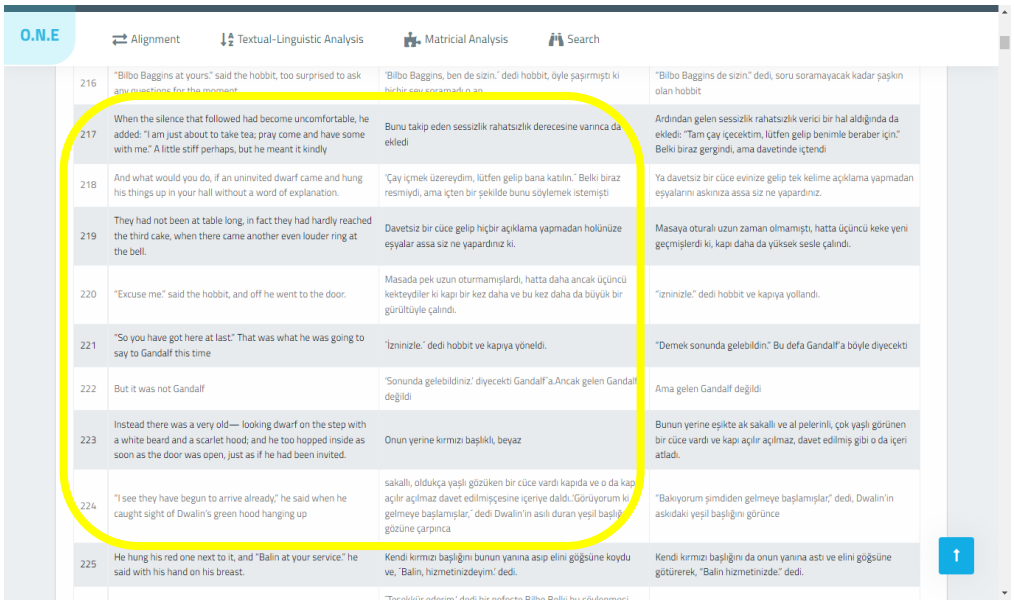
Figure 8. First aligned segment of TT-2.

There are 5,845 segments in ST that were determined by the software. When the results examined, it is seen that segments of TT-1 are successfully aligned between segment 84 (the first aligned segment) and segment 216. However, segments start to disalign and/or partially align in segments 217 (see fig. 9). This partial alignment continues until segment 224 (see fig. 10).



213	It was a dwarf with a blue beard tucked into a golden belt, and very bright eyes under his dark—green hood	Bu, mavi sakalı altın bir kemere sıkıştırılmış ve koyu yeşil başlığının altında gözleri pırıl pırıl parlayan	Gelen altın kemerine sıkıştırdığı mavi bir sakalı ve koyu yeşil başlığının altında çok parlak gözleri olan bir cüceydi
214	As soon as the door was opened, he pushed inside, just as if he had been expected.	bir cüceydi ve kapı açılır açılmaz davet edilmişcesine içeriye daldı. I cut the line to be equal with left	Kapı açılır açılmaz, beklenen bir konuk gibi Bilbo'yu itip içeri girdi.
215	He hung his hooded cloak on the nearest peg, and "Dwalin at your service." he said with a low bow.	Başlıklı pelerinini en yakın askıya asıp, 'Dwalin, hizmetinizdeyim.' dedi yerlere kadar eğilerek.	Başlıklı pelerinini en yakın askıya astı ve yerlere kadar eğilerek, "Dwalin hizmetinizde." dedi.
216	"Bilbo Baggins at yours." said the hobbit, too surprised to ask any questions for the moment	'Bilbo Baggins, ben de sizin.' dedi hobbit, öyle şaşırmıştı ki hiçbir şey soramadı o an	"Bilbo Baggins de sizin." dedi, soru soramayacak kadar şaşkın olan hobbit
217	When the silence that followed had become uncomfortable, he added: "I am just about to take tea; pray come and have some with me." A little stiff perhaps, but he meant it kindly	Bunu takip eden sessizlik rahatsızlık derecesine varınca da ekledi	Ardından gelen sessizlik rahatsızlık verici bir hal aldığında da ekledi: "Tam çay içecektim, lütfen gelip benimle beraber için." Belki biraz gergindi, ama davetinde içtendi
218	And what would you do, if an uninvited dwarf came and hung his things up in your hall without a word of explanation.	'Çay içmek üzereydim, lütfen gelip bana katılın.' Belki biraz resmiydi, ama içten bir şekilde bunu söylemek istemişti	Ya davetsiz bir cüce evinize gelip tek kelime açıklama yapmadan eşyalarını askınıza assa siz ne yapardınız.
219	They had not been at table long, in fact they had hardly reached the third cake, when there came another even louder ring at the bell.	Davetsiz bir cüce gelip hiçbir açıklama yapmadan holünüze eşyalar assa siz ne yapardınız ki.	Masaya oturalı uzun zaman olmamıştı, hatta üçüncü keke yeni geçmişlerdi ki, kapı daha da yüksek sesle çalındı.
220	"Excuse me." said the hobbit, and off he went to the door.	Masada pek uzun oturmamışlardı, hatta daha ancak üçüncü kekteydiler ki kapı bir kez daha ve bu kez daha da büyük bir gürültüyle çalındı.	"izninizle." dedi hobbit ve kapıya yollandı.

Figure 9. Beginning of misalignment for TT-1 on segment 217.



216	"Bilbo Baggins at yours." said the hobbit, too surprised to ask any questions for the moment	'Bilbo Baggins, ben de sizin.' dedi hobbit, öyle şaşırmıştı ki hiçbir şey soramadı o an	"Bilbo Baggins de sizin." dedi, soru soramayacak kadar şaşkın olan hobbit
217	When the silence that followed had become uncomfortable, he added: "I am just about to take tea; pray come and have some with me." A little stiff perhaps, but he meant it kindly	Bunu takip eden sessizlik rahatsızlık derecesine varınca da ekledi	Ardından gelen sessizlik rahatsızlık verici bir hal aldığında da ekledi: "Tam çay içecektim, lütfen gelip benimle beraber için." Belki biraz gergindi, ama davetinde içtendi
218	And what would you do, if an uninvited dwarf came and hung his things up in your hall without a word of explanation.	'Çay içmek üzereydim, lütfen gelip bana katılın.' Belki biraz resmiydi, ama içten bir şekilde bunu söylemek istemişti	Ya davetsiz bir cüce evinize gelip tek kelime açıklama yapmadan eşyalarını askınıza assa siz ne yapardınız.
219	They had not been at table long, in fact they had hardly reached the third cake, when there came another even louder ring at the bell.	Davetsiz bir cüce gelip hiçbir açıklama yapmadan holünüze eşyalar assa siz ne yapardınız ki.	Masaya oturalı uzun zaman olmamıştı, hatta üçüncü keke yeni geçmişlerdi ki, kapı daha da yüksek sesle çalındı.
220	"Excuse me." said the hobbit, and off he went to the door.	Masada pek uzun oturmamışlardı, hatta daha ancak üçüncü kekteydiler ki kapı bir kez daha ve bu kez daha da büyük bir gürültüyle çalındı.	"izninizle." dedi hobbit ve kapıya yollandı.
221	"So you have got here at last." That was what he was going to say to Gandalf this time	'İzninizle.' dedi hobbit ve kapıya yöneldi.	"Demek sonunda gelebildin." Bu defa Gandalf'a böyle diyecekti
222	But it was not Gandalf	'Sonunda gelebildiniz.' diyecekti Gandalf'a.Ancak gelen Gandalf değildi	Ama gelen Gandalf değildi
223	Instead there was a very old— looking dwarf on the step with a white beard and a scarlet hood; and he too hopped inside as soon as the door was open, just as if he had been invited.	Onun yerine kırmızı başlıklı, beyaz	Bunun yerine eşikte ak sakallı ve al pelerinli, çok yaşlı görünen bir cüce vardı ve kapı açılır açılmaz, davet edilmiş gibi o da içeri atladı.
224	"I see they have begun to arrive already," he said when he caught sight of Dwalin's green hood hanging up	sakallı, oldukça yaşlı gözüken bir cüce vardı kapıda ve o da kapı açılır açılmaz davet edilmişcesine içeriye daldı.'Görüyorum ki gelmeye başlamışlar,' dedi Dwalin'in asılı duran yeşil başlığı gözüne çarpınca	"Bakıyorum şimdiden gelmeye başlamışlar," dedi, Dwalin'in askıdaki yeşil başlığını görünce
225	He hung his red one next to it, and "Balin at your service." he said with his hand on his breast.	Kendi kırmızı başlığını bunun yanına asıp elini göğsüne koydu ve, 'Balin, hizmetinizdeyim.' dedi.	Kendi kırmızı başlığını da onun yanına astı ve elini göğsüne götürerek, "Balin hizmetinizde." dedi.

Figure 10. Partial alignment for TT-1 between segments 217-224.

From the point where segments begin to misalign partially on segment 226 again, these shifts from alignment to misalignment -and vice versa- occur occasionally until segment 589 of TT-1. From segment 589 onwards, the shifts affect the alignment for the major part of TT-1 except for the segment ranges presented in table 4.

Table 4

Exact Matches of TT-1 (From Segment 589 Onwards)

Segm.	ST	TT-1
608	There was a fearful mess in the room, and piles of unwashed crocks in the kitchen	Odada ürkütücü bir karmaşa, mutfakta yığınla yıkanmamış çanak çömlek vardı
609	Nearly every pot and pan he possessed seemed to have been used	Nerdeyse sahip olduğu her kap ve her kacak kullanılmıştı
664	Then they came to lands where people spoke strangely, and sang songs Bilbo had never heard before	Sonra insanların garip konuştuğu ve Bilbo´nun hiç duymadığı şarkılar söylediği topraklara geldiler
665	Now they had gone on far into the Lone—lands, where there were no people left, no inns, and the roads grew steadily worse	Sonra taa, hiç kimsenin kalmadığı, hiç han bulunmayan ve yolların gittikçe kötüleştiği Yalnıztopraklar´a kadar gittiler
698	When they had looked at it for some while, they fell to arguing	Işığa bir müddet daha baktıktan sonra tartışmaya başladılar
699	Some said “no” and some said “yes”	Bazıları ‘hayır’ bazıları ‘evet’ dedi.
716	Off Bilbo had to go, before he could explain that he could not hoot even once like any kind of owl any more than fly like a bat	Bilbo nasıl bir kez bile bir yarasa gibi uçmamışsa, herhangi bir baykuş gibi de ötemeyeceğini açıklayamadan gitmek zorunda kaldı
720	So, naturally, he got right up to the fire—for fire it was—without disturbing anyone	Böylece doğal olarak hiç kimseyi rahatsız etmeden doğruca ateşe — çünkü bu bir ateşti — ulaştı
721	And this is what he saw.	Ve gördükleri işte şunlardan ibaretti.
732	William choked	William ağzını tıka basa doldurdu
737	And time’s been up our way, when yer’d have said ‘thank yer Bill’ for a nice bit o’ fat valley mutton like what this is.” He took a big bite off a sheep’s leg he was roasting, and wiped his lips on his sleeve.	Bunun gibi güsel etle semis biy vadi koyunu itin “Sağ ol Bill" demen gerekilken hem de.’ Kızarttığı koyun butundan koca bir ısırık alıp ağzını koluna sildi.
738	Yes, I am afraid trolls do behave like that, even those with only one head each	Evet korkarım ki yalnızca tek kafaları olan troll’ler bile böyle davranırlar
839	In the end they decided to mince them fine and boil them	Sonunda onları kıyma yapıp kaynatmaya karar verdiler
840	So they got a great black pot, and they took out their knives.	Böylece büyük kara bir kazan aldılar ve bıçaklarını çıkarttılar.

For TT-2, alignment is relatively successful compared to TT-1. The software achieves exact matches from segment 102, where the first aligned segment is detected, to segment 285. From segment 285 onwards, shifts from alignment to (partial) misalignment occurs occasionally until segment 776. Starting from segment 777, misalignment becomes consecutive. These shifts from alignment to misalignment are demonstrated in table 5 according the segment range that they occur in. Out of 5,845 segments of ST, 5,414 segments were aligned with TT-2, and total number of exact matches is 476. The remaining 4,938 segments are either partially aligned or completely misaligned.

Table 5

Alignment Results for TT-2 (by Segment Range)

Segment Range	Are they aligned successfully? (Yes/No)	Number of Segments in the Range
102-284	Yes	183
285-297	No	13
298-332	Yes	35
333-335	No	3
336-384	Yes	49
385-386	No	2
387-406	Yes	20
407-435	No	29
436-439	Yes	4
440-449	No	10
450-454	Yes	5
455-457	No	3
458-470	Yes	13
471-493	No	23
494-553	Yes	60
554-567	No	14
568-571	Yes	4
572-577	No	6
578-618	Yes	41
619-697	No	79
698-703	Yes	6
704-720	No	17
721-776	Yes	56
777-5515	No	4739
	Total exact matches	476

As figure 11 shows, alignment ends in segment 5519 for TT-1, and in segment 5515 for TT-2. Factors leading to the perpetual misalignment from a certain segment might be due to the structural differences between TT-1 and TT-2. For example, there are images and text boxes containing captions in TT-1 while no image is included within the main body of TT-2. Another factor is related to the translators' decisions to divide or merge sentences. Although O.N.E is programmed to take action for dividing merged segments or merging divided segments wherever it detects such difference in segmentation, even one misaligned line led to a chain reaction for certain segment ranges until the software fixed it through division or merge. In addition, some chapters include intentional misspellings like the dialogue between the trolls in TT-1 (Tolkien 1996, 46-53) and TT-2 (Tolkien 2015, 26-28), or the lines of Gollum e.g. "*kıymetlimss.*" causing failure to match the source segment with target segments.

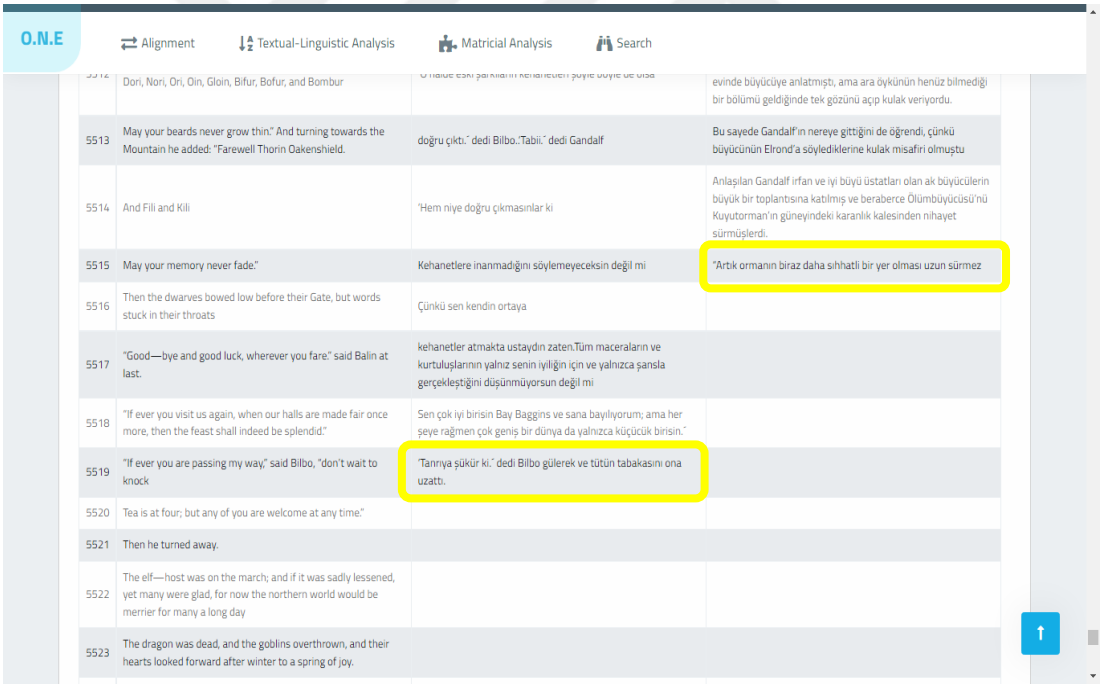


Figure 11. End of alignment for TT-1 and TT-2.

4.2 Matricial Analysis

Tasks categorized under 'Matricial Analysis' examine the texts for the completeness, segmentation, and distribution of linguistic material. It must be noted that the analysis

includes all the content parts such as book cover, title page, copyright page, author's biography etc. They should be kept in mind as factors affecting the results.

4.2.1 Word Count

Compared to the method that Duymaz (2020) used in her research, length of the examined texts are measured differently in the current research. Duymaz measures the length based on the number of pages in the texts, which changes depending on the stylistic features of a document such as margins, font size, and inclusion of paratexts. It should be added that Duymaz herself is also aware of this factor as she explains that the longest text is the one which was published using the largest font (38). The present research, on the other hand, takes word count as a criterion to measure the length. The results obtained by this task can help the researcher observe the variation in text length and attempt to justify the reason behind it.

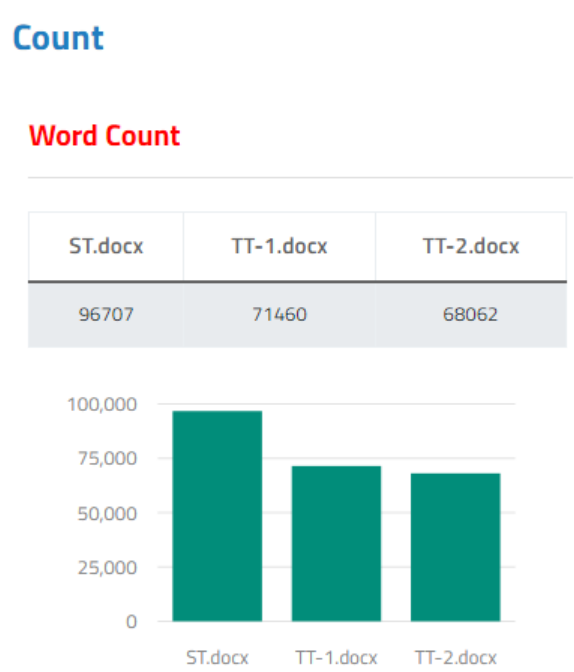


Figure 12. Results for word count.

As figure 12 shows, the software counts 96,707 words in ST, while the number is 71,460 for TT-1 and 68,062 for TT-2.The results indicate a considerable difference in the word count between ST and TTs. One might explain this difference by the structural

difference between the source and target languages, or the discrepancy in the word count of TTs may signal possible omission of some elements in TT-2. In order to confirm such predictions and reach a more accurate conclusion, one should compare the word count results with other the number of sentences and average paragraph length.

4.2.2 Sentence Count

O.N.E identifies sentences based on punctuation. That is, a string is considered as a sentence if it ends in any of these punctuation marks: full stop (excluding personal titles like ‘Mr.’, ‘Mrs.’), exclamation mark, question mark, and ellipsis. Figure 13 shows the total number of sentences in ST (5,844), TT-1 (5,839) and TT-2 (5,604). It can be seen that the gap between the word counts of ST and TT-1 is closed here. In other words, the difference in word counts may be related to the linguistic differences between English and Turkish. However, there is still difference in the number of sentences between ST and TT-2, and between TT-1 and TT-2. The reason may be due to the omission of some paragraphs or chapters. The researcher can compare the sentence count results with the results of average sentence length, paragraph length and confirm if any chapter was removed or added via ‘Content’ task.

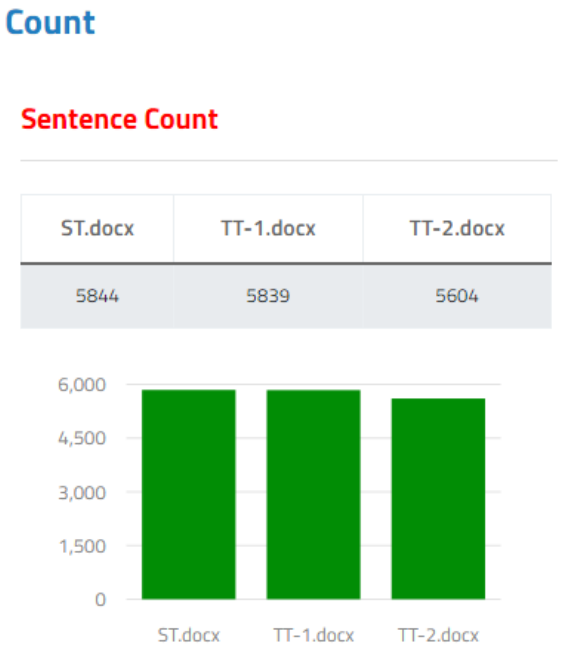


Figure 13. Results for sentence count.

4.2.3 Paragraph Count

O.N.E counts the paragraphs that are already identified in the related database table. As seen in figure 14, O.N.E detects 2,065 paragraphs in ST, while this number is 1,950 for TT-1 and 1,888 for TT-2. It is apparent from the results that there are 115 more paragraphs in ST compared to TT-1, and 177 paragraphs for TT-2. It may be due to the decisions that the editor/translator made in distribution of the text, e.g. some paragraphs may be merged for publishing purposes. Another possibility is related to the difference in content. For instance, there may be some chapters in ST that was not translated in TT-1 and/or TT-2, or it may be related to difference in publishing i.e. some parts like copyright page, author's note etc. might be different or omitted from translation. The same may be true for the difference between paragraph counts of TT-1 and TT-2 because there are 62 more paragraphs in TT-1. As in the cases of word count and sentence count, further analysis is required in terms of average length and parts of content.

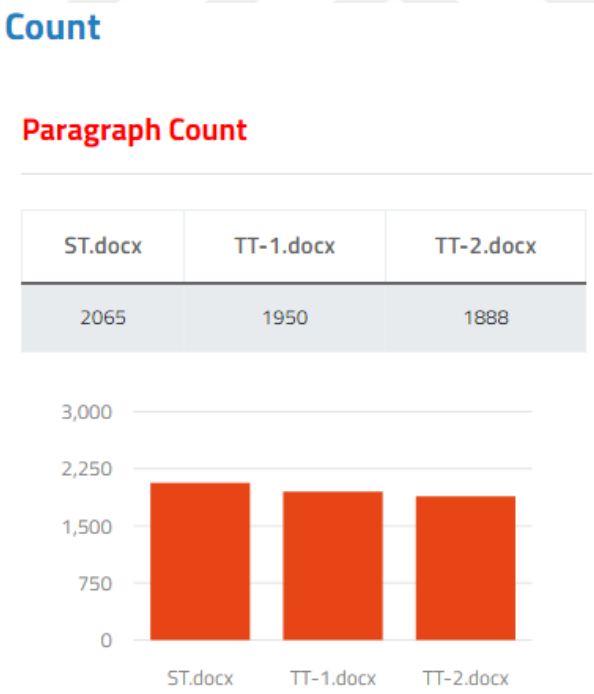


Figure 14. Results for paragraph count.

4.2.4 Footnote/Endnote Count

O.N.E searches for text that is tagged as footnote and/or endnote in ST and TTs and then displays their content side-by-side if they exist in the documents. Figure 15 shows that there 2, 2 and 3 endnotes in ST, TT-1 and TT-2 respectively. Endnotes in ST are consistent with those of TT-1, whereas discrepancy is observed in TT-2. Only the endnote "*Azog'un oğlu (ç.n.)*" matches with the content in ST while the other endnotes "*Apatura iris (ç.n.)*" and "*İğne (ç.n.)*) provide additional information which does not exist in ST or TT-1.

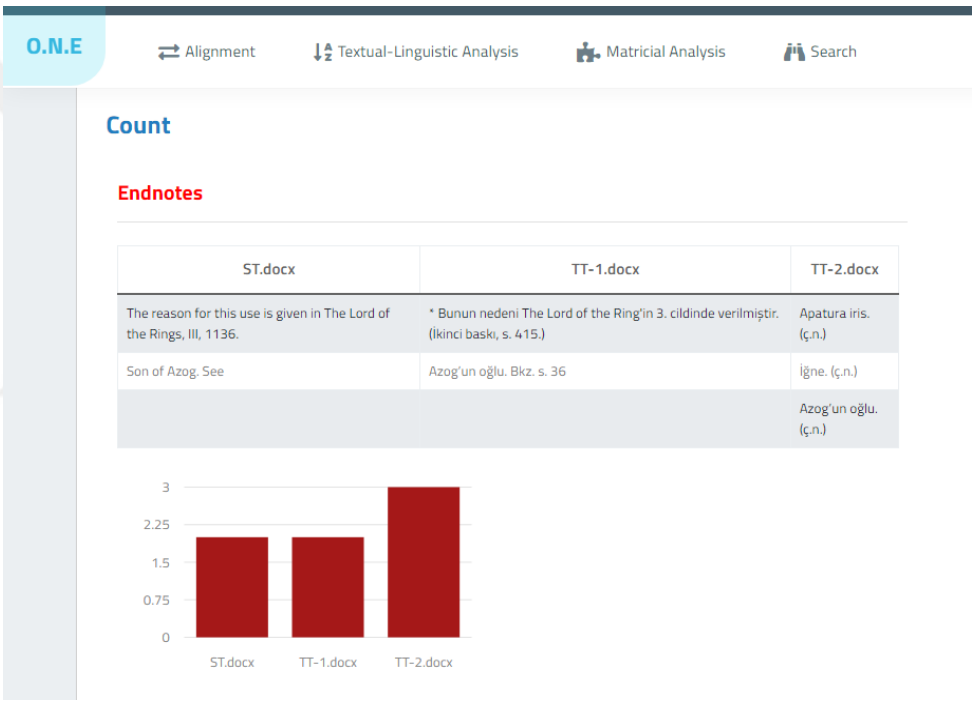


Figure 15. Results for footnote/endnote count.

4.2.5 Image Count

O.N.E searches the source and target text documents for objects that are tagged as image and provides the total number of images in each text as shown in figure 16. Both ST and TTs contain images as the novel *The Hobbit* contains several maps and illustrations, but number of images varies among them. TT-1 has a close number of images (21 images) compared to ST (24 images), yet the number is still fewer by 3 images.

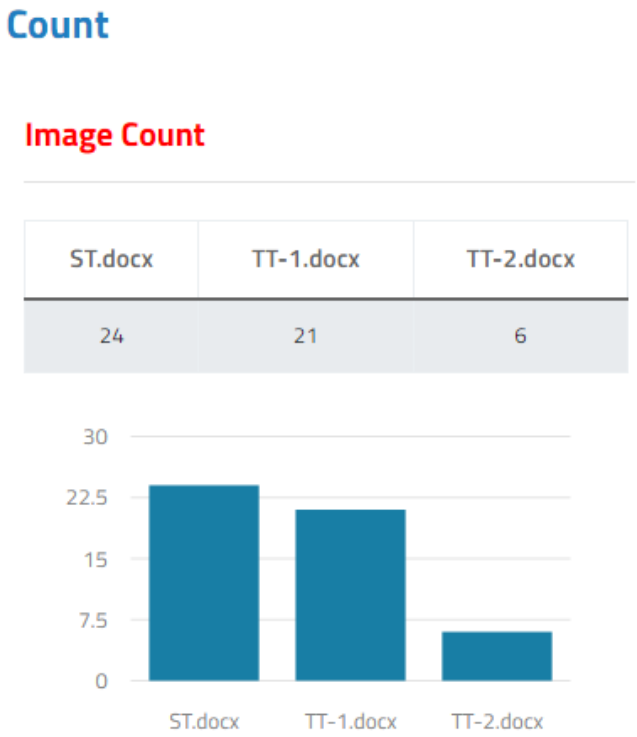


Figure 16. Results for image count.

The difference is rather dramatic in TT-2 where only 6 images are found. A researcher may question the reason behind this difference: Why are there less images in TT-2? Did the editor/publisher actually remove images from TT-2? If so, what might be the reason? Or are there any images in ST or TT-1 that are not related to the translated content and hence cause such difference e.g., logo of the publisher company, book cover images etc.?

When we examine the content of TT-2 manually, it can be seen that the exclusion of "Author's Note" section of ST leads to a major decrease in image count because there are 8 images included within the sentences to introduce the "runes" (e.g. ᛉ, or ᛞ) to the reader. It is certain that further analysis will reveal other details regarding this issue.

4.2.6 Average Word Length

Average word length is calculated dividing the total number of characters -spaces excluded- by the number of words in a text. Figure 17 shows that the average word length is 3.99 for ST, and 6.3 for TT-1 and 6.63 for TT-2. The fact that words in TTs are longer

compared to ST can be explained by the agglutinative nature of Turkish language. The difference between TT-1 and TT-2, on the other hand, is not significant.

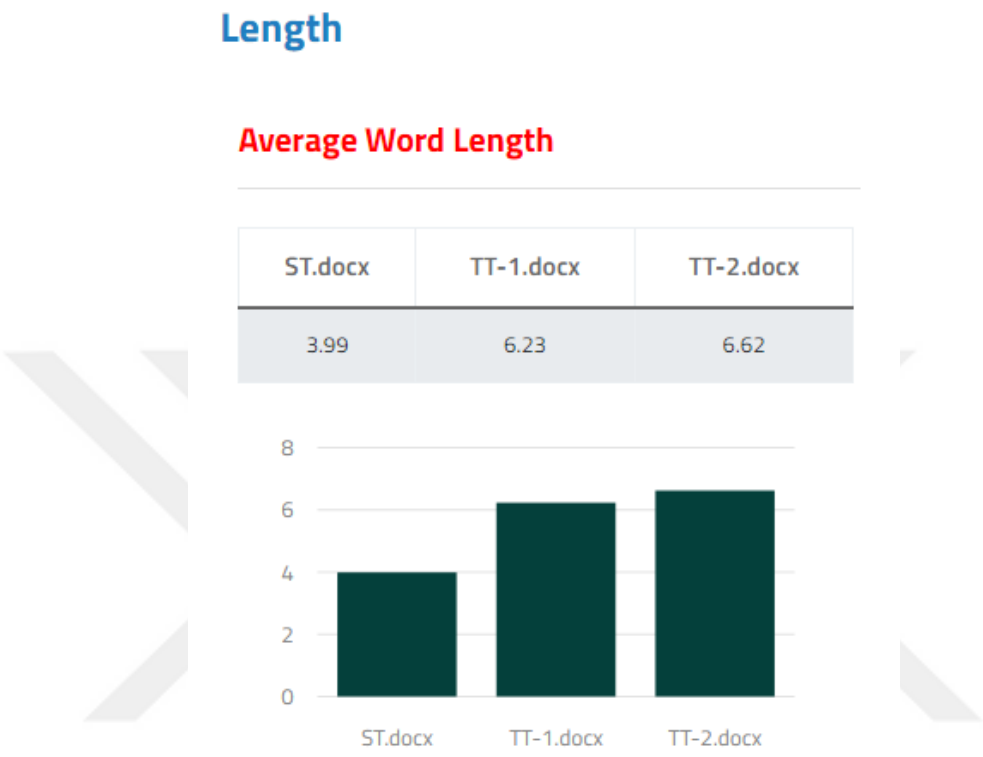


Figure 17. Results for average word length.

4.2.7 Average Sentence Length

O.N.E calculates the average sentence length dividing total number of words by total number of sentences. Figure 18 provides the average sentence length in ST and TTs: On average, there are 16.55 words in the sentences of ST, while the number is 12.24 for TT-1, and 12.15 for TT-2. First thing we notice is the difference between ST and TTs. Sentences in TT-1 and TT-2 contain fewer words, which can be related to the agglutination in Turkish. When TT-1 and TT-2 are considered, only a minor difference is observed; sentences in TT-1 are slightly longer.

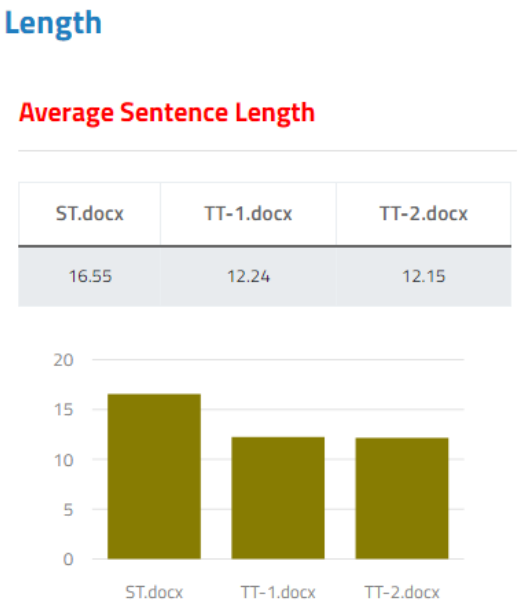


Figure 18. Results for average sentence length.

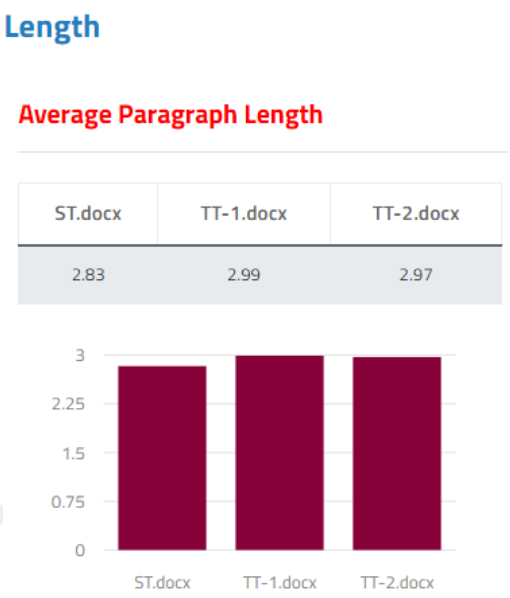


Figure 19. Results for average paragraph length.

4.2.8 Average Paragraph Length

When average paragraph length is calculated, total number of sentences are divided by the number of paragraphs in a text. As shown in figure 19, on average, there are 2.83 sentences in the paragraphs of ST, while the number is 2.99 for TT-1 and 2.97 for TT-2. Average sentence length in each text is close to one another, yet it can be said that there are more sentences in the paragraphs of TTs compared to ST. It may be related to a possible tendency to divide embedded clauses of ST in translation; however, this prediction must be tested through manual analysis on the texts. It should be noted that each image is considered a paragraph due to the paragraph breaks they are followed by.

When deemed necessary, all these results regarding average length can be compared in general in order to discover the correlation between word, sentence and paragraph lengths through the interface of O.N.E (see fig. 20).

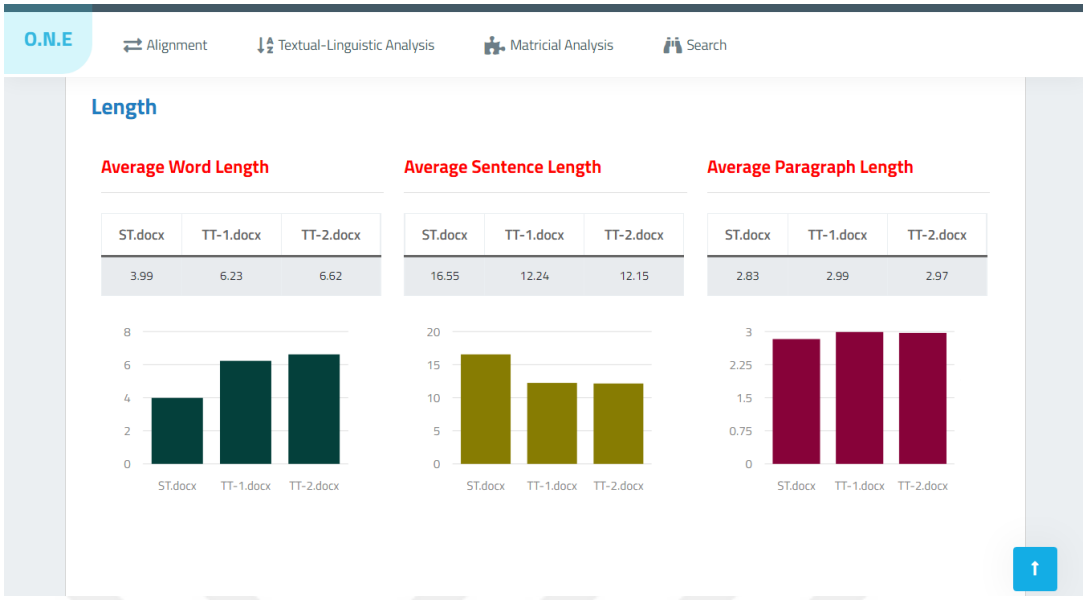


Figure 20. Results for length (side-by-side).

4.2.9 Content

Duymaz (2020, 35-38) describes the parts of content of source and target texts in detail. She sketches out the following parts of the texts: front cover, title page, (author's) notes, copyright page, table of contents, the number of chapters, and back cover. In this research, however, this process takes place within the bounds of the capabilities of O.N.E, which ultimately relies on the quality of the data provided by the user. In this specific case, the queried parts (see chapter 3.2.4) can be found by the software only if they are marked properly in a word processor or contain the exact keywords that are prespecified in the software. Keeping this in mind, content parts that O.N.E found in ST and TTs are given in table 6 (For screenshot from O.N.E, see appendix A). It can be seen that the software found more content parts in ST as they were marked in the document. Only book cover and table of contents were found in TT-1, and no content was found in TT-2.

Table 6

Content of ST and TTs

	ST		TT-1		TT-2	
Content	***Was it found?***	***Title***	***Was it found?***	***Title***	***Was it found?***	***Title***
Book Cover	Yes	Coverpage	Yes	Kapak Tasarımı	No	-
Title Page	Yes	Title Page	No	-	No	-
Copyright Page	Yes	Copyright	No	-	No	-
Dedication Page	No	-	No	-	No	-
Foreword	No	-	No	-	No	-
Preface	No	-	No	-	No	-
Prologue	No	-	No	-	No	-
Acknowledgements	No	-	No	-	No	-
Author's Notes	No	-	No	-	No	-
Author's Bio	No	-	No	-	No	-
Translator's Note	No	-	No	-	No	-
Table of Contents	Yes	Contents	Yes	İçindekiler	No	-
Introduction	No	-	No	-	No	-
Epilogue	No	-	No	-	No	-
Appendix	No	-	No	-	No	-
Bibliography	No	-	No	-	No	-
Glossary	No	-	No	-	No	-
Index	No	-	No	-	No	-

4.3 Textual-Linguistic Analysis

The tasks performed under the category of 'textual linguistic analysis' deal with the selection of linguistic material in source and target texts. The software examines the texts by means of a group of datasets which are presented in chapter 3.2.3.

4.3.1 Most Frequent Words

Analysis of 'most frequent words' are not included in Duymaz's research. They are added into the scope of textual-linguistic analysis here to achieve a purpose. Determining most frequent words in source and target texts can help researchers to ask further questions or make interpretations based on the data provided by the software. For instance, inferences can be made regarding the theme of the texts and whether the themes match in their translations.

In the case of our corpus, the software looks for the words that are used most frequently in source and target texts (For screenshot from O.N.E, see appendix B). There is an important point about this task that should be considered before analysing the results. Lemmatization or stemming is not applied for this task. That is, derivational affixes or inflectional endings are not removed to get the stem of a word (stemming) or return the base or dictionary form of a word (lemmatization), e.g. does – do, or saw – see. The results are provided in table 7.

Table 7

Most Frequent Words (First 20 Entries)

ST		**TT-1**		**TT-2**	
Word	**Count**	**Word**	**Count**	**Word**	**Count**
Bilbo	558	Bilbo	592	Bilbo	597
dwarves	286	Thorin	255	Thorin	252
Thorin	252	Gandalf	197	Gandalf	194
time	225	iyi	183	zaman	189
back	213	dek	172	büyük	186
long	211	uzun	165	uzun	183
Gandalf	187	doğru	148	iyi	172
great	181	büyük	146	in	153
good	165	geri	131	doğru	151
hobbit	161	küçük	123	Dağ	113
dark	142	in	122	küçük	113
goblins	142	Dağ	118	son	106
Mountain	133	yalnızca	108	süre	101
made	131	zaman	105	Baggins	96
thought	126	üzerine	105	üzerine	92
light	121	aşağı	98	Gollum	92
door	120	Baggins	98	geçmeden	91
round	110	cüceler	96	hâlâ	90
dragon	104	süre	95	Smaug	87
Baggins	103	hobbit	91	başka	87

Most frequent words can give the researcher an idea on the theme of *The Hobbit*. For example, the results for ST informs about the main characters of the novel (‘Bilbo’, ‘Thorin’, and ‘Gandalf’) and reveals a thematic framework that is established around the words like ‘Mountain’, ‘back’, ‘dark’, ‘dwarves’, ‘hobbit, ‘dragon’, and ‘goblins’ implying the journey of the main character Bilbo Baggins. The same applies for TTs of which the most frequent words give the researcher a clue about the main characters and

the plot of the novel via the words ‘Bilbo,’ ‘Thorin,’ ‘Gandalf,’ ‘Gollum,’ ‘Smaug,’ and ‘*in*,’ ‘*geri*,’ ‘*Dağ*,’ ‘*uzun*,’ and, ‘*zaman*.’ A researcher may also want to explore why the translator(s) prefer specific lexical items where other options may be equally available in the target language.

4.3.2 Personal Titles

This task searches for the personal titles in source and target texts. Personal titles are not investigated in Duymaz’s study. They are included in this research because a researcher can analyze them for various purposes e.g. to compare denotation of social classes via personal titles in a source text and its translation(s), or to investigate translation of personal titles in terms of “adequacy” or “acceptability” (see Toury 2012). Figure 21 shows that 8 different personal titles were found in ST, while 7 words were recognized as personal title in TT-1 and TT-2 for each text separately. Users can also see how many times each personal title is used.

Personal Titles

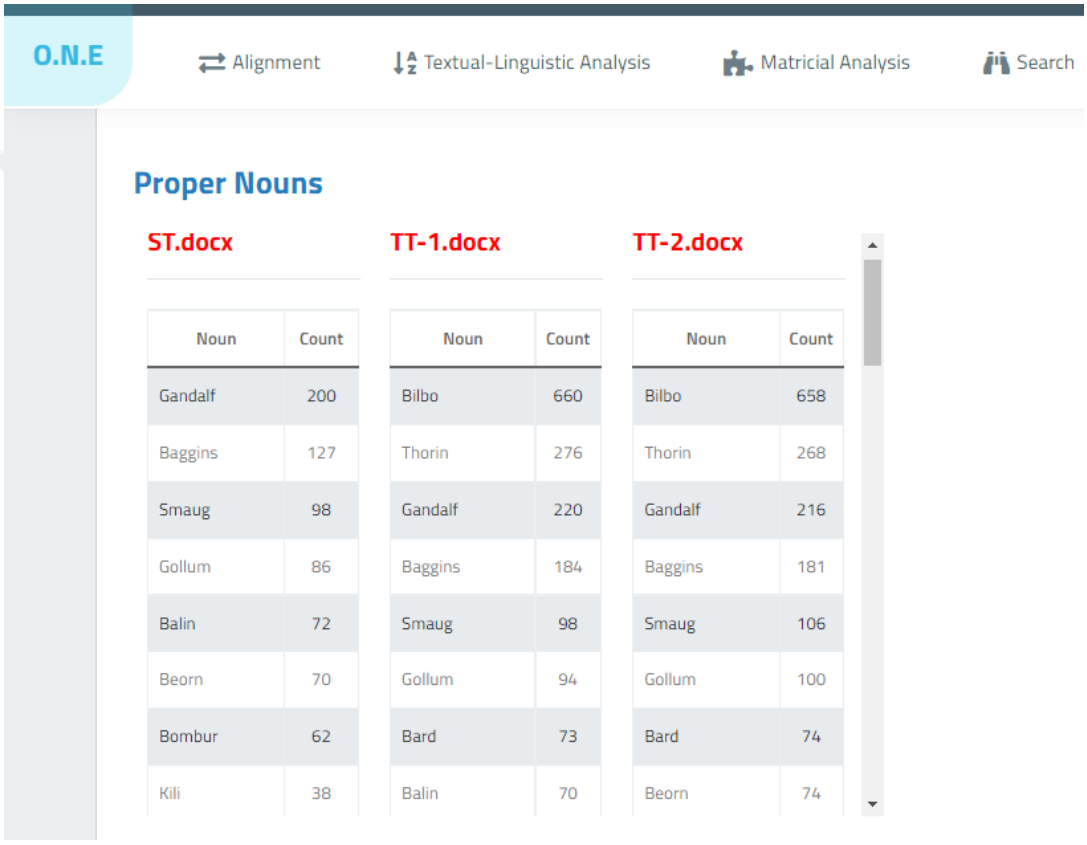
ST.docx		TT-1.docx		TT-2.docx	
Mr	69	Bay	70	Bay	68
Master	36	Efendi	30	Kral	43
Lord	18	Kral	29	Efendi	30
Esquire	2	Bayan	1	Saygıdeğer	2
Sir	2	Saygıdeğer	1	Bayan	1
Professor	1	Sayın	1	Lord	1
Mrs	1	Beyefendiler	1	Beyefendi	1
Messrs	1				

Figure 21. Results for personal titles.

4.3.3 Proper Nouns

This task searches for the proper nouns in source and target texts based on capitalization and position of the word in a sentence. Capitalization of the first letter does not suffice to

identify a word as proper noun since all the first letters in the beginning of a sentence are capitalized, including common words, adjectives or other types of words. Therefore, O.N.E also refers to the lexical database to exclude common words and other words from results. Figure 22 shows the first eight entries of proper nouns that were recognized by the software. O.N.E displays maximum 50 entries in the results (For the list of results, see appendix C).



Noun	Count
Gandalf	200
Baggins	127
Smaug	98
Gollum	86
Balin	72
Beorn	70
Bombur	62
Kili	38

Noun	Count
Bilbo	660
Thorin	276
Gandalf	220
Baggins	184
Smaug	98
Gollum	94
Bard	73
Balin	70

Noun	Count
Bilbo	658
Thorin	268
Gandalf	216
Baggins	181
Smaug	106
Gollum	100
Bard	74
Beorn	74

Figure 22. Results for proper nouns (first 8 entries).

It must be noted that this process primarily depends on the inclusiveness of the lexical database used for the task. Otherwise, words which are not proper nouns may be identified falsely as proper noun. For instance, ‘Eggses’ and ‘Ssss’ are intentionally misspelled in ST, yet they are listed as proper noun as they start with a capital letter and they are not included in the dataset. Complications may still arise even with a comprehensive and well-defined database. For example, a common noun, or an adjective or a number may be used as a proper noun in a text such as “Five Armies” in ST.

4.3.4 Exclamations

Exclamations are defined as "a sharp or sudden utterance" or "vehement expression of protest or complaint."[5] According to the table of results provided by O.N.E (see fig. 23), 302 word units in ST are confirmed to be an exclamation, while the number is 373 for TT-1 and 381 for TT-2. Due to lack of space, only the first 50 entries for each text are provided in appendix D. As explained in chapter 3.2.3, only the sentences consisting of maximum three words are listed as entries to rule out the exclamatory sentences. Yet, exclamatory sentences containing three words still appeared in the results, e.g. '*Şimdi zaman yok!*' or 'You may indeed!'

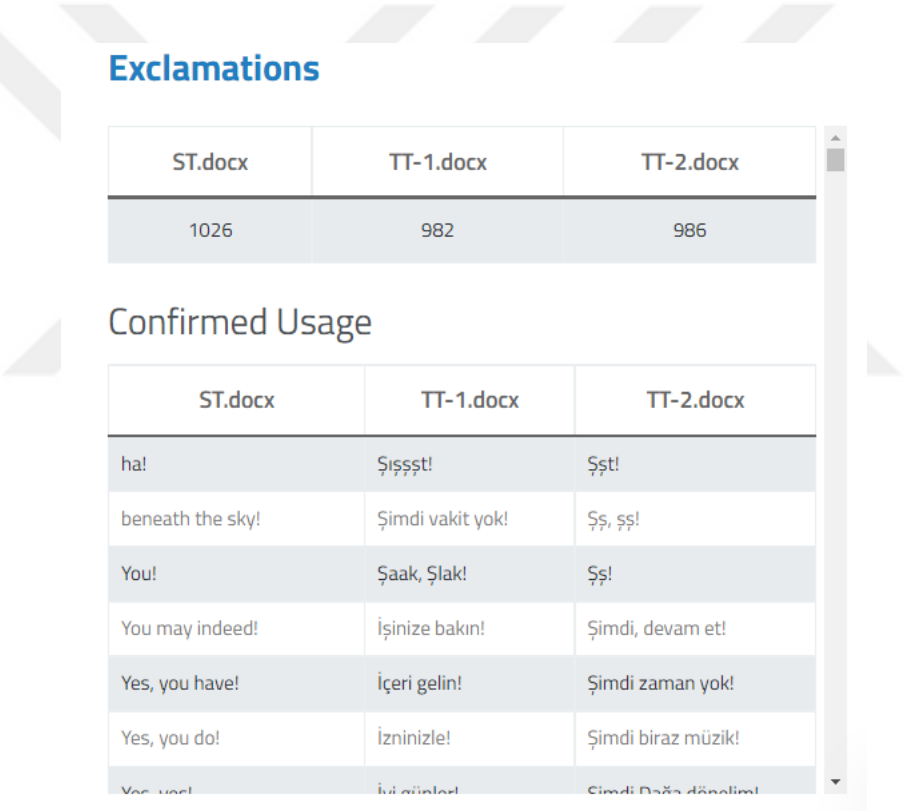


Figure 23. Results for exclamations.

4.3.5 Non-defined Words

Tolkien (2009) coins new words in *The Hobbit* and Duymaz (2020, 44-60) analyzes their translation under the title of "translation of neologisms." In order to perform the same

[5] *Merriam-Webster*, s.v. "exclamation," accessed December 28, 2021, https://www.merriam-webster.com/dictionary/exclamation.

analysis, the software follows a process which is similar to the one in 'Proper Nouns.' Words that are not found in the datasets are identified as non-defined words.

The drawback of this process is that O.N.E retrieves both coined words and other words like deliberately or unintentionally misspelled ones, or proper nouns. For instance, the character named 'Gollum' in ST speaks in a distinctive way, duplicating plurals ("preciousss") and wrongly using the third person singular ("we eats it"). The software therefore recognized most of his lines as non-defined words. Therefore, the researcher should count out the irrelevant words to obtain neologisms.

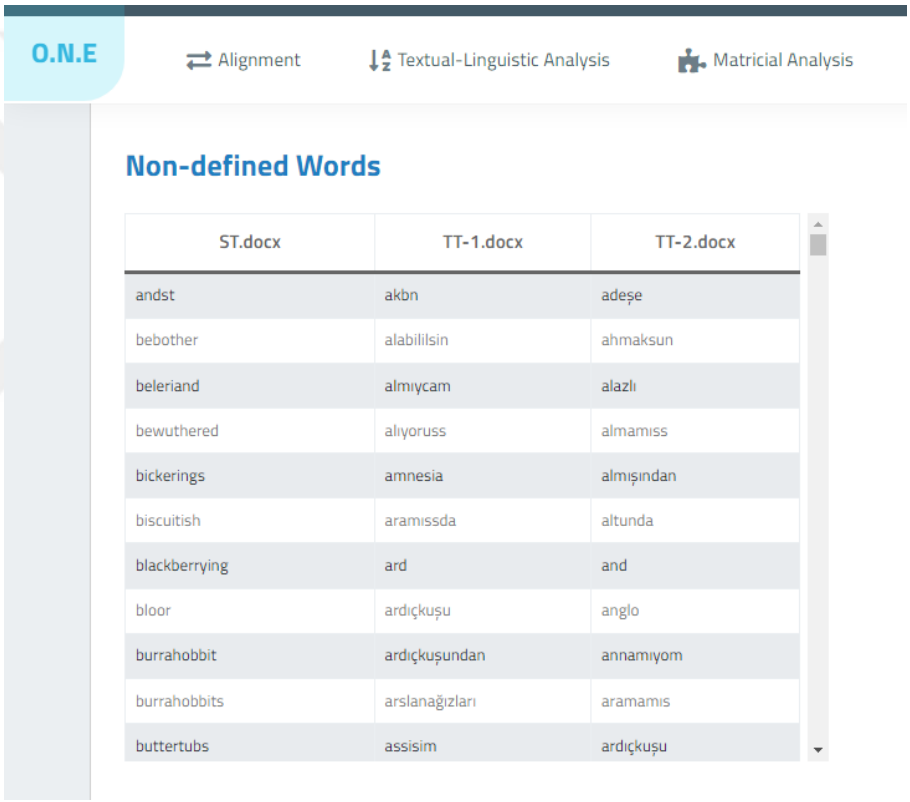


Figure 24. Results for non-defined words (first 11 entries).

Figure 24 present the results for non-defined words detected by O.N.E (For the first 50 entries, see appendix E). 72 words for ST, 299 words for TT-1 and 293 words for TT-2 were found in total. The difference in total numbers between ST and TTs can be related to agglutination in Turkish language. When all the entries of TT-1 and TT-2 are compared manually, it is seen that the results in TT-1 for the word 'goblin' exemplifies this issue better. Fourteen entries were listed among the entries: *'goblin', 'goblini', 'goblinle', 'goblinler', 'goblinlerce', 'goblinlere', 'goblinleri', 'goblinlerin',*

'goblin<u>lerinden</u>', *'goblin<u>lerine</u>'*, *'goblin<u>lerinin</u>'*, *'goblin<u>lerle</u>'*, *'goblin<u>lerler</u>'*, '*goblin<u>lerlerden</u>*.'

4.3.6 Measurement Units

O.N.E searches for the measurement units in source and target texts. Then, they are presented along with their context, e.g. 'a hundred **yards** through the' (see fig. 25). The software found 46, 49 and 39 items for ST, TT-1, and TT-2 respectively. The complete list is given in appendix F. Examining measurement units can be useful as the researcher may wish to determine whether they conform to the source norms or target norms. Users must note that the software does not perform a context-dependent search. As a consequence, it may show false positives such as '*dek **dört ayak** üzerinde emekledi* (crawled on all fours).'

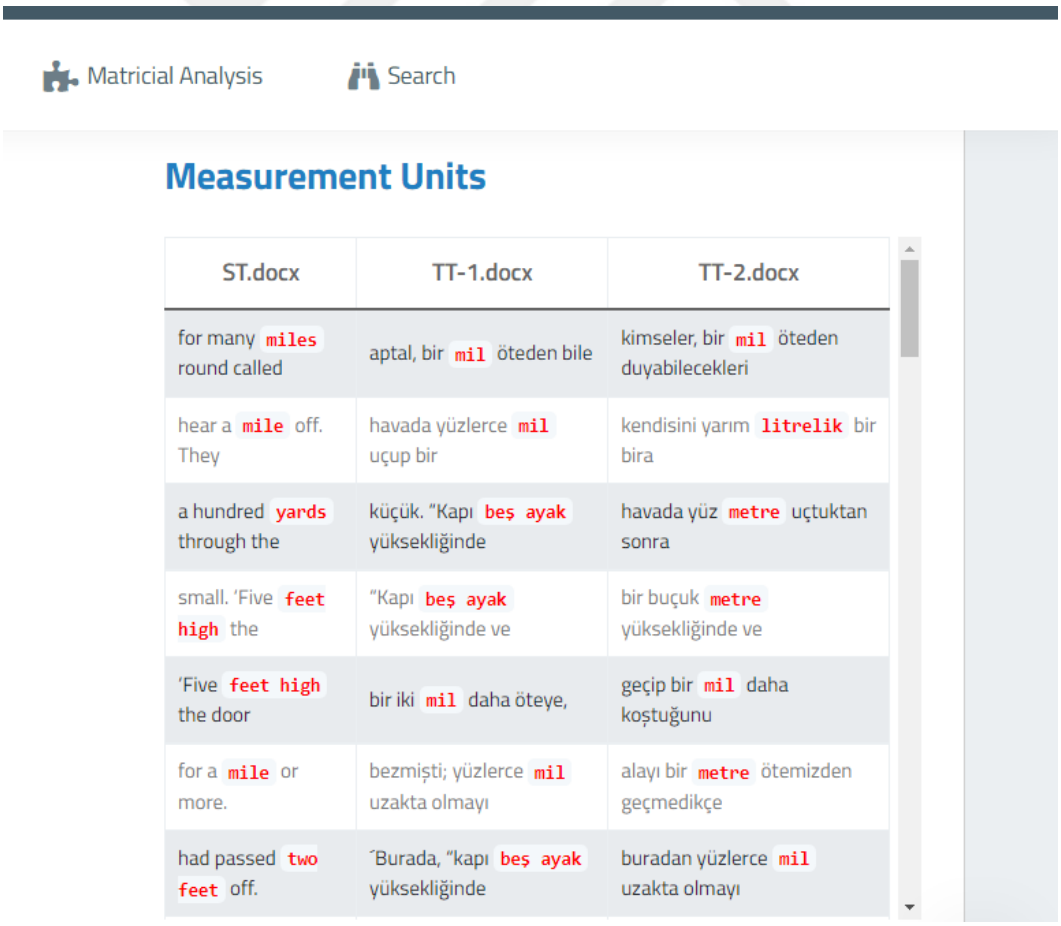


Figure 25. Results for measurement units (first 7 entries).

4.3.7 Idioms

In Duymaz's (2020, 65-71) research, translation of idioms are examined through six examples, and the following phrases in ST are identified as idioms (Underlined idioms were detected by O.N.E as well):

- lend a hand
- All in good time!
- you walked right in and put your foot in it
- out of the frying-pan into the fire
- before he could say knife
- I don't like the sound of that.
- keep your heart up!
- You may come through all right.
- While there's life there's hope!
- Third time pays for all.
- Escaping goblins to be caught by wolves!

In this study, the software identifies idioms in source and target texts according to a dataset (see chapter 3.2.3). For Turkish idioms, a dataset is compiled from an online open-content collaborative dictionary as it was the most comprehensive source that is digitally available and computer-readable[6]. However, there are some entries that are falsely-identified as idioms (such as '*zifiri karanlık* (pitch black),' and '*rica etmek* (to request)' in the dictionary. When presenting the results (see appendix G), we excluded false results manually according to the dictionary of proverbs and idioms of Turkish Language Institution.[7] Consequently, we identified 19 idioms out of 59 entries for TT-1, and 18 idioms out of 57 entries for TT-2.

Detected idioms are highlighted in each sentence as it can be seen in Figure 26. The results show that there are 54 segments in ST where an idiom is used, while this number is 59 for TT-1, and 57 for TT-2. It should be added that results are not aligned, so the translations are not aligned with ST. Also, users should keep in mind that the search is not context-dependent. That is, the software may find false-positives as in '*Şimdi boş midelerinin üzerine* ***kemerleri sıkmak****[tan ... başka çareleri kalmamıştı]*' where '*kemerleri sıkmak* (to tighten [one's] belts)' is used in the literal sense instead of denoting 'to spend less money than one did before because one has less money.'

[6] Wiktionary, "Kategori:Türkçe deyimler."
[7] *Türk Dil Kurumu Atasözleri ve Deyimler Sözlüğü,* last modified May 29, 2009, https://www.sozluk.gov.tr.

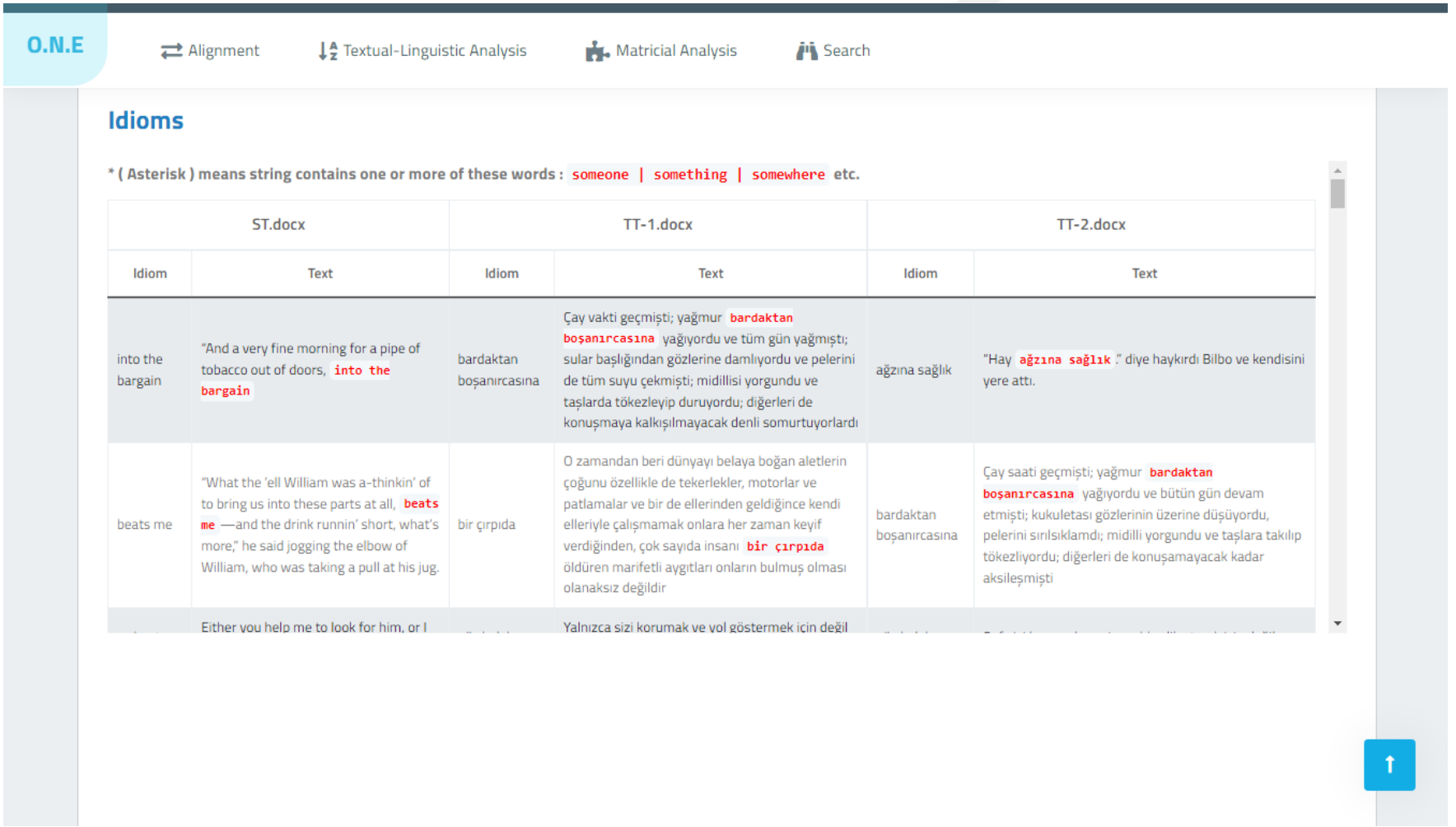


Figure 26. Results screen for idioms.

4.3.8 Slang

The last task performed under textual-linguistic analysis look for the slang in source and target texts. As in the task of 'idioms,' O.N.E searches for the word(s) identified as slang in the dataset and displays the results for each text separately (see fig. 27). According to this process, there are 3,481 segments in ST where slang is recognized; while 1,470 segments were detected in TT-1, and 1,481 segments in TT-2. However, not all of these results actually contain slang for the reasons explained below.

First of all, the reason why the total number of entries is so high is connected to the comprehensiveness of the slang dataset. In addition, recognizing slang in texts requires context-dependent interpretation of words, which poses a problem for this task causing false identification. For instance, 'bare' denotes 'a lot of, significant amount of' in the slang dataset. However, the word means 'not covered by grass, trees, or plants' in ST: "[...] nor yet a dry, bare, sandy hole with nothing in it to sit down on or to eat" (Tolkien 2009, 12). As a matter of fact, the same applies for almost all the results (both in English and Turkish) obtained by this task. Therefore, a feature is added to the software which enables users to filter out specific word(s) from the search. Even with the filter option, elimination process is still too time-consuming and not practical for this specific corpus. Nevertheless, the present researcher endeavored to eliminate falsely-identified results and finally obtained the results given in table 8.

Table 8

List of Slang

ST	
Slang	***Context***
'ell	"What the **'ell** William was a—thinkin' of to bring us into these parts at all, beats me—and the drink runnin' short, what's more," he said jogging the elbow of William, who was taking a pull at his jug.
blighter	"Poor little **blighter**," said William
blimey	"Mutton yesterday, mutton today, and **blimey**, if it don't look like mutton again tomorrer," said one of the trolls.
blimey	"**blimey**, Bert, look what I've copped." said William.

Table 8 -continued

List of Slang

TT-1	
Slang	***Context***
sıçan	Evet ne söyleyeceksin bakalım seni **sıçan** soyundan yaratık.´
sıçan	Cidden**, sıçan** soyundan bir yaratık ha
zırıl	Hem ağaçlar arasından bir ateş görür gibi olmuştu ve bu da, sular damlayan**, zırıl zırıl** kıyafetleri üzerine soğuk ve nemli yapışırken onu iyice cezbetmişti.
palas	Söylerken de arkasını döndü ve iki uzun ağaç arasındaki son boşluğun bir örümcek ağıyla — ama şans eseri bu düzgün bir örümcek ağı değil, yalnızca **palas** pandıras ağaçtan ağaca örülmüş, çift kat koca teller halinde bir örümcek ipiydi yalnızca
helal	'Dün koyun, bugün koyun ve eğer yarın da koyun değilse **helal** olsun,' dedi troll'lerden biri.
TT-2	
Slang	***Context***
sıçan	Ne söyleyeceksin, **sıçan** soyu."
harbiden	"Tartışmayı baştan başlatma, Bili," dedi, "yoksa **harbiden** de bütün gece sürer."
bodoslama	Karanlığın içinde çılgınca sendeledikten, ağaçlara **bodoslama** daldıktan ve ormanın içinde millerce uzakta ne varsa, her şeyi uyandırana kadar bağırıp birbirlerine seslendikten sonra nihayet bir yığın halinde toplanmayı ve el yordamıyla birbirlerini saymayı başardılar

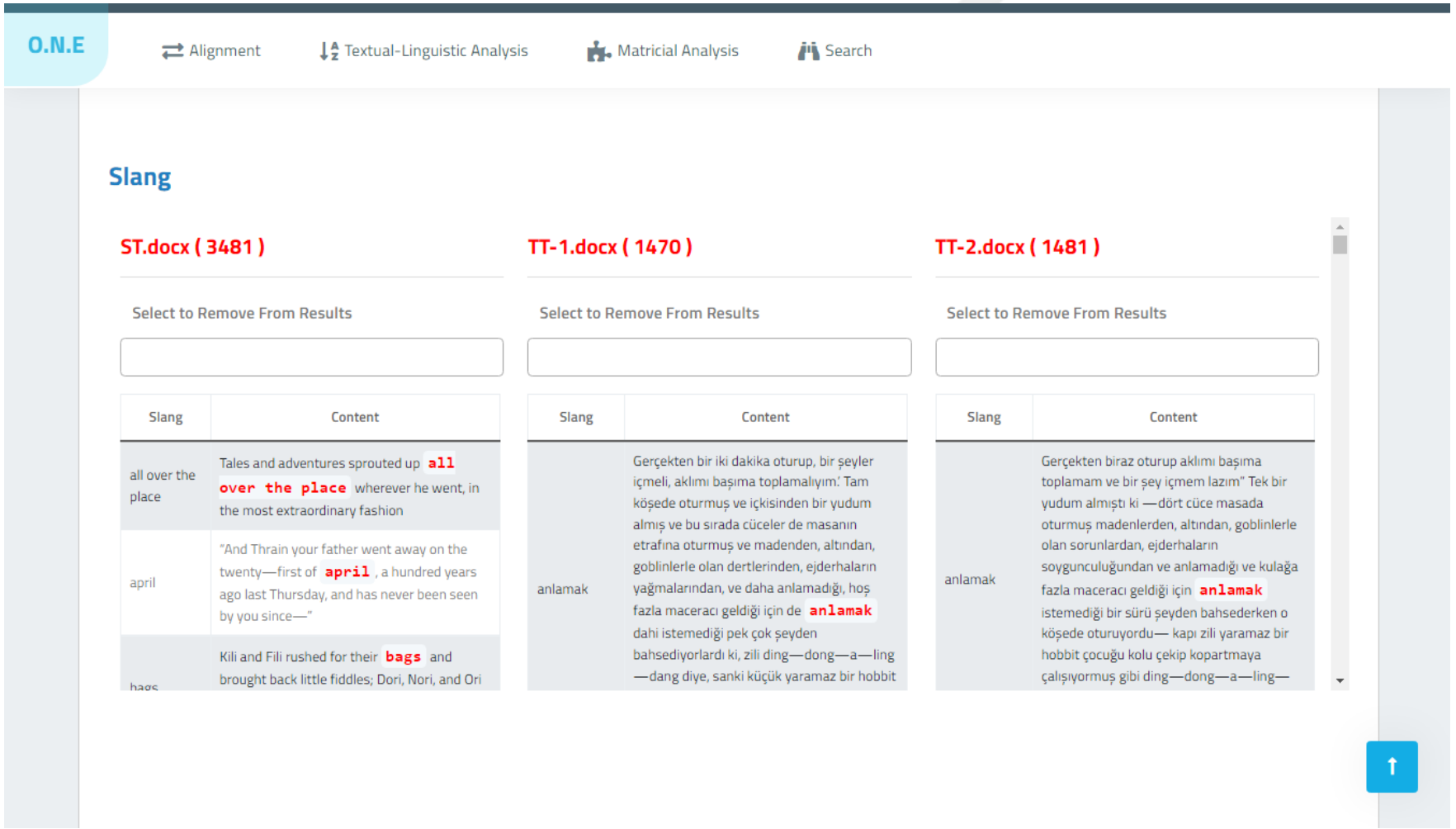


Figure 27. Results screen for slang.

5. DISCUSSION

The hypothesis of the research was that text mining would facilitate exhaustive analysis to extract operational norms. In order to test this hypothesis, a voluminous text -*The Hobbit* by J.R.R Tolkien- and its two translations into Turkish were chosen as the corpus. These texts have already been analyzed in terms of 'translational norms' by another researcher, which offered the opportunity to compare the results to those obtained by human effort only.

The reason for limiting the scope to 'operational norms' is because the techniques or functions in text mining are particularly well-suited to identify operational norms. Although Toury's aim was not to "point to strict statistical methods for dealing with translational norms" or "supply sampling rules", the results of this study affirm Toury's perspective on the possibility of exhaustive analysis through "many more and much bigger samples" with the help of advances in computing world (2012, 91-92).

The findings suggest that text mining can make it possible and easier to analyze these specific types of norms thoroughly, which would not be possible by sole human effort particularly on large corpora. The results generated by the alignment feature of O.N.E revealed the structural differences between the target texts on the macroscale. For instance, more segments in TT-2 were aligned accurately with the segments of ST compared to TT-1, which may be related to the similarity in sentence formation between TT-2 and ST.

There have been previous efforts by the present author to facilitate exhaustive analysis through text mining. In comparison with those researches, a progress has been made in terms of encompassing the investigation on translation of expressive language and cultural elements. For instance, recognition of slang and idioms in source and target texts were added in this research. Yet, further development is needed to avoid false-identification so that the software can interpret slang and idioms within their context. I believe it will be achievable with more training data and through the advances in NLP.

There have been two limitations in the tasks performed under 'textual-linguistic analysis' category. Firstly, accuracy of the outputs was fundamentally dependent on the quantity and quality of the data, i.e. thoroughness of the datasets. The scope of datasets can be extended to achieve more successful results. The second limitation is related to the

processing capacity or computational power. A better server could help us process more data and run tests in less time. In that sense, possibilities are endless given the fact that technologies are developing each day.

CONCLUSION

This research was carried out to facilitate exhaustive analysis on operational norms as introduced by Gideon Toury because a thorough examination is not possible due to human limitation. A tool was developed by a software developer and translations of the novel *The Hobbit* by J.R.R Tolkien were examined. The corpus was chosen specifically to evaluate the effectiveness of the tool in comparison to an analysis made by a human. The results were presented under two norms: 'matricial norms' and 'textual-linguistic norms.' The results of matricial analysis revealed statistics about the distribution of linguistic material in the source and target texts, while the results of textual-linguistic analysis offered information on the selection of linguistic material. In the light of the findings, one can suggest that the aim of this research has been fulfilled as the software provided significant amount of quantitative data allowing for further insights on operational norms.

Certain limitations to this study need to be acknowledged. First limitation is related to the preparation of the corpus for processing by the software. The efficacy of the software was limited to the quality of the input data. Therefore, pre-editing was required for scanned documents, which demanded a great deal of time and effort. Secondly, it should be noted that performance of the software is limited to the capacity of the servers. In other words, even though it is theoretically correct that the software can process as many texts as a researcher wishes to examine, the time required to process them depends on the computational power that the user benefits from. Third limitation deals with the textual-linguistic analysis. Accuracy of the results for this category is essentially dependent on the comprehensiveness of the datasets.

Despite all the limitations, it can be claimed that the expected outcome was achieved in this research. Now, further research and effort are needed to develop the software in order to test its viability for other text types. In that regard, this study has the potential for being a notable contribution to the field of Descriptive Translation Studies.

REFERENCES

Altıntaş, Kemal, Fazlı Can, and Jon M. Patton. 2007. "Language Change Quantification Using Time- Separated Parallel Translations." *Literary and Linguistic Computing* (Oxford University Press) 22, no. 4: 375-393. https://doi.org/10.1093/llc/fqm026.

Baker, Mona. 1993. "Corpus Linguistics and Translation Studies - Implications and Applications." In *Text and Technology: In Honour of John Sinclair*, edited by Mona Baker, Gill Francis and Elena Tignini-Bonelli, 233-250. Philadelphia: John Benjamins. https://doi.org/10.1075/z.64.15bak.

———. 2000. "Towards a Methodology for Investigating the Style of a Literary Translator." *Target* 12, no. 2: 241-266. https://doi.org/10.1075/target.12.2.04bak.

Bengi-Öner, Işın. 1999. "Türk Edebiyatında La Dame Aux Camélias Çevirileri." In *Çeviri Bir Süreçtir… Ya Çeviribilim?*, 45-67. İstanbul: Sel Yayıncılık.

Bilgin, Orhan. 2016. "Frequency Effects in the Processing of Morphologically Complex Turkish Words." Master's Thesis, Boğaziçi University.

Bozan, Nalan. 2021. "A Text Mining Approach to 'Operational Norms'." *transLogos Translation Studies Journal* 4, no. 1: 57–80. https://doi.org/10.29228/transLogos.31.

Cai, Yanli, and Jian-Tao Sun. 2009. "Text Mining." In *Encyclopedia of Database Systems*, edited by Ling Liu and M. Tamer Özsu, 3061-3065. Boston, MA: Springer. https://doi.org/10.1007/978-0-387-39940-9_418.

Can, Fazlı, and Jon M. Patton. 2012. "Determining translation invariant characteristics of James Joyce's Dubliners." In *Quantitative Methods in Corpus-Based Translation Studies: A Practical Guide to Descriptive Translation Research*, edited by Michael P. Oakes and Meng Ji, 209-230. Amsterdam: John Benjamins. https://doi.org/10.1075/scl.51.

Can, Fazlı, Ethem F Can, and Ceyhun Karbeyaz. 2010. "Translation Relationship Quantification: A Cluster-Based Approach and its Application to Shakespeare's Sonnets." In *Computer and Information Sciences*, edited by Erol Gelenbe, Ricardo Lent, Georgia Sakellari, Ahmet Sacan, Hakki Toroslu and Adnan Yazici, 117-120. Netherlands: Springer. https://doi.org/10.1007/978-90-481-9794-1_24.

Cheesman, Tom, et al. 2017. "Multi-Retranslation corpora: Visibility, variation, value, and virtue." *Digital Scholarship in the Humanities* 32, no. 4 (December): 739–760. https://doi.org/10.1093/llc/fqw027.

Çalışkan, Sevilay. 2020. "Text Mining Analysis of Translation, Social Communication and Literary Writing for Turkish." Master's Thesis, Bilkent University.

Çetin, Özge. 2009. “Çeviride İnsan Zekâsı ve Yapay Zekâ.” Master's Thesis, Muğla University.

Duymaz, Zeynep. 2020. “John Ronald Reuel Tolkien’in The Hobbit Adlı Eserinin Türkçe Çevirilerinin Yeniden Çeviri Varsayımı ve Çocuk Edebiyatı Kapsamında İncelenmesi.” [An analysis of the Turkish translations of John Ronald Reuel Tolkien’s The Hobbit within the framework of retranslation hypothesis and children’s literature.] Master's Thesis, Istanbul Okan University.

Fielding, Henry. 1749. *The History of Tome Jones, a Foundling*. Londra: Andrew Miller.

———. 1990. *Tom Jones*. Translated by Mina Urgan. İstanbul: İletişim Yayınları.

———. 1945. *Tom Jones: Sokakta Bulunmuş Bir Çocuğun Hikayesi*. Translated by Mina Urgan. İstanbul: Tan Basımevi.

Jackson, Peter, and Isabelle Moulinier. 2002. *Natural Language Processing for Online Applications*. Amsterdam: John Benjamins.

Ji, Heng. 2009. “Information Extraction.” In *Encyclopedia of Database Systems*, edited by Ling Liu and M. Tamer Özsu, 1476-1481. Boston, MA: Springer. https://doi.org/10.1007/978-0-387-39940-9_204.

Kit, Chunyu, and Nie Jian-Yun. 2015. “Information Retrival and Text Mining.” In *The Routledge Encyclopedia of Translation Technology*, edited by Chan Sin-wai, 494-535. New York, NY: Routledge.

Li, Defeng. 2017. “Translator style: A corpus-assisted approach.” In *Corpus Methodologies Explained: An empirical approach to translation studies*, edited by Meng Ji, Michael Oakes, Defeng Li and Lidun Hareide, 103-136. Abingdon, Oxon: Routledge.

Li, Hua. 2009. “Text Clustering.” In *Encyclopedia of Database Systems*, edited by King Liu and M. Tamer Özsu, 3044-3046. Boston, MA: Springer. https://doi.org/10.1007/978-0-387-39940-9_415.

Merriam-Webster, s.v. “exclamation,” accessed December 28, 2021, https://www.merriam-webster.com/dictionary/exclamation.

Monti, Johanna, Annibale Elia, Alberto Postiglione, Mario Monteleone, and Federica Marano. 2012. “In search of knowledge: text mining dedicated to technical translation.” *Proceedings of ASLIB 2011 Translating and the Computer Conference*. London.

Oi Yee, Olivia Kwong. 2015. “Natural Language Processing.” In *The Routledge Encyclopedia of Translation Technology*, edited by Chan Sin-wai, 563-577. New York, NY: Routledge.

Paice, Chris D. 2009. “Stemming.” In *Encyclopedia of Database Systems*, edited by Ling Liu and M. Tamer Özsu, 2790-2793. Boston, MA: Springer.

Princeton University. *WordNet 3.1 Database*. WordNet. 2010.

Shen, Dou. 2009. “Text Categorization.” In *Encyclopedia of Database Systems*, edited by Ling Liu and M. Tamer Özsu, 3041-3044. Boston, MA: Springer, 2009. https://doi.org/10.1007/978-0-387-39940-9_414.

———. 2009. “Text Summarization.” In *Encyclopedia of Database Systems*, edited by Ling Liu and M. Tamer Özsu, 3079-3083. Boston, MA: Springer. https://doi.org/10.1007/978-0-387-39940-9_424.

Tolkien, John Ronald Reuel. 1996. *Hobbit: Oradaydık ve Şimdi Buradayız*. Translated by Emel İzmirli. İstanbul: Altıkırkbeş / Mitos.

———. 2009. *The Hobbit*. London: HarperCollins.

———. 2015. *Hobbit*. Translated by Gamze Sarı. İstanbul: İthaki.

Toury, Gideon. 1978. “The Nature and Role of Norms in Translation.” In *Literature and translation: new perspectives in literary studies : with a basic bibliography of books on translation studies*, edited by James S Holmes, José Lambert and Raymond van den Broeck, 83-100. Leuven: Acco.

———. 1995. *Descriptive Translation Studies - and beyond*. Amsterdam: John Benjamins.

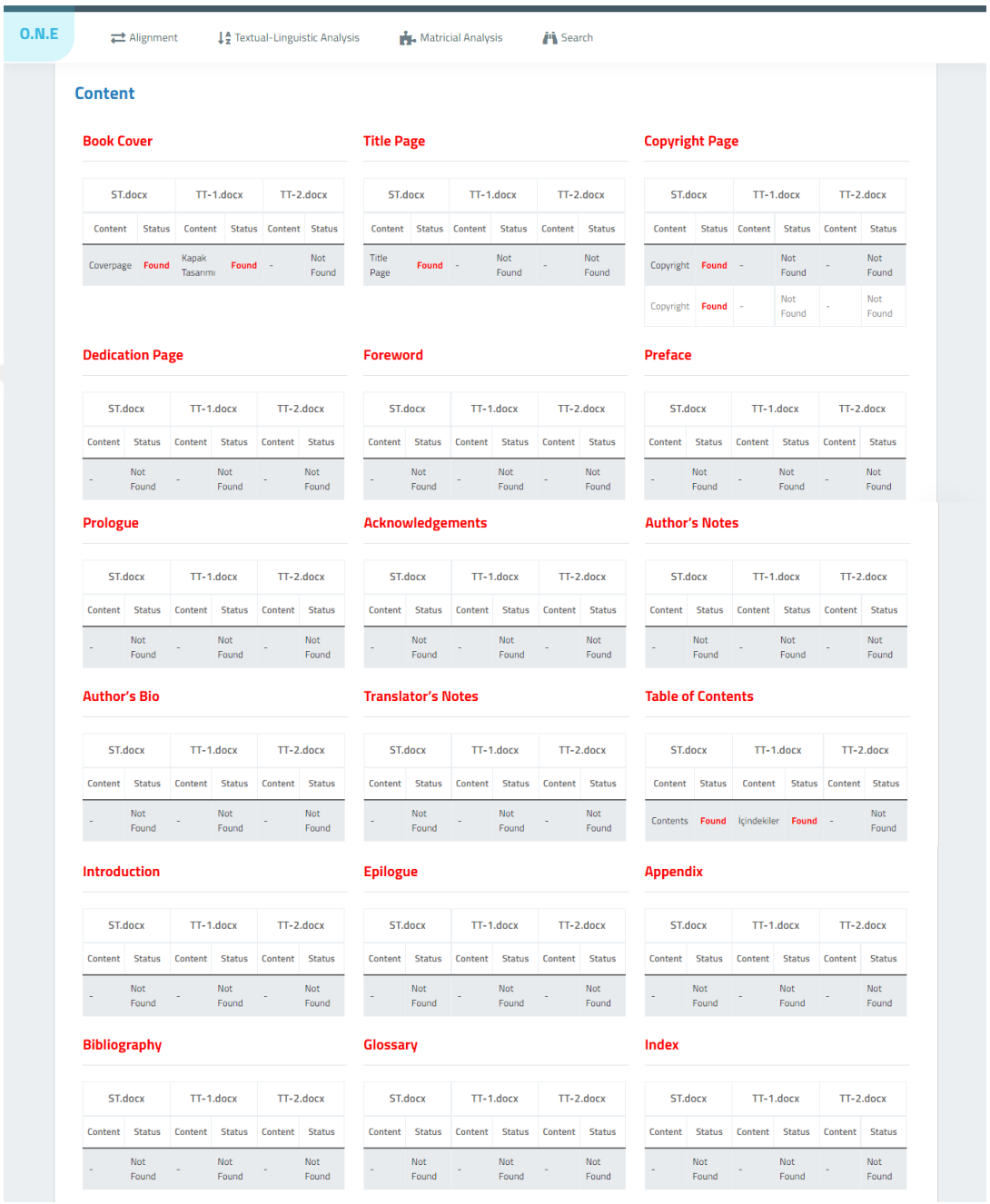
———. 2012. *Descriptive Translation Studies – and beyond*. rev. ed. Amsterdam: John Benjamins.

Urgan, Mina. 2013. *Bir Dinozorun Anıları*. İstanbul: Yapı Kredi Yayınları.

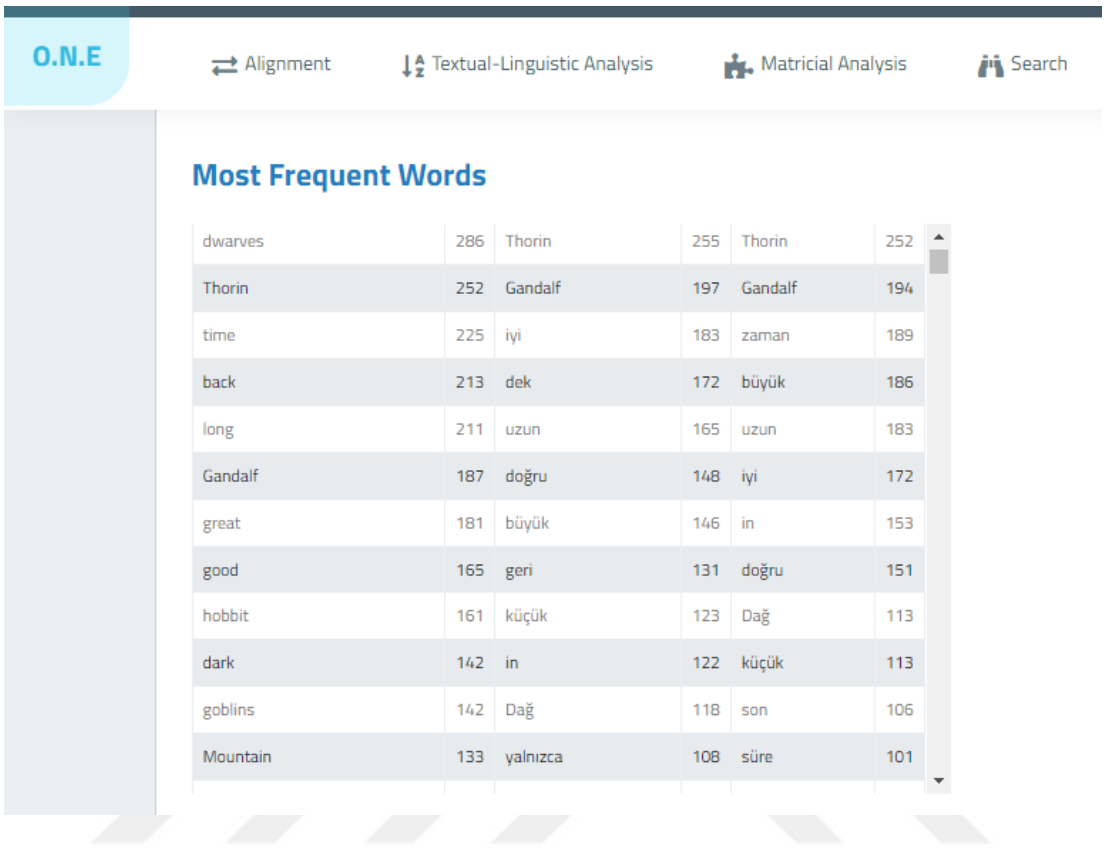
Zanini, Nadir, and Vikas Dhawan. 2015. “Text Mining: An introduction to theory and some applications.” *Research Matters: A Cambridge Assessment Publication*, no. 19 (Winter): 38-44.

APPENDICES

Appendix A. Screenshot of Content Results on O.N.E



Appendix B. Screenshot of Most Frequent Words Results



Appendix C. List of Proper Nouns

ST		**TT-1**		**TT-2**	
Proper Noun	**Count**	**Proper Noun**	**Count**	**Proper Noun**	**Count**
Gandalf	200	Bilbo	660	Bilbo	658
Baggins	127	Thorin	276	Thorin	268
Smaug	98	Gandalf	220	Gandalf	216
Gollum	86	Bilbo	660	Baggins	181
Balin	72	Thorin	276	Smaug	106
Beorn	70	Gandalf	220	Gollum	100
Bombur	62	Baggins	184	Beorn	74
Kili	38	Smaug	98	Bard	74
Elvenking	32	Gollum	94	Balin	71
Mirkwood	28	Bard	73	Bombur	56
Oakenshield	26	Beorn	70	Dori	40
Elrond	26	Balin	70	Goblin	40
Wargs	24	Bombur	60	Dale	39
Gloin	23	Dale	40	Kuyutorman	32
Dwalin	22	Goblin	40	Took	30
Bofur	22	Dori	38	William	30
Arkenstone	21	Kasvetormanı	36	Goblinler	26

Thror	21	William	28	Elrond	25
Bifur	19	Took	27	Dağaltı	25
Carrock	13	Meşekalkan	26	Meşekalkan	24
Tolkien	11	Elrond	26	Gloin	23
Esgaroth	10	Thror	26	Bofur	23
Girion	10	Goblinler	25	Thrain	23
Durin	10	Bofur	24	Hobbit	23
Ravenhill	7	Elfkral	24	Thror	21
Orcrist	6	Gloin	23	Arkentaşı	21
Gondolin	6	Dwalin	23	Dwalin	21
Tookish	5	Thrain	22	Bifur	20
Bolg	5	Bifur	21	Vagonkaya	15
Bagginses	5	Arkentaşı	20	Nori	14
Carc	5	Hobbit	14	Girion	14
Rivendell	4	Nori	14	Esgaroth	11
Thief	3	Girion	13	Tolkien	10
Glamdring	3	Esgaroth	11	Durin	10
Roac	3	Warglar	11	Moria	8
Hobbits	3	Durin	10	Kuzguntepesi	7
Bywater	2	Elfler	9	Roac	7
Warg	2	Moria	8	Bolg	6
Wilderland	2	Troll	8	Orcrist	6
Tookishness	2	Kuzguntepesi	7	Gondolin	6
Bladorthin	2	Gondolin	7	Elfler	6
Unwin	2	Bolg	7	Carc	5
Eggses	2	Hobbitin	6	Şimdiyse	5
Ssss	2	Kafadanbacaklı	6	Nain	5
Goblinses	2	Tookça	6	Hobbitin	5
Dorwinion	2	Wargların	5	Eklembacak	5
Bloor	1	Güzellergüzeli	5	Boğakükreten	4
Gundabad	1	Orcrist	5	Çatırt	4
Pymble	1	Carc	5	Warglar	4

Appendix D. List of Exclamations (First 50 Entries)

ST	TT-1	TT-2
ha!	Şışşşt!	Şşt!
beneath the sky!	Şimdi vakit yok!	Şş, şş!
You!	Şaak, Şlak!	Şş!
You may indeed!	İşinize bakın!	Şimdi, devam et!
Yes, you have!	İçeri gelin!	Şimdi zaman yok!
Yes, you do!	İzninizle!	Şimdi biraz müzik!
Yes, yes!	İyi günler!	Şimdi Dağa dönelim!
Yes please!	İntikam!	Şerefim üzerine!
Ya-harri-hey!	Özür dilerim!	Şendir Mayıs mevsimi!
Ya hoy!	Öyleydi!	Şak!
Ya hey!	Öldü!	İşinizi görün!

Wrong!	Çok haklısın!	İçeri gelin!
Work, work!	Çok güzel!	İyi!
What!	Çimdikleyip durma!	İyi sabahlar!
Well, well!	Çekiçle maşa!	İntikam!
Well, thief!	Çarpış çarpış!	Öğ!
Well, let him!	Çalış durma çalış!	Öyleydi!
Well, hurry up!	işte aşağıdaki vadide!	Öyle mi!
Well, Merry People!	ha!	Öldü mü!
Well!	gökyüzünün altında!	Çok haklısın!
Well done!	fSss, sss, gollum!	Çimdik atma!
We are trapped!	Zırvalamayı kes!	Çekiçle maşa!
Wait!	Zaman!	Çatırt, patırt!
Very well!	Yolunuz açık olsun!	Çalış, çalış!
Very pretty!	Yok ol!	izninizle!
Upon my word!	Yo, yo, hayır!	ha!
Um!	Yo, hayır!	Zamanı geldi!
Ugh!	Yaşasın, duyuyorum!	Yuvarlan-yuvarlan-yuvarlan delikten aşağıya!
Twelve yards!	Yarım dakika!	Yavaş!
To the Mountain!	Yanlış!	Yarım saniye!
To me!	Yakala, gaspet!	Yarım dakika!
Times up!	Yaa!	Yanıyor tüm yıldızlar!
Time!	Ya-harri-hey!	Vira salpa!
Throng!	Ya!	Vay vay!
Those were crows!	Ya hoy!	Vay canına!
Thorin!	Ya hey!	Vay başıma gelenler!
This way!	Vur!	Vay başıma gelen!
Thieves!	Vira et, vira!	Vadiye hoş geldiniz!
Thief, thief, thief!	Ve az yemeğimiz!	Uh!
They have come!	Vadiye hoşgeldiniz!	Troller!
Then baa—baa—baa!	Uff!	Thorin!
The key!	Trolller!	Thorin Meşekalkan, hizmetinizde!
The Eagles!	Thorin!	Teşekkür ederim!
The Arkenstone!	Thorin!	Ssss, sss, gollum!
Thats enough!	Thorin Meşekalkan, hizmetinizdeyim!	Sonunda geldik işte!
Thats better!	Teşekkür ederiz!	Sensin dangalak!
Thank you!	Teşekkür ederim!	Sen!
Swish, smack!	Taşıyor işte nehrimiz!	Sen tasalanma!
String, or nothing!	Tastamam!	Selam, Thorin!
Stop!	Tanrıya şükür ki!	Selam sana, Thorin!

Appendix E. First 50 Entries of Non-defined Words

ST	TT-1	TT-2
andst	akbn*	adeşe
bebother	alabililsin	ahmaksun
beleriand	almıycam	alazlı
bewuthered	alıyoruss	almamıss
bickerings	amnesia	almışından
biscuitish	aramıssda	altunda
blackberrying	ard	and
bloor	ardıçkuşu	anglo
burrahobbit	ardıçkuşundan	annamıyom
burrahobbits	arslanağızları	aramamıs
buttertubs	assisim	ardıçkuşu
bywater	assiss	arplara
carrocks	assissim	assla
coalmining	assissini	ayakaltından
coverpage	asssisssm	ağzunu
creepsy	assından	bah
diningroom	atmalıyıss	bakmamıss
dursn	ayakparmakları	bakınmaksınız
dwarvish	ayakparmaklarımda	balmumundan
ebook	ayrıldımki	barnahlarını
elven	azalıyo	başımısı
emptyhanded	bayats	beklemeliyiss
exceptfor**	başavuran	beklemessi
eyeses	başhizmetkârının	berü
frontdoor	beklemeliyiss	bidikteydi
gettin	bekleyemessin	biras
goblinses	bilemeyiss	bis
greybeards	biliyoruss	bisde
guesseses	bilşey	bisden
hadn	biss	bise
halffinished	bisse	bisi
handses	bissi	biss
harpercollins	bissim	bisse
harpercollinspublishers	bisten	bissim
helter	bitilemelecegiz	bissimle
hobbitlike	biy	bissoru
hobbits	butuk	bissürü
hurryings	böyek	bitiremeyüz
ivthe	büktüğümüssde	bos
jibbering	ceplerins	buççuk
jusst	ceplerinsde	buğün
losst	cücemsi	bööle
manflesh	değiştokuş	cad
miserabler	ding	camlaşana
nassty	diyebilmez	canuna
ork	dulmasını	canımısa

pocketses	dwarf	cebümde
praps	dwarfish	cessaretimiss
preciouss	dwarfs	copyright
preciousss	dwarwes	dahaki

*This entry is presented as it appears in TT-1 (except that it is written in capital letters in TT-1). It is an acronym used by the publisher of TT-1 which stands for 'altı kırkbeş book number.'

**This typo is due to ST because there is an image (of a rune) in between the words 'except' and 'for' without surrounding white spaces i.e. 'except[image]for'. O.N.E removes images before performing the task of 'Non-defined Words,' that's why, it appears as 'exceptfor.'

Appendix F. List of Measurement Units

ST	TT-1	TT-2
for many **miles** round called	aptal, bir **mil** öteden bile	kimseler, bir **mil** öteden duyabilecekleri
hear a **mile** off. They	havada yüzlerce **mil** uçup bir	kendisini yarım **litre**lik bir bira
a hundred **yards** through the	küçük. "Kapı beş **ayak** yüksekliğinde	havada yüz **metre** uçtuktan sonra
small. 'Five **feet** high the	"Kapı **beş** **ayak** yüksekliğinde ve	bir buçuk **metre** yüksekliğinde ve
'Five **feet high** the door	bir iki **mil** daha öteye,	geçip bir **mil** daha koştuğunu
for a **mile** or more.	bezmişti; yüzlerce **mil** uzakta olmayı	alayı bir **metre** ötemizden geçmedikçe
had passed **two feet** off.	´Burada, "kapı **beş** **ayak** yüksekliğinde	buradan yüzlerce **mil** uzakta olmayı
passed **two feet** off. As	"kapı **beş** **ayak** yüksekliğinde ve	'bir buçuk **metre** yüksekliğinde ve
a hundred **miles** away, and	Evcil Ev´i **millerce** geride bıraktıktan	ancak bir **metre** yanından geçip
say 'five **feet high** the	değene dek **dört** **ayak** üzerinde	Gün kavuşmadan **millerce** uzakt a olmalıyız.
'five **feet high** the door	dek **dört ayak** üzerinde emekledi;	hayvan üç **metre** havaya zıpladıktan
Homely House **miles** behind, they	kafasının üzerinden, **üç** **ayak** havaya	bile bir **mil** aşağısındaki toprakta
feet, many **many feet** which	üzerinden, **üç ayak** havaya ve	dev tarafından **millerce** yaylaya savrulan
many **many feet** which seemed	havaya ve **yedi ayak** ileriye;	patikanın birkaç **mil** kuzeyindesiniz. Buraya

only a **yard** to one	ve **yedi ayak** ileriye; geçitin	evinin yüz **mil** yakınandayken bir
he jumped, **seven** **feet** forward	tam zamanında, **dört** **ayak** üzerinde	az bir **litre** bal likörü
jumped, **seven feet** forward and	zamanında, **dört** **ayak** üzerinde güçlükle	bulmak için **millerce** yürümem gerekti;
must be **miles** on before	Karanlık basmadan **millerce** yol almış	geriye doğru **millerce** yürüdüm Artık
have gone **miles** and miles,	iki üç **metre** havaya zıplayıp	yürütüp Vagonkaya'nın **millerce** gü neyinde büyük
ground a **mile** below even	ve bir **mil** aşağıdaki toprakta	Vagonkaya'nın yüz **mil** kuzeyinden bile
piece cast **miles** into the	ovaya doğru **millerce** fırlatılmış devasa	pek çok **mil** koyduklarında konuşmaya,
still some **miles** north of	patikanın birkaç **mil** kuzeyindeyiz. Buralara	sırtında bununla **millerce** ayak sürümek
a hundred **miles** of his	birkaç yüz **mil** civarında olduğunuz	ayrılıp yüz **mil** kadar kuzeye,
to walk **miles** before I	az çeyrek **litre** bal likörü	On iki **metreden** fazla olduğunu
and then **miles** back again	bulabilmem için **millerce** ve sonra	"On iki **metre** mi! Ben
great river **miles** south of	için tekrar **millerce** geriye yürümem	on iki **metre** epey uzun.
a hundred **mile**s north of	ve Taşıtkaya´nın **millerce** güne yindeki büyük	Bilbo. "Yarım **metre** daha uzağa
put many **miles** between them	Taşıtkaya´nın yüz **mil** kuzeyine geçmeye	ormanın içinde **millerce** uzakta ne
trudging for **miles** and miles	ile aralarını **millerce** açınca konuşmaya,	duyuldu ve **bir** **ayak** parmağı
miles and **miles** with all	onca şeyle **millerce** yorgun argın	ve **bir ayak** parmağı hızla
two hundred **miles** or so	çıkıp ikiyüz **mil** ya da	sınırının birkaç **mil** içinde, doğu
Still twelve **yards** is as	değil. Oniki **metreden** fazla olduğunu	bir iki **metre** aşağıdaki soğuk
had only **one foot** on	'Oniki **metre** mi? Ben	nehrin birkaç **metre** yukarısında sol
only **one foot** on the	azından otuz **metre** derdim, ama	kırk beş **metrelik** bir uçurum
cave some **miles** within the	de oniki **metre** şimdi bir	yaklaşık bir **metre** yukarısında bir

day, some **miles** up the	şimdi bir **mil** gibi geliyor.	Bir buçuk **metre** yüksekliğinde ve
hundred and **fifty feet** beside	ve ormanda **millerce** uzaktaki her	ve bir **metre** genişliğinde bir
and **fifty feet** beside them	yükseldi ve **bir ayak** olanca	iyeliklerindeyse, son **gramına** kadar bilirler;
suddenly about **three feet** from	ve **bir ayak** olanca gücüyle	nehre giden **millerce** açık arazide
about **three feet** from the	bir iki **mil** dahilindeki büyük	
A door **five feet** high	bir buçuk **metre** aşağıdaki soğuk	
door **five feet** high and	nehrin birkaç **mil** yukarısında sola,	
five **feet high** and three	birden yerden **üç ayak** yükseklikte	
to an **ounce** as a	yerden **üç ayak** yükseklikte bir	
long open **miles** back to	ayak yüksekliğinde **üç ayak** eninde	
was several **inches** longer, and	yüksekliğinde **üç ayak** eninde bir	
	de her **gramını** kanun gibi	
	ve açık **millerce** yolu yayan	
	altınımızın bir **gramını** bile, ne	

Appendix G: List of Idioms

ST	
Idiom	**Context**
into the bargain	"And a very fine morning for a pipe of tobacco out of doors, **into the bargain**
beats me	"What the 'ell William was a—thinkin' of to bring us into these parts at all, **beats me**—and the drink runnin' short, what's more," he said jogging the elbow of William, who was taking a pull at his jug.
as best you can	Either you help me to look for him, or I go and leave you here to get out of the mess **as best you can** yourselves
born and bred	As all things come to an end, even this story, a day came at last when they were in sight of the country where Bilbo had been **born and bred**, where the shapes of the land and of the trees were as well known to him as his hands and toes
the comings and goings	Though all the old adornments were long mouldered or destroyed, and though all was befouled and blasted with **the comings and goings** of the monster, Thorin knew every passage and every turn
to death	Bilbo was dreadfully frightened, but he had the sense to understand what had happened and to sneak behind a big barrel which held drink for the

	goblin—guards, and so get out of the way and avoid being bumped into, trampled **to death**, or caught by feel.
every now and again	Nothing moved in the waste, save the vapour and the water, and **every now and again** a black and ominous crow
every now and again	The dwarves could not, of course, compare with the hobbit in real stealth, and they made a deal of puffing and shuffling which echoes magnified alarmingly; but though **every now and again** Bilbo in fear stopped and listened, not a sound stirred below
every now and again	The air grew warmer as they got lower, and the smell of the pine—trees made him drowsy, so that **every now and again** he nodded and nearly fell off, or bumped his nose on the pony's neck
every now and again	but **every now and again** he would open one eye, and listen, when a part of the story which he did not yet know came in.
every now and again	The only sound was the sound of the stony water, and **every now and again** the harsh croak of a bird
every now and then	The great bunch seemed very heavy to his arms, and his heart was often in his mouth, in spite of his ring, for he could not prevent the keys from making **every now and then** a loud clink and clank, which put him all in a tremble.
every now and then	There were passages leading off to the side **every now and then**, as he knew by the glimmer of his sword, or could feel with his hand on the wall
keep an eye on	Not only to guard you and guide you, but to **keep an eye on** the ponies too
after a fashion	"Then the prophecies of the old songs have turned out to be true, **after a fashion**." said Bilbo.
play with fire	Also naughty little boys that **play with fire** get punished." He said it to make them angry, and to show them he was not frightened of them—though of course he was, wizard though he was
put your foot in it	Many a time afterwards the Baggins part regretted what he did now, and he said to himself: "Bilbo, you were a fool; you walked right in and **put your foot in it**."
the four corners of the world	He is an enemy far beyond the powers of all the dwarves put together, if they could all be collected again from **the four corners of the world**
out of the frying—pan into the fire	CHAPTER VI: **out of the frying—pan into the fire**
out of the frying—pan into the fire	Chapter VI: **out of the frying—pan into the fire**
out of the frying—pan into the fire	"Escaping goblins to be caught by wolves." he said, and it became a proverb, though we now say "**out of the frying—pan into the fire**" in the same sort of uncomfortable situations.
near at hand	"We will have King Bard." the people **near at hand** shouted in reply.
lend a hand	"Why don't they come and **lend a hand**." Lo and behold
be hard put	Dwarf—mail may be good, but they will soon **be hard put** to it
late in the day	After a while, however, the river took a more southerly course and the Mountain receded again, and at last, **late in the day** the shores grew rocky, the river gathered all its wandering waters together into a deep and rapid flood, and they swept along at great speed.

lo and behold	“Why don’t they come and lend a hand.” **lo and behold**
cut loose	First of all a barrel was **cut loose** by Bilbo and pushed to the shore and opened
loud and clear	But, why O People.”—and here the Master rose to his feet and spoke very **loud and clear**—“Why do I get all your blame.
loud and clear	Below it many elves were singing **loud and clear** on the banks of the stream.
like mad	“That’ll teach ’em,” said Tom; for Bifur and Bombur had given a lot of trouble, and fought **like mad**, as dwarves will when cornered.
a mile off	There is little or no magic about them, except the ordinary everyday sort which helps them to disappear quietly and quickly when large stupid folk like you and me come blundering along, making a noise like elephants which they can hear **a mile off**
make a move	There they sat for a long while and did not dare to **make a move**.
in the nick of time	“And what brought you back **in the nick of time**.”
in the nick of time	So they too had gathered in great numbers, under the great Eagle of the Misty Mountains; and at length smelling battle from afar they had come speeding down the gale **in the nick of time**
very much	After all the others had ordered their breakfasts without so much as a please (which annoyed Bilbo **very much**), they all got up
a parting shot	And so do burglars,” he added as **a parting shot**, as he darted back and fled up the tunnel.
at a pinch	But he felt he could not trust this slimy thing to keep any promise **at a pinch**
in a pinch	“Gets funny queer fits, but he is one of the best, one of the best—as fierce as a dragon **in a pinch**.”
in a pinch	If you have ever seen a dragon **in a pinch**, you will realize that this was only poetical exaggeration applied to any hobbit, even to Old Took’s great— grand—uncle Bullroarer, who was so huge (for a hobbit) that he could ride a horse
read the runes	“They were not made by any troll, nor by any smith among men in these parts and days; but when we can **read the runes** on them, we shall know more about them.”
shades of	They would have been surprised, if they could have seen what happened down by the shore, after they had gone and the **shades of** night had fallen
on the shelf	He was feeling very queer indeed in his head by this time after the dreadful adventures of the last three days with next to nothing to eat, and he found himself saying aloud: “Now I know what a piece of bacon feels like when it is suddenly picked out of the pan on a fork and put back **on the shelf**.”
on the shelf	Later on you might have seen a bright fire **on the shelf** of rock and the figures of the dwarves round it cooking and making a fine roasting smell
out of sight	The truth was he had been lying quiet, **out of sight** and out of mind, in a very dark corner for a long while.
blow smoke	“But I have no time to **blow smoke**—rings this morning
on the spot	“Wind, wind of course,” he said, and he was so pleased that he made up one **on the spot**
under the table	He was blowing the most enormous smoke—rings, and wherever he told one to go, it went—up the chimney, or behind the clock on the

	mantelpiece, or **under the table**, or round and round the ceiling; but wherever it went it was not quick enough to escape Gandalf
under the table	Beside them a pony pushed two low—seated benches with wide rush—bottoms and little short thick legs for Gandalf and Thorin, while at the far end he put Beorn's big black chair of the same sort (in which he sat with his great legs stuck far out **under the table**)
tell tales	When dinner was over they began to **tell tales** of their own, but Beorn seemed to be growing drowsy and paid little heed to them
a long time ago	"We have, **a long time ago**; but what has he got to do with us." asked the king.
lose track	"You **lose track** of time inside goblin—tunnels
not care tuppence	They did **not care tuppence** about the butterflies, and were only made more angry when he told them of the beautiful breeze, which they were too heavy to climb up and feel.
not believe a word of it	"Do you now." said the dragon somewhat flattered, even though he did **not believe a word of it**.
have the last word	But he could not resist the temptation to **have the last word**
TT-1	
Idiom	**Context**
bardaktan boşanırcasına	Çay vakti geçmişti; yağmur **bardaktan boşanırcasına** yağıyordu ve tüm gün yağmıştı; sular başlığından gözlerine damlıyordu ve pelerini de tüm suyu çekmişti; midillisi yorgundu ve taşlarda tökezleyip duruyordu; diğerleri de konuşmaya kalkışılmayacak denli somurtuyorlardı
göz kulak olmak	Yalnızca sizi korumak ve yol göstermek için değil aynı zamanda midillilerine **göz kulak olmak** için peşimizde
kapağı atmak	Eğer yolunu bulursa davet edilmediği bir akşam yemeğine **kapağı atmakta** bir an olsun duraksamayacaktı; o kadar uzun süredir bunu yapmaya zorlanmıştı ki; onu ilgilendiren yalnızca tıka basa dolu bir kilerin lezzeti değildi, gerçekten aç kalmanın ne demek olduğunu biraz fazlasıyla iyi öğrenmişti
kemerleri sıkmak	Şimdi boş midelerinin üzerine **kemerleri sıkmak** ve boş çantalarını yüklenip, uzanıp açlıktan ölmeden önce bu ormandan çıkmaya dair öyle büyük umutlar da taşımadan yol boyunca ağır ağır ilerlemekten başka çareleri kalmamıştı
kulak misafiri olmak	Asıl sürpriz bir gün muhafızlardan bazılarının konuşmasına **kulak misafiri olmak** ve hapiste, özellikle derin karanlık bir yerde, bir cüce daha olduğunu öğrenmesi oldu
paha biçilmez	Smaug kanatları ölçüp biçilemez büyüklükte bir yarasaymışçasına kıvrılmış ve hafifçe bir tarafına abanmış yatıyordu; ejderha **paha biçilmez** yatağında boylamasına yattığından hobbit, altın parçacıkları ve değerli taşlarla kaplanmış uzun soluk karnını ve alt kısımlarını görebildi
paha biçilmez	Ey **paha biçilmez** derece varlıklı Smaug, başarınızın size bazı amansız düşmanlar kazandırdığını fark etmiş olmalısınız.´
şimşek gibi	'Durun.' diye bağırdı gök gürlemesi gibi bir sesle ve asası **şimşek gibi** bir ışık saçtı
şimşek gibi	Ve tam yukarıdan goblinlerin mızraklarının ortasına atlamaya hazırlandığı sırada asasından **şimşek gibi** bir parlaklık patladı
şimşek gibi	'İntikam.´ diye gürledi burnundan soluyarak ve gözlerinin ışıkları salonu yerden tavana dek kızıl bir **şimşek gibi** aydınlattı

şöyle bir göz atmak	Bilbo **şöyle bir göz atmak** için gözünün tekini araladı ve kuşların çoktan iyice yükseldiğini, dünyanın çook uzaklarda kaldığını ve dağların arkalarındaki uzaklıklara karıştığını gördü
yalayıp yutmak	Onların orman insanlarının dostları olduklarını ve buraya casusluk yapmaya geldiklerini ve planlarının haberini vadilere götüreceklerini, sonra da goblinlerle kurtların, tutsak yakalamak ve uykularından aniden uyandırılmış insanları **yalayıp yutmak** yerine korkunç bir çarpışmada savaşmak zorunda kalacaklarını düşünüyorlardı
yıldırım gibi	Bir **yıldırım gibi** hızla aşağı indiğinde pek çoğunu öldürecek olsa da bu onun da sonu olacaktı
zerre kadar	Bilbo bunca yolu bir sürü macera yaşayarak onu görmek için gelmişti, ama şimdi görünüşünü **zerre kadar** beğenmemişti.
buyur etmek	Bu arada güneş de yükselmişti; ve ön kapı da içeriye ılık bir bahar meltemini **buyur etmek** üzere açık duruyordu
idare etmek	Daha yolculuğun sonu gelmeden cep mendilleri ve daha bir sürü şey olmadan **idare etmek** zorunda kalacaksın
idare etmek	Birden keklerinin yetmeyebileceği gibi korkunç bir düşünce geldi aklına, — gerçi evsahibi olarak görevini bilir ve ne kadar rahatsız edici olursa olsun yerine getirirdi — ama keksiz **idare etmek** zorunda da kalabilirdi.
kaşlarını çatmak	Thorin ise gerçekten kibirli birisiydi ve hizmet konusunda hiçbir şey söylemedi; ama zavallı Bay Bilbo Baggins öyle çok özür diledi ki sonunda, 'lütfen, rica ederim' diye homurdandı ve **kaşlarını çatmaktan** vazgeçti.
ne pahasına olursa olsun	Bilbo sefil yaratığı neyin çıldırttığını bulamadı, ama her şeyin mahvolduğunu ve Gollum´un **ne pahasına olursa olsun** onu öldürmek niyetinde olduğunu gördü
TT-2	
Idiom	**Context**
ağzına sağlık	"Hay **ağzına sağlık**." diye haykırdı Bilbo ve kendisini yere attı.
bardaktan boşanırcasına	Çay saati geçmişti; yağmur **bardaktan boşanırcasına** yağıyordu ve bütün gün devam etmişti; kukuletası gözlerinin üzerine düşüyordu, pelerini sırılsıklamdı; midilli yorgundu ve taşlara takılıp tökezliyordu; diğerleri de konuşamayacak kadar aksileşmişti
göz kulak olmak	Sırf sizi korumak ve size rehberlik etmek için değil, midillilere **göz kulak olmak** için de
pılı pırtı	Bilbo bu sözleri duydu ve taşın bulunması durumunda —yastık olarak kullandığı **pılı pırtı**dan oluşan eski bir bohçaya sarılmıştı— ne olacağını merak ederek korktu
su yüzüne çıkmak	Yine de tek oğlu olan Bilbo'nun, güvenilir ve rahatına düşkün babasının kopyası gibi görünmesine ve davranmasına karşın, Took tarafından, **su yüzüne çıkmak** için fırsat kollayan tuhaf bir şey almış olması muhtemeldir
şimşek gibi	Ses Gandalf'ı saniyenin binde biri kadar sürede uyandırıp ayıltmıştı ve onu yakalamaya gelen ve mağarada **şimşek gibi** çakan, korkunç bir ışıkla ve barutu andıran bir kokuyla karşılaşan goblinlerin birkaçı ölmüştü.
zerre kadar	Korkuları yerindeydi, zira fikrinden **zerre kadar** hoşlanmadılar ve içinde bulundukları tehlikeye rağmen yüksek sesle homurdanmaya başladılar.
zerre kadar	Bilbo onu görmek için çok uzaklardan gelip pek çok serüvenden geçmişti, ama dağın görüntüsünden **zerre kadar** hoşlanmamıştı.

açık seçik	"Yine de her şeyin **açık seçik** ortaya konulmasını istiyorum," dedi Bilbo inatla, iş bilir tavrını (genellikle kendisinden borç isteyen kimselere sakladığı) takınarak ve bilge, ileri görüşlü, profesyonel görünüp, Gandalf'ın tavsiyesine layık olmak için elinden geleni yaparak
alnının akıyla	İşten **alnının akıyla** çıkman mümkün
bereket versin	Ama bütün ceplerine vurup her yanını yoklayarak kibrit ararken eli küçük kılıcının kabzasına değdi — trollerden aldığı ve aklından hepten çıkmış olan küçük hançere; **bereket versin**, silahı pantolonunun içinde taşıdığından onu goblinler de fark etmemişti.
bıyık altından gülmek	Artık zihninde berbat bir şüphe büyümeye başlamıştı —bu önemli noktayı cüceler de mi unutmuştu, yoksa baştan beri ona **bıyık altından gülmek**te miydiler
göz açıp kapayıncaya kadar	Elbette sizin cevabı bildiğinizi veya **göz açıp kapayıncaya kadar** bilebileceğinizi tahmin ediyorum, evinizde, düşüncelerinizi bulandıran yenme tehlikesi olmadan rahatça oturduğunuz için
hatırı sayılır	Güzelim kapıda **hatırı sayılır** bir göçük açmış, bu arada da önceki sabah oraya koyduğu gizli işareti silivermişti.
hatırı sayılır	Çok geçmeden nehir kenarındaki köyde **hatırı sayılır** bir kargaşa çıktı; ama Bilbo elinde kendisine ait olmayan bir somun ekmek, deriden yapılma bir şişe dolusu şarap ile bir turta ile ormana kaçtı
kolaçan etmek	Bu mağaralardan bazıları da goblinlerin öncesine, çağların başlangıcına dayanır; goblinler mağaraları genişletip geçitlerle birbirine bağlamıştır ve mağaraların ilk sahipleri hâlâ ırak köşelerde sinip etrafı **kolaçan etmek**tedir.
ne çıkar	Bazıları şimdiden fıçıları hepsi birden su yüzüne **çıkar** çıkmaz dışarı salmak üzere su kapısındaki ızgaralı kapıyı tutan ipleri çekmeye gitmişti.
saman gibi	Yine de midem bulanıyor, bacaklarım da **saman gibi**."

ÖZGEÇMİŞ			
Adı, Soyadı	Nalan		BOZAN
Doğum Yeri ve Yılı			
Bildiği Yabancı Diller	İngilizce		Almanca
ve Düzeyi	İleri		Başlangıç
Eğitim Durumu	**Başlama - Bitirme Yılı**		**Kurum Adı**
Lise	2009	2013	Özel Aziziye Anadolu Lisesi
Lisans	2013	2017	İstanbul Üniversitesi Mütercim Tercümanlık (İngilizce)
Yüksek Lisans	2019	2022	İstanbul 29 Mayıs Üniversitesi Çeviribilim (İngilizce)
Çalıştığı Kurum/lar	**Başlama - Ayrılma Yılı**		**Kurum Adı**
1. Serbest Çevirmen	2017	2018	Linguist Medikal Dil ve Danışmanlık Hizmetleri
2. Öğretim Görevlisi	2018	2020	İstanbul Gelişim Üniversitesi
3. Serbest Çevirmen	2019	-	MioMotto Dil ve Danışmanlık Hizmetleri
Üye Olduğu Bilimsel ve Mesleki Kuruluşlar	Tekom Türkiye		
Katıldığı Proje ve Toplantılar			
Yayınlar	Bozan, N. "A Text Mining Approach to 'Operational Norms'". *transLogos Translation Studies Journal* 4 (2021): 57-80		
Diğer			
İletişim (E-Posta)			
Tarih **İmza** **Adı Soyadı**	01/02/2022 Nalan BOZAN		