EVALUATION OF LEXICO – SEMANTIC FEATURES OF KHAYYAM'S QUATRAINS: COMPARATIVE ANALYSIS OF HUMAN TRANSLATION, GOOGLE TRANSLATE, AND TARGOMAN

MOHAMMAD AMIN EBRAHIMI FARD SHARABIANI

UNIVERSITI SAINS MALAYSIA

2024

EVALUATION OF LEXICO – SEMANTIC FEATURES OF KHAYYAM'S QUATRAINS: COMPARATIVE ANALYSIS OF HUMAN TRANSLATION, GOOGLE TRANSLATE, AND TARGOMAN

by

MOHAMMAD AMIN EBRAHIMI FARD SHARABIANI

Thesis submitted in fulfilment of the requirements
for the degree of
Doctor of Philosophy

December 2024

ACKNOWLEDGEMENT

I would like to express my deepest gratitude to my chief academic supervisor, Associate Professor Dr. Manjet Kaur Mehar Singh, for her invaluable guidance, unwavering support, and remarkable patience throughout my academic journey. I also wish to extend my thanks to my former supervisor, Professor Tengku Sepora Binti Tengku Mahadi, for her insightful advice until her retirement. Their extensive expertise and considerable experience have been a continuous source of encouragement throughout my research. Furthermore, I am grateful to Dr. Mohamed Abdou Moindjie and Dr. Paramaswari Jaganathan for their valuable feedback and technical support. Finally, I wish to acknowledge my parents and sibling for their unwavering support, which has helped me maintain my morale during this challenging period. The successful completion of my academic pursuits would have been impossible without the profound understanding and support I have received over the past few years.

TABLE OF CONTENTS

ACKNOWLEDGEMENT ..... **ii**
TABLE OF CONTENTS ..... **iii**
LIST OF TABLES ..... **viii**
LIST OF FIGURES ..... **ix**
LIST OF ABBREVIATIONS ..... **xi**
LIST OF APPENDICES ..... **xii**
ABSTRAK ..... **xiii**
ABSTRACT ..... **xiv**
CHAPTER 1 INTRODUCTION **1**
1.1 Overview 1
1.2 Background 1
1.2.1 Lexico Semantics Features 4
1.3 Overview of Analysis Parameters and Variables 7
1.3.1 Analysis Parameters and Variables 8
1.4 Statement of the Problem 9
1.5 Objective of the Study 12
1.5.1 Objective One 12
1.5.2 Objective Two 12
1.5.3 Objective Three 12
1.6 Research Questions 13
1.6.1 Resaerch Question Onex 13
1.6.2 Research Question two 13
1.6.3 Research Question Three 13
1.7 Significance of the Study 14
1.8 Limitations of the Study 16
1.9 Definition of Key Terms 17
1.9.1 Literary Translation 17
1.9.2 Human Translation 18
1.9.3 Machine Translation 18
1.9.4 Quatrains 19

1.9.5 Lexico Semantic Features……..…….…….....…….…..19
1.9.6 Google Translate……...……20
1.9.7 Targoman…….....……..…20
1.9.8 Errors……..…….……..……..……...20
1.10 Hypotheses of the Study.....21
1.11 Organization of the Study……..…..…..……..……...…….22
1.12 Summary……..……..…….……...….24
CHAPTER 2 LITERATURE REVIEW 26
2.1 Overview 26
2.2 Translation and its History 26
2.3 Translation Studies and its History 28
2.3.1 Evolution of Translation Studies..... 33
2.3.2 Historical Context and Influential Figures……...……..33
2.3.3 Modern Translation Studies…….……...……34
2.3.4 Relevance Evolution of Translation Studies to the Thesis…...…..34
2.4 Human Translation..... 34
2.5 Machine Translation and its History 36
2.5.1 Evaluation of Machine Translation 53
2.5.2 Difficulties with MT Systems 59
2.6 Literary Translation..... 61
2.6.1 Classical Persian Poetry and World Literature 66
2.7 Challenges in Literary Translation..... 69
2.7.1 Multiplicity of Meaning 70
2.7.2 Integration of Content and Form..... 70
2.7.3 The Style in Literary Translation 73
2.7.4 Translation of Literary Styles..... 74
2.8 Poetry Translation 73
2.9 Poetry (UN) Translatability..... 78
2.10 Semantics Definition and Translation 82
2.10.1 Collocation 83
2.10.2 Idiom…….…… ……..….....…….84
2.10.3 Literary Figure …….……..……...….....85
2.10.4 Phrasal Verb……...…….....86
2.10.5 Proverb ……..…….…….....…... ….87

2.10.6 Semantic Level (Sense and Reference) ……..…....…..….. 90
2.10.7 The Descriptive Scope of Lexical Semantics…….93
2.10.8 Polysemy and Vagueness ……..……..…….94
2.10.9 Semantic Relations……..97
2.10.10 Lexical Fields and Componential Analysis……99
2.10.11 Lexical Relations...…...……...…….…….100
2.10.12 Lexico-Semantics and its Ambiguity in Translation…….…....101
2.11 Introduction of several well-known English to Persian MT systems 110
2.11.1 Google Translate 110
2.11.2 Babylon Translator..... 115
2.11.3 Pars Translator 115
2.11.4 Padideh..... 115
2.11.5 Targoman 116
2.12 Research on Translation Error Type 119
2.13 Translation error analysis..... 122
2.14 Steps to Error Analysis 122
2.15 Error Taxonomies..... 124
2.15.1 The Major Error Taxonomies Proposed by Albir (1995)..... 124
2.15.2 The major error taxonomies proposed by the American (ATA).. 125
2.15.3 Dulay, Burt and Krashen's Error Taxonomies..... 125
2.15.4 Liao's (2010) Classification of Errors 127
2.16 Classification of Error Types Adopted from Liao (2010)..... 129
2.16.1 Rendition Errors..... 129
2.16.2 Language Errors..... 131
2.16.3 Miscellaneous Errors..... 134
2.17 Julian House Covert/Overt Model 138
2.18 Nida Equivalency Theory…….....144
2.19 Omar Khayyam & quatrains (Rubiaiyat)..... 147
2.20 Edward Fitz Gerald 149
2.21 Theoretical Perspective (House's Overt/Covert Model (1997) & Liao's Taxonomy of Errors (2010) And Their Intersection with Equivalence Theory ……..……..…..……..153
2.22 Related Studies 158

2.23 Relevance of the Models to the Research Topic……..…….162
2.24 Summary 164
CHAPTER 3 METHODOLOGY 167
3.1 Overview 167
3.2 Research Design 167
3.3 Corpus of the Study 169
3.3.1 Omar Khayyam and Quatrains (Ruba'i) 170
3.3.2 Justification for Selecting the Corpora 171
3.3.3 Data Collection in Source Language 173
3.3.3(a) Inclusion Criteria 174
3.3.3(a)(i) Sampling Technique 174
3.3.3(a)(ii) Rationale for Sample Size……..175
3.3.3(a)(iii) Data Analysis Framework……175
3.3.3(b) The selected quatrains in Persian to answer the research questions 177
3.3.3(b)(i) Selected Quatrains for Research Question One 178
3.3.3(b)(ii) Selected Quatrains for Research Question Two 179
3.3.3(b)(iii) Selected Quatrains for Research Question Three 181
3.3.4 Data Collection in Target Language 183
3.3.5 Units of Analysis 185
3.4 Theoretical Framework 186
3.4.1 Theoretical Framework and Methodological Integration…..187
3.4.2 The Methodological Integration of House's Overt/Covert Model and Liao's Error Analysis Model 188
3.5 Procedure 190
3.6 Data Analysis 195
3.7 Summary 200
CHAPTER 4 FINDINGS AND DISCUSSION 201
4.1 Introduction 201
4.2 The first research question 201
4.3 The second research question 307

4.4 The third research question…….....367
4.5 Summary……..….425
CHAPTER 5 CONCLUSION AND FUTURE RECOMMENDATIONS....426
5.1 Introduction……...….426
5.2 Major Findings and Their Implications 426
5.3 Discussion……..435
5.4 Conclusion…….....435
5.5 Contribution of the Research……438
5.6 Implication of the Study……440
5.6.1 Implication for Teaching……...……440
5.6.2 Implication for Translation……....441
5.7 Recommendations for further study…….…446
REFERENCES.....448
APPENDICES

LIST OF TABLES

Page

Table 2.1 The Descriptive Scope of Lexical Semantics ……..……94
Table 2.2 Classification of Error Types Adopted from Liao (2010)…… 128
Table 2.3 Classification of Error Types Adopted from Liao (2010)…… 154
Table 4.1 Liao's errors classification applied to GT and Targoman……208
Table 4.2 House TQA model applied to GT and Targoman……. 210
Table 4.3 Liao's errors classification applied to GT and Targoman……. 213
Table 4.4 House TQA model applied to GT and Targoman……218
Table 4.5 Liao's errors classification applied to GT and Targoman……225
Table 4.6 House TQA model applied to GT and Targoman……. 227
Table 4.7 Liao's errors classification applied to GT and Targoman……232
Table 4.8 House TQA model applied to GT and Targoman……235
Table 4.9 Liao's errors classification applied to GT and Targoman……239
Table 4.10 House TQA model applied to GT and Targoman……..245
Table 4.11 Liao's errors classification applied to GT and Targoman……250
Table 4.12 House TQA model applied to GT and Targoman……..253
Table 4.13 Liao's errors classification applied to GT and Targoman……258
Table 4.14 House TQA model applied to GT and Targoman……..268
Table 4.15 Liao's errors classification applied to GT and Targoman……270
Table 4.16 House TQA model applied to GT and Targoman……..275
Table 4.17 Liao's errors classification applied to GT and Targoman……283
Table 4.18 House TQA model applied to GT and Targoman……384
Table 4.19 Liao's errors classification applied to GT and Targoman……290
Table 4.20 House TQA model applied to GT and Targoman……..294
Table 4.21 The aggregate table of error numbers for GT and Targoman……304
Table 4.22 The percentage of Liao's error type GT and Targoamn……304
Table 4.23 The percentage of House's error type in GT and Targoamn…….305

LIST OF FIGURES

Page

Figure 2.1 Graphical User Interface of Google Translate …….…....…114

Figure 2.2 Graphical User Interface of Targoman …….…… 119

Figure 2.3 House Revisited Overt Covert Model……..…… 159

Figure 3.1 Theoretical Framework Figure …….208

Figure 3.2 House TQA Translation model ……...……. 192

Figure 3.3 Liao's Errors Classification ……...…….……. 192

Figure 3.4 The Data Collection Procedure in Brief …….….….218

Figure 3.5 The Process of Data Anlysis Based on Two Models…….225

LIST OF ABBREVIATIONS

APA	American Psychological Association
API	Application Programming Interface
BLEU	Bilingual Evaluation Understudy
CS	Conceptual structures
DQF	Dynamic Quality Framework
GT	Google Translate
GTM	Google Translate Machine
GTT	Google Translate Toolkit
HT	Human Translation
IOS	Iphone Operating System
LS	Lexico Semantics
TT	Target Text
TQA	Translation Quality Assessment

LIST OF APPENDICES

Appendix A The quatrains of Khayyam & The Translation (English version)

Appendix B The quatrains of Khayyam (Persian version)

PENILAIAN CIRI-CIRI LEKSIKO – SEMANTIK DALAM RUBAIYAT KHAYYAM: ANALISIS PERBANDINGAN TERJEMAHAN MANUSIA, GOOGLE TRANSLATE, DAN TARGOMAN

ABSTRAK

Kajian ini menjalankan analisis perbandingan dan deskriptif terhadap kesalahan terjemahan leksiko-semantik dalam Rubaiyat Khayyam (1120) menggunakan dua mesin terjemahan dalam talian, Targoman dan Google Translate. Terjemahan manusia oleh Fitzgerald (1859) berfungsi sebagai penanda aras. Kajian ini menggunakan pendekatan campuran, melibatkan analisis kualitatif dan kuantitatif. Objektif kajian adalah untuk menilai kualiti Rubaiyat Khayyam (1120) yang diterjemahkan oleh Google Translate dan Targoman menggunakan model overt/covert House (1997) dan klasifikasi jenis kesalahan Liao (2010), serta membandingkan kekerapan kesalahan leksiko-semantik dengan terjemahan manusia. Reka bentuk eksperimen melibatkan analisis perbandingan dan deskriptif. Penyelidikan deskriptif digunakan untuk menganalisis terjemahan Rubaiyat Khayyam (1120), mengenal pasti spesifikasi dan ciri-ciri data yang dipilih. Kaedah penyelidikan perbandingan digunakan untuk menilai dan menerangkan persamaan dan perbezaan dalam terjemahan oleh Google Translate, Targoman, dan Fitzgerald (1859). Analisis kualitatif melibatkan pemeriksaan terperinci terhadap kesalahan terjemahan tertentu, manakala analisis kuantitatif mengukur kekerapan dan jenis kesalahan yang dikenal pasti. Rangka kerja terjemahan terang/tersembunyi House (1997) digunakan untuk menilai ketepatan dan kesetiaan terjemahan. Kajian ini mengkaji kesalahan dalam terjemahan mesin bagi rubaiyat Parsi terpilih. Penemuan menunjukkan bahawa, berdasarkan rangka kerja House (1997), Google Translate menunjukkan prestasi lebih

baik daripada Targoman, menghasilkan terjemahan yang lebih tepat dengan lebih sedikit kesalahan. Bagi Targoman, kesalahan paling kerap adalah kesalahan rendisi, terutamanya dalam kolokasi (48 kejadian) dan kata kerja frasa (35 kejadian), dengan kesalahan bahasa terjemahan dalam peribahasa (8 kejadian) sebagai kesalahan yang paling kurang. Penyimpangan makna kecil (43.9%) adalah kesalahan yang paling kerap menurut model House (1997). Bagi Google Translate, kesalahan yang paling biasa adalah dalam kolokasi (32 kejadian) dan kata kerja frasa (23 kejadian). Mengikut model House (1997), penyimpangan makna kecil (48.1%) adalah kesalahan yang paling kerap, dengan kedua-dua tiada terjemahan dan penapisan budaya sebagai kesalahan yang paling kurang (3.7%). Google Translate menghasilkan terjemahan dengan kesetiaan yang lebih tinggi kepada teks asal berbanding Targoman. Kajian ini mencadangkan penyelidikan lanjut mengenai kualiti terjemahan mesin dan menggalakkan kaedah untuk mengurangkan kesalahan terjemahan.

EVALUATION OF LEXICO – SEMANTIC FEATURES OF KHAYYAM'S QUATRAINS: COMPARATIVE ANALYSIS OF HUMAN TRANSLATION, GOOGLE TRANSLATE, AND TARGOMAN

ABSTRACT

This research conducted a comparative and descriptive analysis of lexico-semantic translatability errors in Khayyam's Quatrains (1120) using two online translation machines, Targoman and Google Translate. Fitzgerald's (1859) human translation served as a benchmark. The study adopted a mixed-methods approach, employing both qualitative and quantitative analyses. The objectives were to evaluate the quality of Khayyam's quatrains (1120) translated by Google Translate and Targoman using House's (1997) overt/covert model and Liao's (2010) classification of error types and to compare the frequency of lexico-semantic errors with human translations.The experimental design involved a comparative and descriptive analysis. Descriptive research was used to analyze the translation of Khayyam's quatrains (1120), identifying specifications and features of the selected data. Comparative research methods were employed to evaluate and explain the similarities and differences in translations by Google Translate, Targoman, and Fitzgerald (1859). The qualitative analysis involved a detailed examination of specific translation errors, while the quantitative analysis measured the frequency and types of errors identified. House (1997) overt/covert translation framework was utilized to assess the accuracy and fidelity of the translations. The research examined errors in machine translations of selected Persian quatrains. Findings revealed that, based on House's (1997) frame work, Google Translate performed better than Targoman, producing more accurate translations with fewer errors. For Targoman, the most frequent errors were rendition

errors, particularly in collocations (48 occurrences) and phrasal verbs (35 occurrences), with the least frequent being translation language errors in proverbs (8 occurrences). Minor deviation in meaning (43.9%) was the most frequent error according to House's overt/covert model (1997). The most common errors for Google Translate were rendition in collocations (32 occurrences) and phrasal verbs (23 occurrences). According to House's model (1997), minor deviation in meaning (48.1%) was the most frequent error, with both no translation and cultural filtering being the least frequent errors (3.7%). Although Google Translate produced translations with higher fidelity to the original text compared to Targoman, this research suggests further research on machine translation quality and advocates for methods to reduce translation errors.

CHAPTER 1

INTRODUCTION

1.1 Overview

This chapter provides the background for the research by introducing the main parameters and variables of the study. It will then present detailed information about the study's aims and objectives, as well as the significance of conducting this research. Additionally, it will address the research problems from various perspectives, defining the issues, challenges, and gaps associated with online machine translation in rendering Persian classical poetry, specifically Khayyam, in detail. The research questions and possible hypotheses of the study will also be discussed. Finally, the key terms, along with the specific limitations and delimitations of the study, will be presented.

1.2 Background

Translation plays a significant role in bridging the differences among various cultures and nations. Literary translations, in particular, help these diverse nations find common ground and contribute to a universal culture. A literal translation can be described, according to linguistic definitions, as a translation made according to the standards of the target language (TL) at a level insufficient to convey the content in its original form (Barkhudarov, 1969). Utami and Hikmaharyanti (2019) referred to Catford (1965), who expressed a similar view, explaining that the unit of translation must conform to TL grammar from the outset. Translating literary works is one of the most important subfields of translation research, where conceptual practices have sparked debates

among researchers. Traditionally, literary translation studies have focused on source-target text relations and conceptual issues, emphasizing two closely linked topics: equivalence and communicative purpose. Additionally, concerns about translating style have also been addressed (Parks & Boase-Beier, 2019).

The translation of literary works, particularly poetry, presents a unique set of challenges due to the intricate interplay of form, meaning, and cultural nuances. Omar Khayyam’s quatrains (1120), renowned for their philosophical depth and linguistic beauty, epitomize this complexity. The comparative evaluation of lexico-semantic features in translations of Khayyam’s quatrains (1120) is crucial for understanding how different translation approaches affect the fidelity and interpretive quality of the texts. This research critically examines the comparative performance of Google Translate, Targoman, and human translators in capturing these features (Karimi-Hakkak, 2019).

Sofo (2024) refers to previous research that has extensively explored the nuances of literary translation, emphasizing the challenges posed by poetry. Venuti (2017) highlights the translator's invisibility and the need to balance fidelity and fluency, which is particularly relevant for poetic texts where literal translation often fails to capture the original's aesthetic and emotive impact. Similarly, Jakobson's (1959) framework of intralingual, interlingual, and intersemiotic translation underscores the transformative process involved in translating poetry, where semantic and stylistic elements must be carefully negotiated.

Studies focusing on the translation of Khayyam’s quatrains (1120) have underscored the difficulties in preserving the original’s philosophical undertones and rhythmic

structure. FitzGerald's (1859) translation is often cited for its liberal approach, prioritizing poetic form over literal accuracy, which has sparked considerable debate about the trade-offs in poetic translation. More recent analyses, such as those by Karimi-Hakkak (2019), argue for a balance between form and content, suggesting that modern translation technologies need to address these dual aspects to be effective.

Li et al. (2024) refer to Wu et al. (2016) in their research, noting that the advent of machine translation (MT) has revolutionized the field of translation studies. Google Translate, powered by neural machine translation (NMT) models, represents a significant advancement in MT, capable of producing fluent translations that closely approximate human output. However, research by Toral and Way (2018) indicates that while NMT systems excel in general text translation, they struggle with literary texts due to the subtle semantic and stylistic nuances involved.

Targoman, an MT system specifically tailored for Persian-English translation, claims to offer better contextual understanding and handling of cultural references (Alizadeh et al., 2020). Despite these advancements, a critical gap remains in evaluating these systems against human translators, especially concerning the translation of complex literary texts like Khayyam's quatrains (1120).

Comparative studies provide valuable insights into the efficacy of different translation approaches. For instance, Munday (2016) emphasizes the importance of comparative analysis in identifying strengths and weaknesses across various translation methodologies. Comparative evaluations of Google Translate and other MT systems often reveal limitations in handling idiomatic expressions and cultural subtleties, which are crucial in literary translation (de Souza, 2024).

The present research aims to bridge this gap by systematically comparing the lexico-semantic features of Khayyam's quatrains (1120) across translations produced by Google Translate, Targoman, and human translators. This comparative analysis will focus on fidelity to the original text, preservation of poetic elements, and representation of cultural context.

Building on the foundation laid by previous studies, this research addresses the critical need for a nuanced comparative evaluation of MT systems in literary translation. By focusing on Khayyam's quatrains (1120), the study contributes to the broader field of translation studies and provides specific insights into the capabilities and limitations of contemporary translation technologies. The findings are expected to inform future developments in MT systems, particularly in enhancing their application to literary texts.

1.2.1 Lexico Semantic Features

The study of lexico-semantic features in translation involves analyzing how words and their meanings are structured and conveyed across different languages. In the context of translating Khayyam's quatrains (1120), these features are crucial for preserving the poetic and cultural essence of the original text. The increasing use of machine translation systems, such as Google Translate and Targoman, necessitates a closer examination of how these systems handle complex linguistic features compared to human translators. Recent studies highlight the challenges and advancements in translating collocations, idioms, literary figures, phrasal verbs, and proverbs (Munday, 2020).

One of the significant concepts in semantic relations is the lexical field, which represents a set of lexical items that share a common intangible association, either according to a field of knowledge or habitual co-occurrence in real contexts. Current notions of word formation derive from Trier (1931), who built on Saussure's (2011) ideas, arguing that the meaning of a word cannot be understood in isolation from its relationship to other meaningfully connected words in the same lexicon. Decades of research have confirmed that words within the same semantic field have strong semantic relationships. This can be illustrated when two different words share the same phonology or spelling. For example, the word "post," meaning a long wooden or metal stick, is related to "tent," while another meaning of "post" refers to either the north or south poles of the Earth's rotation (or another celestial body), suggesting a connection to the word "north." These two meanings of "post" (identical phonology, different semantics) are in separate lexical domains (Stringer, 2019).

While this general associative concept is valuable, more specific word formations are often used to describe the relationships between particular words, such as polysemy, homonymy, synonymy, hyponymy, antonymy, and meronymy. These relationships are explored in depth. Homonyms are words with different meanings that are either phonetically similar (homophones) or spelled similarly (homographs). For instance, the homophonic occurrences of "right" /raɪt/ illustrate relationships between simplified phonological structures (PS), syntactic structures (SS), and conceptual structures (CS) (Stringer, 2019).

Lexical units play a significant role in the creativity of a writer's language. They are crucial for clearly conveying intended meanings and expressing opinions and emotions in ways that resonate with readers. These elements contribute to the appeal of the

language. The lexical level includes a wide range of items, such as synonyms and antonyms, collocations, adverbs, phrasal verbs, registers, adjectives, and prepositions (Stringer, 2019).

Smith (2020) refers to Geeraerts (2013), who explored how semantics is used to study meaning in language. Specifically, semantics aims to clarify how speakers and listeners encode and decode literal meanings. Other aspects of meaning are examined in relation to semiotics and context, which involve the study of how both phonetic and non-linguistic symbols are interpreted. For example, when someone gives another person a ring, saying "Here is a ring," several layers of meaning can be identified. On the surface level, the deictic pronoun "here" indicates a proximate location, the verb denotes presence, the determiner suggests that both the giver and the receiver are aware of the ring, and the word "ring" represents a specific object (Smith, 2020).

At the pragmatic level, depending on the context, the listener might infer that this gesture signifies a marriage proposal, a divorce petition, or a request for a magical quest. From a semiotic perspective, the ring itself can be understood as a symbol of bond, union, or promise, potentially representing marriage, loyalty to an institution, or religious commitment. Semantics is narrower compared to pragmatics and semiotics because it focuses specifically on the linguistic aspects of meaning. Different theoretical approaches to semantics exist, including formal semantics, which applies propositional logic to understand the relationships between linguistic terms and their referents, and cognitive semantics, which views meaning as arising from shared cognitive processes (Geeraerts et al., 2024).

One significant concept in semantic relations is the lexical field, which represents a set of lexical items that share an intangible association or are related by

knowledge or common co-occurrence in real contexts. Current theories of word formation trace back to Trier (1931), who expanded on Saussure's (2011) ideas, arguing that a word's meaning can not be understood in isolation but must be considered in relation to other meaningfully connected words within the same lexicon. Decades of research have confirmed that words within the same semantic field have strong semantic relationships. This is evident when two words share the same phonology or spelling but differ in meaning. For example, the word "post," meaning a long wooden or metal stick, is related to "tent," while another meaning of "post" refers to either the north or south pole of the Earth (or another celestial body), suggesting a connection to "north." These two meanings of "post" (identical phonology, different semantics) belong to distinct lexical domains (Stringer, 2019).

Liu et al. (2024) highlight in their paper that various studies have explored lexico-semantic patterns in translation. For instance, Ahmadi and Hashemi (2018) examined lexico-semantic patterns in English-Persian translation, while Huang (2016) investigated lexico-semantic features in Chinese-English translation. Jiang and Nida (2012) conducted a lexico-semantic analysis of translation in a case study involving Chinese and English. Additionally, Mossop (2013) and Zhang (2013) focused on lexico-semantic issues in translation, particularly emphasizing culture-specific items.

1.3 Overview of Analysis Parameters and Variables

This research evaluates the lexico-semantic translatability errors in Khayyam's Quatrains (1120) using different translation methods. The key analysis parameters include House's (1997) overt/covert translation quality assessment framework and

Liao's (2010) classification of error types. The main variables analyzed are the types and frequencies of errors in translations produced by Google Translate, Targoman, and a human translator. Detailed definitions and descriptions of these parameters and variables are provided in Chapter 3 (Methodology).

1.3.1 Analysis Parameters and variables

The analysis parameters for this study are structured around two primary models. First, House's (1997) overt/covert translation model, which is instrumental in assessing the accuracy and fidelity of translations. Second, the Error Classification System follows Liao's (2010) methodology, which categorizes and quantifies lexico-semantic errors within the translations.

In terms of variables, the study contrasts different translation methods, including Human Translation, represented by Fitzgerald's (1859) translation of Khayyam's Quatrains, and Machine Translations, specifically Google Translate and Targoman. The study identifies several error types: Rendition Errors, such as incorrect word choice, improper collocations, and inaccurate phrasal verbs; Translation Language Errors, which pertain to errors in proverbs, idioms, and literary figures; Miscellaneous Errors, involving slight deviations from the original meaning; and Overt/Covert Errors, as classified by House (1997). The study also tracks the frequency of errors by counting the occurrences of each error type across the translations.

Additionally, the research focuses on specific lexico-semantic features such as Collocations, which refer to frequently occurring word combinations (Mulina & Hurtado, 2020); Phrasal Verbs, where verbs combined with prepositions or adverbs generate new meanings (Chen & Li, 2020); Idioms, which are phrases with non-literal

meanings (Ebrahimkhani, 2023); Proverbs, defined as short, generally known sentences that convey wisdom, truth, morals, and traditional views in a metaphorical, fixed, and memorable form (Tang, 2019); and Literary Figures, including metaphors, similes, and rhetorical devices (Farghal & Shunnug, 2018).

1.4 Statement of the problem

The significance of poetry and Persian literature, especially classical poetry, in the Western world and its audience cannot be overstated. The works of Khayyam (1120), particularly his quatrains, occupy a notably important place within this context. For Khayyam's messages and philosophy to be effectively understood by his audience, the translation of these poems is crucial. Nations prioritize the recognition of their exceptional literary figures and strive to disseminate their works globally. In the Iranian context, Khayyam is widely recognized as a poet of great prominence and international renown. Therefore, it is essential for scholars to conduct a thorough examination of how Khayyam's poetry is rendered by machine translation algorithms available on the internet (Kamali et al., 2019).

Despite the limited number of studies on this topic (Bourdieu, 1993; King, 2021), a comprehensive comparison of virtual translations of Khayyam's poetry remains relatively rare in the literature. Thus, investigating the translation of poetry is imperative. In the field of Artificial Intelligence (AI), computer-based machine translation presents a significant challenge. This challenge is due to the substantial differences between translations produced by human translators and those generated by translation machines (King, 2021). As conceptualized by Jakobson, translation intricately intertwines meaning and form. Meaning, or content, is one aspect, while form, or style, contributes to shaping the meaning of the text. Although meaning may

be transferable between languages, form often presents challenges in translation, particularly in verse. This difficulty arises from the diverse linguistic characteristics among languages, leading to the potential untranslatability of poetry, as well as songs, advertisements, puns, and similar forms of communication. Phonetic devices and double meanings are often challenging, if not impossible, to replicate accurately in the target language (Hatim & Munday, 2019).

The translation of poetry presents unique challenges compared to the translation of ordinary language and prose. Poetry stands as an independent case within literary translation due to the indivisible interrelation of form and content. The form of poetry cannot be recreated in the target language (TL) in the same way as in the source language (SL). The difficulties in translating poetry stem from its deviant nature, which is intensely connotative rather than denotative. Poetry is often condensed and encompasses specific characteristics such as musical mode, formal meter, and rhyming patterns, which are intrinsic to its form. Translators must not only reflect on these characteristics but also reproduce a text that functions as poetry in the TL. As a result, some argue that poetry cannot be fully translated. Nabokov, who strongly believed in the impossibility of translating poetry, advocated for translations accompanied by extensive glosses and footnotes and argued that poetry should be restricted to literal rendering (Hatim & Munday, 2019). Moreover, evaluating poetry in translation requires consideration of not only meaning but also prosody and accuracy. This issue is further complicated when the original poetry, such as classical Persian poems, is unique and contains a variety of stylistic, rhythmic, and linguistic features (Kadkani, 2010; Hatim & Munday, 2019).

This research occupies a unique and pivotal position within the scholarly landscape, addressing a critical gap in the existing literature on the translation of poetic works, particularly within the realm of machine translation (MT) systems. While earlier studies have explored various aspects of translation quality assessment, such as similes, metaphors, and stylistic nuances (Dunder et al., 2020; Hatim & Mason, 1990), a notable gap remains in the comprehensive comparative evaluation of lexico-semantic features—such as collocations, phrasal verbs, literary figures, idioms, and proverbs—in poetic translations. Previous studies have often focused on isolated elements, overlooking the holistic analysis necessary to address the myriad challenges and complexities inherent in translating literary texts. Additionally, existing research has predominantly concentrated on single MT platforms or specific translation errors, failing to provide a comprehensive comparative analysis that encompasses multiple MT systems and a diverse range of translation errors.

This research aims to address the existing gap by conducting a thorough examination of lexico-semantic features in translations of Khayyam's quatrains (1120). It employs both Persian-based (Targoman) and English-based (Google Translate) machine translation (MT) systems, comparing these against human translation as a benchmark. This comparative approach provides a nuanced understanding of the strengths and limitations of each translation method, highlighting the intricate interplay between linguistic and cultural nuances inherent in the translation process. Additionally, the study integrates House's (1997) overt/covert model and Liao's (2010) taxonomy of lexico-semantic errors, offering a comprehensive framework for scrutinizing and categorizing specific errors within the translated texts. By leveraging these models, the research not only addresses the shortcomings of previous studies but

also provides a novel and holistic perspective on the challenges and opportunities associated with translating poetic works using contemporary computational tools.

1.5 Objective of the Thesis

1.5.1 Objective 1

To identify and categorize the types and frequencies of lexico-semantic errors in the translations of Khayyam's Quatrains (1120) produced by Google Translate and Targoman. This will be done using House's (1997) overt/covert model and Liao's (2010) error classification framework. This objective aligns with RQ1 by focusing on the specific error types and their frequencies across different translation systems.

1.5.2 Objective 2

To compare the lexico-semantic errors found in Targoman's translations of Khayyam's Quatrains (1120) with those found in the human translation by Fitzgerald (1859). This comparison will be based on Liao's (2010) error classification and House's (1997) overt/covert model. This objective directly addresses RQ2 by examining how machine translation errors in Targoman differ from those in human translation.

1.5.3 Objective 3

To compare the lexico-semantic errors in Google Translate's translations of Khayyam's Quatrains (1120) with those in Fitzgerald's human translation. This

comparison will use Liao's (2010) classification and House's (1997) model. This objective relates to RQ3 by contrasting the performance of Google Translate with human translation.

1.6 1.6 Research Questions

In the present thesis, three inquiries are posited concerning the outcomes of the comparative analysis. The responses to these questions serve as crucial inputs in elucidating the treatment of lexico-semantic characteristics in the rendering process, thereby making a valuable contribution to the domain of research.

1.6.1 Research Question 1

What are the most and the least frequent errors in lexico-semantics features that are collocation, phrasal verb, literary figure, idiom, and proverb in the *Google Translate* and *Targoman* translation of Khayyam's Quatrains (1120) based on the House overt/covert model (1997) and Liao's classification of errors (2010)?

1.6.2 Research Question 2

How does the Targoman translation machine differ from human translation in terms of lexico-semantic errors, as defined by Liao's model (2010) and House overt/covert model (1997)?

1.6.3 Research Question 3

How does Google Translate machine differ from human translation in terms of lexico-semantic errors, as defined by Liao's model (2010) and House overt/covert model (1997)?

1.7 1.7 Significance of the Study

The significance of this research lies in its potential to advance the understanding and application of machine translation (MT) in the literary domain, specifically for translating Persian poetry into English. One key factor underscoring the importance of this study is the underexplored language pair. Despite extensive research on MT, there is a noticeable lack of studies focusing on Persian (Farsi) to English translations. This research addresses this gap by providing a detailed analysis of MT performance in translating Khayyam's Persian quatrains (Alizadeh & Habibi, 2020).

Another crucial aspect of this research is its focus on quality comparison. The study evaluates the translation quality of two MT systems—Google Translate and Targoman—against a human translation benchmark provided by Fitzgerald (1859). This comparative analysis is essential for discerning the strengths and weaknesses of each MT system in capturing the lexico-semantic nuances inherent in poetic texts (Toral & Way, 2020).

Furthermore, the research emphasizes error analysis by identifying and categorizing lexico-semantic errors in the translations. This provides valuable insights into the specific challenges that MT systems face in literary translation. The findings aim to inform the development of more accurate and contextually aware MT models, contributing to advancements in the field (Castilho et al., 2018).

The study also holds practical implications for developers and engineers working on MT technologies. The results highlight the need for IT system developers to be aware of client requirements and the gaps in current MT capabilities. The insights

gained from this research can drive improvements in MT software, enhancing its ability to translate complex poetic texts more effectively (Bentivogli et al., 2020).

Moreover, this research contributes to advancing MT research by serving as a foundation for future studies, particularly those evaluating MT systems' capacity to handle intricate lexico-semantic elements in poetry. It underscores the importance of semantic features in translation studies and the need for ongoing advancements in MT technology (Popović, 2019).

The significance of this research extends to several key beneficiaries. For Translation Studies scholars and researchers, it provides an in-depth comparative analysis of human and machine translation methods, enriching the academic discourse on the effectiveness and limitations of current translation technologies. These findings allow scholars to explore further the dynamics between human creativity and machine efficiency in translation. Professional translators can gain valuable insights into the strengths and weaknesses of machine translation tools like Google Translate and Targoman. This knowledge helps them make informed decisions about when to rely on these tools and how to enhance their translations by understanding the nuances captured by human translators.

Developers and engineers working on machine translation technologies will benefit from the detailed analysis of lexico-semantic features captured by their algorithms. This feedback is essential for improving the accuracy and cultural relevance of machine-generated translations. Educators and students in translation and linguistics can utilize the findings to better understand the comparative strengths of human and machine translations, using real examples from Khayyam’s quatrains. This aids in appreciating the complexities involved in translation and highlights the

importance of cultural and contextual knowledge in achieving high-quality translations.

Additionally, scholars interested in Persian literature and the works of Omar Khayyam will find value in understanding how different translation methods impact the interpretation and reception of literary texts. This comparative analysis emphasizes the significance of preserving the poetic and philosophical essence of Khayyam's quatrains across languages. Lastly, readers and enthusiasts of Khayyam's poetry can gain a deeper appreciation of how translations can vary significantly depending on the method used. This awareness leads to a more informed and enriched reading experience, recognizing the subtleties and beauty of the original work as conveyed through different translation lenses. By addressing the needs and interests of these diverse groups, this research aims to foster a more comprehensive understanding of translation practices and encourage ongoing discourse in the field.

1.8 1.8 Limitations of the Study

The limitations of this research include two main factors. First, the findings are specific to the Persian-English language pair, which limits their generalizability to other language pairs. Each language has distinct syntactic and semantic features that influence translation outputs, and these features can vary significantly even when using the same machine translation system (Ding et al., 2021). Consequently, the insights gained from this study may not be applicable to translations involving different language pairs.

Second, technical constraints of the Targoman translation machine also posed a significant limitation. The machine's restriction to translating no more than 200

words at a time at its highest quality necessitated that the researcher input the text in smaller units—specifically, one quatrain at a time. This segmentation could potentially impact the coherence and overall quality of the translations (Wu et al., 2019). These constraints should be considered when interpreting the results of this research.

1.9 1.9 Definition of Key Terms

1.9.1 Literary Translation

According to Venuti (2018), literary translation involves the intricate process of transferring literary texts from one language to another while retaining their artistic and cultural nuances. Venuti highlights the translator's role in mediating between different linguistic and cultural systems, emphasizing the challenges of preserving the foreignness of the original text in the translation. He defines literary translation as a transformative act that entails not only linguistic transfer but also cultural and ideological negotiation, underscoring the translator's creative agency. Similarly, Baker (2021) describes literary translation as a dynamic process involving linguistic, cultural, and socio-political factors.

In the context of this research, literary translation specifically refers to translating Khayyam’s Quatrains (1120) from Persian to English. The study examines how effectively both human and machine translators preserve the poetic and cultural essence of these quatrains. It focuses on lexico-semantic features such as collocations, phrasal verbs, literary figures, idioms, and proverbs, which are crucial to maintaining their artistic integrity.

1.9.2 Human Translation

Wu et al. (2019) describe human translation as a complex process involving mediation between different linguistic and cultural systems, underscoring the translator's creative agency. Bassnett (2020) defines it as a dynamic and creative act that goes beyond mere linguistic transfer, emphasizing the role of the translator in navigating cultural and socio-political factors. Baker (2021) views human translation as a transformative process involving cultural negotiation.

In this research, human translation specifically refers to the translation of Khayyam’s Quatrains (1120) by an experienced human translator, notably FitzGerald’s 1859 translation. This human translation serves as a benchmark for comparing the quality of translations produced by machine translation systems (Kamali & Mortazavi, 2020).

1.9.3 Machine Translation

Way (2020) defines machine translation as the automated process of translating text from one language to another using computer algorithms. Cho (2019) explores recent advancements in neural machine translation models, highlighting their impact on translation quality. Special (2021) describes machine translation as a rapidly evolving field that leverages computational methods to enhance translation efficiency and accuracy.

In this research, machine translation specifically refers to the automated translations of Khayyam’s Quatrains (1120) produced by Google Translate and Targoman. The study comparatively evaluates the lexico-semantic accuracy of these machine-generated translations against the human benchmark provided by Fitzgerald’s translation.

1.9.4 Quatrains

Lee (2020) defines quatrains as poetic forms consisting of four lines, typically with a specific rhyme scheme and meter. Morley (2019) highlights their role in various literary traditions. Weitz (2021) describes quatrains as a versatile poetic form.

In the context of this research, quatrains refer to the specific four-line verses from Khayyam's collection (1120) that are being translated and analyzed for their lexico-semantic accuracy.

1.9.5 Lexico-Semantic Features

Rojo (2020) defines lexico-semantic features as the lexical and semantic aspects of language that contribute to meaning construction. Sullivan (2019) explores their role in language processing, highlighting how these features influence comprehension and interpretation. Włodarczak (2021) examines their importance in translation studies, emphasizing how accurate representation of these features is crucial for maintaining the integrity of the translated text.

In this research, lexico-semantic features refer to the lexical and semantic elements of Khayyam's Quatrains (1120) that are essential for accurate translation. The study focuses on identifying and analyzing errors in these features within machine translations, comparing them with human translations to assess how well each approach preserves the original meaning and nuances of the text.

1.9.6 Google Translate

Google Translate employs cutting-edge neural machine translation (NMT) technology to facilitate translations across various languages. Cho et al. (2020) and

Zhang et al. (2021) highlight the advancements in neural machine translation models, noting their ability to produce fluent and contextually appropriate translations by leveraging deep learning techniques.

In this research, Google Translate is one of the machine translation systems evaluated for its capability to accurately translate the lexico-semantic features of Khayyam's Quatrains (1120) from Persian to English. The study assesses how well Google Translate captures the intricate lexical and semantic elements of the quatrains compared to human translations, focusing on its effectiveness in preserving the poetic and cultural essence of the original text.

1.9.7 Targoman

Targoman is a machine translation system similar to Google Translate. Studies by Rahimi et al., (2021) and Sadeghi et al., (2020) explore its performance and limitations.

In the context of this research, Targoman is the domestic machine translation system evaluated alongside Google Translate. The research assesses its effectiveness in translating the lexico-semantic features of Khayyam's Quartrains (1120).

1.9.8 Errors

Translation errors encompass linguistic, cultural, and contextual inaccuracies. Gao et al., (2020) and Zhang et al., (2021) focus on identifying and categorizing different types of errors. Liu et al., (2019) and Chen et al., (2021) explore using machine learning techniques to mitigate translation errors.

In this research, *errors* refer to inaccuracies found in the translations of Khayyam's Quatrains (1120) by machine translation systems. The research classifies these errors according to Liao's (2010) model and House's overt/covert (1997) model, focusing on their impact on the accuracy of lexico-semantic features.

1.10 Hypotheses of the Study

Hypothesis 1: There will be significant differences in the frequency and types of lexico-semantic errors between the translations produced by Google Translate and Targoman. Specifically, Google Translate will demonstrate a different error pattern compared to Targoman, which can be analyzed using House's (1997) overt/covert model and Liao's (2010) error classification. This hypothesis addresses RQ1 by hypothesizing about the differences in error patterns between the two machine translation systems.

Hypothesis 2: The human translation of Khayyam's Quatrains (1120) by Fitzgerald (1859) will exhibit a lower frequency of lexico-semantic errors compared to the translations produced by both Google Translate and Targoman. This hypothesis aims to compare the accuracy of human translation with that of machine translations, corresponding to RQ2 and RQ3.

Hypothesis 3: Both Google Translate and Targoman will exhibit higher rates of errors in translating idiomatic expressions and cultural references compared to other lexico-semantic elements, such as collocations, phrasal verbs, and literary figures. This hypothesis explores the difficulties machine translation systems face with idiomatic and culturally specific content, aligning with RQ1.

Hypothesis 4: Notable differences will be observed in the frequency and types of lexico-semantic errors between Google Translate and Targoman, reflecting their distinct algorithms and training data. This hypothesis evaluates the comparative performance of the two machine translation systems, corresponding to RQ1.

Hypothesis 5: More complex and metaphorical quatrains in Khayyam's work will show a higher error rate across all translation methods compared to simpler quatrains. This hypothesis investigates the impact of quatrain complexity on translation accuracy, addressing the broader implications of error types and frequencies as outlined by House's (1997) model and Liao's (2010) classification. This hypothesis indirectly relates to all research questions by exploring the influence of text complexity on translation errors.

These hypotheses will be tested through a comparative analysis of the translations, focusing on the identification and categorization of lexico-semantic errors as outlined by Liao's (2010) model and House's (1997) model. The outcomes of this analysis will provide insights into the capabilities and limitations of both machine and human translation methods for literary texts.

1.11 Organization of the Study

This thesis is organized into seven main sections: five chapters, a reference section, and appendices.

Chapter One, titled *Introduction*, introduces the study by discussing the primary issues addressed by the research and the significance of the study. It outlines the research objectives, questions, and hypotheses and provides definitions of key terms. Additionally, the limitations and delimitations of the study are discussed.

Chapter Two, *Literature Review*, presents a comprehensive review of the relevant theories and literature on key terms pertinent to the study. It includes a detailed analysis of recent case studies, providing context for the research and highlighting the gaps that the current study aims to address.

Chapter Three, *Methodology*, is dedicated to the methods employed in this research. It describes the theoretical framework and research design, detailing the procedures, tools, sampling methods, and corpus. This chapter also explains the data collection and data analysis methods used in the study.

Chapter Four, *Findings and Discussion*, analyzes and discusses the data obtained from the research. It aims to answer the research questions and evaluate the extent to which the results support the hypotheses. The chapter provides a thorough discussion of the findings in relation to the theoretical framework and literature reviewed in Chapter Two, and it also discusses the implications of the study's findings.

Chapter Five, *Conclusion*, summarizes the main findings of the study, drawing conclusions based on the analysis presented in Chapter Four. It discusses the implications of the findings and offers suggestions for further research. Practical recommendations based on the study's results are also provided.

The *References* section lists all sources cited throughout the thesis, following the appropriate academic citation style (American Psychological Association - APA).

The *Appendices* include additional material that supports the main text of the thesis. Appendix A presents Khayyam’s quatrains from Kazzazi (1994), providing the original Persian quatrains used for analysis. Appendix B includes Edward Fitzgerald's translations (1859) of Khayyam’s quatrains, which serve as a benchmark for evaluating the accuracy and fidelity of machine translations in this research.

1.12 Summary

Chapter one of this thesis serves as the introduction, setting the stage for the study by providing a comprehensive framework. It includes a background of the study, bridging it to the current research, and contextualizes the research problem or topic within existing literature, explaining its relevance and importance. Additionally, the section on Lexico-Semantic Features discusses specific linguistic and semantic aspects relevant to the study.

Furthermore, the Overview of Analysis Parameters and Variables describes the variables, factors, or criteria that will be examined during the study. The Objective of the Study clearly states the primary purpose and aim of the research. In contrast, the Purpose of the Study expounds on the broader intention and rationale behind conducting the research. The Aim of the Study specifies the specific goals the research intends to achieve.

The Research Questions presented in this chapter outline the central questions that the research seeks to answer, guiding its focus. The Significance of the Study articulates the potential contributions, implications, and applications of the research findings. The Limitations of the Research acknowledge the constraints, boundaries, and restrictions that may affect the study's scope or conclusions. The Definitions of