



Hacettepe University Graduate School of Social Sciences  
Department of Translation and Interpretation

**CONSISTENCY IN THE EVALUATION METHODS OF MACHINE  
TRANSLATION QUALITY**

Özden ŞAHİN

Master's Thesis

Ankara, 2015



# CONSISTENCY IN THE EVALUATION METHODS OF MACHINE TRANSLATION QUALITY

Özden ŞAHİN

Hacettepe University Graduate School of Social Sciences  
Department of Translation and Interpretation

Master's Thesis

Ankara, 2015

## KABUL VE ONAY

Özden ŞAHİN tarafından hazırlanan "Consistency in the Evaluation Methods of Machine Translation Quality" başlıklı bu çalışma, 3 Haziran 2015 tarihinde yapılan savunma sınavı sonucunda başarılı bulunarak jürimiz tarafından yüksek lisans tezi olarak kabul edilmiştir.



Prof. Dr. Ayfer ALTAY (Başkan)



Prof. Dr. Aymil DOĞAN (Danışman)



Yrd. Doç. Dr. Elif ERSÖZLÜ



Yrd. Doç. Dr. Zeynep DOYURAN



Yrd. Doç. Dr. İsmail ERTON

Yukarıdaki imzaların adı geçen öğretim üyelerine ait olduğunu onaylım.

Prof. Dr. Yusuf ÇELİK

Enstitü Müdürü

## BİLDİRİM

Hazırladığım tezin/raporun tamamen kendi çalışmam olduğunu ve her alıntıya kaynak gösterdiğimi taahhüt eder, tezimin/raporumun kağıt ve elektronik kopyalarının Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü arşivlerinde aşağıda belirttiğim koşullarda saklanmasına izin verdiğimi onaylarım:

- Tezimin/Raporumun tamamı her yerden erişime açılabilir.
- Tezim/Raporum sadece Hacettepe Üniversitesi yerleşkelerinden erişime açılabilir.
- Tezimin/Raporumun ..... yıl süreyle erişime açılmasını istemiyorum. Bu sürenin sonunda uzatma için başvuruda bulunmadığım takdirde, tezimin/raporumun tamamı her yerden erişime açılabilir.

3 Haziran 2015



Özden ŞAHİN

## ACKNOWLEDGMENTS

Machine translation research contains lots of howlers and misconceptions, but the classics of all time, and my personal favorite is as follows: Once an MT program translated “The spirit is willing, but the flesh is weak” into Russian and the back translation was “The vodka is good, but the meat is rotten”.

I have been studying on this thesis for the last two years, and there were lots of times when neither my spirit nor my flesh was willing, but thanks to some special people, this research is finally put into these pages. This page is dedicated to thank them.

First and foremost, I would like to express my gratitude to my thesis advisor, Prof. Dr. Aymil Doğan, with whom studying was an amazing chance for me. Her advisor role was never limited to this thesis and her life energy has always inspired me.

I would like to express my gratitude to Prof. Dr. Ayfer Altay for her support. Working as a research assistant in Hacettepe University was a great experience, and I am grateful to her for this working environment.

I would like to express my gratitude to Prof. Dr. Asalet Erten, for her care and support all through these last three years. I would never forget her kindness, and I am deeply grateful for her encouragement and helping hand.

I would like to express my gratitude to my thesis committee members Assist. Prof. Dr. Elif Ersözlü, Assist. Prof. Dr. Zeynep Doyuran and Assist. Prof. Dr. İsmail Erton for their invaluable feedbacks and contributions.

I would like to thank my dear friends Aslı Polat and Nazan İşi, for turning Ankara and my office into an enjoyable place, and for taking care of me whenever I needed. I would never be able to put my feelings into words to describe how I am happy to have met them and how I am grateful to them. I would also like to thank my friend and office mate Didem Kaya, I am really happy to have met her; she has enriched my Hacettepe experience.

I would like to thank 30 students who have participated to human evaluation part of this study. I am deeply grateful for their cooperation and meticulous work.

I would like to express my gratitude to TUBITAK for granting me a scholarship.

I would like to thank my family for their support all through these years. Without their constant encouragement, I would have never come this far.

Thanks to all these people, my spirit has remained and will always remain willing, regardless of my flesh.

## ÖZET

ŞAHİN, Özden. *Bilgisayar Çevirisi Kalitesinin Değerlendirmesi Yöntemlerinde Tutarlılık*. Yüksek Lisans Tezi, Ankara, 2015.

Bilgisayar çevirisi, insan yardımına çeşitli derecelerde başvurarak otomatik çeviri yapan programlar için kullanılan genel terimdir. Bilgisayar bilimi, dilbilim ve çeviribilimi bir araya getiren bu disiplinler arası alanda henüz yanıtlanmamış pek çok soru vardır. Bu sorulardan biri de bilgisayar çevirisi kalitesinin değerlendirmesidir. Buna ek olarak, çeviri teknolojileri söz konusu olduğunda genellikle BÇA'ya yönelen çeviribilim literatüründe bilgisayar çevirisi konusunda bir boşluk yer almaktadır. Bu çalışma İngilizce ve Türkçe arasında çeviri yapan Google Translate, Proçeviri ve Sametran bilgisayar çevirisi programlarının kalitelerini değerlendirerek literatürdeki bu boşluğu doldurmayı hedeflemektedir.

Bu çalışma İngilizce'den Türkçe'ye bilgisayar çevirisinin kalitesinin değerlendirmesinde kullanılan iki farklı yöntem olan mikro değerlendirme ve makro değerlendirme arasındaki tutarlılığı araştırmayı amaçlamaktadır. Buna ek olarak, bilgisayar çevirisinin farklı metin türlerinde gösterdiği farklılıklar ve hataların insanların bilgisayar çevirisi kalitesi konusundaki algısı üzerindeki etkisi de araştırılmıştır. Bu sorulara dört farklı metin türü için örnek metinleri çeviren üç farklı bilgisayar çevirisi programlarının çıktıları üzerinde bir inceleme yapılarak cevaplar aranmıştır.

Bu çalışmada betimleme yöntemi kullanılmıştır. Yapılan karşılaştırmalı ve karşıtsal inceleme iki basamaklıdır. Önce, bilgisayar çevirisi programlarının çıktıları için Flanagan tarafından tanımlanan hata sınıflandırması çerçevesinde hata incelemesi yapılmıştır. Daha sonra 20 tercümanlık öğrencisinin katılımıyla bir insan değerlendirmesi gerçekleştirilmiştir. Yorumcular çıktıları anlaşılabilirlik, sadakat ve genel kaliteye göre incelemiştir.

Gerçekleştirilen incelemelerden elde edilen sonuç, bilgisayar çevirisini tutarlı bir şekilde değerlendirmenin mümkün olduğudur; hata sayıları ve yorumcuların derecelendirme ve sıralamaları arasında benzerlikler vardır. Buna ek olarak, anlaşılabilirlik ve sadakat dereceleri ile genel kalite sıralamaları arasında da benzerlikler vardır. İncelemeden elde



edilen en önemli sonuçlardan biri de bilgisayar çevirisi programlarının kalitesinin temel olarak cümlenin uzunluğuna bağlı olduğudur.

Çalışmanın hem insan hem de bilgisayar çevirisi araştırmasına, tercümanlık öğrencilerinin katılımıyla gerçekleşen kapsamlı bir değerlendirme sunarak İngilizceden Türkçeye çalışan bilgisayar çevirisi programlarının çıktıları üzerinde bir hata sınıflandırması sunarak katkıda bulunması beklenmektedir.

### **Anahtar Kelimeler**

Bilgisayar çevirisi, bilgisayar çevirisi değerlendirmesi, çeviri teknolojileri, insan değerlendirme, hata sınıflandırması, anlaşılabilirlik, sadakat.

## ABSTRACT

ŞAHİN, Özden. *Consistency in the Evaluation Methods of Machine Translation Quality*. Master's Thesis, Ankara, 2015.

Machine translation is the general term for the programs concerning the automatic translation with or without human assistance. It is also an interdisciplinary research area with different questions yet to be answered. One of the fundamental questions of the area is related to the quality assessment of machine translation outputs. In addition, despite to its interdisciplinary nature, machine translation is rarely the research topic for Translation Studies, which focuses more on CAT tools. This study intends to fill this gap in the literature by focusing the evaluation of quality of MT programs available between English and Turkish, namely Google Translate, Proçeviri and Sametran.

The study aims at exploring the consistency among two evaluation methods of machine translation quality from English into Turkish: micro evaluation and macro evaluation. In addition, the differences between MT quality for different text types and the impact of errors on human perception of MT quality are also sought. These questions are answered by conducting an analysis of outputs of three machine translation programs translating samples for four different text types.

Descriptive method is adopted in the study. The comparative and contrastive analysis conducted is two-fold. Firstly, an error analysis is carried out on the outputs of machine translation programs within the framework of error categorization defined by Flanagan. Then, a human evaluation is conducted with the participation of 20 annotators who are trainee translators. The annotators have evaluated the outputs in terms of intelligibility, fidelity and general quality.

The conclusion derived from the analyses carried out is that it is possible to evaluate machine translation consistently; there are similarities between error numbers and the rankings and ratings of human annotators. In addition, there are also similarities between intelligibility and fidelity ratings and general quality ratings. One of the most important results of the analysis is that the quality of machine translation programs depends mainly on the length of sentence.

It is expected that the study will contribute both to human translation and machine translation research by providing a comprehensive evaluation by human annotators and by providing an error categorization on the outputs of machine translation programs from English into Turkish.

**Key words**

Machine translation, machine translation evaluation, translation technologies, human evaluation, error categorization, intelligibility, fidelity.

## TABLE OF CONTENTS

|  |              |
|--|--------------|
| <b>KABUL VE ONAY</b> .....                     | <b>i</b>     |
| <b>BİLDİRİM</b> .....                          | <b>ii</b>    |
| <b>ACKNOWLEDGEMENTS</b> .....                  | <b>iii</b>   |
| <b>ÖZET</b> .....                              | <b>v</b>     |
| <b>ABSTRACT</b> .....                          | <b>vii</b>   |
| <b>TABLE OF CONTENTS</b> .....                 | <b>ix</b>    |
| <b>LIST OF TABLES</b> .....                    | <b>xiv</b>   |
| <b>LIST OF FIGURES</b> .....                   | <b>xvi</b>   |
| <b>LIST OF APPENDIXES</b> .....                | <b>xvii</b>  |
| <b>LIST OF ABBREVIATIONS</b> .....             | <b>xviii</b> |
| <br>   |              |
| <b>CHAPTER 1: INTRODUCTION</b> .....           | <b>1</b>     |
| 1.1. GENERAL REMARKS .....                     | 1            |
| 1.2. IMPORTANCE OF THE RESEARCH .....          | 6            |
| 1.3. AIM OF THE RESEARCH .....                 | 7            |
| 1.4. RESEARCH QUESTIONS AND SUBQUESTIONS ..... | 7            |
| 1.5. LIMITATIONS .....                         | 7            |
| 1.6. ASSUMPTIONS .....                         | 8            |
| 1.7. DEFINITIONS .....                         | 8            |
| 1.8. OUTLINE OF THE STUDY .....                | 9            |
| <br>   |              |
| <b>CHAPTER 2: HISTORICAL BACKGROUND</b> .....  | <b>11</b>    |
| 2.1. BEFORE COMPUTERS .....                    | 11           |
| 2.2. EARLY STAGES OF MT .....                  | 12           |

|  |           |
|--|-----------|
| 2.3. AFTERMATH OF ALPAC .....                    | 18        |
| 2.4. RENAISSANCE OF MT .....                     | 19        |
| 2.5. MT IN THE NEW MILLENNIUM .....              | 21        |
| 2.6. TURKISH MT SYSTEMS .....                    | 22        |
| 2.6.1. Proçeviri .....                           | 23        |
| 2.6.2. Sametran Sametech .....                   | 23        |
| 2.6.3. Google Translate .....                    | 24        |
| 2.6.4. Bing Translators .....                    | 25        |
| <b>CHAPTER 3: STATE OF ART .....</b>             | <b>27</b> |
| 3.1. LINGUISTIC BACKGROUND .....                 | 27        |
| 3.1.1. Morphology .....                          | 27        |
| 3.1.2. Lexicography .....                        | 28        |
| 3.1.3. Syntax .....                              | 29        |
| 3.1.4. Semantics .....                           | 29        |
| 3.1.5. Pragmatics .....                          | 31        |
| 3.1.6. Discourse Analysis .....                  | 32        |
| 3.2. PROCESSES, METHODS AND RESOURCES .....      | 33        |
| 3.2.1. Text Segmentation .....                   | 33        |
| 3.2.2. Part of Speech Tagging .....              | 35        |
| 3.2.3. Parsing .....                             | 37        |
| 3.2.4. Word-Sense Disambiguation .....           | 39        |
| 3.2.5. Anaphora Resolution .....                 | 41        |
| 3.2.6. Controlled Language and Sublanguage ..... | 42        |
| 3.2.7. Artificial Intelligence in MT .....       | 44        |
| 3.3. STRATEGIES .....                            | 46        |

|  |           |
|--|-----------|
| 3.3.1. Rule-Based Systems.....                               | 47        |
| 3.3.1.1. Direct Translation Strategy.....                    | 47        |
| 3.3.1.2. Interlingua Translation Strategy .....              | 49        |
| 3.3.1.3. Transfer Translation Strategy .....                 | 50        |
| 3.3.2. Corpus-Based Strategies.....                          | 52        |
| 3.3.2.1. Example-Based Machine Translation .....             | 53        |
| 3.3.2.2. Statistical Machine Translation .....               | 54        |
| 3.3.3. Hybrid Methods .....                                  | 55        |
| 3.4. MILESTONE MT SYSTEMS .....                              | 56        |
| 3.4.1. Systran.....  | 56        |
| 3.4.2. Meteo.....  | 57        |
| 3.4.3. Eurotra.....  | 57        |
| 3.4.4. Google Translate.....                                 | 58        |
| <br>   |           |
| <b>CHAPTER 4: TEXT TYPES AND EVALUATION TYPES IN MACHINE</b> |           |
| <b>TRANSLATION .....</b>                                     | <b>60</b> |
| 4.1. EVALUATION TYPES .....                                  | 60        |
| 4.2. TEXT TYPES BY REISS.....                                | 66        |
| <br>   |           |
| <b>CHAPTER 5: METHODOLOGY.....</b>                           | <b>70</b> |
| 5.1. DESIGN OF THE STUDY .....                               | 70        |
| 5.2. TEXT SAMPLES .....                                      | 70        |
| 5.3. DATA COLLECTING METHODS AND MATERIALS .....             | 71        |
| 5.4. TEST DESIGN .....                                       | 72        |
| 5.4.1. Micro Evaluation .....                                | 72        |
| 5.4.1.1. Grammatical Error Analysis.....                     | 72        |

|  |            |
|--|------------|
| 5.4.2. Macro Evaluation .....  | 73         |
| 5.4.2.1. Human Evaluation .....                                      | 73         |
| <b>CHAPTER 6: FINDINGS.....</b>                                      | <b>79</b>  |
| 6.1. CONTENT-FOCUSED TEXT TYPE .....                                 | 95         |
| 6.2. FORM-FOCUSED TEXT TYPE .....                                    | 117        |
| 6.3. APPEAL-FOCUSED TEXT TYPE .....                                  | 131        |
| 6.4. AUDIO-MEDIAL TEXT TYPE .....                                    | 152        |
| 6.5. DISCUSSIONS AND COMMENTS ON MACRO EVALUATION .....              | 154        |
| 6.6. DISCUSSIONS AND COMMENTS ON MICRO EVALUATION .....              | 158        |
| <b>CHAPTER 7: CONCLUSIONS AND RECOMMENDATIONS .....</b>              | <b>158</b> |
| 7.1. CONCLUSIONS REGARDING THE RESEARCH QUESTIONS .....              | 158        |
| 7.1.1. Main Research Question .....                                  | 158        |
| 7.1.2. Sub Research Question .....                                   | 161        |
| 7.2. ADDITIONAL CONCLUSIONS, OBSERVATIONS, RECOMMENDATIONS .....     | 173        |
| 7.2.1. Additional Conclusions and Observations .....                 | 174        |
| 7.2.2. Recommendations .....   | 175        |
| 7.2.3. Future Research .....   | 176        |
| <b>BIBLIOGRAPHY .....</b>  | <b>178</b> |
| <b>Appendix 1: TEZ ÇALIŞMASI ETİK KURUL İZİN MUAFİYETİ FORMU ...</b> | <b>188</b> |
| <b>Appendix 2: ORJİNALLİK RAPORU .....</b>                           | <b>190</b> |
| <b>Appendix 3: CD.....</b>   | <b>192</b> |

## LIST OF TABLES

|   |     |
|---|-----|
| Table 1: Error classification by Flanagan (1994) .....  | 65  |
| Table 2: Text samples and their features .....  | 71  |
| Table 3: The macro evaluation and micro evaluation of MT programs for content-<br>focused text type ..... | 92  |
| Table 4: The macro evaluation of each machine translation program for content-focused<br>text type .....  | 93  |
| Table 5: The micro evaluation of each machine translation program for content-focused<br>text type .....  | 95  |
| Table 6: The macro evaluation and micro evaluation of MT programs for form-focused<br>text type .....     | 114 |
| Table 7: The macro evaluation of each machine translation program for form-focused<br>text type .....     | 114 |
| Table 8: The micro evaluation of each machine translation program for form-focused<br>text type .....     | 116 |
| Table 9: The macro evaluation of each machine translation program for appeal-focused<br>text type .....   | 128 |
| Table 10: The macro evaluation of each machine translation program for appeal-focused<br>text type .....  | 129 |
| Table 11: The micro evaluation of each machine translation program for appeal-focused<br>text type .....  | 130 |
| Table 12: The macro evaluation and micro evaluation of MT programs for audio-medial<br>text type .....    | 149 |
| Table 13: The macro evaluation of each machine translation program for audio-medial<br>text type .....    | 150 |



|  |     |
|--|-----|
| Table 14: The micro evaluation of each machine translation program for audio-medial text type .....  | 151 |
| Table 15: Total errors of MT systems for each text type .....  | 156 |
| Table 16: The added up numbers of sentences which received two highest ratings from human annotators in terms of fidelity and intelligibility and the times of translations chosen best in ranking ..... | 160 |
| Table 17: Total errors of MT systems for each text type .....  | 173 |

## LIST OF FIGURES

|  |    |
|--|----|
| Figure 1: The places of MT in the map of Translation Studies (Holmes, 1988)..... | 2  |
| Figure 2: Machine and human translation .....                                    | 4  |
| Figure 3: Classification of different machine translation approaches .....       | 5  |
| Figure 4: The user interface of Proçeviri 3.2.....                               | 23 |
| Figure 5: The user interface of Sametech 1.04.....                               | 24 |
| Figure 6: The user interface of Google Translate in February 2015 .....          | 25 |
| Figure 7: The user interface of Bing Translator in March 2015.....               | 36 |
| Figure 8: Parse tree representation .....  | 38 |
| Figure 9: Direct machine translation work flow .....                             | 48 |
| Figure 10: Interlingua machine translation work flow .....                       | 50 |
| Figure 11: Transfer machine translation work flow .....                          | 51 |
| Figure 12: Example-based machine translation work flow .....                     | 54 |
| Figure 13: The test design of the research .....                                 | 75 |
| Figure 14: Results of three separate tests for a sentence .....                  | 78 |

## LIST OF APPENDIX

|   |     |
|---|-----|
| Appendix 1: TEZ ÇALIŞMASI ETİK KURUL İZİN MUAFİYETİ FORMU ..... | 188 |
| Appendix 2: ORJİNALLİK RAPORU .....                             | 190 |
| Appendix 3: CD.....   | 192 |

## LIST OF ABBREVIATIONS

|        |  |
|--------|--|
| AI     | : Artificial intelligence                          |
| ALPAC  | : Automatic Language Processing Advisory Committee |
| BLEU   | : Bilingual evaluation understudy                  |
| CAT    | : Computer assisted translation                    |
| EBMT   | : Example -based machine translation               |
| FAHQMT | : Fully automatic high quality machine translation |
| HAMT   | : Human-aided machine translation                  |
| MAHT   | : Machine-aided human translation                  |
| MT     | : Machine translation                              |
| SMT    | : Statistical machine translation                  |

## CHAPTER 1

### INTRODUCTION

This chapter dwells on a short introduction to machine translation as well as a general framework of the thesis, importance and aim of the research, problem statement, and research questions. The assumptions, limitations and abbreviations are also given.

#### 1.1. GENERAL REMARKS

Machine translation (MT), as a term, refers to computerized systems responsible for the production of translation with or without human assistance. Being one of the earliest applications for computers, MT is today used by more than 200 million people daily (Och, 2012). However, MT has its own peculiarities and difficulties for each language pair. MT between Turkish and English has proven to be challenging for users especially in commercial and academic world for many reasons, such as syntactic differences between the languages or lack of linguistic and financial resources for Turkish MT, etc.

Machine translation is an interdisciplinary research area in the intersection of computer science, linguistics and Translation Studies.

Machine translation has its own places in the map of Translation Studies, which was drawn by Holmes (1988), who has been credited with laying the general framework for Translation Studies. Holmes has categorized Translation Studies under two general branches; pure and applied. Pure Translation Studies describe the translation phenomena and try to establish general principles to understand and predict that phenomena. It has two subbranches; theoretical and descriptive. Theoretical branch is also divided into two subbranches; general and partial. General branch tries to account for every type of translation theory while partial branch tries to establish principles by limiting itself to certain factors as medium, area, rank, text type, time and problem. Descriptive branch has three focus points; product, process and function, enabling Translation Studies to establish

general principles. Applied Translation Studies focus on the translator training, translation aids and translation criticism.

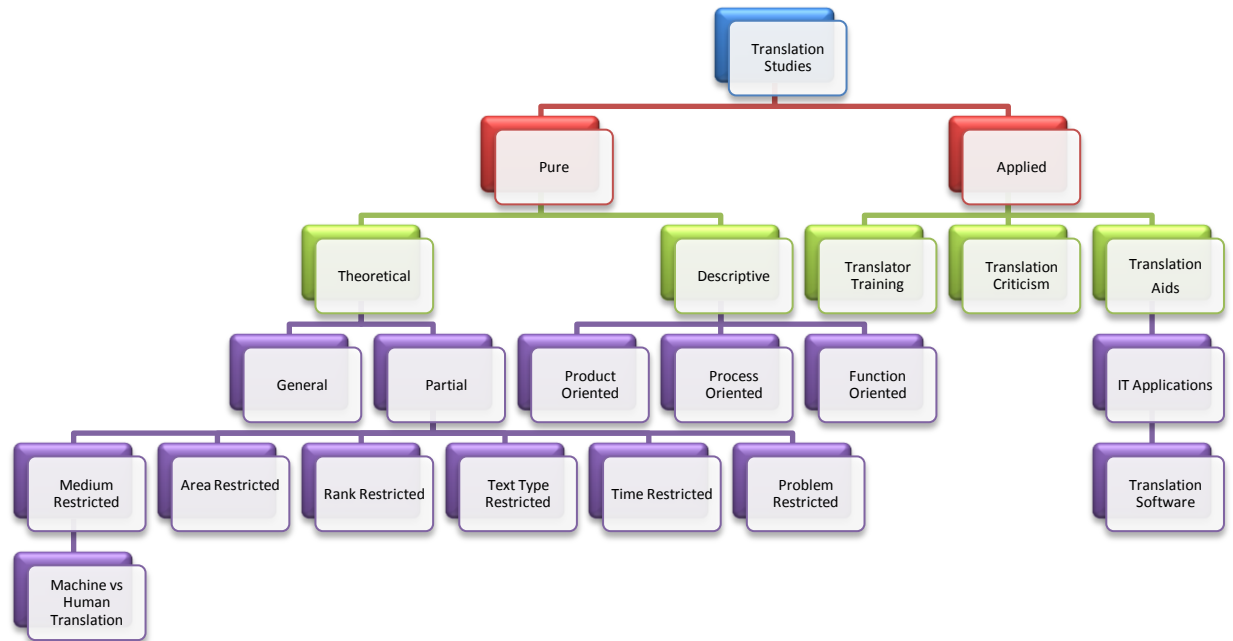


Figure 1: The places of MT in Translation Studies map (Holmes, 1988).

Machine translation has been categorized under both in pure and applied Translation Studies. In pure Translation Studies, machine translation falls under the medium-restricted theoretical branch. In applied Translation Studies, it is a research topic under the translation aids.

The academic interest in machine translation was first aroused by a mathematician, Warren Weaver, with his famous memorandum written to Rockefeller Foundation (1949): “When I look at an article in Russian, I say ‘This is really written in English, but it has been coded in some strange codes. I will now proceed to decode.’” (p. 14). This memorandum has served as a spark for the academic interest and state and private sector funding in the USA.

Today, MT has many commercial applications, which are available for many language pairs.

MT has proved to be of social, political, scientific and philosophical importance. Social and political importance emerges from the necessity to understand the other. Binational or multinational countries and organizations need to translate great volumes of texts into many languages in a very limited time. For instance, European Union allocates around €330m a year to translate from and into 23 official languages. In addition, Union allocates nearly % 1 of the annual budget for all the language services (DG Translation official website, 2014). European Union uses an internal machine translation engine, which has shifted from rule - based to statistical MT system in the recent years. Commercial importance emerges from the fact that for each step in international markets, from business agreements to instruction manuals, translation is a requirement for people to interact with each other. The delays in translation can be costly, so using MT can help translators and trading parties in the most efficient ways. Scientifically, MT is an interdisciplinary area at the intersection of computer science, linguistics and artificial intelligence. It is known to be one of the earliest non-numerical applications for computers. Philosophically, MT is basically automation of translation, which requires complex language and world knowledge. The automation of world knowledge, including common sense, can mark a new epoch in computer science and computer-human interaction.

Machine translation systems are developed either for a single language pair, called bilingual systems (for example, SAMETRAN) or for more than 2 languages, called multilingual systems (for example, Google Translate). They can work either into one language, which is unidirectional, or into many languages, which is multidirectional.

As far as MT is concerned, there are many different practices and approaches. The different degrees of human-computer cooperation have resulted in the classification of machine translation technology. Machine translation (MT) refers to the full automation of translation process, without human intervention. Machine-aided human translation (MAHT) refers to the computer -based translation tools, such as term banks, online dictionaries, translation

memories. MAHT is generally referred as CAT, which is computer assisted translation tools. Human-aided machine translation (HAMT) refers to the machine translation production with human intervention, i.e. pre-editing and post-editing. These are two notions which are as old as machine translation itself. They were first proposed in 1950, by E. Reifler of Washington University (Buchmann, 1987:6). Pre-editing is the process of ambiguity resolution and determining a single meaning for words with multiple meanings before the text is submitted to machine translation. Post-editing is the revision of machine output by a human expert before the distribution of the translated material. Human translation (HT) refers to the traditional human translation, without any computerization of translation.

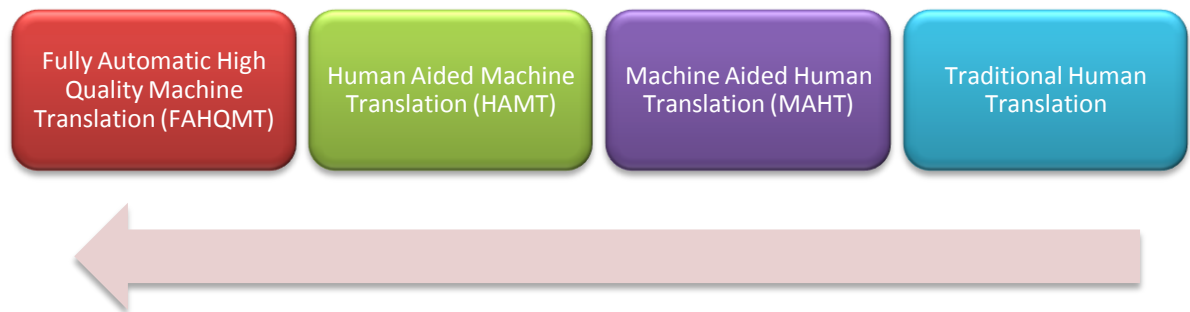


Figure 2: Machine and human translation. Mechanization increases in the direction of arrow (Somers & Hutchins, 1992).

The following figure illustrates different approaches to machine translation:



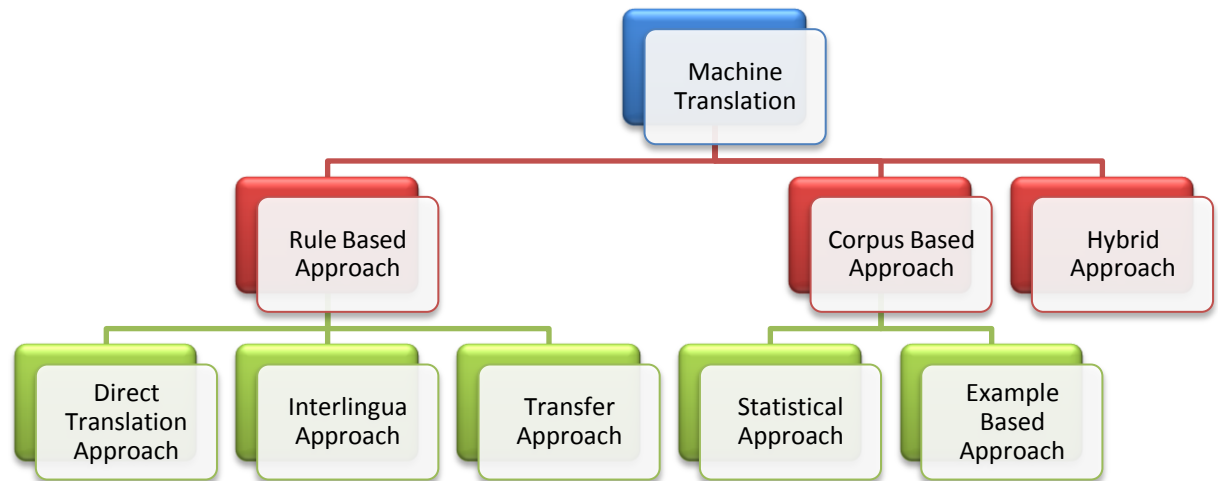


Figure 3: Classification of different machine translation approaches.

As seen in the above figure, there are different approaches to MT which can be classified under three main branches; rule -based approach, corpus -based approach and hybrid approach. Rule -based approaches, which are mainly governed by linguistic rules, include direct translation approach, interlingua approach and transfer approach. In direct translation approach, translation is direct from source text to target text, with the least possible syntactic and semantic analysis. The quality of translation mostly relies on a large bilingual dictionary, and output of the process is usually a word-for-word translation. In interlingua approach, MT translates source texts into abstract descriptions which are believed to be language independent and common to more than one language. From these interlingual representations, texts are generated into other languages. For instance, Esperanto, being an artificial language, bears more common points with European languages which are marked by gendered parts of speech, and serves the purpose of establishing an interlingual language better than natural languages and thus treated within the interlingual approach. Google Translate uses English as interlingua or bridge language between distant language pairs, e.g. between Turkish and Swahili; the Turkish input is first translated into English and, the translated text is then translated into Swahili. Alternatively, logical artificial languages can also be employed as in the case for Stanford MT Project, which employed predicate

calculus for an interlingual MT system for English – French language pair (Wilks, 2003: 387). Transfer approach has a three-step working flow. Firstly, the texts are converted into the intermediate representations such as Chomskian tree representations, followed by disambiguation. Then, these representations are converted into the representations of the target language. And finally, the target text is generated. Corpus -based approach has two main applications. In statistical MT large bilingual text corpora are analyzed and parameters are set for the translation. This approach requires large parallel corpora for higher quality. Example-based MT (EBMT) assumes that translation involves finding or recalling how a particular source language expression or a similar expression has been translated before. And lastly there are hybrid machine translation systems making use of both statistical and rule-based methods.

Machine translation evaluation is a crucial component of the whole process. According to Arnold et al., there are various factors to keep in mind while evaluating machine translation, such as technical environment, organizational changes, engine performance in terms of quality and speed, etc. (1994: 157). The quality of output can be evaluated in various ways, either by human or automatic evaluation methods. Human evaluation methods mainly consist of ranking or rating the output sentences in terms of their usability, intelligibility, fluency, etc. Automatic evaluation is usually done by text similarity programs, or by measuring translation editing effort.

## **1.2. IMPORTANCE OF THE RESEARCH**

Machine translation is an old academic discipline. Despite the fact that it has been studied a lot in computer science departments, specific emphasis to the evaluation of the output has not been encountered among the studies carried out in translation departments in Turkey. Thus, this thesis is expected to fill in this gap in this field.

### **1.3. AIM OF THE RESEARCH**

The aim of this research is to see if consistency can be achieved in the evaluation methods of machine translation quality by comparing the performance of 3 machine translation programs working from English into Turkish for translating four different text types categorized by Reiss.

### **1.4. RESEARCH QUESTION AND SUBQUESTIONS**

Research Question:

**Is it possible to evaluate MT quality consistently with different evaluation methods for different text types?**

Subquestions:

1. Is there a difference between micro evaluation and macro evaluation of MT performance for different texts in Reiss' typology exemplified with abstract, short story extract, online advertisement and subtitle?
2. To what extent is the output of an MT system fluent (intelligibility) and faithful (fidelity) for the human annotators?
3. Which types of MT errors have the highest impact on the human perception of translation quality?
4. Does MT output quality change for different text types?

### **1.5. LIMITATIONS**

1. The research is confined to machine translation programs working from English into Turkish.

2. The research employs three machine translation programs, Google Translate (February 2015), Proçeviri (version 3.2) and SameTech (version 1.04).
3. Human annotators are 20 senior students of English Translation and Interpreting Department at Hacettepe University in 2014-2015 academic year.

## **1.6. ASSUMPTIONS**

1. Human annotators are assumed to participate in the questionnaires with their full concentration.
2. Human annotators are assumed to answer questions in the questionnaire and evaluate their own performances sincerely.

## **1.7. DEFINITIONS**

In this study, the following concepts and terms are used within the frame of their definitions.

**BLEU:** Bilingual evaluation understudy, an algorithm for evaluating the quality of machine translation output.

**Consistency:** The agreement between different methods in evaluating the same thing.

**Fidelity:** The degree to which the meaning of source language translation unit is preserved in the target language.

**Fully automatic high quality translation:** A machine translation without any human intervention, the output of which cannot be distinguished from human translation.

**Human-aided machine translation:** Translation where human user may post or pre-edit the machine translation output.

**Human translation:** Translation performed solely by human, without any computer.

Intelligibility: The degree to which the translated text is read grammatical and natural in the target language.

Machine-aided human translation: Translation where human user may use computer supporting tools.

Machine translation: Automatic translation between human languages.

## **1.8. OUTLINE OF THE STUDY**

The outline of the study is as follows:

In this first chapter, an introduction to machine translation and research thereof is presented. General framework of the study encompassing problem statement, research questions and importance of the research assumptions as well as limitations and abbreviations pertaining to this thesis is given.

In second chapter, historical background of machine translation is presented, mostly from a European point of view, as the literature of machine translation has been mostly provided by European and American scholars. Turkish MT history and MT systems are also detailed in this chapter.

Third chapter dwells on the state of art to provide a solid basis to develop an understanding on the functioning of machine translation. The framework includes linguistic background as well as processes, methods and resources and basic strategies of machine translation.

Fourth chapter is devoted to the evaluation types and methods of MT. Being one of the central research areas of machine translation, evaluation is detailed in terms of history, methods, variables, etc. In addition, text typology outlined by Reiss is explained in this chapter.

The fifth chapter constitutes the methodology of the research. This chapter includes samples, design of the research, data collecting materials and data analysis techniques.

Case studies constitute the sixth chapter. Case studies include the comparison of translations of 4 texts chosen in accordance with the text typology outlined by Reiss in 1976. Comparison between the translations produced by three different machine translation systems is drawn which paves the way to develop an understanding for the best machine translation system for Turkish-English language pair.

Last chapter constitutes the conclusion of the research, where the applicability of hypotheses is discussed. Some suggestions for training, future action and research are argued accordingly.

In this chapter, definitions of basic terms and relevant approaches to machine translation have been introduced. Preliminary information on the research such as research question and sub questions, importance of research, limitations and assumptions are also elucidated.

In the following chapter, history of machine translation is recounted, including the the history and present status of MT in Turkey.

## **CHAPTER 2**

### **HISTORICAL BACKGROUND**

Long history of machine translation, with specific reference to important papers, books, experiments and researchers in the area, as well the beginnings and recent developments in Turkish MT, are the main focus points of this chapter.

#### **2.1. BEFORE COMPUTERS**

Machine translation was one of the first applications of the computers, and was first envisaged even before the invention of computers (Hutchins, 1986: 21). The fall of Latin as the universal scientific language and the supposed inability of natural languages to express thought unambiguously led thinkers such as Descartes and Leibniz to come up with the idea of numerical codes for languages. Descartes, in a letter dated 1692, described a universal language cipher, where the lexical equivalents of the all known languages would be given the same code (Hutchins, 1986: 21). Such dictionaries were actually published by three people; by Cave Beck in 1657, by Athanasius Kircher in 1663 and by Johann Joachim Becker in 1661 (Hutchins, 1986: 22). However, the first involvement of construction of machines was first proposed in 1933, by two different inventors living in different countries. The first inventor was a French engineer, Georges Artsrouni. He was granted a patent for what he called “Mechanical Brain” on 22 July 1933 (ibid.). It was a device “worked by electric motor for recording and retrieving information on a broad band of paper which passed behind a keyboard... Each line of broad tape would contain the entry word (SL word) and equivalents in several other languages (TL equivalents).” (ibid.). However, the start of Second World War prevented the installation of this device. The second patent was granted to Petr Petrovich Smirnov-Troyanskii in Moscow on 5 September 1933. What makes this patent interesting for MT is his proposed “logical

analysis”. In logical analysis, all inflected words were to be transformed into their base forms and they were to be ascribed their syntactic functions in the sentence. It is Troyanskii’s foresight on the mechanization of logical analysis that makes him truly precursor of machine translation (Hutchins, 1986: 23). However, his proposal couldn’t get the support of Soviet scientists, the unavailability of necessary computer facilities and technology also resulted in the omission of this patent.

## **2.2. EARLY STAGES OF MT**

The well - documented history of machine translation can be said to have started after the Second World War. The creation of computers and computer science and the development of cryptography during the War paved way for machine translation. The first applications of computers were naturally numerical, such as the calculation of ballistic firing tables, but after the War, the limits of electronic brains were tested with non-numerical applications.

The use of computers for translation was first suggested by Andrew D. Booth and Warren Weaver in 1946 (Hutchins, 1986: 26). This is the generally accepted birth date of MT. Andrew D. Booth was a British scientist interested in constructing computational facilities in the University of London. He obtained funds from Rockefeller Foundation to visit the USA in 1946. Warren Weaver was the vice president of Rockefeller Foundation at that time. Booth explained his intention for building a machine for University of London based on American Experience. Weaver, according to Booth (Hutchins 1986: 25), saw little chance in Americans financing British for a computer in numerical applications. Then Weaver suggested a translating machine, and treating translation as a cryptographic problem. However, this was not the first time that Weaver proposed mechanization of translation. In a letter to Norbert Wiener, one of the pioneers of mathematical theory of communication, Weaver raised his ideas on mechanical translation (Weaver, 1947, quoted in Weaver, 2003):

I have wondered if it were unthinkable to design a computer which would translate. Even if it would translate only scientific material, and even if it did



produce an inelegant (but intelligible) result, it would seem to me worthwhile. Also knowing nothing official about, but having guessed and inferred considerable about, powerful new mechanized methods in cryptography—methods which I believe succeed even when one does not know what language has been coded—one naturally wonders if the problem of translation could conceivably be treated as a problem in cryptography. When I look at an article in Russian, I say: “this is really written in English, but it has been coded in some strange symbols. I will now proceed to decode.” (p. 14).

However, Wiener, as a linguist, was well aware of the vastness of differences between languages: “At the present time, the mechanization of language, beyond such a stage as the design of photoelectric reading opportunities for the blind, seems very premature. . . .” (Wiener, 1947, quoted in Weaver, 2003). After another failure in interesting another linguist, Ivor Richards, Weaver wrote a memorandum on 15 July 1949 to some 200 of his acquaintances who might have an interest in “mechanical translation”. The memorandum was a success; it launched machine translation as a scientific enterprise in the USA and consequently other countries.

In the memorandum, Weaver raised four main problems, which should be resolved. The first problem is meaning and context, which is the problem of multiple meaning. He suggested that this problem can be solved if a sufficient amount of immediate context (one or two words before and/or after) is taken into account. The second problem is finding the logical basis of language. Weaver was optimistic on this issue. After recounting a theorem which states that a computer is capable of deducing any legitimate conclusion from a finite set of premises, Weaver states that the problem of translating with a computer is at least formally solvable. The third issue he raised was on the applicability of communication theory and cryptographic techniques. Weaver recounted an anecdote by R.E Gilman of Brown University Mathematic Department. Gilman was given a 100-word coded text in Turkish. He didn’t know Turkish, moreover he had no idea in which language the text had been coded. He successfully decoded message without even having knowledge of the language (Weaver, 1949). By recounting that anecdote, Weaver put forward his belief that translation could largely be solved by “statistical semantic studies” (Hutchins, 1986: 29). The last point Weaver raised was the language universals, or invariants. He presented an analogy to make his point more clear (Weaver, 1949):

Think, by analogy, of individuals living in a series of tall closed towers, all erected over a common foundation. When they try to communicate with one another, they shout back and forth, each from his own closed tower. It is difficult to make the sound penetrate even the nearest towers, and communication proceeds very poorly indeed. But, when an individual goes down his tower, he finds himself in a great open basement, common to all the towers. Here he establishes easy and useful communication with the persons who have also descended from their towers. Perhaps the way is to descend, from each language, down to the common base of human communication—the real but as yet undiscovered universal language—and then re-emerge by whatever particular route is convenient (p. 17).

Weaver's memorandum received mixed reactions. It was received well by computer scientists as a new and unexplored study area for computers. However, linguists found its assumptions on the formalization of language and translation process very naïve. Linguists' insights were found solid when Weaver's approach to translation as a cryptography problem was widely recognized mistaken. Deciphering is based on the frequency of letters, pair of letters which can be found in a text of whom context is well-known by the person who is deciphering; the place of attack or the date of a landing. And after a text is deciphered, it can be translated to another language (Hutchins, 1986:31).

Although Weaver was mistaken in approaching the problem, his memorandum started MT as a serious research area and study groups began to be formed with the funding by the United States government. Weaver tried to take this interest a step further and convened a meeting at Massachusetts Institute of Technology in January 1950 with scientists from different backgrounds. The meeting was a success, one year later, in 1951 Yehoshua Bar-Hillel became the first full-time researcher on MT with two specified tasks: the possibility of MT and the planning of future research on MT. After studying the current position, Bar-Hillel presented a paper. The paper was very influential, the topics mentioned in the paper set the agenda for the following years in MT research.

The growing interest in MT resulted in a public demonstration on 7 January 1954 (Hutchins, 1986:37). Leon Dostert of Georgetown University collaborated with IBM for an "actual experiment rather than resolving the problem theoretically" (Dostert, 1955: 125). A text with a restricted vocabulary of 250 Russian words were translated from Russian into

English with just 6 grammar rules, without any pre or post-editing. The translated text was intelligible, which was proved to be enough to convince the general public and the government to believe that MT was feasible. This experiment started the official support by the government and also started the exaggerations, which would stop all the funding one decade later. However, it further stimulated MT research in the United States and it started MT research in the Soviet Union.

After the successful Georgetown experiment, the optimism was widespread. For instance, Delavenay went on to claim that machine translation programs would even translate poetry: “Will machine translate poetry? To this there is only one possible reply – why not?” (Delavenay, 1960, cited in Hutchins, 1986: 151).

Meanwhile, the first academic journal devoted to MT was published by Locke and Yngve in 1955, under the name of Mechanical Translation. Following Georgetown experiment, 12 research groups established only in the USA. Different research groups approached the problems differently; but there was a problem which they shared. The computer hardware was not enough; many groups had no access to computers, and those who had access to the computer faced with storage problem. The storage available was far away from being adequate; large dictionaries, obvious prerequisites for even word-for-word translation, were hard to create and maintain. In addition, the lack of generalised and formalised linguistic theories was also meant that the research groups were approaching problem still as an engineering problem. However, the general optimism and the trust in the developments in the computer science led the MT researchers to think that optical character readers and large storage capacity computers would be invented soon.

The advent of formalised linguistic theories, such as Chomsky’s formal models, Harris’ transformational grammar, etc., led the MT researchers to take linguistics aspects of MT more seriously. The obvious poor quality of word-for-word translation could only be improved with the syntactic and semantic analysis for the FAHQT. This meant longer theoretical research for higher quality. However, once widespread optimism started to fade.

The first person to be appointed as a full time researcher of MT was the first person to reveal doubts on it. Bar-Hillel's paper entitled *Report on the state of machine translation in the United States and Great Britain* was published in 1959. It was a compilation of his criticisms of MT research groups. In his paper, he argued that FAHQT was not impossible just in near future, but it was impossible altogether. His famous example was a short sentence: "The box was in the pen." in the context: "Little John was looking for his toy box. Finally, he found it. The box was in the pen." (Bar-Hillel, 1959). Bar-Hillel argued that the homonymy problem in this sentence required world knowledge that "box" had to be bigger than writing tool "pen", so the "pen" must have referred to the "playpen". In order to solve this problem, "translation machine should not only be supplied with a dictionary but also with a universal encyclopaedia." (Bar-Hillel, 1959), which, according to Bar-Hillel, was hardly deserving of any further discussion. He criticised many of the MT research groups on their adherence to the aforementioned impossible goal. The optimism spread by the Georgetown experiment had been criticised for convincing both MT community and public to believe that operational systems would be on the market in just a few years. However, this would not happen, so Bar-Hillel advised the MT researchers to attain "either fully automatic, low quality translation or partly automatic, high quality translation". Due to strong adherence to "impossible goal" of FAHQT, many research groups, in Bar-Hillel's opinion, wasted the national funds. In many resources the funding is stated to be around £ 20 000 000 (Hutchins, 1986: 167; Arnold et al., 1994: 13). Despite the harsh criticism of multi-million affair voiced by Bar-Hillel, MT received another funding from US House of Representatives, it has been stated in a report that the funding was given for "the overall importance of MT to intelligence and scientific effort of the Nation, for the translation of English text for the exchange of cultural, economic, agricultural, technical, and scientific documents that will present the *American way* of life to people throughout the world" (US House of Representatives, 1960). The developments in MT in USSR, and the current situation of MT in the USA were included in the report. The report also foresaw a national center for machine translation, and a national machine translation program. Apart from its monetary importance, the funding was considered to be "official" acknowledgment of national importance of MT efforts.

After the official acknowledgment, national and international conferences were held, making MT a more mature science. In addition, as an indication of scientific maturity, Association for Machine Translation and Computational Linguistics was established in 1962. The MT researchers were still quite optimistic for the future, despite the harsh criticisms of Bar-Hillel.

However, the public started to get impatient. After all the funding and a decade of research, there was no operative system, not only in the market, but even in research laboratories. A book by Mortimer Taube entitled *Computers and Common Sense* (Taube, 1961) put forward the general perception of MT in the eyes of public. Taube criticised MT researchers for failing to produce any actual achievements. He held the same strong ideas with Bar-Hillel about the impossibility of FAHQMT. Taube insisted that the formalisation of language was impossible, and computers demanded precise, formalised information. The work of Noam Chomsky on formal language analysis was criticised as “an *aberration, which cast a mystique over the whole field of MT.*” (Taube, 1961, cited in Hutchins, 1986: 162). He saw no point in continuing MT research, for which there wasn’t even a feasibility study. Taube insisted that the dehumanising venture was doomed to fail. The book had considerable effect on public perception, which was already impatient due to the slow progress. The growing impatience led the funding agencies to conduct a survey on the current MT research.

The director of National Science Foundation requested National Academy of Sciences to establish a committee to survey MT and advise funding agencies for the future of MT research and funding. The committee is known as Automatic Language Processing Advisory Committee and the resulting report has been known as ALPAC report (1966). The committee included two linguists, one psychologist, two MT specialists and one AI researcher. MT specialists were David G. Hays and Anthony Oettinger, the writer of the first doctoral dissertation on MT. The committee mainly approached MT with economic considerations. It undertook studies on the current market of translation only from Russian into English, and evaluated the output of MT for just from Russian into English texts. The committee concluded that supply exceeded the demand in translation market and what the

translation market was in need was improving speed, quality and consistency. The answer was not MT, but machine aids for translation (CAT). The committee emphasized the unavailability of any machine translation for scientific texts, both in the present and in the future. Raw MT product was in need of post-editing, which was seen as a failure of MT. In conclusion, ALPAC report advised the funding agencies to reduce the funding because “there is no immediate or predictable prospect of useful machine translation.” (ALPAC, 1966). Instead, computational linguistics and machine aids for translation were advised to be supported.

### **2.3. AFTERMATH OF ALPAC**

The ALPAC report started what is known in MT as “dark ages”. The funding stopped in the United States and the number of research groups immediately decreased. The loss of status was so severe that the Association of Machine Translation and Computational Linguistics had to remove “Machine Translation” from title in 1986. However, even an insider, Hutchins found the reduction of funding quite rightful: “The ALPAC report may have dismissed solid achievements too readily, but it was quite right to call a halt to indiscriminate and wasteful expenditure on unrealistic projects.” (Hutchins, 1986:169). Other researchers disagreed with the report and prepared another report to invalidate it. *Commentary on the ALPAC report* approached to MT from a more scientific point. The main disagreement was over the absence of operational MT systems. Two MT systems were actually in regular use, those were in IBM Mark II (used by USAF) and Georgetown System (used by EUROTAM) (King, 1987). Although the quality was poorer than expected, they were still operational. However, this report failed to invalidate the ALPAC report. The damage was done in the United States. The effect was not limited to the United States, all three British MT research groups ceased their studies, Japan and USSR MT research groups suffered from reduced funding. However, research continued in the continent Europe, the result of which would be called as “second generation MT”.

#### **2.4. RENAISSANCE OF MT**

In the aftermath of ALPAC report, research and funding for MT stopped sharply. However, one decade later the picture was much brighter. The channelling of funding from MT into computational linguistics and artificial intelligence, led to the improvements in theoretical issues which were closely relevant to MT. According to Warwick, by 1975, enthusiasm renewed in MT due to the matured linguistic theories, growing interest from commercial sector and success of tentative result such as TAUM-METEO project, a FAHQT project, which was considered impossible by Bar-Hillel due to the lack of computers understanding “meaning” and having “universal encyclopaedia” one decade ago (1987). In addition, the need to keep up with international developments, especially in technological and military areas, the need to disseminate the information properly as a result of increasing international trade and the emphasis on multilingualism especially in the Commission of European Communities and bilingualism in Canada resulted in a new future for MT globally. After one decade of “dark ages” (Hutchins, 1986:174), MT enjoyed a new status with a more realistic view of situation, which can be called a sort of renaissance.

The most visible sign of this renaissance was SYSTRAN, an MT system developed in Georgetown University. A more developed version of SYSTRAN was sold to Commission of European Communities for English- French translation in 1976. In addition, the fifth generation project of Japan helped to restore MT’s status a lot. According to Pugh (1992: 26), MT has a privileged status in Japan, due to the widespread perception of language technologies as the key technology of the new century. In addition, government, universities and private sector regards MT as an essential part of information-based society. Apart from long term benefits, economic considerations also play an important role in the status of MT. According to a report by Japanese Electronic Industry Development Association in 1989, the translation market from Japanese into English was around 800 million yen, and of total market, English was either source or target language in 90% of translated texts. Large Japanese companies such as Mitsubishi, Toshiba, Hitachi, etc. all invested in MT. In Europe, another sign of MT renaissance was multinational investment of Commission of European Communities to EUROTRA project, a multilingual MT for each

of Commission's member countries' languages. Studies started in 1979 (Hutchins, 1986:179), and due to being a multinational investment, a national university of each member country in Commission had been assigned the responsibility of developing that country's language system. Being a multinational project from its inception, the studies covering the languages of member states started simultaneously for EUROTRA (King, 1987: 373). The Commission granted about 12 million dollars according to Hutchins (1986: 264) and demanded "the creation of a machine translation system of advanced design capable of dealing with all the official languages of the Community" (King, 1987: 374).

When research projects in the 1990s are inspected, the emergence of new techniques can easily be seen. One technique is incorporating Artificial Intelligence (AI) into MT systems in order to resolve semantic problems, such as anaphora resolution. Those systems using AI techniques are known as "knowledge-based MT" (KBMT) (Somers, 1992: 192). Using sublanguage is another technique, which has proven itself in the most successful MT system, METEO. METEO is an MT system which was in use starting from 1981 to 2001, and it was specifically developed to translate weather bulletins of Canada from English into French, as a result of country's language equality policy (Melby, 1992: 147). Corpus-based MT systems use statistical and probabilistic methods to decide the best equivalent for a ST, by investigating large pre-translated bilingual or multilingual corpora (Somers, 2003: 8).

The quality of MT products has been another important problem of 1990s MT research. The lack of theories regarding the quality of human translation compelled MT quality researchers to find their own criteria for quality assessment. For instance, intelligibility, fidelity, acceptability, revision time are frequently mentioned variables of quality measurement. In addition, the emergence of personal computers at accessible prices led to the widespread use of computer assisted translation tools (CAT) by professional translators (Somers, 2003: 93).



## **2.5. MT IN THE NEW MILLENNIUM**

New millennium witnessed the use of translation technologies spreading worldwide. Commercialization of MT systems, and more importantly, online MT systems, thus, became a reality of life. Online MT systems such as Babelfish, Google Language Tools, etc. use system design developed by SYSTRAN, but they incorporate it with statistical approaches (Cancedda, Dymetman, Foster & Goutte, 2009). Recently, Google Inc. announced that Google Translate, online and free machine translation service, is used by more than 200 million people each month. This equals to 1 million books each day (Och, 2012). In addition, a wider coverage of languages has made online MT systems more attractive enabling MT to translate between not only the commercially important languages but also other less commercially attractive languages. However, according to Somers, the MT systems for some of world's top 20 most spoken languages such as Hindi, Urdu, Telegu, Tamil, Cantonese, etc. are still either under developed or have never developed (2003: 87).

Speech translation (abbreviated as S2S) has also become a more matured research area since the beginning of the millennium. A speech translation consists of roughly three modules: speech recognition, machine translation and speech synthesis. However, there are various problems to overcome for spoken language translation to be widely used. The spoken language contains ungrammatical sentences which abound in proper names and colloquial expressions. In addition, the speaker dependency of speech recognition systems also limits the usage of speech translation technologies. However, there are some websites and programs capable of speech translation; such as Skype, YouTube (speech recognition and machine translation for generating subtitles), Bing Translator and Google Translate.

Advent of smart phone technology and emergence of applications for smart phones have also provided a new ground for machine translation technologies. Some of the applications for machine translation are Google Translate, Bing Translator, Auto Translation, Babylon, inter alia for Google Play and Translate Voice Free, iTranslate and Speak and Translate for

Apple Store. Google Translate, Bing Translator, Auto Translation and some other applications use voice, writing, photograph as input.

The increasing number of European Union and European Commission languages has led European politicians to resort to machine translation systems more in order to protect the multilingual nature of organizations. There are various machine translation systems and projects supported and used by European Union and European Commission. Both organizations use an internal machine translation system, which are statistical. MT@EC, the translation sytem used by Commission can translate between 552 language pairs ("Machine translation service," 2015).

Apart from these developments, new millennium has also witnessed an increasing number of corpora for many different and distant language pairs. The multilingual websites such as Wikipedia, have also served as a corpora for many languages. All these developments have contributed to machine translation, especially to statistical systems.

## **2.6. TURKISH MT SYSTEMS**

Machine translation systems for Turkish have been developed more recently than many other languages. Agglutinative morphology and lack of parallel corpora can be the main reasons for the delay in the development of systems. Both commercial and experimental systems have been developed in the last two decades. The compilation of METU Turkish Corpus has resulted in many theses and articles related to MT and computational linguistics between English and Turkish. In addition, commercial systems have been in the market since 2000. However, only two commercial systems have been developed so far; these are Proçeviri marketed in 2000, and Sametran in 2006. Google Translate and Bing Translator are two online MT systems incorporating Turkish.

### 2.6.1. Proeviri

Proeviri is the first MT system developed between English and Turkish. The system translates from English into Turkish, not vice-versa, which makes it a unidirectional system. The user can improve the dictionary of system. In addition, by enabling “select right function”, the system asks for disambiguation in terms of unknown words and their morphological categories. The system also calculates translating time.

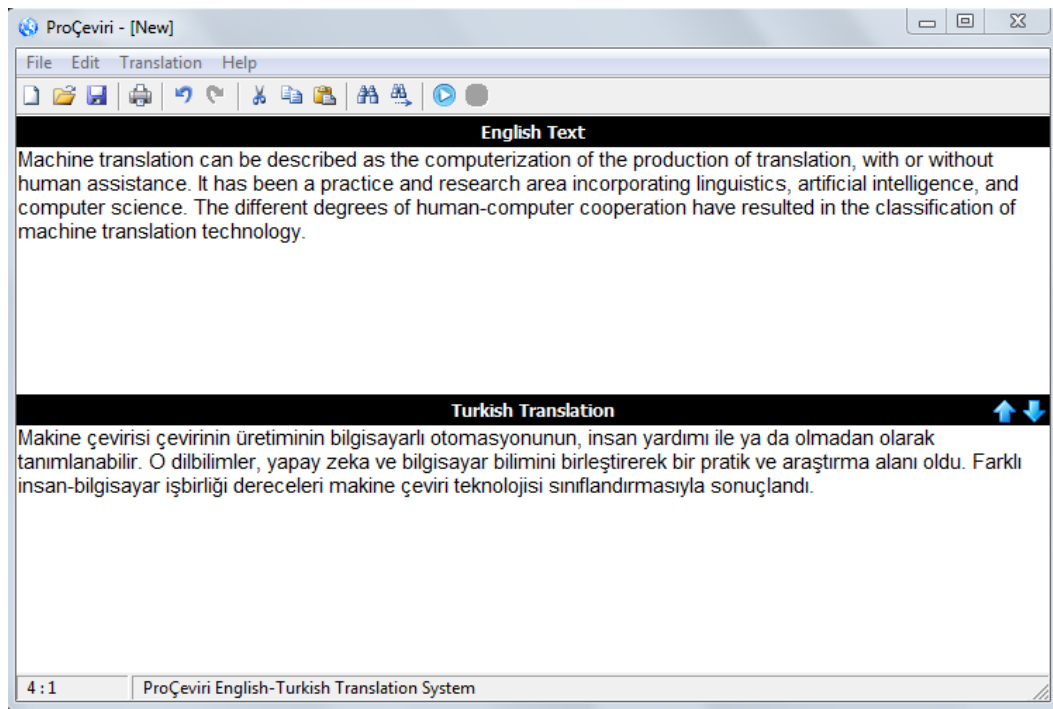


Figure 4: The user interface of Proeviri 3.2.

### 2.6.2. Sametran Sametech

Sametran Sametech 1.04 is the second MT system which can translate from English into Turkish. The system is unidirectional, taking English as the source language. The system can recognize 1.200.000 words and many more can be added to the dictionary. The system enables the user to analyze the aligned source and target texts.

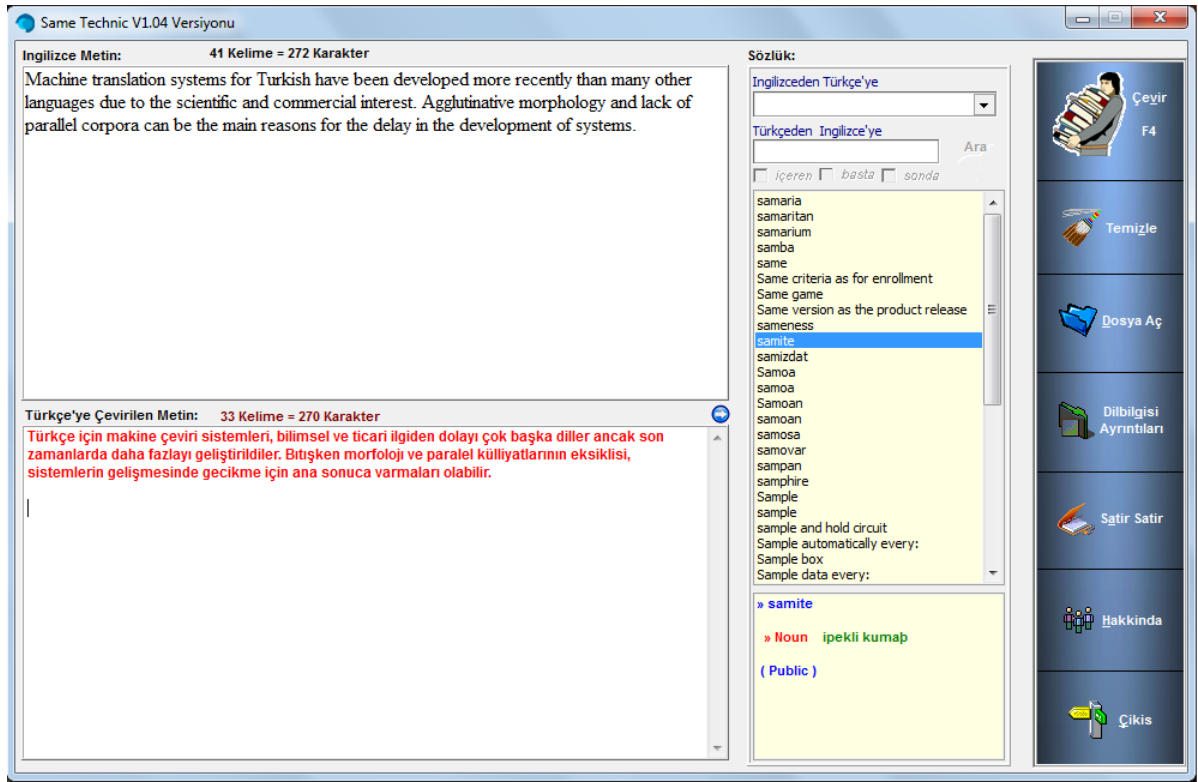


Figure 5: The user interface of Sametech 1.04.

### 2.6.3. Google Translate

Google Translate is the multilingual and bidirectional statistical online machine translation system developed by Google Inc. Since 2009, the system supports Turkish. The system enables user to edit the translation. In addition, by Google Translator Toolkit, the user can develop translation memory, specialized dictionaries and can edit and invite others to edit the raw translations. The system learns from the previously translated parallel corpora and the dictionaries are constantly updated.

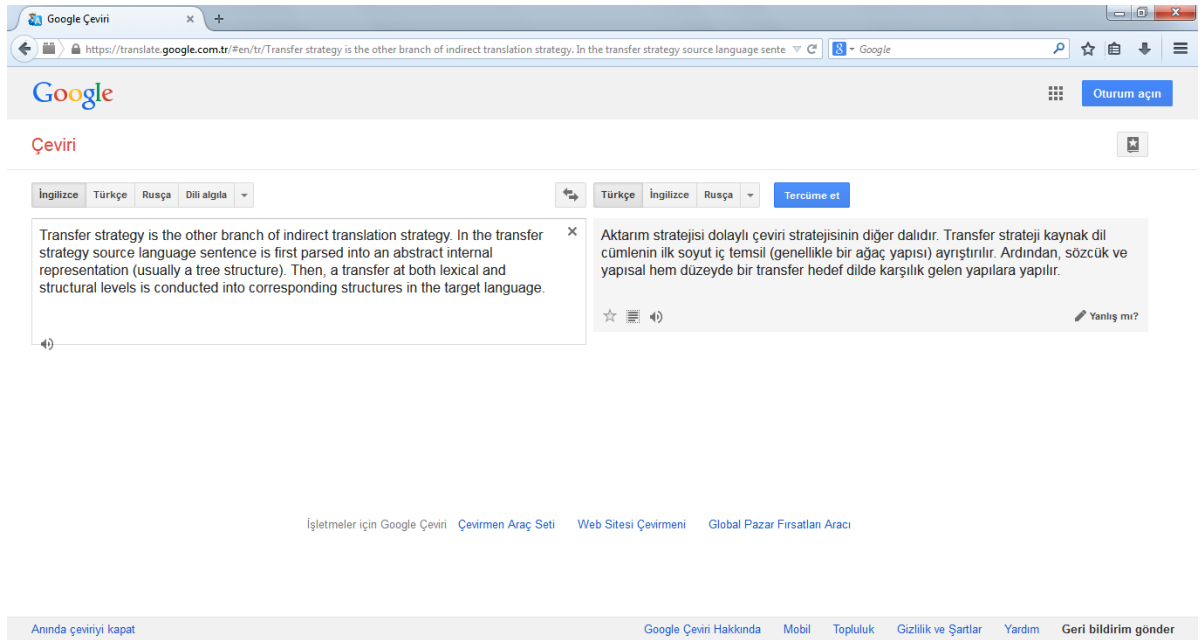


Figure 6: The user interface of Google Translate in February 2015.

#### 2.6.4. Bing Translator

Bing Translator is an online MT system launched by Microsoft Translator. The system can make translation between 45 languages. Likewise, Google Translate, Bing Translator is also statistical. Thus, Bing Translator is not included in the evaluation in this study.

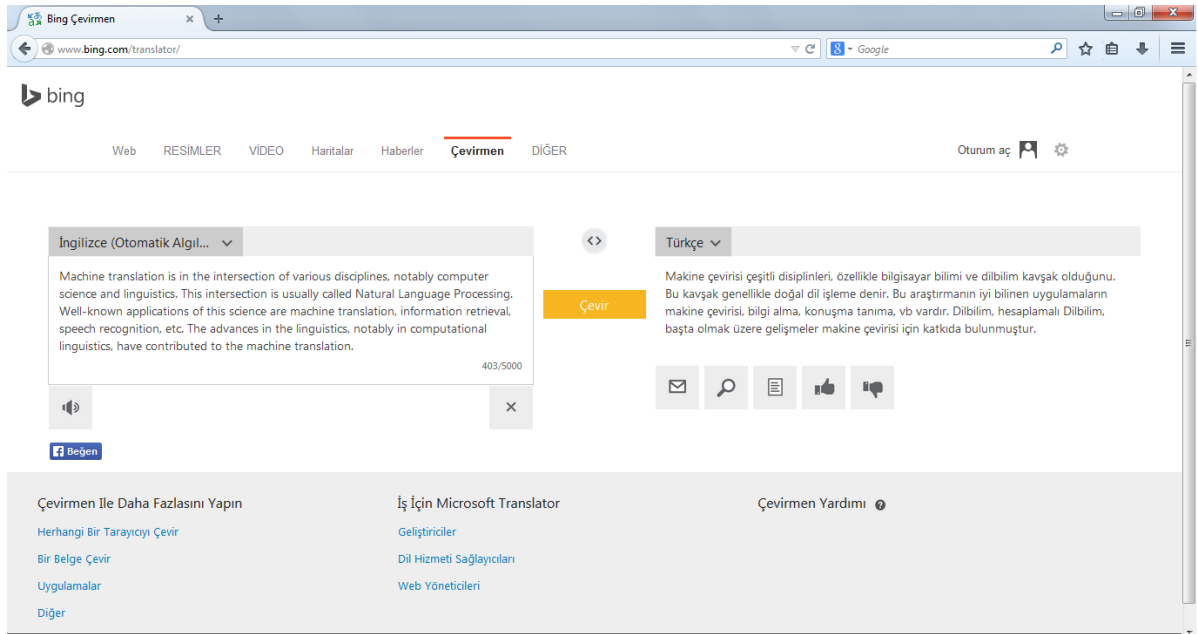


Figure 7: The user interface of Bing Translator in March 2015.

This chapter has dwelled on the historical development of machine translation, recent problems and issues in the area, together with the developments in Turkish MT. In the following chapter, state of art in machine translation is presented. The linguistic background as well as computational processes and methods are introduced as the theoretical framework of the research.

## **CHAPTER 3**

### **STATE OF ART**

This chapter lays the theoretical framework of the research. The linguistic background of machine translation is explained. Then the chapter dwells on the state of art solutions. In this context, processes, methods and other kinds of technological developments are explained.

#### **3.1. LINGUISTIC BACKGROUND**

Machine translation is in the intersection of various disciplines, notably computer science and linguistics. This intersection is usually called Natural Language Processing. Well-known applications of this science are machine translation, information retrieval, speech recognition, etc. The advances in the linguistics, notably in computational linguistics, have contributed to the machine translation.

Machine translation requires developing the understanding of the problems which can occur at all the levels of languages and producing appropriate solutions for them. The problems may be on word, sentence or text level. Below, these levels of language are explained starting from word level (morphology) to the text level (discourse).

##### **3.1.1. Morphology**

Morphology is the very first level of analysis for machine translation programs. It is the study of internal structure of the forms.

Computational morphology has many notable applications used daily. For instance, spelling correction in word processing programs such as Microsoft Word is a low-level

computational morphology application. Spelling correction uses root lexicon, as comparing input to a list of words would create a never ending job with the coinage of words and would take up so much memory space.

Morphological analysis has been a component of machine translation systems from the very beginning. It is considered easy for most of European languages but when agglutinative languages are involved in the system, the processes get complex. Inflected forms of the same words (paradigms) can be retrieved very easily by means of morphological analysis, which reduces the dictionary size and accelerates the computing time.

Words constitute the first analysis level for machine translation programs. The more words are compiled for that program, the higher the quality is achieved. Word compilation is thus a very important step of analysis for machine translation programs.

### **3.1.2. Lexicography**

Lexicography is the activity of compiling dictionaries. Lexicon is a list of the lexical items in that language. Dictionaries have been traditionally the largest component of the machine translation systems. The scope and coverage of the dictionaries directly limits the quality of machine translation output. For instance, absent words in the dictionary may lead to the untranslated words, which is very undesirable for commercial machine translation systems. However, end users may contribute to the machine translation systems by adding new words to the dictionary, which is also the case for CAT programs.

Creating automatic dictionaries is the first and most obvious task in a machine translation project. In the first machine translation systems, dictionary compilation had crucial importance. This is due to the fact that in the early days of machine translation, a mechanical replacement of source language words in the text with target language words was considered to be enough. However, even this compilation process was cumbersome due to the limited internal memory space of the computers of the time led machine translation researchers to look for external memory space, such as magnetic drum,



dielectric stores, vacuum tubes, and various other ways (Hutchins, 1986: 41). In order to reduce the dictionary size, early researchers included only stems and endings in the dictionary. In current machine translation systems, thanks to the advent of computers, the space problem has been resolved.

The quality of machine translation programs depends on the quality of their dictionaries. However, the extent of dictionaries is not the only deciding factor. In order to produce intelligible outputs, machine translation programs have to be supplemented with syntactic rules for analyzing the input sentences.

### **3.1.3. Syntax**

Syntax is the study of the rules or principles by which words can combine and form sentences. Syntax tries to account for all the grammatically correct phrases and sentences in whatever language it is analyzing.

Syntax has been one of the core research areas in the MT. First direct systems' failure is directly related to the lack of syntactic analysis of the SL (Hutchins, 2003: 163). However, in parallel with the developments in syntactic theory, MT has also developed and undertook a more comprehensive approach to syntactic analysis.

A qualified analysis of syntax enables machine translation systems to produce outputs which are of higher quality than word-for-word translation. However, this is not enough for many consumers of MT output. In order to attain FAHQMT goal, machine translation researchers have long understood the importance of "understanding meaning", which falls into the domain of semantics.

### **3.1.4. Semantics**

Semantics is the study of the meaning. It aims to model the way in which the meanings of lexical items contribute to the meanings of the phrases and sentences in which they appear.

Semantic features are analyzed to understand the meaning. For instance, a common feature for “boy”, “girl”, “woman” and “man” is “human”. This feature distinguishes “boy” from “lion”. Another feature like “male” distinguishes “boy” from “girl” or “man” from “woman”. By analyzing these features, a semantic feature hierarchy can be drawn. This type of hierarchy allows for generalizations. For instance, “human” and “animal” are categorized under “animate” heading. So, it can be inferred that “woman” under the heading of “human” is also “animate”. This generalizations can be extended to the verbs. For instance, the verb “talk” is specific to “human”. So, a “teacher” under the heading of “human” can “talk”, but a “bird” under the heading “animal” cannot.

A semantic representation of “boy”, “girl”, “man”, and “woman” in terms of semantic features can be (taken from Arnold et al. (1994).):

*man* = (+HUMAN, +MASCULINE, +ADULT)

*woman* = (+HUMAN, -MASCULINE, +ADULT)

*boy* = (+HUMAN, +MASCULINE, -ADULT)

*girl* = (+HUMAN, -MASCULINE, -ADULT)

This association also contributes to the syntactic analysis by imposing semantic constraints to the words. For instance the verbs “eat” and “drink” can only take “animate” agents and “edible” or “drinkable” patients. So, one cannot “eat” or “drink” “keys” or “honesty” because they are not edible or drinkable. In addition, another constraint for the context can be added for higher quality translation. For instance, adding a constraint like “economics” to the verb “supply” will direct the analysis process to the relevant dictionary term “arz etmek (supply)” in Turkish, rather than “sağlamak (provide)” which is also an equivalent but not for the field “economics”. By imposing these kinds of constraints, ungrammatical or unwanted analyses are dismissed. However, this imposition also rules out the metaphorically right analyses such as “This car drinks gasoline.”. In order to avert this problem, the compilation of a separate dictionary for metaphors and other figures of speech has been implemented as a solution (Hutchins, 1986: 42).

Semantics has been seen as the solution to the grave problems in first direct systems in MT. In order to move beyond the word for word translation, semantic-based MT projects were developed in parallel with the developments in Artificial Intelligence (Hutchins, 1986: 19). Interlingual MT systems made use of the analysis of semantic relations, which later became a standard procedure for MT systems. However, understanding semantic relations is not enough for FAHQMT, as texts are not just consisted of linguistic relations. Pragmatics, the study of meaning in context, is thus also an important part of MT research.

### **3.1.5. Pragmatics**

Pragmatics studies the meaning of linguistic messages in terms of their context. It is the study of how people comprehend and produce speech act in a concrete situation. According to Yule, “pragmatics is the study of “invisible” meaning, or how we recognize what is meant even when it isn’t actually said or written.” (2006: 128).The difference between semantics and pragmatics is their focus point. Semantics studies the linguistic meaning while pragmatics focuses on the contextual meaning. One famous example to distinguish semantics from pragmatics is the utterance “It is cold here”. In terms of semantic analysis, it states the low temperature in the given environment. In terms of pragmatic analysis, it may be a request for increasing the temperature by, for instance, closing the window or turning the heater on.

Pragmatics has been a core study area for computational linguistics with the advent of computational dialogue systems, especially spoken dialogue systems (SDSs). A very successful example of computational dialogue system is ELIZA, developed in 1960s. Using keywords, ELIZA program was able to simulate human-machine interaction to the point that some users took “DOCTOR” (a script in the program which simulates a psychologist) seriously (Melby & Warner, 1995: 147).

In the history of MT, pragmatic problems were once considered insurmountable, and FAHQMT was considered to be unreachable both in theory and practice, even for technical

texts. In his influential paper, entitled *The Present Status of Automatic Translation of Languages*, Bar-Hillel states the idea to supply MT systems with a universal encyclopedia is not something that should be taken seriously. Inference “is not at the disposal of electronic computer” (Bar-Hillel, 1959). The success of human translator lies in ability to understand the inferences as well as analyse the relations between sentences. Thus, MT systems also try to understand these relations by means of discourse analysis.

### **3.1.6. Discourse Analysis**

Discourse analysis studies above the sentence level; it focuses on language in use, the relation between language and context; and relationship between sentences in a text. Yule states “When we were concentrating on linguistic description, we were concerned with the accurate representation of the forms and structures. However, as language-users, we are capable of more than simply recognizing correct versus incorrect forms and structures.” (2006: 142). As language users, people are able to create discourse interpretations from fragmented linguistic messages, which cannot be understood by computers.

Discourse creates fundamental problems for MT. Without a proper anaphora resolution, MT cannot go beyond the boundary of single sentences. Anaphora resolution is especially a great problem for languages marking gender in pronouns. In addition, when Bar-Hillel argued that FAHQMT was not only practically, but also theoretically impossible (1959), he highlighted the real world knowledge, which cannot be understood by machines. However, with the advent of statistical machine translation and related corpora studies, many problems related to the discourse can be solved with statistical algorithms.

Machine translation is established on the foundations of linguistics, computer science and artificial intelligence. The researchers have been developing different processes, methods and creating resources to improve MT systems, which are explained below.

### **3.2. PROCESSES, METHODS AND RESOURCES**

Various processes, methods and resources are employed in the MT systems to produce high quality translations. These processes, methods and resources include text segmentation, part of speech tagging, parsing, word-sense disambiguation, anaphora resolution, controlled languages and sublanguages and corpus linguistics.

Like the first part of this chapter, second part also starts with word-level analysis and gradually continues to text-level analysis.

Let us first define some preliminary concepts which are going to be referred frequently. A computer, or electronic brain, consists of hardware and software. Hardware is physical units, such as screen, memory. It also includes input and output devices by means of which computer can communicate the outside world and humans. These input and output devices include scanners, keyboards, optical character readers, printers, etc. Software is a type of data, stored in the memory, which tells computer what and how to do. Software includes many programs, such as Microsoft Word, prepared for a special type of activity. These programs are written in logical programming languages. These languages enable human-computer communication. Programs can be written for various activities, such as word recognition, misspelling correction, etc. MT is a computer program which consists of many different programs for segmenting, tagging, parsing, etc. These programs are combined for the special needs of the MT program.

This part is devoted to explain these programs, starting from the first level of analysis, which is text segmentation.

#### **3.2.1. Text Segmentation**

Text segmentation is the initial stage of MT. Understanding and segmenting the sequence of characters into linguistics units such as numbers, words, punctuation, etc. is essential for the quality of MT. Errors in this stage may lead to more errors at the later stages of MT.

The segmentation process is called tokenization and segmented units are called tokens. Segmenting tokens is relatively an easier task in designing MT systems for most of the western languages as the tokens are delimited by blank spaces and punctuation, unlike many Oriental languages such as Chinese and Arabic, in which there are no explicit boundaries of tokens.

There are different ways to segment words and segment sentences. Blank space between tokens is the easiest way to understand the word boundary. Exclamation marks such as a period or question mark signals the sentence boundary. However, abbreviations and acronyms may also have periods between each letter. In addition, the different punctuation systems and hyphenated words may also contribute to the problems.

Abbreviations and acronyms with periods between each letter need to be taken as a token for the accurate translation. However, if the program accepts the period as the sentence boundary, the abbreviation and acronym in question are segmented as sentences. In order to overcome this problem, researchers in MT usually maintain a list of known abbreviations and acronyms (Mikheev, 2003: 205). Thus, the success of translation of abbreviations and acronyms is directly determined by the length and coverage of the list.

The main problem with hyphenated words is the ambiguity problem, that is, whether hyphenated segment is one word or two words. For instance, self-confidence should be segmented as a single token, but “Ankara-based” should be segmented as “Ankara” and “based”. If “Ankara-based” is segmented as “Ankara-based”, then at the later stages of MT, the system would be unable to find “Ankara” in the dictionary. Another problem is the end of line hyphenated words, which occur due to the formatting of the document. The problem is solved by reuniting the hyphenated parts and removing the hyphen, then looking up in the dictionary for the word. If the word is found in the dictionary as a single token, then it is segmented as reunited.

Problems with numerical expressions and other special expressions such as telephone numbers, dates, measures, punctuations are also handled at the text segmentation process. There are many cases where languages use different punctuation. For instance, in order to

indicate decimal point, Turkish uses comma (3,14) while English uses period (3.14). Another example is the different date formats of American English and Turkish. American English uses mm/dd/yy or mm/dd/yyyy format while Turkish uses dd.mm.yy or dd.mm.yyyy format (d: day, m: month, y: year). In addition to different places for day and month, two systems also use different punctuation for date format. In the text segmentation, these differences can be handled by writing a rule to convert them easily.

Sentence segmentation, or sentence boundary disambiguation (SBD), is the process of determining the sentence boundary accurately. As stated above, the period usually signals the end of the sentence. However, there are many cases where period may be the part of an abbreviation, an acronym or a series of numbers. In order to solve this, local context around periods and other punctuations are analyzed. In addition, determining end of the sentence may also contribute to identify proper nouns and common nouns. For instance, if a capitalized word is not preceded by period, then it is a high probability that the word is a proper noun.

The success and quality of MT is indisputably relies on the correct segmentation of words and sentences. English is the most researched language in terms of text segmentation, many SBD programs are written for English. Turkish, on the other hand, cannot enjoy the same status. However, within the framework of TUBITAK (Turkish Scientific and Technological Research Foundation) project 105E020 “Building a Statistical Machine Translation for Turkish and English” segmentation methods for Turkish are analyzed.

After the text segmentation has been completed, the second step in MT analysis is part of speech tagging.

### **3.2.2. Part of Speech Tagging**

Parts of speech have been first documented by Dionysius Thrax of Alexandria (c. 100 B.C.) (Jurafsky & Martin, 2006: 137) in Greek grammar book (techne). The parts he proposed were noun, verb, pronoun, preposition, adverb, conjunction, participle, and article. These

parts turned into the fundamental description categories for many natural languages. With the advent of computers and natural language processing applications, these parts started to be employed by researchers in computer science. The computational application for describing parts of speech in an electronic document is tagging. A tagger is software that associates the each word in the corpus with its appropriate part of speech (Bowker, 2003: 60). The tags given by tagger are described as tag sets and are tailored accordingly for the corpus. For instance, Brown corpus, assembled at Brown University, consists of 1 million words. An 87-item tag set was employed to tag the corpus, including punctuation, particle, modal, symbols, copula verb, etc. (Jurafsky & Martin, 2006: 148).

Part of speech tagging serves many important ends. First of all, the reliability of higher levels depends on tagging. Large tagged corpora are necessary for NLP applications. Many terms can be extracted from these tagged corpora. Once the tagging of corpus is finished, a special term extraction software matches tagged items with the specified combinations, such as noun+noun (Bowker, 2003: 60). By doing so, potential terms can be extracted from the corpus and dictionary compiling can be fastened. In addition, part of speech tagging can also contribute to ambiguity resolution. For instance, an article is more likely to precede a noun than a verb. Thus, when the tagger encounter the token “a convict”, “convict” is more likely to be tagged as noun. Tagged corpora also have a vital importance for developing statistical machine translation systems. Statistical knowledge derived from parallel corpora serve as the basis for this type of MT.

For Turkish, same ends and problems are true for taggers. Agglutinative nature of language allows for many interpretations of the same lexical item. For instance; “evin” can be interpreted as “(your) house)”or “of the house”. Thus, in order to determine the part of speech, tagger needs to look for contextual clues, such as a preceding possessive pronoun or following noun.

A special tagger for Turkish has been developed by Oflazer and Kuruöz (1994).

The quality of tagging mostly relies on the extent of the dictionary and training corpus of MT system. The tagged sentences are the input of the next analysis level, which is parsing.



### 3.2.3. Parsing

Parsing can be described as the process of assigning syntactic relations between lexical items (Melby & Warner, 1995: 25). Without syntactic knowledge representation, MT systems cannot go beyond word for word translation. Parser is the software that accepts sentences as input, analyzes them by means of grammar, and produces representations of syntactic knowledge. These representations can be in the form of brackets, charts and trees, and they are used for the subsequent processes (Petitpierre, 1987: 111). Grammar can be defined as the computational linguistic theory of language which indicates the acceptable sentences and/or phrases. Many systems use phrase structure rules as grammar. For instance, a simple phrase structure rule for English such as the following one will parse sentences which have lexical items to fill the specified categories:

$S \rightarrow NP VP$

$NP \rightarrow (DET) (ADJ) N$

$VP \rightarrow V (NP)$

Let us take “Jennifer loves cats.” and parse it by using the above phrase structure rule as grammar.

1.  $S \rightarrow NP VP$

2.  $NP \rightarrow (DET) (ADJ) N$

3.  $VP \rightarrow V (NP)$

4.  $N \rightarrow Jennifer$

5.  $V \rightarrow loves$

6.  $N \rightarrow cats$

This can be interpreted as follows:

1. A sentence consists of a noun phrase and a verb phrase.

2. A noun phrase consists of a noun, which may be preceded by a determiner and an adjective.
3. A verb phrase consists of a verb, which may be followed by a noun phrase.
4. An instance of noun for the noun phrase is “Jennifer”.
5. An instance of verb for the verb phrase is “loves”.
6. An instance of noun for the noun phrase is “cats”.

This can be schematically represented by the following parse tree:

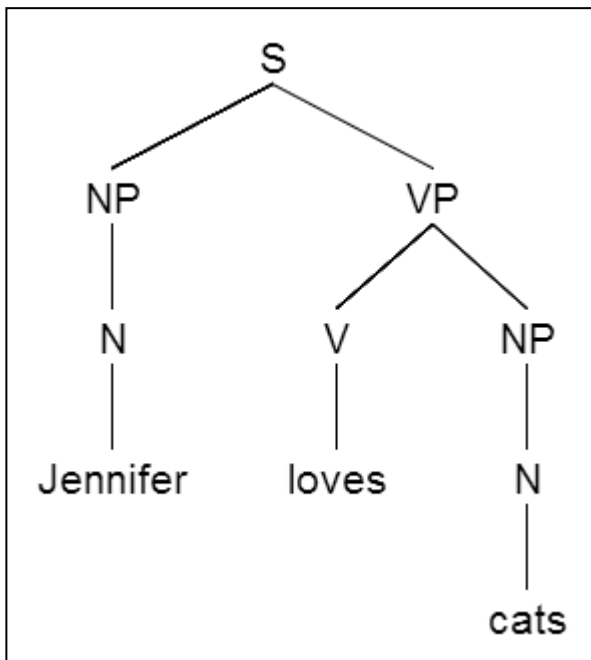


Figure 8: Parse tree representation.

Scientists working on developing Turkish parser have frequently mentioned free word order with explicit case marking and complex agglutinative morphology as the main problems (Bozşahin, 2002; Güngördü & Oflazer, 1995). However, these problems are overcome by the implementation of lexical functional grammar parser, which is stated to parse 82% of the sentences correctly (Güngördü & Oflazer, 1995).

After solving syntactic relations, MT systems analyze sentences for semantic relations, especially for solving ambiguity problems.

### 3.2.4. Word-Sense Disambiguation

Ambiguity can be defined as the situation in which a word, phrase or sentence conveys more than one meaning. This situation creates a bottleneck for many NLP applications. For MT, this situation becomes even a much graver problem, as two languages mean much more effort for researchers.

Process of identifying the meanings of words, phrases or sentences in a context is called word-sense disambiguation. It was first emerged as a distinct task in MT research, when the quality of first systems didn't meet the expectations and when research, according to Melby and Warner crushed into an insurmountable semantic wall (1995: 44).

However, there are some approaches for disambiguation. One of the earliest attempts for disambiguation was proposed by Wilks (1972). This approach is called preference semantics. In this approach, each sense in the lexicon has a formula associated with it which expresses its meaning. Some of the senses are HUMAN, ADULT, ABSTRACT, AGRICULTURAL PROCESS, etc. Disambiguation is carried out by choosing a formula for each ambiguous word. For instance, a dictionary entry for ball would resemble the following:

ball → concrete noun → SOCIAL ACTIVITY → ASSEMBLY → DANCE

ball → concrete noun → PHYSICAL OBJECT → SPHERE

ball → concrete noun → PHYSICAL OBJECT → CANNONBALL (Shann, 1987: 72).

In this approach, verbs in the lexicon are also attached the context expectations in terms of syntactic and semantic features. For instance, the verb “buy” requires a subject which is

ANIMATE and HUMAN, and an object which is not HUMAN. By this way, meaningful reading of a sentence is determined and ambiguity is resolved.

Another approach for word-sense disambiguation derives from the dictionary definition overlaps. This approach proposed by Lesk (1986) makes use of machine readable dictionaries. Dictionary definitions of neighboring or close words share similar meanings. For instance in the phrase “pine cone”, “pine” has two senses (evergreen tree and waste away from sorrow) and “cone” has three senses (solid body which narrows at a point, shape and fruit of evergreen tree). The sense both “pine” and “cone” share is “evergreen tree”. Thus, with a simple dictionary lookup, ambiguity is resolved.

A machine learning technique was also used for word-sense disambiguation. In order to resolve ambiguities for French-English statistical MT, Brown et al. (1991) studied on the Hansard, English-French parallel corpus consisting of proceedings of Canadian Parliament. Different senses of the same word are observed to be translated differently. For instance, “duty” is translated as “droit” when it means “tax” and “devoir” when it means “obligation”. In their research, Brown et al. (ibid) analyzed first and second words in the left and right of the ambiguous words. By doing so, a new way of disambiguation was created without the cost of hand tagging. However, this approach requires a well-aligned, bilingual corpus, which is not available for most of the language pairs.

Although English is a well studied language for word-sense disambiguation, with lots of suitable tools and corpora, Turkish cannot enjoy the same situation. Orhan and Altan list the agglutinative morphology and lack of resources such as language processing tools and annotated corpora as the main problems for Turkish word-sense disambiguation (2006). They studied on METU Corpus Project focusing on frequently used verbs such as “al-, bak-, çalış-, git-, gir-, çık- (take, look, study / work, go, enter, exit)” and nouns such as “ara, baş, el, sıra, yan, yüz (interval, head, hand, line, side, face)”. The accuracy of their disambiguator for verbs is 23 and 62 per cent and for nouns 35 and 75 per cent, for minimum and maximum values, respectively (ibid).

After solving problems within the sentence, MT systems then proceed to analyze and solve problems beyond the sentence. Anaphora, cataphora and other discourse features have to be understood in order to produce higher quality MT translation output.

### **3.2.5. Anaphora Resolution**

Anaphora is the act of referring back to a previously mentioned item in the text. By using anaphora, the author creates coherence. Understanding and translating anaphora are important in NLP applications, especially in MT, as without a proper understanding of the text; the quality of output is low. This is true especially in languages which mark gender and number such as French and Russian. Knowledge obtained from previous steps in the MT analysis, such as parsing, word-sense disambiguation is combined to resolve anaphora. For instance, antecedent can be determined from lexical information such as gender and number. By this way, many unsuitable candidates for antecedent are eliminated.

Pronominal anaphora is the most used anaphora type (Mitkov, 2003: 268). Personal, possessive, reflexive and demonstrative pronouns are varieties of pronominal anaphora in English. Turkish has six types of pronouns: personal, demonstrative, reflexive, possessive, interrogative and indefinite. It should be noted that Turkish marks plural in 2<sup>nd</sup> personal pronoun (sen, siz) and it doesn't mark gender in 3<sup>rd</sup> personal pronoun, which is the case for English (he, she, it).

An automatic anaphora resolution has three main stages. First of all, anaphors are detected. In this stage, non-anaphoric occurrences such as idiomatic expressions are detected and eliminated. Some of these idiomatic expressions are "it must be stated / underlined / etc." for English and "sözüm ona (seemingly), saat onda (at ten o'clock), etc." for Turkish. After these occurrences are eliminated, remaining anaphoric expressions are analyzed for locating antecedents. In this stage, all noun phrases in a certain scope are regarded as antecedent. The scope can range from one or two preceding sentences to previous paragraphs. Lastly, the program will attempt to compare anaphor and identified candidate

antecedents. In this stage, constraints and preferences are taken into considerations. Gender and number agreement are two obligatory constraints in selecting possible antecedents. In addition, semantic constraints such as animate, human, etc. are also obligatory constraints, which derive from the dictionary look-up. Preferences include recency (the fact that most recent noun phrase is most likely to be antecedent), emphasis, theme relations, etc.

Many programs and algorithms were developed for the anaphora resolution in English. Anaphora resolution for Turkish has also been studied and algorithms have been developed. In a study conducted on METU Corpus, Küçük and Turhan Yöndem reported to automatically identify pronominal anaphora with 98 per cent accuracy (2007).

Anaphora resolution has been one of the fundamental problems of MT, especially in the first systems. The success of these systems couldn't go beyond the isolated sentences. Most of the first systems were developed for Russian-English or English-French language pairs, and due to the gender and number marking in Russian and French, the quality of output was very low. However, with new algorithms developed for different languages, including Turkish, the success of anaphora resolution, thus overall MT, has increased. Nevertheless, before the advent of these algorithms, the researchers had to find other ways to resolve ambiguity and to increase the quality of MT systems. As a solution to the problems created by the natural languages, controlled languages and sublanguages have been proposed and employed successfully in MT research.

### **3.2.6. Controlled Language And Sublanguage**

A controlled language is a set of pre-defined restrictions of a natural language that imposes some constraints on lexicon, grammar and style. Several different controlled languages (CL) were developed for making technical language accessible to both non-experts and non-native speakers (Kittredge, 2003: 441). Some areas where controlled languages are extensively used are aerospace industry and telecommunication.

Controlled languages are employed to improve quality and uniformity of documentation. Reduced number of words and one to one correspondence between words lead to unambiguous texts, which are easier to understand for MT systems. Lexicon of controlled languages involves approved words, their syntactic category and examples of their use (Nyberg, Mitamura & Huijsen, 2003: 245). In addition, by limiting the number of words in a sentence and noun phrases, controlled languages eliminates complex sentence structures. Another advantage of controlled language is that the texts can easily be reused when appropriate due to the uniformity of style and lexicon.

However, there are some disadvantages of controlled languages both for technical writers and translators. Writers may find it difficult to conform to the rules of controlled language. In addition, writing with controlled language may reduce the power of expression. Translators may feel limited by the controlled language constraints. In the long run, however, the advantages of controlled language such as consistency, uniformity and reusability outweigh these disadvantages.

Sublanguage is another subpart of language. In contrast to controlled language, sublanguage is natural. Sublanguage is not imposed by a higher authority, but it occurs naturally. Sublanguage arises when experts communicate among themselves (Somers, 2003: 283). Lexicon of sublanguage is highly specialized. Apart from technical terms, everyday words may have different and specialized meaning, such as “mouse” as in computer science. In addition, sublanguage is consistent and complete in expressing the statements in the domain. Syntax can also be different from standard language. For instance, continuing in computer science, “Windows” is singular despite the plural allomorph “-s”, so, it behaves as a singular noun: “Windows is the best operating system.”. Many domains have their own sublanguage, spoken between experts who share common knowledge about domain, such as facts, assumptions, etc. Medicine, engineering, economics, etc. can be given as examples.

The applicability of controlled language and sublanguage to MT has been recognized by researchers as a solution to ambiguity resolution. Due to restricted lexicon and one to one

correspondence between words, many fundamental problems of MT such as word-sense disambiguation and ambiguity resolution, were easily overcome. In addition, according to Kittredge, languages share a greater similarity in sentence structure and text structure in scientific and technical writings than standard language (2003: 439). Thus, a greater success can be achieved when MT systems are designed for sublanguages or controlled languages. The reusability of texts in controlled languages also leads to a higher success in MT and CAT programs.

MT system with the highest success rate so far has been developed as sublanguage MT for the translation of weather reports from English into French. METEO system, developed by University of Montreal in 1974, and began full-time operation in May 1977 (Somers, 2003: 289). Since then, more than 30 million words translated and less than 5 per cent post-editing is required (Arnold et al., 1994: 150). METEO is a proof of the success that can be achieved when MT systems are tailored to the needs and peculiarities of languages and domains.

### **3.2.7. Artificial Intelligence in MT**

Artificial intelligence can be defined as a research area within computer science which aims to imitate intelligent human behaviour in computers. The application of artificial intelligence (henceforth AI) techniques to MT emerged from an influential report entitled *The present status of automatic translation of languages* published in 1959 by Yeshoshua Bar-Hillel, who was appointed to plan the future of MT at MIT (Hutchins, 1986: 33). In this report, Bar-Hillel pointed out the impossibility of FAHQMT even for scientific texts. He states (1959):

A human translator, in order to arrive at his high quality output, is often obliged to make intelligent use of extralinguistic knowledge which sometimes has to be of considerable breadth and depth. Without this knowledge he would often be in no position to resolve semantical ambiguities. At present no way of constructing machines with such a knowledge is known, nor of writing programs which will ensure intelligent use of this knowledge.



His famous example is a short sentence:

The box was in the pen.

The context is

Little John was looking for his toy box. Finally, he found it. The box was in the pen. John was very happy.

The two meaning of pen are “writing utensil” and “enclosure where small children can play”. According to Bar-Hillel, no program can understand and solve this homonym. He argued that semantic problems can be solved if computers have access to large encyclopaedias of general knowledge, which was, then, out of question.

After this influential report, MT researchers tried to incorporate AI techniques into MT systems to cope with semantic problems. Semantic parsing and script theory are two of these techniques used in Stanford University and Yale University systems.

Semantic parsing was developed by Yorick Wilks for Stanford University French-English MT system between 1970 and 1974 (Hutchins, 1986: 273). It is based on the recognition of semantic features either in patterns or in conceptual frameworks. An example of a pattern is MAN HAVE THING. So, “Ken has a car.” sentence will be analyzed semantically in this pattern, with semantic implication of OWNERSHIP (Shann, 1987: 77). In conceptual framework approach, system automatically looks up for semantically compatible items when a lexical item is occurred. For instance, conceptual framework for “drink” is as follows:

(((\*ANI SUBJ)((FLOW STUFF)OBJE)((\*ANI IN)((THIS(\*ANI(THRU PART)))TO)BE CAUSE))))))

This is interpreted as “an action, preferably done by animate things to liquids, of causing the liquid to be in the animate thing and via a particular aperture of animate thing; mouth of course” (Hutchins, 1986: 274).

Research in Yale University was based on the assumption that “Modelling human understanding of language requires the representation of meaning in terms of primitive semantic relationships which express not only what is explicit in the surface forms but what is also implied or can be inferred.” (Hutchins, 1986: 276). This requirement was fulfilled by means of “scripts” about what happens stereotypically in a car accident, in restaurants, in hospitals, during shopping etc.

The fundamental limitation of AI approaches to MT is extendibility. Application of AI to full-scale MT systems would require the formation of thousands of script or pattern, the number of which would be ever-expanding.

This part has been devoted to explaining the processes, methods and resources related to the MT systems. The next part dwells on the general system design of MT programs.

### **3.3. STRATEGIES**

MT has a long history with different approaches. These approaches are in parallel with the developments in computer science and computational linguistics. An inclination among MT researchers to refer to generations of programs started from the beginning of the research. However, the boundaries of generations are not clear cut. There are three generations for now. First generation of systems refers to historically oldest, word-for-word translation systems. Second generation consists of systems which incorporate analysis stages. Third generation refers to systems with semantic analysis stage. The last generation also consists of AI approaches (Hutchins, 1986: 334). In addition, it is possible to categorize the system according to their core technologies. Two core technologies have been developed so far; rule-based and corpus-based. Some MT systems can make use of these two core technologies at the same time, thus making themselves hybrid. In this part, these systems are analyzed in a chronological order to understand the problems and solutions in MT.

### 3.3.1. Rule -Based Systems

Rule -based systems systems are the very first MT systems to be developed. They require explicit formulation of grammatical rules and large dictionaries. It can be said that the better the grammatical analysis is conducted on the source and target languages and the larger the dictionary is, the higher the quality of machine translation output. There are three machine translation strategies developed under this core technology. These are direct, transfer and interlingua translation strategies.

#### 3.3.1.1. Direct Translation Strategy

Direct translation is historically the oldest strategy. In this strategy, each sentence of the source text is passed through a series of principal stages. The output of the previous stage is the input of the next stage. In this strategy, the aim is to go directly from source language to target language, with the minimal analysis of linguistic structure (Arnold, 2003: 123). The quality of output is at a word-for-word translation level. Tucker summarizes the stages of direct translation systems as follows:

1. Source text dictionary lookup and morphological analysis
2. Identification of homographs
3. Identification of compound nouns
4. Identification of noun and verb phrases
5. Processing of idioms
6. Processing of prepositions
7. Subject-predicate identification
8. Syntactic ambiguity identification
9. Synthesis and morphological processing of target text
10. Rearrangement of words and phrases in target text (1987: 23).

Many of the direct translation systems relied on large dictionaries and extensive post-editing for acceptable results (Hutchins, 1989: 4). The dictionary comprised of all lexical and syntactic information required including a table of all the specific syntactic ambiguities (Vauquois, 2003: 334). The quality of translations produced by these systems is limited. Most direct systems were firstly developed for Russian-English language pair due to the political reasons. Classical example for these systems is Georgetown system (for further information see 3.4.1). It is the first operational system which has translated hundreds of pages from Russian into English since 1970 (Tucker, 1987: 29). However, the limitations of direct translation systems, such as high dependence of post-editing and inability to go beyond word-for-word translation, have resulted in the emergence of transfer systems.

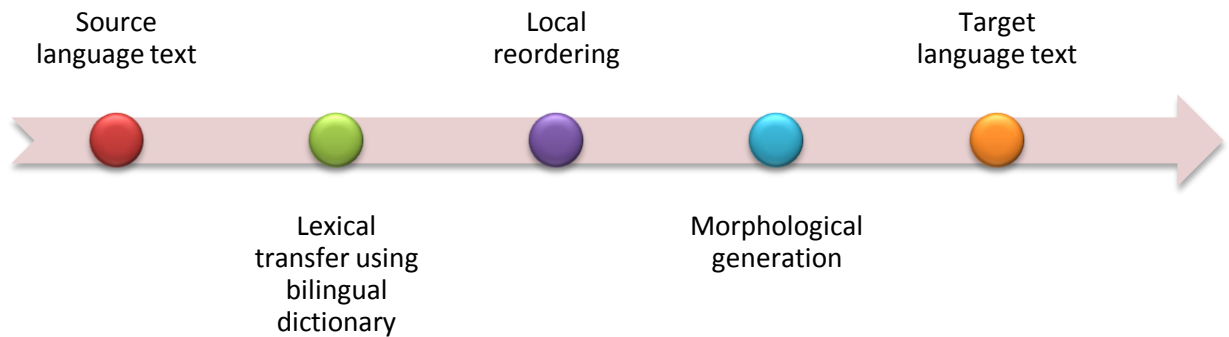


Figure 9: Direct machine translation work flow.

Due to the limited quality of direct translation strategy, the researchers tried to separate dictionary and grammar data. The resulting strategy has been known as indirect translation strategy which has two branches, interlingua and transfer.

### 3.3.1.2. Interlingua Translation Strategy

In interlingua translation strategy, source language texts are converted into semantic-syntactic representations common to more than one language. Interlingua is an old notion dating back to 17<sup>th</sup> century. Descartes and Leibniz have suggested creating dictionaries based on universal numerical codes (Somers & Hutchins, 1992: 5). Universal language was seen as a tool of communication eliminating the misunderstandings. In the later years, artificial languages for international use were developed, of which Esperanto is the best known. Warren Weaver, who has been credited as the founder of MT, has also suggested the idea of universal language in his famous memorandum, which has initiated the research and funding in MT. He writes (Weaver, 1949):

Think, by analogy, of individuals living in a series of tall closed towers, all erected over a common foundation. When they try to communicate with one another, they shout back and forth, each from his own closed tower. It is difficult to make the sound penetrate even the nearest towers, and communication proceeds very poorly indeed. But, when an individual goes down his tower, he finds himself in a great open basement, common to all the towers. Here he establishes easy and useful communication with the persons who have also descended from their towers (p. 17).

These ideas were later supported by Chomsky's deep vs. surface structure notions, according to which languages share a common "deep structure" while differing greatly on "surface structure" (Hutchins, 1986: 176).

In the strategy, translation is in two stages, first from source language to interlingua, second from interlingua to target language. The advantage of this approach is in the fact that the procedures of source language are not oriented towards any target language, but only towards interlingua. Thus, in multilingual MT systems, adding a new language would take less effort.

The main drawback of this approach is the complexity of interlingua itself. In addition, interlingua approach requires the resolution of all ambiguities of SL text, which may not be necessary for translation (Hutchins, 1996: 56). For instance, Turkish makes a distinction

between elder brother and little brother, as well as between elder sister and little sister. While translating from English into Turkish, this ambiguity has to be solved. However, French doesn't make such a distinction, but the interlingua still requires this ambiguity resolution for English-French translation. In addition, due to the limited expressive power of representation theories, interlingua translation strategy was later found too ambitious for its time (Knowles, 1982: 29).

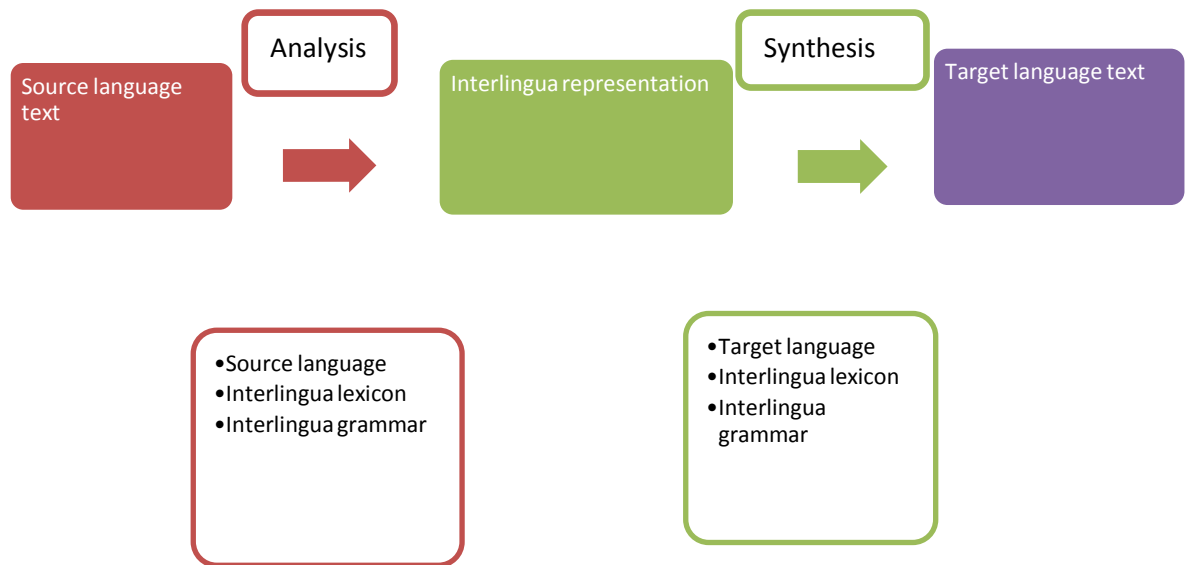


Figure 10: Interlingua machine translation work flow.

After seeing the low-prospect of interlingua approach, the researchers has been prompted by the other indirect strategy, which is transfer strategy.

### 3.3.1.3. Transfer Translation Strategy

Transfer strategy is the other branch of indirect translation strategy. In the transfer strategy, source language sentence is first parsed into an abstract internal representation (usually a tree structure). Then, a transfer at both lexical and structural levels is conducted into corresponding structures in the target language. Then, the translation is generated. A source language dictionary, a bilingual transfer dictionary and a target language dictionary are

compiled for transfer strategy. The difference between direct and transfer translation strategies is that, the previous one employs no structural analysis while the latter uses heavy structural and lexical analysis for higher quality MT output. The strategy has been adapted by many important MT projects, such as EUROTRA developed for European Commission, and TAUM METEO which is one of the most successful MT systems.

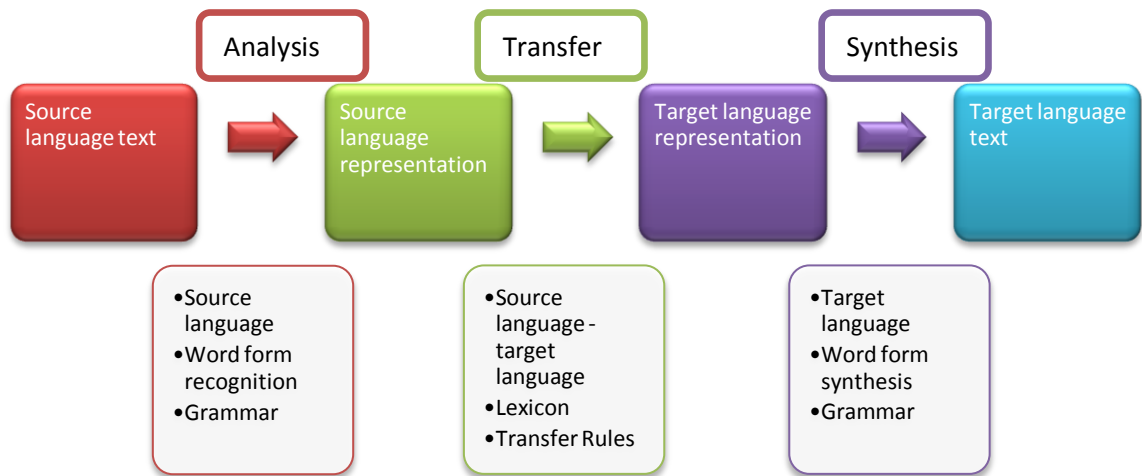


Figure 11: Transfer machine translation work flow.

The advantage of transfer strategy over interlingua strategy is the difference of ambiguity resolution requirements of the strategies. Transfer strategy requires ambiguity resolution only for languages involved. However, there are also some shortcomings of the strategy. A bilingual source language and target language dictionary has to be compiled for each new language added to the system, thus requiring more effort and investment. In addition, analysis and synthesis are limited to single sentences. As a result, semantic and pragmatic analysis is not available. The lack of these levels of analyses has given rise to new strategies which use texts previously translated by human translators. By doing so, the researchers try to move beyond the sentence and syntax level.

### 3.3.2. Corpus-Based Strategies

Two new strategies have recently emerged in an attempt to overcome the shortcomings of aforementioned strategies. With the advent of statistical approaches in NLP and the availability of bilingual corpora, research in MT has focused on using statistical approaches on these corpora to take advantage of previously translated material. These new strategies are Example-Based Machine Translation and Statistical Machine Translation.

Before explaining these approaches, it is necessary to define and describe corpus. Corpus is a large body of linguistic evidence of language use. By large, it is meant to be over several millions of words. A corpus may consist of everyday conversations, news, etc. It needs to be representative; samples in the corpus should include many different text types by different language users. (McEnery, 2003: 449).

Corpora may be compiled as either monolingual, comparable or parallel. Monolingual corpus represents one language. Comparable corpora are two or more monolingual corpora with a similar sampling frame, including same text types. Parallel corpora are the ones used for corpus -based machine translation. They include texts in one language with their translations in either one or more languages. Hansard corpus, consisting of documentation Canadian parliament proceedings in English and French, and Europarl corpora consisting of translations of Union's documents in all official languages of the European Union, are examples of parallel corpora. For Turkish, monolingual METU Turkish Corpus and National Turkish Corpus have been compiled. METU Turkish Corpus consists of over 2 million words. It represents 10 different genres. One part of the Corpus has been annotated for further analysis ("Metu turkish corpus" ). National Turkish Corpus consists of nearly 50 million words. It represents 5 different text type and 9 subject areas. Transcribed speeches are also included in the corpus ("Amaç," 2015).



### 3.3.2.1. Example-Based Machine Translation

Example -based machine translation (henceforth EBMT) is a corpus -based approach to MT. For developing an EBMT system, a bilingual corpus consisting of domain specific texts is necessary. This corpus constitutes the knowledge base of system. Work flow of an EBMT has three stages. In the first stage, a matching algorithm looks up in the bilingual corpus for the most similar examples to the input sentence. Then, by means of an alignment algorithm, input sentence is rebuilt in the target language in accordance with the most similar example. In the last step, input sentence is recombined in the target language in terms of syntax. In other words, bilingual corpus gives a translation template, which can be filled in by word-for-word translation (Arnold et al., 1996: 196).

An example can illustrate the process better.

*Input: I eat spaghetti.*

*Matches: I eat waffle.*

*Ben waffle yerim.*

*She cooks spaghetti.*

*O spagetti pişirir.*

*Result: Ben spagetti yerim.*

This approach has many advantages. For instance, the system is bidirectional in nature; same algorithms are used for the translation from and into the both languages. In addition, no dictionary compilation is necessary, which is mostly considered to be the most expensive and time consuming step in MT system design (Somers & Hutchins, 1992: 320). Furthermore, bilingual corpus can be compiled in accordance with the specific needs of users, thus a higher quality of output can be achieved.

The principal downside of the approach is the lack of bilingual corpora for most of the languages. In addition, system requires the annotation of bilingual corpus, which is also

difficult for many languages due to the unavailability of parsers or low quality of existing ones.

Different EBMT systems have been developed for different languages, such as Japanese and English. In addition, an EBMT system was developed for Turkish by N.Deniz ÖZTÜRK in 2007. ORHUN EBMT System (named after the first Turkish epigraph) uses MS Office Help Documents as bilingual corpus.

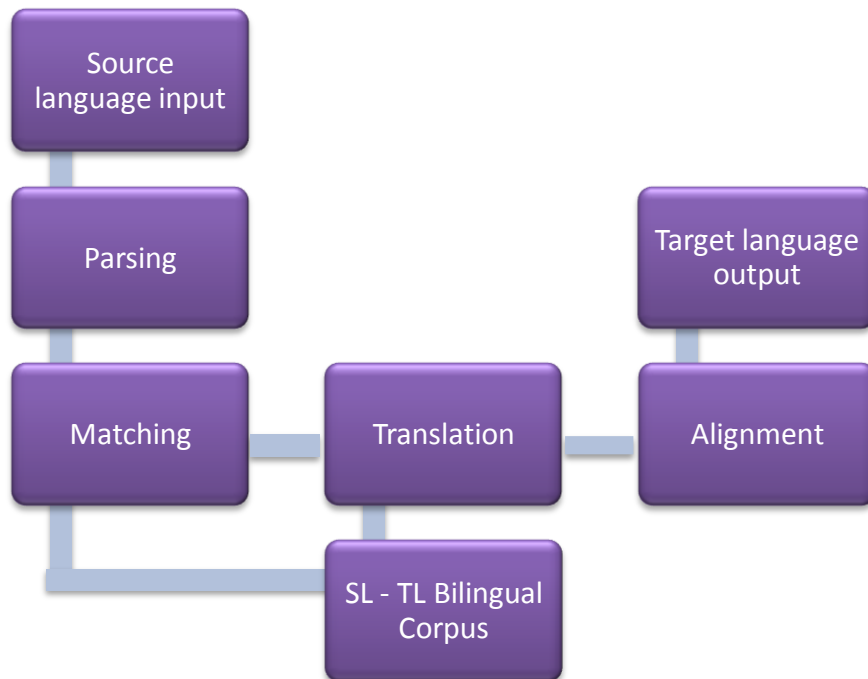


Figure 12: Example-based machine translation work flow.

The advent of statistical methods and large bilingual corpora has led to another corpus - based approach, which is statistical machine translation.

### 3.3.2.2. Statistical Machine Translation

Statistical machine translation (henceforth SMT) is a rather new approach to MT. By using large bilingual corpora, this approach aims to translate automatically without linguistic data. An SMT system has two stages. First, system aligns every lexical item (from words to

sentences) in the parallel bilingual corpus, and then calculates the probabilities of correspondence between aligned words, phrases and sentences. The approach was devised by IBM research group in early 1990s (Somers, 2003: 516). Machine translation systems, which have access to large multilingual or bilingual corpora, such as Google Translate, have implemented statistical methods as their basic strategy. The increasing number of parallel, bilingual or multilingual corpora has contributed to the success of SMT. In addition; multilingual webpages are another source for SMT. It can be said that the more bilingual, aligned texts are fed into an SMT, the higher the quality of output gets. This is the reason for the higher quality of SMT between many European languages even in free online SMT systems such as Google Translate. The multilingual policy of many European and American organization has resulted in many parallel multilingual or bilingual corpora such as Hansard (Canadian Parliament Proceedings in French and English), Europarl (consists of official languages of European Union) and many others. However, bilingual corpora for non-European languages, such as Turkish, are hard to find and compile. This results in lower quality in SMT for these languages.

### **3.3.3. Hybrid Methods**

Hybrid methods emerged when single approaches failed to give required MT output quality. The systems combine rule-based, example -based and artificial intelligence methods to solve problems which cannot be solved with a single approach.

This part has explained the overall design of MT systems. In the next part, milestone MT systems, which have been considered as the most successful ones, are presented with their historical and computational backgrounds.

### **3.4. MILESTONE MT SYSTEMS**

In the long history of MT, many research groups have been set up in universities and private laboratories. However, most systems couldn't go beyond experimental stage. The successful ones, on the other hand, are employed by the large international organizations and companies. In this section, MT systems, which have been historically and commercially proven successful, are analyzed in terms of their emergence, system structure, end-users, etc.

#### **3.4.1. Systran**

Systran has been regarded as the living proof of possibility of MT (Wilks, 1992: 166). It is one of the oldest and commercially most successful systems.

Systran was developed by Peter Toma, who was principal programmer of Georgetown University Experiment, of which success started the flow of funding to MT research (Somers & Hutchins, 1992: 175). After setting up his own company, Toma developed and sold an MT system (SYSTRAN) to the USA Air Force to translate from Russian into English (Wheeler, 1987: 192). After this success, Toma sold a version of SYSTRAN to Commissions of European Communities. New pairs of languages were incrementally added to the system, from English into French, German, Italian, Russian, Spanish, Portuguese, Chinese, Korean, Arabic, Danish, Dutch, Finnish, Norwegian, Swedish, Greek, and Polish ("Systran 7 premium translator," ). Other important users of Systran are NASA, American Navy, General Motors, Xerox Corporation. In addition, Systran is available online for many language pairs. The majority of translated text is for information gathering and the reported estimated error is 5 per cent for all the system (Aref, Al-Mulhem & Al-Muhtaseb, 1995).

In the beginning, SYSTRAN was described as a direct translation system, which was highly criticized for lacking a linguistic theory (Wheeler, 1987: 192). After more than 40 years, SYSTRAN is now a hybrid MT system, incorporating statistical methods ("Corporate profile,").

### **3.4.2. Meteo**

METEO is an MT system developed by University of Montreal to translate weather bulletins from English into French. It is considered as “advertisement” of the success of MT (Somers & Hutchins, 1992: 220).

The development of METEO emerged from the bilingual policy of Canada which requires every official document to be published in two official languages, English and French (Nirenburg, 1987: 12). The system began full time operation in 1977, and continued to translate weather bulletins until it was replaced by another MT system in 2001.

METEO system is usually described as a direct translation system. The lexicon of system consists of 1, 500 entries, half of which are city and geographic names. The most striking feature of the system is its use of sublanguage. With comparatively low semantic and syntactic ambiguity, system has a success rate of 97 per cent in translating five million words annually (Tucker, 1987: 31). The system has been developed with close cooperation with the translators of Canadian Weather Service, who used to have low job satisfaction due to the repetitive and boring nature of translating weather reports (Somers & Hutchins, 1992: 220).

METEO is now working on an SMT system by experimenting on the large corpus of previously translated weather bulletins (Gotti, Langlais & LaPalme, 2014).

### **3.4.3. Eurotra**

Eurotra is considered as a milestone MT project by many researchers in the area (Hutchins, 1986: 271; Tucker, 1987: 34; King and Porschke, 1987, 373). It is a multilingual MT system developed for European Commission. The multilingualism policy of European Commission created huge problems in the administrative processes, each document needed to be translated into official languages of EC, which were nine (Spanish, Danish, German,

Greek, English, French, Italian, Dutch and Portuguese) in 1978. The translation cost is reported to be 35 to 65 per cent of operational expenditure (Eurotra, 1990). The inability to improve Systran into a multilingual MT system led to the development of Eurotra project. Eurotra had two important aims. First and foremost was the creation of a prototype machine translation system capable of dealing with all Community languages (Hutchins, 1986: 264). Other aim was the creation and support of expertise in MT across Europe (Eurotra, 1990).

Starting from the inception, Eurotra was a collaborative project. The research was spread among universities all across Europe. Countries were responsible for the development of linguistic and computational processes relating to their own language.

The project discontinued in 1992. The result was a state-of-art prototype MT system. In addition, the project contributed to MT and NLP research in European languages. The compiled terminologies and corpora, together with the language-specific parsers, taggers, etc., have greatly contributed to other nascent MT and NLP projects for European languages (Maegaard, 1995).

#### **3.4.4. Google Translate**

Google Translate is a free, online translation service developed by Google Inc. Founded on the rule-based technology of Systran; Google Translate later prepared its own SMT system in 2007 and the service is now available in 80 languages (Mauser, 2014). Turkish was added to the supported languages in 2009. The system uses multilingual parallel corpora consisting of previously translated books, web pages, UN documents, etc. to extract translation (Tanner, 2007). Google Translate uses bridge languages for higher quality word alignment. English is the bridge language for most of the languages, that is; most languages are first translated into English, then into required output language. For some languages, system uses other European languages as a bridge before the translation into English. For instance, Belarusian input is first translated into Russian, then into English and then into the

required output language. Google Translate is also available for smart phones. The official application can use photograph, voice or handwriting as input.

Apart from Google Translate, Google Inc. also provides a free, online CAT tool, Google Translator Toolkit. Toolkit uses Google Translate SMT system. The translations completed in Toolkit are automatically fed into the translation system to improve overall quality (Wu, 2012).

This chapter has presented the state of art in MT research. Linguistic problems, together with their solutions in computational linguistics have been outlined to understand how an MT system works. In addition, different MT system designs are introduced to explain the workflow of MT. Lastly; MT systems considered as “milestone” in the MT history are given in a chronological order. One of the central research topics of both machine and human translation, that is the evaluation of translation output, is elucidated in the next chapter for developing an understanding of the subject which has proven itself complex.

## **CHAPTER 4**

### **TEXT TYPES AND EVALUATION TYPES**

#### **IN MACHINE TRANSLATION**

This chapter presents different evaluation types used in machine translation. In addition, text types by Katherina Reiss (2000) are explained in order to provide theoretical background of text sampling used in the study.

#### **4.1. EVALUATION TYPES**

Evaluation of MT has been one of the central problems in the discipline. It is a *sine qua non* for researchers, funders, buyers and end-users. The problem of defining a “good” translation is even a problem *per se* for translation scholars. Thus, defining a “good” MT output is even more problematic.

Historically, evaluation of MT was a newsworthy event. In the very first days of MT, the capability of computers to do human activity was both scary and miraculous for people. Amazing promises of computers to improve lives of human were published and broadcasted nationally in the USA (White, 2003: 212). After Georgetown University experiment conducted in 1954 raised public awareness of MT and started public funding, a large scale evaluation activity was organized in 1966. This first evaluation, known as Automatic Language Processing Advisory Committee (ALPAC) report was a comparative evaluation of then available, publicly funded MT systems. The report used intelligibility and fidelity as variables and made a number of suggestions related to the future of MT and all NLP studies. The suggestion of Committee to shift MT funding to development of AI and CAT tools ended MT research for nearly three decades (Melby & Warner, 1995: 31).



Commercial importance of MT evaluation emerges from the fact that developing, maintaining, and procuring an MT system is both expensive and time-consuming. Researchers, funders and end-users need to evaluate the system to understand whether their money, time and effort worth the end product. Researchers also need to evaluate the system constantly before moving it beyond the laboratory.

The lack of consensus about the ideal translation is a problem per se for MT research. Same text can be translated differently by different human translators. There are variables such as fidelity and intelligibility which can help compare the end products. MT researchers also take advantage of these and many other variables, such as engine performance, error analysis, etc.

White outlines six types of MT evaluation, which address to different needs of MT buyers, users and funders (White, 2003: 222).

Feasibility test makes the evaluation as to whether the realization of the idea is possible or not. Georgetown Experiment (1954) is an example of feasibility test for overall MT in the USA. By implementing a new linguistic or computational approach on a small scale experiment, researchers try to provide funding for their project. This type of evaluation is important for investors, who will financially support the project, and for other researchers, who will analyze and compare new approach with their own approach.

Internal evaluation is conducted during the research and development stage of a project. The experimental systems are usually tested on a small corpus, the errors in structural and lexical levels are amended, and then the system is tested on a larger corpus. By doing so, researchers test the extendibility of their system.

Declarative evaluation, on the other hand, deals with the overall performance of system. By employing different variables, such as intelligibility and fidelity, the evaluation measures the ability of MT system to handle different, unconstrained types of texts. This evaluation is especially important to investors, end users and developers. ALPAC report is also an

example of declarative evaluation, the first evaluation activity of MT output, as mentioned above.

Usability evaluation measures to ability of MT system to handle the needs of end-users. Two important variables, utility and satisfaction of user groups are measured.

Operational evaluation is the measurement of cost and benefits of MT system. Cost-effectiveness is measured with the overall cost of human translation. This evaluation is important for buyers of the MT system.

Comparison evaluation measures the performance of an MT system against other MT systems. By comparing different MT systems, buyers can purchase the one which fits to their special needs before procurement, as most MT systems have demo versions, which can be downloaded for free.

Another important distinction is made between glass-box and black-box evaluation, which are used synonymously with micro evaluation and macro evaluation.

Black-box or macro evaluation focuses on the output of MT systems, without taking the translation engine or design into consideration. It aims to describe the performance of MT systems by analyzing output in terms of adequacy (White, 2003, 217). In addition, it is employed in comparison of quality between different MT systems. Black-box or macro evaluation needs to employ different criteria to be adequate and powerful. Various criteria have been developed for the macro evaluation of MT systems. Some of these criteria are intelligibility, fidelity, coherence, usefulness, acceptability, reading time, correction time and translation time as well as automatic metrics. Intelligibility and fidelity have been widely used in many comparative evaluation and various methods have been developed for measuring them (Van Slype, 1979: 57; Callison-Burch et al., 2008). They have been proven the most cost-effective way to analyze and measure MT output quality. Intelligibility has been defined in various ways. It has been defined by Halliday as the ease with which a translation can be understood (Halliday in Van Slype, 1979: 62). A more objective and widely-recognized definition states that intelligibility is related to the grammatical errors,

mistranslations and untranslated words (White, 2003: 216). Higher intelligibility means less post-editing, while lower intelligibility scores mean rewriting the sentence for post-editors. Fidelity has been defined as measurement of meaning preservation in the output sentence (White, 2003: 216). It is the measurement of the correctness of the information transferred from the source language to the target language. Fidelity and intelligibility are closely correlated. Various methods have been proposed for measuring intelligibility and fidelity such as cloze-tests, noise tests, multiple choice tests, rating and ranking. The most cost-effective ones have been proven to be ranking on a five-point scale and simply rating them (Van Slype, 1979: 106; Callison-Burch et al., 2008).

In addition to fidelity, intelligibility, and other criteria, automatic evaluation metrics have been developed to overcome human subjectivity and non-reusability as well as to reduce the cost of large scale evaluation schemas. Various evaluation metrics have been developed for measuring quality. The common point of them can be said to be their dependence on reference translations produced by human translators. One of the most frequently used automatic evaluation metric is BLEU (Bilingual Evaluation Understudy) developed by Papineni et al. (2002). Its rationality is “The closer a machine translation is to a professional human translation, the better it is.” (ibid.).

Glass-box or micro evaluation focuses on improvability; it tries to understand how an MT system can be improved by analyzing the outputs. Grammatical error analysis, calculating post-editing rates and analyzing the causes of errors are the methods used in micro evaluation (White, 2003: 216; Van Slype, 1979: 116).

Grammatical error analysis is the identification and classification of individual errors in a machine translated text. This type of analysis reveals the strong and weak areas in the machine translation system. It can be used as a descriptive and comparative method for MT evaluation. Many large MT evaluation schemas, such as the evaluation of SYSTRAN by Commission of European Communities (Chaumier in Van Slype, 1979: 118), have employed grammatical error analysis. Different classifications have been proposed for different language pairs (Vilar et al., 2006 (for Chinese-English); Llitjos, Aranovich &

Levin, 2005 (for English-Spanish)); however, classifications independent of language pairs have also been proposed, such as Flanagan (1994). The categories and descriptions are presented below:

| <b>Error Category</b> | <b>Description</b>   |
|-----------------------|--|
| Spelling              | Misspelled word  |
| Not-found-word        | Word not in dictionary   |
| Capitalization        | Incorrect upper or lower case  |
| Elision               | Wrong elision or elision not made  |
| Verb inflection       | Incorrectly formed verb, or wrong tense  |
| Noun inflection       | Incorrectly formed noun  |
| Other inflection      | Incorrectly formed adjective or adverb   |
| Rearrangement         | Sentence elements ordered incorrectly  |
| Category              | Category error (e.g. noun vs. verb)  |
| Pronoun               | Wrong, absent or unneeded pronoun  |
| Article               | Absent or unneeded article   |
| Preposition           | Incorrect, absent or unneeded preposition  |
| Negative              | Negative particles not properly placed or absent   |
| Conjunction           | Failure to reconstruct parallel constituents after conjunction, or failure to identify boundaries of conjoined units |
| Agreement             | Incorrect agreement between subject-verb, noun-adjective, past   |

|                 |   |
|-----------------|---|
|                 | participle agreement with preceding direct object, etc.                     |
| Clause boundary | Failure to identify clause boundary, or clause boundary unnecessarily added |
| Word Selection  | Word selection error (single word)  |
| Expression      | Incorrect translation of multi-word expression                              |

Table 1: Error classification by Flanagan (1994).

Post-editing is the revision of machine output by a human expert before the distribution of the translated material. Higher intelligibility means less post-editing, while lower intelligibility scores mean rewriting the sentence for post-editors. Thus, rating post-editing has been an important part of MT evaluation (Van Slype, 1979: 25). Calculation of post-editing rate has been used extensively in MT evaluation schemas such as evaluation of SYSTRAN, ALPAC, etc. Various software and metrics have been developed for measuring correction rate. One of these measures, HTER (Human-mediated Translation Error Rate) has been proven to have a high correlation with human annotators (Snover et al., 2006). It is a measurement of edit-distance; that is, the fewest edits required to the MT output so that it can give the complete meaning of ST in the most fluent way. Each addition, deletion, modification to the MT output is counted as one edit.

As different evaluation methods have been developed for different needs and different stakeholders, quality expectations of the MT users may be different for various kinds of text. Below, a typology for text types in translation is presented.

## 4.2. TEXT TYPES BY REISS

Various types of texts are submitted to MT programs and the features of these texts directly affect the quality of MT systems. As a result of a survey for developing a multilingual corpus for machine translation evaluation, Elliott, Hartley & Atwell have found that texts which are most frequently submitted to MT systems are web pages, academic abstracts, newspaper articles, e-mails, tourist information, scientific and medical documents, business letters, user manuals and instruction booklets (2003). Many large scale evaluation schemas such as DARPA and annual EuroMatrix project have used newspaper articles as text sample for quality evaluation (Callison-Burch et al., 2012). Moreover, METU Turkish Corpus, which has over 2 million words chosen from 10 genres, include samples from news as well as novels, short stories, interviews, travel information and memoirs. In addition, subtitles have been previously studied and used as samples by many researchers such as Flanagan (2009) and Etchegoyhen, Bywood, Fishel, Georgakopoulou, Jiang, Van Loenhout, del Pozo & Sepesy Maucec (2014). The texts submitted to MT systems are usually 250-500 words-length (Knowles in Van Sylpe, 1979: 162).

Various text typologies have been proposed for categorization of translation strategies for different texts. Text types by Katherina Reiss (2000) have been widely accepted in the Translation Studies, thus this typology is chosen to provide the theoretical background for the evaluation of MT output.

In her book entitled *Translation Criticism – The Potentials & Limitations* (2000), Katherina Reiss focuses on the concept of equivalence. The foremost important aim of categorization is providing a framework for translation criticism. Just like translator, a critic also has to know what type of text in question before embarking upon criticism. The typological framework should be same for both the translator and the critic. She takes the three-way categorization of language functions by Bühler and develops her own theory focusing on the relationship between language functions and their corresponding language dimensions. Naturally, language functions are found in combination with each other in the text. However, the dominant one will be the deciding factor in determining the text type and

appropriate translation method. Reiss has a fourfold text typology which differentiates her work from the previous threefold typology frameworks (2000: 24):

First category of her typology is the content-focused text. In this text type, information is transmitted through logical or referential language dimension. It focuses on the plain communication of facts. Content or topic is the most important point in communication. News, lecture, reports exemplify this category. These text types are usually anonymous. Their aim is generally to provide information rapidly, accurately and comprehensively. Content-focused texts are assessed in terms of their semantic, grammatical and stylistic features, which are also reflected in their translation. Translation method should transmit referential content. "Plain prose" is the recommended translation method. In addition, additional information may be needed to make the translated text clearer and more target reader-oriented. For the translation to be regarded successful, the topic should be fully represented in the translation. In this research, the type of text chosen as sample for this category is academic abstract. Due to the globalization and the rise of English as the lingua franca, many researchers are now writing in English in order to attain a wider readership. In addition, machine translation has been promoted to produce high quality results when submitted technical texts consisting of little ambiguity. Thus, we have chosen academic abstract by Oulton (2013) on economy to observe whether machine translation can produce high quality results for unambiguous technical texts.

The second category of the typology is the form-focused text. Form is related to how an author expresses himself or herself, in contrast to the content, which is related to what an author says. In this text type, phonological as well as figurative elements, such as manners of speaking, proverbs, metaphor are frequently observed. The language dimension of this type is aesthetic, and it is form-focused. It expresses sender's attitude. All texts which express more than they state are in this category. Poem, play and biography are among the examples of this type. Recommended translation method is identifying method, translator should adopt the perspective of source text author. The translator should create equivalents through new forms, by which the text can achieve a similar effect. By doing so, translator can transmit the aesthetic form. Literary texts have always been a controversial category for

machine translation researchers. Some researchers such as Delanevay have mentioned their high hopes for literary machine translations (1960, cited in Hutchins, 1986: 151), while many others have stated that machine translation is not suitable for literary texts. However, although limited, there is a growing interest in machine translation community to go beyond the limits of technical texts, as indicated by the increasing literature related to the automatic poetry translation (Genzel, Uszkoreit & Och, 2010; Jiang and Zhou, 2008). The sample chosen for this text type is an extract from the famous story “Rocking-Horse Winner” by D. H. Lawrence. This short story extract is chosen in order both to see the quality of machine translation when a literary text is submitted as input and to compare the quality of different text types.

Third category of Reiss is the appeal-focused texts. These texts not only convey information in a linguistic form, but also present the information with a particular perspective, an explicit purpose, and it aims to result in a non-linguistic result. In other words, it aims to appeal to or persuade the reader to act in a certain way. The linguistic form is less important than the non-linguistic purpose of the message. Language function and the text focus are both appellative. Electoral speeches, advertisements, propaganda, publicity and sermons exemplify this type. Translator should create an equivalent effect by adapting. Target text should produce the desired response in the reader. The type of text chosen as sample for this category is tourist information (hotel advertisement). The language of hotel advertisements is characterized by the frequent use of adjectives and long sentences to describe the hotel and its facilities. Websites of many international hotels are translated into different languages such as Arabic, Russian, French, German, Chinese, Japanese, etc. in order to attract more tourists and to expand market potential. However, Turkish is frequently not included in the language options of websites, thus, Turkish speakers usually need to translate and understand English version of the website. The machine translation programs can be used for understanding these websites.

Last category of the typology is the audio-medial text. In audio-medial text, above three text types are supplemented with visual or audio texts, such as pictures or music. Audio-medial texts are dependent on non-linguistics media and on graphic, acoustic, and visual



kinds of expressions. Films, radio newscasts, advertisements with visual or audio embeddings are examples of this type. Translation of audio-medial texts requires supplementing written words with visual images and music. Translation must preserve the same effect on the hearer that the original has in the source language hearer. The type of text chosen as sample for this category is subtitle. Due to the globalization and language policy of European Union, many projects such as EU-funded SUMAT (an online service for subtitling by machine translation) have focused on machine translation for producing multilingual subtitles (Fraser, 2013). The language of subtitles are very distinctive from the previous text samples in that subtitles frequently involve proper names, colloquial language, and exclamations. Many popular British and American TV series have a high number of audience in Turkey. Different platforms on internet such as dizimag.com, dizist.com, etc. provide these TV series with translated subtitles. The episodes of many popular TV series such as *How I Met Your Mother*, *Dexter*, *Breaking Bad*, etc. are translated and subtitled in a very short time. Machine translation can be helpful in translating these texts. The subtitles we have chosen for audio-medial machine translation is from *The Big Bang Theory*, Season 13, Episode 1 “The Locomotion Interruption”.

The importance and difficulty of machine translation evaluation have been long acknowledged and resulted in various studies in the field. In addition, text typology has been studied in both translation theory and machine translation as an attempt to establish quality assessment principles. The following chapter presents the methodology of the research, which is itself another attempt to pin down the “quality assessment” concept for the machine translation between Turkish and English.

## CHAPTER 5

### METHODOLOGY

The methodology used in this study to evaluate machine translation output is further explained in this chapter by presenting information about the samples, annotators, test design, data collection instruments, test procedure, software and techniques used to analyze the related data.

#### 5.1. DESIGN OF THE STUDY

Descriptive and comparative methods are used in this study. Two different evaluation methods are conducted in order to get data related to the MT performances, which then are compared with each other.

#### 5.2. TEXT SAMPLES

Text samples are chosen in accordance with the text types by Reiss. The main criterion for choosing the samples is their “relevance” to the MT end-users (see 4.2).

| Text Type<br>Text Feature | Content-focused | Form-focused | Appeal-focused | Audio-<br>medial |
|---------------------------|-----------------|--------------|----------------|------------------|
| <b>Language Function</b>  | Informative     | Expressive   | Appellative    | Mixed            |
| <b>Language Dimension</b> | Logical         | Aesthetic    | Dialogic       | Mixed            |

|  |                                     |                                   |                                 |          |
|--|-------------------------------------|-----------------------------------|---------------------------------|----------|
| <b>Main Aim of Translation</b>           | Transmission of referential content | Transmission of aesthetic content | Elicitation of desired response | Mixed    |
| <b>Types of text chosen as sample</b>    | Academic abstract                   | Short Story Extract               | Online Advertisement            | Subtitle |
| <b>Number of sentences in the sample</b> | 12                                  | 25                                | 9                               | 28       |
| <b>Number of words in the sample</b>     | 265                                 | 260                               | 255                             | 259      |

Table 2: Text samples and their features.

### 5.3. DATA COLLECTING METHODS AND MATERIALS

Data of this study are collected from different sources.

#### A) Evaluation Sources

1) One source is human annotators, who rate and rank sentences translated by MT programs. For understanding and rating MT outputs, annotators need to have a high level command of English and Turkish. 20 senior students from Hacettepe University English Translation and Interpretation Department have participated in a survey which constitutes human evaluation part of the study.

#### B) Materials

A questionnaire which consists of three separate tests is prepared by means of onlineanketler.com, which is a web service that provides online surveys. Two of the tests are 5-likert type. One of the tests is 3-likert type. The questionnaire is sent to 20 students and the responses are comparatively analyzed.

## **5.4. TEST DESIGN**

The study provides a comparative macro evaluation and micro evaluation of three machine translation programs between English and Turkish to evaluate the quality of MT output from different aspects.

### **5.4.1. Micro Evaluation**

Micro evaluation focuses on improvability; it tries to understand how an MT system can be improved by analyzing the outputs. Grammatical error analysis, calculating post-editing rates and analyzing the causes of errors are the methods used in micro evaluation (White, 2003: 216; Van Slype, 1979: 116). This study has employed grammatical error analysis as the method for micro evaluation.

#### **5.4.1.1. Grammatical Error Analysis**

Grammatical error analysis is the identification and classification of individual errors in a machine translated text. This type of analysis reveals the strong and weak areas in the machine translation system. It can be used as a descriptive and comparative method for MT evaluation.

This study analyzes grammatical errors in the MT outputs within the framework of classification by Flanagan (1994) and compares MT systems in question. Proper names are analyzed and categorized under “not-found words” error category, as their translation indicates a problem with the dictionary features.

The test design for the analysis of grammatical errors is as follows:

- 1) Sample texts are translated by 3 different MT programs.
- 2) Outputs of MT programs are analyzed for grammatical errors within the classification of Flanagan (1994).
- 3) An error profile is drawn for each MT system.

4) Error profiles of each MT system is compared and contrasted.

### **5.4.2. Macro Evaluation**

Macro evaluation aims to describe the performance of MT systems by analyzing output and it has been used extensively in comparative MT evaluation schemes for measuring system quality. Various methods and criteria have been developed to measure system performance. Some of these are intelligibility, fidelity, usefulness, translation time, etc. In addition, in order to prevent human subjectivity and reduce the cost of human annotators, various automatic evaluation metrics have been developed.

This study employs intelligibility and fidelity criteria as well as ranking of sentences for human evaluation.

#### **5.4.2.1. Human Evaluation**

Human evaluation has been the main method for measuring MT output quality. The evaluation is conducted by means of tests which aim to measure one aspect of quality such as intelligibility, fidelity, usefulness, low translation time, etc.

Intelligibility has been defined by Halliday as the ease with which a translation can be understood (Halliday in Van Slype, 1979: 62). A more objective and widely-recognized definition states that intelligibility is related to the grammatical errors, mistranslations and untranslated words (White, 2003: 216).

Fidelity has been defined as measurement of meaning preservation in the output sentence (White, 2003: 216). It is the measurement of the correctness of the information transferred from the source language to the target language. Fidelity and intelligibility are closely correlated.

The human evaluation follows below steps:

- 1) Sample texts are translated by 3 different MT programs.
- 2) An online survey is prepared by using onlineanketler.com. Survey is composed of following parts:
  - a) Rating: The source text sentence is given together with the target text sentences produced by MT systems. Annotators are instructed to assign points in terms of how intelligible and faithful the target translation sentences produced by MT programs.

(Instruction for fidelity: Please compare the target text sentences with the source text sentence and assign how much of the meaning in the source text sentence is preserved in the target text sentences: 5 = All      4 = Most      3 = Much      2 = Little      1 = None).

(Instruction for intelligibility: Please compare the target text sentences with the source text sentence and assign how fluent the translation is:

5 = Flawless

4 = Good

3 = Non-native

2 = Disfluent

1 = Incomprehensible).

- b) Ranking: The source text sentence is given together with the target text sentences produced by MT systems. Annotators are instructed to rank the target text sentences (Instruction for ranking: Please rank each whole sentence translations from Best to Worst relative to the other choices.)

The 8<sup>th</sup> and 9<sup>th</sup> sentences of subtitle are omitted from the survey as they have been translated identically by three machine translation programs. (See Appendix for the complete survey).

- 3) Annotators are given a short training on intelligibility and fidelity.
- 4) A pilot test has been conducted with 10 annotators between 27 and 30 March, 2015.
- 5) After the analysis of results of pilot test, the survey for the real test has been prepared.
- 6) The survey for the real test has been sent to 20 annotators between 1 and 7 April, 2015.
- 7) The results are compared and contrasted.

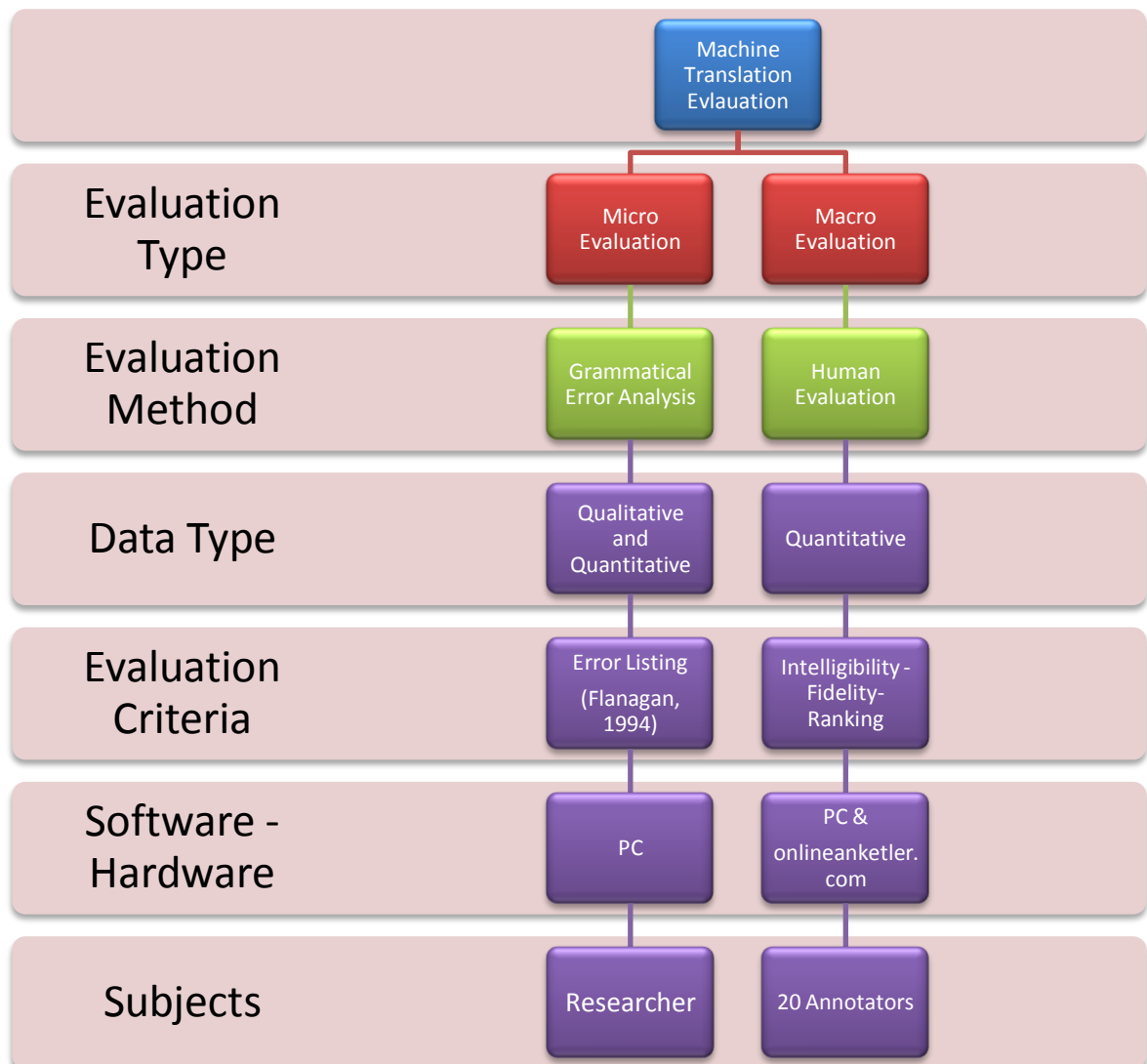


Figure 13: The test design of the research.

The methodology of the research, including research questions, samples and test design has been presented in the present chapter. The next chapter dwells on the presentation and discussion of findings via tables and charts.



## **CHAPTER 6**

### **FINDINGS**

In this chapter, the samples which are translated by three different machine translation programs are analyzed and compared in accordance with the methodology presented in the previous chapter. Micro and macro evaluations are carried out and findings are presented in tables and charts. The results are discussed in view of the research questions of this study.

In micro evaluation, the source text and three different translations produced by machine translation programs are analyzed within the framework of grammatical error analysis. The analysis is carried out on sentence level.

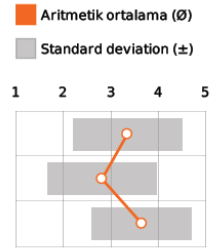
Macro evaluation of machine translation is conducted via a survey, which consists of three separate tests. The first and second tests consist of five-likert type questions and aim to understand the meaning preserving and grammaticality of machine translation outputs, respectively. The last test consists of three-likert type questions, which aims to find the ranking of machine translation outputs. The survey was completed by 20 senior students, who have cumulatively spent 25 hours. The survey and the results can be seen at CD enclosed at the end of thesis.

The results of three separate tests for a sentence are presented below:

2. The productivity performance of the UK economy in the period 1990-2007 was excellent.

Katılımcıların sayısı: 20

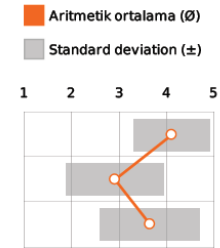
|                                   | All<br>(1) |      | Most<br>(2) |       | Much<br>(3) |       | Little<br>(4) |       | None<br>(5) |       | Ø    | ±    |
|-----------------------------------|------------|------|-------------|-------|-------------|-------|---------------|-------|-------------|-------|------|------|
|                                   | Σ          | %    | Σ           | %     | Σ           | %     | Σ             | %     | Σ           | %     |      |      |
| döneminde 1990-2007 yılında ...   | 1x         | 5,00 | 3x          | 15,00 | 8x          | 40,00 | 4x            | 20,00 | 4x          | 20,00 | 3,35 | 1,14 |
| Dönem 1990-2007'i'nde BK eko...   | 1x         | 5,00 | 10x         | 50,00 | 3x          | 15,00 | 4x            | 20,00 | 2x          | 10,00 | 2,80 | 1,15 |
| Belirli bir döneme ait 1990-20... | 1x         | 5,00 | 2x          | 10,00 | 3x          | 15,00 | 11x           | 55,00 | 3x          | 15,00 | 3,65 | 1,04 |



74. The productivity performance of the UK economy in the period 1990-2007 was excellent.

Katılımcıların sayısı: 20

|                                | Flawless<br>(1) |      | Good<br>(2) |       | Non-Native<br>(3) |       | Disfluent<br>(4) |       | Incomprehensible<br>(5) |       | Ø    | ±    |
|--------------------------------|-----------------|------|-------------|-------|-------------------|-------|------------------|-------|-------------------------|-------|------|------|
|                                | Σ               | %    | Σ           | %     | Σ                 | %     | Σ                | %     | Σ                       | %     |      |      |
| döneminde 1990-2007 yılın...   | -               | -    | 1x          | 5,00  | 2x                | 10,00 | 11x              | 55,00 | 6x                      | 30,00 | 4,10 | 0,79 |
| Dönem 1990-2007'i'nde BK ...   | 1x              | 5,00 | 6x          | 30,00 | 9x                | 45,00 | 2x               | 10,00 | 2x                      | 10,00 | 2,90 | 1,02 |
| Belirli bir döneme ait 1990... | -               | -    | 3x          | 15,00 | 6x                | 30,00 | 6x               | 30,00 | 5x                      | 25,00 | 3,65 | 1,04 |



146. The productivity performance of the UK economy in the period 1990-2007 was excellent.

Katılımcıların sayısı: 20

|  | Best<br>(1) |       | Moderate<br>(2) |       | Worst<br>(3) |       | Ø    | ±    |
|--|-------------|-------|-----------------|-------|--------------|-------|------|------|
|  | Σ           | %     | Σ               | %     | Σ            | %     |      |      |
| döneminde 1990-2007 yılında İngiltere ...  | 4x          | 20,00 | 8x              | 40,00 | 8x           | 40,00 | 2,20 | 0,77 |
| Dönem 1990-2007'i'nde BK ekonomisinin...   | 13x         | 68,42 | 6x              | 31,58 | -            | -     | 1,32 | 0,48 |
| Belirli bir döneme ait 1990-2007'de UK ... | 2x          | 10,00 | 6x              | 30,00 | 12x          | 60,00 | 2,50 | 0,69 |

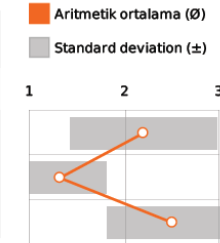


Figure 14: Results of three separate tests for a sentence.

The macro evaluation and micro evaluation of texts are pieced together, analyzed and summarized for each sentence via charts. Each chart presents, on the columns, original text and translations produced by Google Translate, Proçeviri and Sametran, respectively. Then, on the rows, the intelligibility, fidelity ratings, the rank of the sentence, the macro evaluation (summary for intelligibility, fidelity and ranking) and micro evaluation (grammatical error analysis) are presented.

Each text type is analyzed under a separate title. After macro evaluation and micro evaluation are carried out for all sentences in the text, the summary of two evaluations are presented.

### 6.1. CONTENT-FOCUSED TEXT TYPE

For the content-focused text type, the abstract of an academic article entitled “Medium and Long Run Prospects for UK Growth in the Aftermath of the Financial Crisis” (Oulton, 2013) is chosen. The 265 word-length text includes financial terminology, percentages and abbreviations.

The macroevaluation and micro evaluation of outputs of machine translation programs for each sentence are presented below:

| No.                     | Original   | Google Translate  | Proçeviri  | Sametran   |
|-------------------------|--|---|--|--|
| 1                       | The productivity performance of the UK economy in the period 1990-2007 was excellent.  | döneminde 1990-2007 yılında İngiltere ekonomisinin verimlilik performansı mükemmel. | Dönem 1990-2007'i'nde BK ekonomisinin üretkenlik performansı mükemmeldi. | Belirli bir döneme ait 1990-2007'de UK ekonomisinin üretkenlik yerine getirmesi, mükemmeldi. |
| <b>Intelligibility</b>  |  | Disfluent (55%)   | Non-Native (45%)   | Disfluent (30%)  |
| <b>Fidelity</b>         |  | Much (40%)  | Most (50%)   | Little (55%)   |
| <b>Rank</b>             |  | Moderate (40%)  | Best (65%)   | Worst (60%)  |
| <b>Macro evaluation</b> | The translation of Google Translate was considered to be disfluent. However, as the meaning was preserved more in this translation than in |   |  |  |

|                         |   |   |  |  |
|-------------------------|---|---|--|--|
|                         | <p>Sametran's translation, it was considered to be moderate.</p> <p>The translation produced by Proçeviri was ranked the best.</p> <p>The disfluency and low meaning-preserving of the translation by Sametran made it rank worst.</p>  |   |  |  |
| <b>Micro evaluation</b> | <p>In translating this short sentence, Google Translate made a capitalization mistake, in addition to verb inflection and arrangement mistakes.</p> <p>Furthermore, the abbreviation UK was translated as "İngiltere", which doesn't denote the same political entity. This can be regarded as a word selection error.</p> <p>Proçeviri produced the best translation. The sentence needs a rearrangement.</p> <p>Translation produced by Sametran also needs a rearrangement. In addition, the word selection for "performance" is wrong. The abbreviation "UK" was not translated, which is a not-found word error.</p> |   |  |  |
| <b>No.</b>              | <b>Original</b>   | <b>Google Translate</b>   | <b>Proçeviri</b>   | <b>Sametran</b>  |
| 2                       | <p>Based entirely on pre-crisis data, and using a two-sector growth model, I project the future growth rate of GDP per hour in the market sector to be 2.61% p.a.</p>   | <p>Tamamen kriz öncesi verileri ve iki sektör büyüme modelini kullanarak dayanarak, ben 2.61% pa olmak üzere piyasa sektöründe saat başına GSYİH gelecekteki büyüme proje</p> | <p>Tamamen önceden-kriz verisinde ve bir iki-sektör büyüme modelini kullanmak temel aldı, ben pazar sektöründe saat başına 2.61%'i p.a. olmak için GDP'in gelecek büyüme oranını tasarlarım.</p> | <p>Bir iki-kesim büyüme modelini tamamen kurulan kullanmak, ve ön-kriz verisinde , ben 2.61% p.a'sı olmak için Pazar kesiminde GDP saat başının gelecek büyüme oranını tasarlarım.</p> |

|                         |  |  |   |  |
|-------------------------|--|--|---|--|
| <b>Intelligibility</b>  |  | Incomprehensible (60%)   | Disfluent (30%)   | Incomprehensible (45%)   |
| <b>Fidelity</b>         |  | Little (40%)   | Little (50%)  | None (45%)   |
| <b>Rank</b>             |  | Worst (80%)  | Moderate (45%)  | Moderate 55%)  |
| <b>Macro evaluation</b> | None of the translations were chosen as the best translation. The ratings of intelligibility and fidelity are all very low for all machine translation programs.   |  |   |  |
| <b>Micro evaluation</b> | <p>Google Translate failed to inflect two verbs; “based” and “project”. In addition, the abbreviation “p.a” which stands for “per annum” was not recognized. Furthermore, the sentence needs a rearrangement. Although Google Translate was the only one MT program to translate the “GDP”, it was considered the worst translation.</p> <p>Proçeviri also made two verb inflection errors; “based” and “using”. The translation needs a rearrangement. In addition, the abbreviations “p.a” and “GDP” were not translated.</p> <p>The abbreviations were not recognized by the dictionary of Sametran, either. In addition, the expression selection for “based on” is wrong. The sentence needs to be arranged. Furthermore, the verb inflections of “be” and “using” are wrong.</p> |  |   |  |
| <b>No.</b>              | <b>Original</b>  | <b>Google Translate</b>  | <b>Proçeviri</b>  | <b>Sametran</b>  |
| <b>3</b>                | But the financial crisis and the Great Recession which began in Spring 2008 have dealt this  | Ancak mali kriz ve 2008 baharında başlayan Büyük Resesyon bu iyimser fotoğrafa bir yıkıcı bir darbe vurmuştur. | Ama bahar 2008i'nde başlayan mali kriz ve büyük durgunluk bu iyimser resme harap edici bir rüzgarı dağıttı. | Fakat hangi büyük Gerileme ve mali kriz Spring'in 2008'ininde, bir yıkıcı darbe bu iyimser resimle ilgilendiğine |

|                             |  |  |  |  |
|-----------------------------|--|--|--|--|
|                             | optimistic<br>picture a<br>devastating<br>blow.  |  |  | başladı.   |
| <b>Intelligibility</b>      |  | Good (45%)   | Non-Native<br>(35%)  | Incomprehensible<br>(60%)  |
| <b>Fidelity</b>             |  | All (45%)  | Little (40%)   | None (70%)   |
| <b>Rank</b>                 |  | Best (95%)   | Moderate (85%)   | Worst (90%)  |
| <b>Macro<br/>evaluation</b> | <p>The translation produced by Google Translate was ranked the best by the majority of annotators. In addition, it has the highest fidelity rating. Proçeviri has lower intelligibility and fidelity, thus it was ranked as moderate.</p> <p>Translation produced by Sametran was ranked the worst. The sentence structure is incomprehensible, and the meaning is not preserved in the translation.</p>                                 |  |  |  |
| <b>Micro<br/>evaluation</b> | <p>The translation of Google Translate can be regarded as the best among the three. There is only one extra article.</p> <p>Proçeviri made one expression selection error in translating “deal a blow”. Sametran has an untranslated word; “spring”. In addition, due to the incorrect attribution of verb “began”, the sentence needs a rearrangement. The clause boundary problem due to the “which” and “began” is another error.</p> |  |  |  |
| <b>No.</b>                  | <b>Original</b>  | <b>Google<br/>Translate</b>  | <b>Proçeviri</b>   | <b>Sametran</b>  |
| <b>4</b>                    | Both GDP and<br>GDP per hour<br>have fallen<br>and are still<br>below the  | Saatte Hem<br>GSYİH ve<br>GSYİH düşmüş<br>ve patlamasının<br>zirvesinde ulaştı | Her iki GDP ve<br>GDP saat başına<br>düştü ve hala<br>artışın zirvesinde<br>ulaşılın düzeyin | Hem GDP ve<br>GDP saat başı,<br>gürlemenin<br>tepesinde<br>uzatılmış düzey |

|                         |  |   |   |  |
|-------------------------|--|---|---|--|
|                         | level reached at the peak of the boom.   | seviyesinin altında hala var.   | aşağısındadır.  | altında hâlâdır düştüler ve.   |
| <b>Intelligibility</b>  |  | Incomprehensible (50%)  | Disfluent (40%)   | Incomprehensible (90%)   |
| <b>Fidelity</b>         |  | None (60%)  | Much (30%)  | None (60%)   |
| <b>Rank</b>             |  | Moderate (75%)  | Best (100%)   | Worst (70%)  |
| <b>Macro evaluation</b> | <p>Google Translate was considered to be moderate by the majority. Although the sentence was not comprehensible and the meaning was not preserved, translation was considered better than Sametran's translation due to the better structure.</p> <p>Proçeviri was considered the best by all the annotators.</p>  |   |   |  |
| <b>Micro evaluation</b> | <p>Google Translate made two verb inflection errors in translating "are" and "reached". In addition, there is a capitalization mistake and sentence needs a rearrangement.</p> <p>Proçeviri failed to translate abbreviations. Apart from these, the translation can be regarded the best among the three.</p> <p>Sametran made two word selection errors "boom" and "peak". In addition, the conjunction error leads to an arrangement problem.</p> |   |   |  |
| <b>No.</b>              | <b>Original</b>  | <b>Google Translate</b>   | <b>Proçeviri</b>  | <b>Sametran</b>  |
| <b>5</b>                | So I discuss a wide range of hypotheses which seek to explain the productivity collapse, including the   | Yani kemer sıkma etkisi de dahil olmak üzere verimlilik çöküşü açıklamak etmeye hipotezler geniş bir yelpazede ele. | Böylece ben, sertlik etkisini dahil ederek, üretkenlik düşüşünü açıklamayı arayan hipotezlerin geniş bir sahasını | Öyleyse sertliğin vuruşunu içermek, üretkenlik çökmesini açıklamayı arayan hipotezlerin birçok çeşitlisi |

|                         |  |   |  |  |
|-------------------------|--|---|--|--|
|                         | impact of austerity.   |   | tartışırım.  | tartışırım.  |
| <b>Intelligibility</b>  |  | Incomprehensible (65%)  | Non-Native (40%)   | Incomprehensible (40%)   |
| <b>Fidelity</b>         |  | Little (50%)  | Little (45%)   | Little (65%)   |
| <b>Rank</b>             |  | Worst (75%)   | Best (80%)   | Moderate (75%)   |
| <b>Macro evaluation</b> | <p>Google Translate was chosen as the worst translation, as the sentence was unintelligible and the meaning was not preserved.</p> <p>The meaning was also not preserved in the translation produced by Proçeviri, however, as it was more intelligible by others, it was ranked the best.</p> <p>Translation produced by Sametran was ranked second. The sentence was less faithful, but as it was more intelligible, it was considered better than Google Translate's translations.</p>                                    |   |  |  |
| <b>Micro evaluation</b> | <p>Google Translate made two verb inflection errors in translating "discuss" and "seek to explain". The word selection for the conjunction is also wrong. Translation produced by Proçeviri had a word selection error; "austerity". In addition, the expression selection for "wide range" is wrong.</p> <p>Sametran made the same word selection errors for "so" and "austerity". In addition, the verb inflection of "including" is wrong. The noun inflection of "wide range" is another problem in the translation.</p> |   |  |  |
| <b>No.</b>              | <b>Original</b>  | <b>Google Translate</b>   | <b>Proçeviri</b>   | <b>Sametran</b>  |
| <b>6</b>                | Most of the conclusions here are negative: the explanation in question   | Burada sonuçların en olumsuz şunlardır: Söz konusu açıklama çalışmıyor. | Sonuçların çoğu burada negatiftir: Söz konusu açıklama çalışmaz. | Burada nihayetlerin en çoğu, olumsuzdur: söz konusu izah çalışmaz. |



|                         |   |  |   |   |
|-------------------------|---|--|---|---|
|                         | doesn't work.   |  |   |   |
| <b>Intelligibility</b>  |   | Disfluent (40%)  | Good (40%)  | Good (30%)  |
| <b>Fidelity</b>         |   | Little (60%)   | Most (55%)  | Much (35%)  |
| <b>Rank</b>             |   | Worst (70%)  | Best (90%)  | Moderate (70%)  |
| <b>Macro evaluation</b> | <p>Google Translate was again ranked the worst.</p> <p>Proçeviri was ranked the best, which was parallel to the intelligibility and fidelity ratings.</p> <p>Sametran was ranked second. The sentence structure rate was very close to Proçeviri's rate, but the fidelity was much lower.</p> |  |   |   |
| <b>Micro evaluation</b> | <p>Google Translate made adjective inflection in translating "most".</p> <p>Proçeviri produced the best translation.</p> <p>Sametran made two word selection errors in translating "conclusions" and "explanation". In addition, the adjective inflection of "most" is also wrong.</p>        |  |   |   |
| <b>No.</b>              | <b>Original</b>   | <b>Google Translate</b>  | <b>Proçeviri</b>  | <b>Sametran</b>   |
| <b>7</b>                | I next turn to the long run impact of financial crises, particularly banking crises, on productivity, capital, TFP and employment.  | Ben bir sonraki verimlilik, sermaye, TFP ve istihdam üzerindeki mali krizler, özellikle bankacılık krizlerinin, uzun dönem etkisinin için. | Ben sonra mali krizlerin uzun vade etkisine dönerim, bilhassa bankacılık krizleri, üretkenlik, sermayede, TFP ve çalışma. | Üretkenlik, belli başlısında, özellikle bankacılık krizleri uzun koşu vuruşuna sonraki dönüş ben, mali krizlerin, TFP ve çalışma. |
| <b>Intelligibility</b>  |   | Incomprehensible (65%)   | Disfluent (30%)   | Incomprehensible (95%)  |

|                         |  |                         |                  |                 |
|-------------------------|--|-------------------------|------------------|-----------------|
| <b>Fidelity</b>         |  | None (55%)              | Little (40%)     | None (70%)      |
| <b>Rank</b>             |  | Worst (55%)             | Best (80%)       | Worst (60%)     |
| <b>Macro evaluation</b> | <p>The intelligibility and fidelity ratings were very low for all machine translation programs.</p> <p>Google Translate and Sametran were both ranked the worst while Proçeviri was ranked the best.</p>   |                         |                  |                 |
| <b>Micro evaluation</b> | <p>Google Translate made one category error, translating the adverb “next” as an adjective. In addition, the abbreviation TFP, which stands for “total factor productivity”, was not recognized and left untranslated. The verb inflection of “turn” is also wrong. Furthermore, the sentence needs to be rearranged.</p> <p>Proçeviri made the same unrecognized abbreviation error. In addition, inflections of nouns “productivity”, “TFP” and “employment” are wrong. The sentence needs rearrangement.</p> <p>Sametran made expression selection error in translating “long run”. In addition, the word selection of “impact” is wrong. The sentence needs rearrangement. Furthermore, the verb “turn” was translated as a noun, which is a category selection error. In addition, the abbreviation “TFP” was not translated by Sametran, either.</p> |                         |                  |                 |
| <b>No.</b>              | <b>Original</b>  | <b>Google Translate</b> | <b>Proçeviri</b> | <b>Sametran</b> |

|                         |   |  |  |   |
|-------------------------|---|--|--|---|
| 8                       | Based on a cross-country panel analysis of 61 countries over 1950-2010, I argue that banking crises generally have a long run impact on the level of productivity but not necessarily on its long run growth rate.  | 1950-2010 aşkın 61 ülkeden bir cross-country paneli analizine dayanarak, ben bankacılık krizleri genellikle ama mutlaka uzun dönem büyüme oranı üzerinde verimlilik düzeyinde uzun dönemli etkisi olduğunu savunuyorlar. | 1950-2010'un üzerinde 61 ülkenin açık alanda bir panel analizinde temel aldı, ben bankacılık krizlerinin genellikle üretkenlik düzeyinde bir uzun vade etkisine sahip olduğunu tartışırım ama muhakkak onunkinde uzun vade büyüme oranı değil. | 1950-2010'un üzerinde 61 ülkenin bir kros panel tahliline dayanarak , ben uzun koşu büyüme oranınınında muhakkak fakat değil üretkenliğin düzeyinde bir uzun koşulan vuruşa sahip olduklarına genellikle crisesin para yatırmak tartışırım. |
| <b>Intelligibility</b>  |   | Disfluent (40%)  | Disfluent (45%)  | Incomprehensible (65%)  |
| <b>Fidelity</b>         |   | Little (45%)   | Little (35%)   | None (10%)  |
| <b>Rank</b>             |   | Moderate (40%)   | Worst (45%)  | Best (40%)  |
| <b>Macro evaluation</b> | <p>Although translation produced by Sametran had the lowest intelligibility and fidelity ratings, it was ranked the best.</p> <p>Google Translate's translation was produced grammatically more acceptable than other translations, but it was ranked as moderate.</p> <p>Proçeviri was ranked the worst in the translation of this sentence.</p> |  |  |   |

|                         |   |  |   |  |
|-------------------------|---|--|---|--|
| <b>Micro evaluation</b> | <p>The long sentence created many problems for machine translation programs. Google Translate has two unrecognized words “cross” and “country”. In addition, the preposition “over” was not correctly translated. Furthermore the negation “not” was not translated. The verb inflection for “argue” is wrong, which resulted in an agreement error. The sentence needs to be rearranged.</p> <p>Proçeviri had one verb inflection error in translating “based”. In addition, preposition “over” was not correctly translated. The sentence had an expression selection error in translating “cross-country”. In addition, the sentence needs to be rearranged.</p> <p>Sametran had one unrecognized word; “crises”. In addition, expression selection for “cross-country” and “long run” are wrong. The word selection for “banking” is not correct. The sentence needs rearrangement.</p> |  |   |  |
| <b>No.</b>              | <b>Original</b>   | <b>Google Translate</b>  | <b>Proçeviri</b>  | <b>Sametran</b>  |
| <b>9</b>                | I therefore predict that the UK will eventually return to the growth rate predicted prior to the crisis.  | Bu nedenle İngiltere sonunda kriz öncesi tahmin büyüme hızına dönecektir tahmin. | Ben BK'in sonunda krizden önce tahmin edilen büyüme oranına dönecek olduğunu bu yüzden tahmin ederim. | Ben bu yüzden sonunda UK'un, krizden önce öngörölmüş büyüme oranına geri döneceğini öngörün. |
| <b>Intelligibility</b>  |   | Disfluent (40%)  | Non-Native (45%)  | Disfluent (45%)  |
| <b>Fidelity</b>         |   | Much (45%)   | Much (30%)  | Most (30%)   |
| <b>Rank</b>             |   | Worst (70%)  | Best (95%)  | Moderate (60%)   |
| <b>Macro</b>            | Google Translate's translation was ranked the worst in each test.   |  |   |  |

|                         |   |  |   |  |
|-------------------------|---|--|---|--|
| <b>evaluation</b>       | Although Sametran's translation had a higher fidelity rating than Proçeşviri's translation, Proçeşviri was given higher intelligibility rating, and thus was ranked the best.   |  |   |  |
| <b>Micro evaluation</b> | Google Translate made a word selection error in translating abbreviation "UK". In addition, the verb inflection for "predict" is wrong.<br>Proçeşviri produced the most successful translation; the sentence just needs a rearrangement.<br>Sametran had an unrecognized abbreviation "UK", in addition to the incorrect verb inflection. |  |   |  |
| <b>No.</b>              | <b>Original</b>   | <b>Google Translate</b>  | <b>Proçeşviri</b>   | <b>Sametran</b>  |
| <b>10</b>               | This prediction is conditional on the UK continuing to follow good policies in other respects, in particular not allowing the government debt-GDP ratio to rise excessively.  | Bu öngörü İngiltere, özellikle yönetim borç-GSYİH oranı aşırı yükselmesine izin vermiyor diğer açılardan iyi politikalar takip devam şartına bağlıdır. | Bu tahmin, özellikle aşırı bir şekilde yükselmek için hükümet borcunu GDP oranına izin vermeyen, diğer hususlarda iyi politikaları izlemeye devam eden BK'de koşullu cümledir | Bu tahmin, aşırı olarak yükseliş hükümet borç-GDP oranına müsaade etmeyen özellikle, başka saygılarda iyi ilkeleri izlemek için devam eden UK'TA koşullara bağlıdır. |
| <b>Intelligibility</b>  |   | Disfluent (30%)  | Incomprehensible (50%)  | Incomprehensible (45%)   |
| <b>Fidelity</b>         |   | Little (30%)   | Little (40%)  | None (45%)   |

| <b>Rank</b>             |  | Moderate (45%)   | Worst (60%)  | Moderate (40%)  |
|-------------------------|--|--|--|---|
| <b>Macro evaluation</b> | The long sentence couldn't get high scores in intelligibility and fidelity ratings. Both Google Translate and Sametran were ranked as moderate while Proçeviri was ranked the worst. It is interesting that while Sametran's fidelity and intelligibility ratings were lower than Proçeviri's, Sametran was ranked higher than Proçeviri.  |  |  |   |
| <b>Micro evaluation</b> | Google Translate made one word selection error in translating "government" and one expression selection error in translating "follow policy". Apart from these, the sentence can be regarded as the most successful one among three.<br>Proçeviri left "GDP" untranslated. In addition, the word selection for "conditional" is wrong. The verb inflection for "rise" is another error. Sametran had two untranslated abbreviations "GDP" and "UK". In addition, the word selection for "respects" is wrong. |  |  |   |
| <b>No.</b>              | <b>Original</b>  | <b>Google Translate</b>  | <b>Proçeviri</b>   | <b>Sametran</b>   |
| <b>11</b>               | Nonetheless the permanent reduction in the level of GDP per worker resulting from the crisis could be substantial, about 5½%.  | Yine kriz kaynaklanan işçi başına GSYİH düzeyinde kalıcı azalma yaklaşık% 5½, önemli olabilir. | Bununla birlikte krizden sonuç veriyor olan işçi başına GDP'in düzeyinde kalıcı azalma önemli olabilirdi, yaklaşık 5 ½%. | Krizden her işçi sonuçlanan GDP'sın düzeyinde daimi azalma yine de, 5½%'in hakkında, sağlam olabilirdi. |
| <b>Intelligibility</b>  |  | Disfluent (60%)  | Disfluent (35%)  | Incomprehensible (60%)  |

|                         |  |   |  |   |
|-------------------------|--|---|--|---|
| <b>Fidelity</b>         |  | Most (40%)  | Much (45%)   | None (55%)  |
| <b>Rank</b>             |  | Best (50%)  | Moderate (55%)   | Worst (65%)   |
| <b>Macro evaluation</b> | <p>Translation produced by Google Translate was ranked the best, mostly due to the fact that it got higher rating in terms of fidelity.</p> <p>Proçeviri was ranked second while Sametran was ranked the worst.</p> <p>Translation produced by Sametran also got the lowest ratings in the intelligibility and fidelity tests.</p>   |   |  |   |
| <b>Micro evaluation</b> | <p>Google Translate produced the most successful translation. Apart from the wrong noun inflection of “crisis”, the sentence is successful.</p> <p>Proçeviri left “GDP” untranslated. The verb inflection of “resulting” is wrong and the sentence needs to be rearranged.</p> <p>Sametran also left the “GDP” untranslated. In addition, the preposition “about” was translated incorrectly. The word selection for “substantial” is wrong. The adverb inflection for “per worker” is also not correct. In addition, the sentence needs to be rearranged.</p> |   |  |   |
| <b>No.</b>              | <b>Original</b>  | <b>Google Translate</b>   | <b>Proçeviri</b>   | <b>Sametran</b>   |
| <b>12</b>               | <p>The cross-country evidence also suggests that there are permanent effects on employment, implying a possibly even larger hit to the level of</p>  | <p>cross-country kanıtlar da yaklaşık% 9 kişi başına düşen GSYİH düzeyine bir olasılıkla daha büyük hit ima istihdam kalıcı etkileri var olduğunu göstermektedir.</p> | <p>Açık alanda kanıt, 9%'den civarı kişi başına GDP'in düzeyine imkan dahilinde hatta daha büyük bir vuruşu ima ederek, çalışmada kalıcı etkiler olduğunu aynı şekilde önerir.</p> | <p>Kros kanıt aynı zamanda 9%un hakkından kişi başına düşen GDP'sin düzeyine daha büyük vurulmuş bir belki hatta ima etmek, çalışmada daimi etkiler olduğunu teklif eder.</p> |

|                         |   |                        |                  |                        |
|-------------------------|---|------------------------|------------------|------------------------|
|                         | GDP per capita of about 9%.   |                        |                  |                        |
| <b>Intelligibility</b>  |   | Incomprehensible (50%) | Non-Native (26%) | Incomprehensible (55%) |
| <b>Fidelity</b>         |   | None (50%)             | Much (40%)       | None (45%)             |
| <b>Rank</b>             |   | Moderate (30%)         | Best (50%)       | Worst (60%)            |
| <b>Macro evaluation</b> | <p>While Google Translate and Sametran got very close fidelity and intelligibility ratings, Google Translate was ranked as moderate and Sametran was ranked the worst.</p> <p>Proçeviri was ranked the best, which was parallel to its intelligibility and fidelity ratings.</p>  |                        |                  |                        |
| <b>Micro evaluation</b> | <p>Long sentence created many problems for machine translation programs. Google Translate made a capitalization error. In addition, there are three untranslated words; “cross”, “country” and “hit”. The verb inflection for “implying” is wrong. In addition, the sentence needs to be rearranged.</p> <p>Proçeviri had an expression selection error in translating “cross-country”. In addition, the abbreviation “GDP” was left untranslated. The word selection for adverb “possibly” is wrong. The sentence needs rearrangement.</p> <p>Sametran also left the abbreviation “GDP” untranslated. The word selections for “possibly” and “suggest” are wrong. In addition, the expression “cross-country” was translated as “kros”, which can be regarded as an expression selection error. The sentence needs to be rearranged.</p> |                        |                  |                        |

Table 3: Macro evaluation and micro evaluation of MT programs for content-focused text type.

Below is presented the table which contains the times a sentence was ranked the best, worst or moderate, the intelligibility and fidelity ratings:



|                        | Google | Proçeviri | Sametran |
|------------------------|--------|-----------|----------|
| <b>Ranking</b>         |        |           |          |
| # Best                 | 2      | 7         | 1        |
| # Moderate             | 5      | 3         | 5        |
| # Worst                | 5      | 2         | 6        |
| <b>Intelligibility</b> |        |           |          |
| Flawless               | 0      | 0         | 0        |
| Good                   | 1      | 1         | 1        |
| Non-Native             | 0      | 5         | 0        |
| Disfluent              | 6      | 5         | 2        |
| Incomprehensible       | 5      | 1         | 9        |
| <b>Fidelity</b>        |        |           |          |
| All                    | 1      | 0         | 0        |
| Most                   | 1      | 2         | 1        |
| Much                   | 2      | 4         | 1        |
| Little                 | 5      | 6         | 2        |
| None                   | 3      | 0         | 7        |

Table 4: The macro evaluation of each machine translation program for content-focused text type.

Content-focused text, which is exemplified with an abstract on economy, was generally best translated by Proçeviri. Number of sentences produced by Proçeviri and was ranked the best by annotators is 7. Google Translate produced 2 sentences which are ranked the best. Sametran produced the most unsuccessful translations, which are ranked the worst 6 times by annotators.

There is no sentence translation which is considered to be flawless. The highest intelligibility rating is “good”, and annotators gave this rating to 1 sentence for each program. 5 sentences translated by Proçeviri were rated as “non-native” and “disfluent”.

“Disfluent” rating was also given to 6 sentences translated by Google Translate. Most of the sentences translated by Sametran were rated as “incomprehensible”, the lowest intelligibility rating.

Only one sentence was given the highest fidelity rating “all”, and it was translated by Google Translate. 2 sentences by Proçeviri and 1 sentence by Google Translate and Sametran were rated as “most”. 4 sentences and by Proçeviri received “much” rating. Most of the sentences were rated “little” or “none”. Out of the 12 sentences, 7 sentences produced by Sametran received lowest fidelity rating “none”. Sentences translated by Proçeviri didn’t receive this rating.

The performance of each machine translation program for micro evaluation (grammatical error analysis) can be seen in the below table:

| <b>Error Category</b>   | <b>Google Translate</b> | <b>Proçeviri</b> | <b>Sametran</b> |
|-------------------------|-------------------------|------------------|-----------------|
| <b>Spelling</b>         | 0                       | 0                | 0               |
| <b>Not-found Word</b>   | 6                       | 8                | 11              |
| <b>Capitalization</b>   | 4                       | 4                | 0               |
| <b>Elision</b>          | 0                       | 0                | 0               |
| <b>Verb inflection</b>  | 11                      | 5                | 4               |
| <b>Noun inflection</b>  | 1                       | 3                | 1               |
| <b>Other inflection</b> | 1                       | 0                | 2               |
| <b>Rearrangement</b>    | 6                       | 6                | 8               |
| <b>Category</b>         | 1                       | 0                | 1               |
| <b>Pronoun</b>          | 0                       | 0                | 0               |
| <b>Article</b>          | 1                       | 0                | 0               |
| <b>Preposition</b>      | 1                       | 1                | 1               |
| <b>Negative</b>         | 1                       | 0                | 0               |
| <b>Conjunction</b>      | 0                       | 0                | 1               |
| <b>Agreement</b>        | 1                       | 0                | 0               |
| <b>Clause boundary</b>  | 0                       | 0                | 1               |

|                       |    |    |    |
|-----------------------|----|----|----|
| <b>Word Selection</b> | 4  | 3  | 12 |
| <b>Expression</b>     | 1  | 4  | 5  |
| <b>Total</b>          | 39 | 34 | 47 |

Table 5: The micro evaluation of each machine translation program for content-focused text type.

While translating content-focused text type, Google Translate made 39 errors. Most of these are verb inflection errors. In addition, the dictionary-related errors, such as not-found words and word selection errors are dominant. Furthermore, most of the sentences need rearrangement. The capitalization mistakes, which are not faced in appeal-focused and audio-medial text types, are high in number.

Proçeviri made the least number of errors, which mostly stemmed from its dictionary. It has 8 not-found words, as well as 7 word and expression selection errors. Apart from these, the sentences produced by Proçeviri need rearrangement.

Sametran made more errors than other MT programs, and the majority of these errors are the result of its dictionary. It made more not-found word as well as word and expression selection errors, which total up to 28 errors. Sametran made less verb inflection errors than others. However, it made more arrangement errors.

Machine translation programs, which were originally created for the translation of foreign language articles, have displayed different performance in translating content-focused text type. In the next section, the performances of machine translation programs in translating form-focused text type, for which the programs were considered unsuitable, are discussed.

## 6.2. FORM-FOCUSED TEXT TYPE

The sample chosen for form-focused text type is an extract from the short story “Rocking-Horse Winner” by D.H. Lawrence. First published in 1926, the story describes the life of a

young boy, Paul, who bets on horses to get rich and earn his mother's love. The 260 word-length extract includes proper names and long sentences. In addition, half of the extract is description while other half is dialogue between protagonist and his mother, which allows us to analyze the machine translation performance better.

The macroevaluation and micro evaluation of outputs of machine translation programs for each sentence are presented below:

| No                      | Original   | Google Translate  | Proçeviri  | Sametran  |
|-------------------------|--|---|--|---|
| 1                       | *It came whispering from the springs of the still-swaying rocking-horse, and even the horse, bending his wooden, champing head, heard it.  | Bu, onun ahşap, champing kafa bükme, hala sallanan sallanan at yaylar ve hatta attan fısıldayan geldi duydum. | O hareketsiz-sallayan sallama-atın baharlarından fısıldayarak geldi ve hatta at, onunkini ahşap bükerek, başı çiğnemek, onu duydu. | Ahşabı kıvırmak, at ve hatta ilkbaharlarından fısıldamaya, imbik-çelen sallanan atının geldi, kafayı çiğnemek, onu duydu. |
| <b>Intelligibility</b>  |  | Incomprehensible (80%)  | Incomprehensible (75%)   | Incomprehensible (90%)  |
| <b>Fidelity</b>         |  | None (75%)  | Little (45%)   | None (75%)  |
| <b>Rank</b>             |  | Moderate (35%)  | Best (65%)   | Worst (65%)   |
| <b>Macro evaluation</b> | <p>* The extract starts with an ellipsis to the previous sentence "There must be more money.", which is repeated for 12 times in the whole story. It is whispered by the house itself, which drives the protagonist Paul to earn money through gambling.</p> <p>The long sentence was given very low intelligibility and fidelity ratings.</p> |   |  |   |

|                         |   |   |   |   |
|-------------------------|---|---|---|---|
|                         | While translation produced Google Translate and Sametran were given same fidelity ratings, due to the lower intelligibility rating, Sametran ranked the worst of three. Proçeviri ranked the best and Google Translate ranked as moderate.  |   |   |   |
| <b>Micro evaluation</b> | <p>The long sentence created many problems for machine translation programs. Google Translate has one untranslated word; “champing”. In addition, “bending” and “heard” were not correctly inflected; this resulted in an agreement problem. The expression selection for “still-swaying” is wrong. The sentence needs rearrangement.</p> <p>Proçeviri made two expression selection errors in translating “still-swaying” and “rocking-horse”. In addition, the word selection for “springs” is wrong. The inflection of the verb “champing” is also wrong. The sentence needs to be rearranged.</p> <p>Sametran made three verb inflection errors in translating “bending”, “whispering” and “champing”. The word selection for “spring” is wrong. In addition, expression selection for “still-swaying” is not correct. This sentence also needs to be rearranged.</p> |   |   |   |
| <b>No.</b>              | <b>Original</b>   | <b>Google Translate</b>   | <b>Proçeviri</b>  | <b>Sametran</b>   |
| <b>2</b>                | The big doll, sitting so pink and smirking in her new pram, could hear it quite plainly, and seemed to be smirking all the more self-consciously  | Büyük bebek, o kadar pembe ve onun yeni bebek arabasında otururken öptüğünde, oldukça açıkça duymak ve daha bilinçli yüzünden öptüğünde gibiydi olabilir. | Büyük bebek, pembeye öyle oturmak ve onda zoraki gülümsemek yeni çocuk arabası, oldukça açıkça onu duyabilirdi ve kendini bilerek ondan dolayı bütün daha çok | Büyük oyuncak bebek, oturmak öyleyse delin ve yeni çocuk arabasında sırtıtmak, açıkça oldukça onu duyabilirdin ve onundan dolayı kendi halini düşünerek bütün |

|                         |   |   |  |   |
|-------------------------|---|---|--|---|
|                         | because of it.  |   | zoraki<br>gülümsüyor<br>olmak için<br>görünüyordu.               | daha fazlayı<br>sırıtmak a<br>benzedin.                                   |
| <b>Intelligibility</b>  |   | Incomprehensible<br>(75%)   | Incomprehensible<br>(55%)  | Incomprehensible<br>(80%)   |
| <b>Fidelity</b>         |   | Little (50%)  | None (60%)   | None (70%)  |
| <b>Rank</b>             |   | Moderate (40%)  | Best (50%)   | Worst (75%)   |
| <b>Macro evaluation</b> | This sentence was also long and it also got low ratings. All of the translations are considered to be incomprehensible, and meaning is also not preserved. Although Google Translate has been rated better in fidelity than Proçeviri, Proçeviri ranked the best.   |   |  |   |
| <b>Micro evaluation</b> | <p>The long sentence was not correctly translated by any of the machine translation programs.</p> <p>Google Translate made a word selection error for “smirking”. In addition, the verb inflections of “hear” and “seemed to be” are wrong. The sentence needs to be rearranged.</p> <p>Proçeviri made the two verb inflection errors while translating “smirking” and “seemed to be”. In addition, the noun inflection of “pram” is wrong. Sametran failed to inflect the verbs “sitting”, “smirking”, “hear”, and “seemed”. There are two agreement problems due to these incorrectly inflected verbs. Noun inflections of “it” and “pram” are also wrong. The word selection for “pink” is not correct. The sentence needs to be rearranged.</p> |   |  |   |
| <b>No.</b>              | <b>Original</b>   | <b>Google Translate</b>   | <b>Proçeviri</b>   | <b>Sametran</b>   |
| <b>3</b>                | The foolish puppy, too, that took the place of the teddy-bear, he   | oyuncak-ayı gerçekleşti çok aptalca yavrusu, o başka bir nedenle bu kadar | Akılsız köpek yavrusu da, oyuncak ayının yerini aldığı, o hiçbir | Aptalca köpek yavrusunda, teddy-ayınının yer aldı, o başka sonuca varması |

|                         |  |  |   |   |
|-------------------------|--|--|---|---|
|                         | was looking so extraordinarily foolish for no other reason but that he heard the secret whisper all over the house:  | olağanüstü aptalca bakıyordu ama o evin her gizli fısıltılarını duydum ki: | diğer sebep olmadığı için öyle alışılmadık biçimde akılsız görünüyordu ama onun evin her yerinde gizli fısıltıyı duyduğu: | yok için fevkalade aptalca öyleye görünmüyordu oysa ev üzerinde bütünü gizli fısıltıyı duydu: |
| <b>Intelligibility</b>  |  | Incomprehensible (70%)   | Incomprehensible (45%)  | Incomprehensible (40%)  |
| <b>Fidelity</b>         |  | None (70%)   | Little (50%)  | Little (40%)  |
| <b>Rank</b>             |  | Worst (70%)  | Best (45%)  | Moderate (45%)  |
| <b>Macro evaluation</b> | The intelligibility results are very low for this sentence. Translations produced by Proçeviri and Sametran were given the same fidelity ratings. Google Translate was ranked the worst among three translations, as its fidelity was also lower than others.  |  |   |   |
| <b>Micro evaluation</b> | <p>Google Translate had arrangement problems. The expression “took place” is not correctly translated. In addition, the negation is missing in the translated sentence. There is an agreement problem, which is the result of incorrectly inflected verb “heard”. The preposition “over” is absent. The noun inflection of “secret” is also wrong.</p> <p>Proçeviri made an expression selection error in translating “teddy-bear”. The inflections of noun “he” and verb “heard” are also wrong.</p> <p>Sametran had one untranslated word “teddy”. The expression “all over” was not correctly translated. In addition, there is a category error in “foolish”, which was translated as an adverb instead of adjective. The word selection for “reason” is another error in the translation. The noun inflections of “puppy” and “bear” are wrong.</p> |  |   |   |

| No.                     | Original   | Google Translate                  | Proçeviri  | Sametran   |
|-------------------------|--|-----------------------------------|--|--|
| <b>4</b>                | "There must be more money!"  | "Daha fazla para olmalı!"         | "Orada daha çok para olmalı"!                          | "orada, daha fazla para olmalı"!                           |
| <b>Intelligibility</b>  |  | Flawless (95%)                    | Good (45%)   | Good (45%)   |
| <b>Fidelity</b>         |  | All (85%)                         | Most (45%)   | Most (35%)   |
| <b>Rank</b>             |  | Best (95%)                        | Moderate (90%)   | Worst (80%)  |
| <b>Macro evaluation</b> | <p>Translation produced by Google Translate was ranked with the highest points in all tests.</p> <p>Translation produced by Proçeviri and Sametran were given the same points for intelligibility and fidelity, but Proçeviri was ranked as moderate while Sametran was ranked the worst among three.</p>  |                                   |  |  |
| <b>Micro evaluation</b> | <p>Google Translate and Proçeviri produced successful translations.</p> <p>Sametran made a capitalization error.</p>   |                                   |  |  |
| No.                     | Original   | Google Translate                  | Proçeviri  | Sametran   |
| <b>5</b>                | Yet nobody ever said it aloud.   | Ancak kimse yüksek sesle söyledi. | Lakin hiç kimse şimdiye kadar yüksek sesle onu demedi. | Henüz hiç kimse şimdiye kadar, yüksek sesle ona söylemedi. |
| <b>Intelligibility</b>  |  | Non-Native (40%)                  | Flawless (40%)   | Non-Native (30%)   |
| <b>Fidelity</b>         |  | None (45%)                        | Most (45%)   | Little (35%)   |
| <b>Rank</b>             |  | Worst (90%)                       | Best (100%)  | Moderate (90%)   |
| <b>Macro evaluation</b> | <p>Translation produced by Google Translate received low ratings in terms of intelligibility and fidelity. It was ranked the worst.</p> <p>Proçeviri was given highest intelligibility rating; in addition, all annotators chose it as the best among three.</p> <p>Sametran was considered to be moderate and its intelligibility and fidelity ratings are parallel to its ranking.</p> |                                   |  |  |
| <b>Micro</b>            | <p>Google Translate made a negation error.</p>   |                                   |  |  |



|                         |   |   |   |   |
|-------------------------|---|---|---|---|
| <b>evaluation</b>       | <p>Proçeviri produced a successful translation.</p> <p>Sametran made a word selection error in translating the conjunction “yet”. In addition, noun inflection of “it” is wrong in this translation.</p>  |   |   |   |
| <b>No.</b>              | <b>Original</b>   | <b>Google Translate</b>                                 | <b>Proçeviri</b>  | <b>Sametran</b>   |
| <b>6</b>                | The whisper was everywhere, and therefore no one spoke it.  | Fısıltı her yerde, ve bu nedenle hiç kimse onu konuştu. | Fısıltı her yerde idi ve bu yüzden hiç kimse onu konuşmadı. | Fısıltı, her yereydi, ve bu yüzden hiç kimse ona konuşmadı. |
| <b>Intelligibility</b>  |   | Disfluent (40%)   | Flawless (50%)  | Non-Native (45%)  |
| <b>Fidelity</b>         |   | Little (40%)  | All (50%)   | Much (25%)  |
| <b>Rank</b>             |   | Worst (90%)   | Best (95%)  | Moderate (85%)  |
| <b>Macro evaluation</b> | <p>Translation produced by Google Translate was again chosen as the worst.</p> <p>Proçeviri’s translation received the highest scores in all three tests.</p> <p>Sametran ranked as moderate.</p>   |   |   |   |
| <b>Micro evaluation</b> | <p>Google Translate made a verb inflection error in translating “was”. In addition, the noun inflection of “it” is wrong.</p> <p>The noun inflection of “it” is also wrong in the translation produced by Proçeviri.</p> <p>Sametran made a word selection error in translating “everywhere”. In addition, noun inflection of “it” is also wrong in this translation.</p> |   |   |   |
| <b>No.</b>              | <b>Original</b>   | <b>Google Translate</b>                                 | <b>Proçeviri</b>  | <b>Sametran</b>   |
| <b>7</b>                | Just as no one ever says:   | Sadece hiç kimse dediği gibi:                           | Tam da hiç kimsenin şimdiye kadar demediği gibi:            | Hiç kimse şimdiye kadar söylemediği gibi şimdi:             |
| <b>Intelligibility</b>  |   | Disfluent (21%)   | Flawless (65%)  | Disfluent (45%)   |

|                         |  |  |  |   |
|-------------------------|--|--|--|---|
| <b>Fidelity</b>         |  | None (45%)   | All (60%)  | Much (50%)  |
| <b>Rank</b>             |  | Worst (85%)  | Best (95%)   | Moderate (75%)  |
| <b>Macro evaluation</b> | <p>Google Translate's translation was ranked the worst, and it was considered to be the most unfaithful translation among three.</p> <p>Translation produced by Proçeviri was given the highest ratings in terms of intelligibility and fidelity.</p>  |  |  |   |
| <b>Micro evaluation</b> | <p>Google Translate made the same negation error. In addition, the word selection for the adverb "just" is wrong.</p> <p>Proçeviri produced a successful translation.</p> <p>Translation produced by Sametran was also very successful, there is only the same word selection error with the Google Translate; "just".</p> |  |  |   |
| <b>No.</b>              | <b>Original</b>  | <b>Google Translate</b>  | <b>Proçeviri</b>   | <b>Sametran</b>   |
| <b>8</b>                | "We are breathing!" in spite of the fact that breath is coming and going all the time.   | "Biz nefes!" nefes geliyor ve her zaman olacak gerçeği rağmen. | "Biz nefes alıyoruz"! Nefesin geliyor olduğu ve daima gidiyor olduğu gerçeğine rağmen. | Nefes, bütün zamana gelmek ve gitmek olduğu halde. "biz soluyoruz". |
| <b>Intelligibility</b>  |  | Incomprehensible (70%)   | Good (40%)   | Disfluent (50%)   |
| <b>Fidelity</b>         |  | None (52%)   | Most (40%)   | Little (40%)  |
| <b>Rank</b>             |  | Worst (60%)  | Best (100%)  | Moderate (60%)  |
| <b>Macro evaluation</b> | <p>Translation produced by Google Translate received the lowest ratings and ranking again.</p> <p>Proçeviri's translation was ranked the best by all annotators.</p> <p>Translation produced by Sametran was again given the second rank.</p>  |  |  |   |
| <b>Micro evaluation</b> | <p>Google Translate made a category error, translating the verb "breathing" as a noun. This resulted in a verb inflection and agreement error. In addition,</p>  |  |  |   |

|                         | <p>there is another category error in translating “is going”, which was mistaken as “be going to”, instead of continuous aspect of the verb “go”. Furthermore, the noun inflection of “fact” is wrong.</p> <p>Proçeviri produced a successful translation, there is one capitalization error. Sametran had an arrangement problem. In addition, the expression “all the time” was not translated correctly. There is also a clause boundary error, as the machine translation program divided the sentence into two.</p> |   |   |  |
|-------------------------|--|---|---|--|
| No.                     | Original   | Google Translate  | Proçeviri   | Sametran   |
| 9                       | "Mother," said the boy Paul one day, "why don't we keep a car of our own? Why do we always use uncle's, or else a taxi?"   | "Anne," Birgün Paul "Biz? Bizim kendi arabamız yok, neden Neden hep amcamın, ya da başka bir taksi kullanabilirim?", Dedi | "Anne", bir gün, "Biz bizim kendimizin bir arabasını neden sürdürmeyiz? Biz her zaman neden amcamın olduğunu kullanıyoruz veya başka bir taksi?", oğlan Paul dedi | Bir gün delikanlı Paul'unu söyledi, Neden daima, ya da bir taksi amcamın kullanırız? "anne" "neden, kendimizin bir arabasını tutarız"? |
| <b>Intelligibility</b>  |  | Incomprehensible (60%)  | Disfluent (55%)   | Incomprehensible (70%)   |
| <b>Fidelity</b>         |  | Little (35%)  | Little (35%)  | None (55%)   |
| <b>Rank</b>             |  | Moderate (55%)  | Best (44%)  | Worst (89%)  |
| <b>Macro evaluation</b> | <p>The translations of the sentence were not given high scores.</p> <p>Translations produced by Google Translate and Sametran were considered to be incomprehensible. But Google Translate ranked as moderate while Sametran ranked the worst, due to Google Translate's better fidelity rating. Proçeviri was ranked the best among three.</p>  |   |   |  |
| <b>Micro evaluation</b> | <p>Long sentence created many problems for machine translation programs.</p> <p>Google Translate had made an arrangement error. In addition, there is no</p>   |   |   |  |

|                         |  |   |   |  |
|-------------------------|--|---|---|--|
|                         | <p>subject-verb agreement in the second sentence. This is the result of a verb inflection error. Furthermore, there is a capitalization error.</p> <p>Proçeviri produced the most successful translation among three. The sentence needs a rearrangement. In addition, the word selection for “keep” is wrong. Sametran made two noun inflection errors in translating “Paul” and “uncle”. The conjunction is not analyzed correctly, which resulted in an arrangement error. In addition, the negation is missing in the translation.</p> |   |   |  |
| <b>No.</b>              | <b>Original</b>  | <b>Google Translate</b>                       | <b>Proçeviri</b>  | <b>Sametran</b>  |
| <b>10</b>               | "Because we're the poor members of the family," said the mother.   | "Biz ailenin fakir üyeleri Çünkü," dedi anne. | "Biz ailenin yoksul üyeleri olduğumuz için", anne dedi. | Anneyi söyledi.<br>"çünkü biz ailenin fakir üyeleriyiz". |
| <b>Intelligibility</b>  |  | Non-Native (45%)                              | Good (45%)  | Disfluent (45%)  |
| <b>Fidelity</b>         |  | Most (40%)                                    | All (50%)   | Little (45%)   |
| <b>Rank</b>             |  | Worst (55%)                                   | Best (90%)  | Moderate (45%)   |
| <b>Macro evaluation</b> | <p>Google Translate's translation received higher scores than Sametran's translation. But Google Translate was ranked the worst among three. Translation produced by Proçeviri had the highest score in fidelity and it was ranked the best.</p>   |   |   |  |
| <b>Micro evaluation</b> | <p>Google Translate made a capitalization mistake. In addition, there is a verb inflection error in translating “are”.</p> <p>Proçeviri produced a successful translation.</p> <p>Sametran made a noun inflection error in translating “mother”.</p>   |   |   |  |
| <b>No.</b>              | <b>Original</b>  | <b>Google Translate</b>                       | <b>Proçeviri</b>  | <b>Sametran</b>  |
| <b>11</b>               | "But why are we, mother?"  | "Ama neden, anne Biz Kimiz?"                  | "Ama biz nedeniz, anne"?                                | "fakat neden anne, bizsin"?                              |

|                         |   |   |  |  |
|-------------------------|---|---|--|--|
| <b>Intelligibility</b>  |   | Incomprehensible<br>(35%)   | Incomprehensible<br>(65%)  | Incomprehensible<br>(75%)  |
| <b>Fidelity</b>         |   | None (60%)  | None (45%)   | None (50%)   |
| <b>Rank</b>             |   | Best (45%)  | Moderate (35%)   | Worst (60%)  |
| <b>Macro evaluation</b> | Three translations of this short sentence received the lowest points in terms of intelligibility and fidelity. Google Translate's translation was ranked as the best among three.   |   |  |  |
| <b>Micro evaluation</b> | <p>The ellipsis between present and previous sentence was not understood by machine translation programs.</p> <p>Google Translate made a word selection error in translating "why". In addition, there are two capitalization errors.</p> <p>Proçeviri produced a successful translation, but it was ranked as moderate.</p> <p>Sametran made verb inflection error in translating "are", which resulted in an agreement error.</p> |   |  |  |
| <b>No.</b>              | <b>Original</b>   | <b>Google Translate</b>   | <b>Proçeviri</b>   | <b>Sametran</b>  |
| <b>12</b>               | "Well - I suppose," she said slowly and bitterly, "it's because your father has no luck."   | O yavaş yavaş ve acı dedi, - "Peki ben, herhalde" "Baban hiç şans var çünkü." | "Şey— ben zannediyorum", o, "O senin baban hiçbir şansa sahip olmadığı için", yavaşça ve acı bir şekilde dedi. | "kuyu". Ben varsayırım ,dİr çünkü babanız şansa sahip olmaz. "o, ağır ağır ve acı olarak söyledi". |
| <b>Intelligibility</b>  |   | Incomprehensible<br>(65%)   | Non-Native<br>(40%)  | Incomprehensible<br>(85%)  |
| <b>Fidelity</b>         |   | None (55%)  | Most (50%)   | None (70%)   |
| <b>Rank</b>             |   | Moderate (75%)  | Best (90%)   | Worst (80%)  |
| <b>Macro evaluation</b> | Although translation produced by Google Translate and Sametran were both given the lowest scores for intelligibility and fidelity, Sametran was ranked the worst.   |   |  |  |

|                         |  |  |   |  |
|-------------------------|--|--|---|--|
|                         | Proçeviri's translation was ranked the best, and its intelligibility and fidelity points are parallel to its rank.   |  |   |  |
| <b>Micro evaluation</b> | <p>Google Translate made an adjective inflection error in translating "bitterly". In addition, there is a negation error. Furthermore, the verb "suppose" was not inflected correctly.</p> <p>Proçeviri produced a successful traslation.</p> <p>Sametran made a category error in translating "well". In addition, there is an agreement error. The sentence also needs to be rearranged.</p> |  |   |  |
| <b>No.</b>              | <b>Original</b>  | <b>Google Translate</b>                      | <b>Proçeviri</b>                                      | <b>Sametran</b>                                      |
| <b>13</b>               | The boy was silent for some time.  | Oğlan bir süre sessiz kaldı.                 | Oğlan biraz zaman boyunca sessizdi.                   | Delikanlı, birkaç zaman için sessizdi.               |
| <b>Intelligibility</b>  |  | Flawless (85%)                               | Good (35%)  | Non-Native (40%)                                     |
| <b>Fidelity</b>         |  | All (95%)                                    | Most (65%)  | Much (30%)   |
| <b>Rank</b>             |  | Best (100%)                                  | Moderate (80%)  | Worst (75%)  |
| <b>Macro evaluation</b> | <p>Translation produced by Google Translate received the highest scores in all three test, it was considered to be the best translation by all annotators.</p> <p>Proçeviri was ranked as moderate while Sametran was ranked the worst.</p>  |  |   |  |
| <b>Micro evaluation</b> | <p>Google Translate produced a successful translation.</p> <p>Proçeviri made a word selection error in translating the adjective "some". Sametran also made the same word selection error for "some". In addition, the preposition "for" is translated incorrectly.</p>  |  |   |  |
| <b>No.</b>              | <b>Original</b>  | <b>Google Translate</b>                      | <b>Proçeviri</b>                                      | <b>Sametran</b>                                      |
| <b>14</b>               | "Is luck money, mother?" he asked, rather timidly.   | "Şans, para mı anne?" o çok çekinerek sordu. | "Şans parası mıdır, anne"? O, oldukça ürkekçe, sordu. | O "anne, şans parası mıdır" sordu , ürkekçe oldukça. |

|                         |   |                         |                  |                  |
|-------------------------|---|-------------------------|------------------|------------------|
| <b>Intelligibility</b>  |   | Good (35%)              | Good (45%)       | Disfluent (45%)  |
| <b>Fidelity</b>         |   | Most (55%)              | Much (25%)       | Little (30%)     |
| <b>Rank</b>             |   | Best (75%)              | Moderate (50%)   | Worst (73%)      |
| <b>Macro evaluation</b> | Google Translate's and Proçeviri's translations were both given the same intelligibility ratings; however, their fidelity results and, as a result, their ranking are different.  |                         |                  |                  |
| <b>Micro evaluation</b> | Google Translate produced a successful translation.<br>Proçeviri made a noun inflection error in translating "luck". In addition, there is a capitalization error.<br>Sametran made the same noun inflection error in translating "luck". In addition, the sentence needs to be rearranged. |                         |                  |                  |
| <b>No.</b>              | <b>Original</b>   | <b>Google Translate</b> | <b>Proçeviri</b> | <b>Sametran</b>  |
| <b>15</b>               | "No, Paul.  | "Hayır, Paul.           | "Hayır, Paul.    | "paul, yok".     |
| <b>Intelligibility</b>  |   | Flawless (100%)         | Flawless (100%)  | Non-Native (35%) |
| <b>Fidelity</b>         |   | All (95%)               | All (95%)        | None (70%)       |
| <b>Rank</b>             |   | Best (100%)             | Moderate (30%)   | Worst (75%)      |
| <b>Macro evaluation</b> | Translations produced by Google Translate and Proçeviri were given the same intelligibility and fidelity points, however, all the annotators considered Google Translate's translation better than Proçeviri's translation. Sametran was ranked the worst.                                  |                         |                  |                  |
| <b>Micro evaluation</b> | Sametran made a word selection error for "No" and a capitalization mistake.   |                         |                  |                  |
| <b>No.</b>              | <b>Original</b>   | <b>Google Translate</b> | <b>Proçeviri</b> | <b>Sametran</b>  |
| <b>16</b>               | Not quite.  | Oldukça değil.          | Tamamen değil.   | Oldukça değil.   |
| <b>Intelligibility</b>  |   | Non-Native (40%)        | Flawless (80%)   | Non-Native (45%) |
| <b>Fidelity</b>         |   | Little (40%)            | All (60%)        | Little (40%)     |
| <b>Rank</b>             |   | Moderate (60%)          | Best (90%)       | Worst (40%)      |

|                         |   |                                 |  |  |
|-------------------------|---|---------------------------------|--|--|
| <b>Macro evaluation</b> | Translation produced by Proçeviri was given the highest points in all three tests. Google Translate was ranked second despite receiving same intelligibility and fidelity points with Sametran.   |                                 |  |  |
| <b>Micro evaluation</b> | The word selections of Google Translate and Sametran for “quite” are wrong.   |                                 |  |  |
| <b>No.</b>              | <b>Original</b>   | <b>Google Translate</b>         | <b>Proçeviri</b>   | <b>Sametran</b>                          |
| <b>17</b>               | It's what causes you to have money."  | Bu para var size neden budur. " | O paraya sahip olman için sana neyin sebebiyet verdiğidir. | Paraya ne sahip olduğunuz sebeplerindir. |
| <b>Intelligibility</b>  |   | Incomprehensible (65%)          | Disfluent (26%)  | Incomprehensible (65%)                   |
| <b>Fidelity</b>         |   | None (70%)                      | Most (31%)   | None (60%)                               |
| <b>Rank</b>             |   | Worst (65%)                     | Best (90%)   | Moderate (65%)                           |
| <b>Macro evaluation</b> | Translations produced by Google Translate and Sametran were both given the lowest scores in terms of intelligibility and fidelity. Google Translate was ranked the worst among three.<br>Proçeviri's translation was given a very low intelligibility point and a very high fidelity point. Thus, it was ranked the best. |                                 |  |  |
| <b>Micro evaluation</b> | Google Translate made two word selections error for “causes” and “have”. Proçeviri also made the same word selection error for “causes”. Sametran made a capitalization error and word selection error for “causes”.  |                                 |  |  |
| <b>No.</b>              | <b>Original</b>   | <b>Google Translate</b>         | <b>Proçeviri</b>   | <b>Sametran</b>                          |
| <b>18</b>               | "Oh!" said Paul vaguely.  | "Ah!" Paul belli belirsiz dedi. | "Oh"! Belirsizce Paul dedi.                                | Belirsizce Paul'u söyledi. "öyle mi?".   |
| <b>Intelligibility</b>  |   | Good (35%)                      | Non-Native (40%)   | Incomprehensible (55%)                   |
| <b>Fidelity</b>         |   | Most (40%)                      | Most (35%)   | None (40%)                               |



| <b>Rank</b>             |   | <b>Best (75%)</b>   | <b>Moderate (70%)</b>  | <b>Worst (95%)</b>   |
|-------------------------|---|---|--|--|
| <b>Macro evaluation</b> | Translation produced by Google Translate was ranked the best. Sametran was given the lowest scores for all three tests.   |   |  |  |
| <b>Micro evaluation</b> | Google Translate produced a successful translation.<br>Proçeviri made a capitalization error and the translation needs to be rearranged.<br>Sametran made a noun inflection error in translating “Paul”. In addition, the word selection for “oh” is wrong.   |   |  |  |
| <b>No.</b>              | <b>Original</b>   | <b>Google Translate</b>   | <b>Proçeviri</b>   | <b>Sametran</b>  |
| <b>19</b>               | "I thought when Uncle Oscar said filthy lucker, it meant money."  | "Ben Oscar Amca pis lucker bu parayı demek dediğinde düşündüm." | "Ben amca Oskar'ın kirli luckere dediği zaman düşündüm, o parayı demek istiyordu". | "ben Uncle'nin Oskar'ıninki zaman, pis luckeri söylediğini düşündüm". ", o parayı kastetti". |
| <b>Intelligibility</b>  |   | Incomprehensible (55%)  | Incomprehensible (40%)   | Incomprehensible (60%)   |
| <b>Fidelity</b>         |   | None (50%)  | Little (45%)   | Little (45%)   |
| <b>Rank</b>             |   | Worst (55%)   | Best (65%)   | Moderate (50%)   |
| <b>Macro evaluation</b> | All translations were given lowest intelligibility points. Translations produced by Proçeviri and Sametran were given the same fidelity points, but Proçeviri was ranked higher than Sametran.  |   |  |  |
| <b>Micro evaluation</b> | The word-play between “filthy lucker” and “luck” is stylistically very important in the story, as it shows the Paul’s innocence as a child and his misunderstanding between dishonest ways of money earning and luck. This word play was not recreated by any of the machine translation programs.<br>Google Translate left “lucker” untranslated. In addition, the sentence needs a rearrangement. |   |  |  |

|                         |   |                                    |                                      |   |
|-------------------------|---|------------------------------------|--------------------------------------|---|
|                         | <p>Proçeviri also left “lucker” untranslated. This translation also needs to be rearranged.</p> <p>Sametran left “uncle” and “lucker” untranslated. In addition, this translation also needs rearrangement.</p>   |                                    |                                      |   |
| <b>No.</b>              | <b>Original</b>   | <b>Google Translate</b>            | <b>Proçeviri</b>                     | <b>Sametran</b>                             |
| <b>20</b>               | "Filthy lucre does mean money," said the mother.  | "Pis lucre para demek," dedi anne. | "Para parayı kastediyor", anne dedi. | Anneyi söyledi.<br>"para, parayı kastetir". |
| <b>Intelligibility</b>  |   | Good (35%)                         | Incomprehensible (50%)               | Incomprehensible (70%)                      |
| <b>Fidelity</b>         |   | Most (55%)                         | None (50%)                           | None (70%)                                  |
| <b>Rank</b>             |   | Best (80%)                         | Moderate (63%)                       | Worst (85%)                                 |
| <b>Macro evaluation</b> | <p>Although there is an untranslated word in it, translation produced by Google Translate was ranked the best among three.</p> <p>Translations produced by Proçeviri and Sametran were given the lowest points in terms of intelligibility and fidelity. Sametran was ranked the worst.</p>   |                                    |                                      |   |
| <b>Micro evaluation</b> | <p>The word-play continues in the dialogue. However, this time, Proçeviri and Sametran translated “lucre” as “para (money).</p> <p>Google Translate left “lucre” again untranslated.</p> <p>Translation produced by Proçeviri needs to be rearranged.</p> <p>Sametran had a noun inflection error in translating “mother”. In addition, the verb inflection of “mean” is wrong.</p> |                                    |                                      |   |
| <b>No.</b>              | <b>Original</b>   | <b>Google Translate</b>            | <b>Proçeviri</b>                     | <b>Sametran</b>                             |
| <b>21</b>               | "But it's lucre, not luck."   | "Ama lucre, değil şans."           | "Ama o, şans değil, paradır".        | "fakat o şans değil, servettir".            |
| <b>Intelligibility</b>  |   | Incomprehensible (35%)             | Flawless (50%)                       | Good (55%)                                  |
| <b>Fidelity</b>         |   | None (55%)                         | Most (30%)                           | Most (25%)                                  |

| <b>Rank</b>             |  | <b>Worst (100%)</b>     | <b>Best (70%)</b> | <b>Moderate (60%)</b>            |
|-------------------------|--|-------------------------|-------------------|----------------------------------|
| <b>Macro evaluation</b> | The untranslated word “lucre” made the translation produced by Google Translate to be ranked the worst by all the annotators. The translation produced by Proçeviri was considered flawless in terms of intelligibility, and it was ranked the best. Translation produced by Sametran was ranked as moderate, and it was rated the most faithful among the three translations. |                         |                   |                                  |
| <b>Micro evaluation</b> | Google Translate left “lucre” untranslated, again. In addition, the sentence needs to be rearranged.<br>Proçeviri produced a successful translation.<br>Sametran made a capitalization error and a word selection error in translating “lucre”.  |                         |                   |                                  |
| <b>No.</b>              | <b>Original</b>  | <b>Google Translate</b> | <b>Proçeviri</b>  | <b>Sametran</b>                  |
| <b>22</b>               | "Oh!" said the boy.  | "Ah!" dedi oğlan.       | "Oh"! Oğlan dedi. | Delikanlıyı söyledi. "öyle mi?". |
| <b>Intelligibility</b>  |  | Flawless (80%)          | Non-Native (45%)  | Incomprehensible (40%)           |
| <b>Fidelity</b>         |  | All (60%)               | Most (25%)        | None (40%)                       |
| <b>Rank</b>             |  | Best (90%)              | Moderate (90%)    | Worst (95%)                      |
| <b>Macro evaluation</b> | The short sentence was best translated by Google Translate, it was rated as flawless and faithful. Translation produced by Proçeviri was ranked as moderate. Sametran’s translation was given the lowest ratings in terms of intelligibility and fidelity.   |                         |                   |                                  |
| <b>Micro evaluation</b> | Google Translate produced a successful translation, without any grammatical errors.<br>Proçeviri made a capitalization error.<br>Sametran made a noun inflection error in translating “boy”. In addition, the word selection for “oh” is wrong.  |                         |                   |                                  |
| <b>No.</b>              | <b>Original</b>  | <b>Google Translate</b> | <b>Proçeviri</b>  | <b>Sametran</b>                  |

|                         |  |                           |   |  |
|-------------------------|--|---------------------------|---|--|
| <b>23</b>               | "Then what is luck, mother?"<br>"Then what is luck, mother?"   | "Sonra şans, anne nedir?" | "O zaman neyin şans olduğu, anne"?                          | "sonra"? "şans ne olduğu, anne"?           |
| <b>Intelligibility</b>  |  | Disfluent (60%)           | Incomprehensible (45%)                                      | Incomprehensible (75%)                     |
| <b>Fidelity</b>         |  | Little (50%)              | None (30%)  | None (55%)                                 |
| <b>Rank</b>             |  | Moderate (52%)            | Best (50%)  | Worst (85%)                                |
| <b>Macro evaluation</b> | The translations of this short sentence were given very low points by annotators. Translations produced by Proçeviri and Sametran were given lowest ratings in terms of intelligibility and fidelity. However, although translation produced by Google Translate was given better ratings than Proçeviri, Proçeviri was ranked the best among three. |                           |   |  |
| <b>Micro evaluation</b> | Google Translate made an arrangement problem. In addition, the word selection for "then" is also wrong.<br>Proçeviri made a word selection for "what".<br>Sametran made the same word selection errors for "then" and "what".  |                           |   |  |
| <b>No.</b>              | <b>Original</b>  | <b>Google Translate</b>   | <b>Proçeviri</b>  | <b>Sametran</b>                            |
| <b>24</b>               | "It's what causes you to have money.   | "Bu para var neden budur. | "O paraya sahip olman için sana neyin sebebiyet verdiğidir. | "paraya ne sahip olduğunuz sebeplerindir". |
| <b>Intelligibility</b>  |  | Incomprehensible (75%)    | Disfluent (30%)   | Incomprehensible (65%)                     |
| <b>Fidelity</b>         |  | None (75%)                | Most (30%)  | None (50%)                                 |
| <b>Rank</b>             |  | Worst (75%)               | Best (90%)  | Moderate (63%)                             |
| <b>Macro evaluation</b> | Translations produced by Google Translate and Sametran were given the lowest ratings for intelligibility and fidelity, and they were ranked the worst and as moderate, respectively.   |                           |   |  |

|                         |   |                             |   |   |
|-------------------------|---|-----------------------------|---|---|
|                         | Proçeviri's was ranked the best for among the three.  |                             |   |   |
| <b>Micro evaluation</b> | <p>The reduplication plays an important role in the story. This sentence is the reduplication of sentence 17. The same mistakes are also repeated by machine translation programs.</p> <p>Google Translate made two word selections error for “causes” and “have”. Proçeviri also made the same word selection error for “causes”.</p> <p>Sametran made a capitalization error and word selection error for “causes”.</p> |                             |   |   |
| <b>No.</b>              | <b>Original</b>   | <b>Google Translate</b>     | <b>Proçeviri</b>                          | <b>Sametran</b>                           |
| <b>25</b>               | If you're lucky you have money.   | Eğer şanslıysanız para var. | Eğer sen şanslıysan, sen paraya sahipsin. | Eğer sen paraya sahip olduğun şanslıysan. |
| <b>Intelligibility</b>  |   | Non-Native (45%)            | Non-Native (40%)                          | Incomprehensible (60%)                    |
| <b>Fidelity</b>         |   | All (30%)                   | All (45%)                                 | None (65%)                                |
| <b>Rank</b>             |   | Moderate (80%)              | Best (80%)                                | Worst (100%)                              |
| <b>Macro evaluation</b> | <p>Translation produced by Google Translate was ranked second among the three translations, it contained all the meaning of the source sentence.</p> <p>Proçeviri was ranked the best translation.</p> <p>All of the annotators gave the translation produced by Sametran the lowest rank. In addition, the sentence was given lowest ratings in terms of fidelity and intelligibility.</p>                               |                             |   |   |
| <b>Micro evaluation</b> | <p>Google Translate made a word selection error for “have”.</p> <p>The short sentence was best translated by Proçeviri. There is an unneeded pronoun.</p> <p>Sametran made an agreement error. In addition, the verb inflection of “are” is wrong.</p>  |                             |   |   |

Table 6: Macro evaluation and micro evaluation of MT programs for form-focused text type.

Below is presented the table which contains the times a sentence was ranked the best, worst or moderate, the intelligibility and fidelity ratings:

|                         | <b>Google Translate</b> | <b>Proçeviri</b> | <b>Sametran</b> |
|-------------------------|-------------------------|------------------|-----------------|
| <b>Ranking</b>          |                         |                  |                 |
| <b># Best</b>           | 8                       | 16               | 0               |
| <b># Moderate</b>       | 7                       | 8                | 10              |
| <b># Worst</b>          | 10                      | 0                | 15              |
| <b>Intelligibility</b>  |                         |                  |                 |
| <b>Flawless</b>         | 4                       | 6                | 0               |
| <b>Good</b>             | 3                       | 5                | 2               |
| <b>Non-Native</b>       | 4                       | 4                | 5               |
| <b>Disfluent</b>        | 3                       | 3                | 4               |
| <b>Incomprehensible</b> | 11                      | 7                | 14              |
| <b>Fidelity</b>         |                         |                  |                 |
| <b>All</b>              | 5                       | 6                | 0               |
| <b>Most</b>             | 4                       | 10               | 2               |
| <b>Much</b>             | 0                       | 1                | 3               |
| <b>Little</b>           | 5                       | 4                | 7               |
| <b>None</b>             | 11                      | 4                | 13              |

Table 7: The macro evaluation of each machine translation program for form-focused text type.

Form-focused text type, which is exemplified with an extract of short story, was generally best translated by Proçeviri. Most of the sentences, which were ranked the best by annotators, were translated by Proçeviri. Google Translate translated 8 of the best ranked

sentences while translations produced by Sametran was not ranked the best, they were ranked as moderate for 10 times and worst for 15 times. In addition, the sentences translated by Proçeviri were never ranked the worst.

In terms of intelligibility, 10 sentences were given the highest rating “flawless”, 6 of them were translated by Proçeviri while rest of them was translated by Google Translate. “Good” rating was given to 10 sentences; half of them were translated by Proçeviri. 4 sentences by Google Translate and Proçeviri and 5 sentences by Sametran were rated as “non-native”. In addition, 3 sentences by Google Translate and Proçeviri and 4 sentences by Sametran were rated as “disfluent”. The lowest intelligibility rating “incomprehensible” was given to 11 sentences by Google Translate, 7 sentences by Proçeviri and 14 sentences by Sametran.

The highest fidelity rating “all” was given to 11 sentences, 6 of them was produced by Proçeviri while rest of them was translated by Google Translate. 10 sentences by Proçeviri, 4 sentences by Google Translate and 2 sentences by Sametran were rated as “most”. 3 sentences by Sametran and 1 sentence by Proçeviri were rated as “much”. Most of the sentences produced by Sametran were rated either as “little” or “none”. 13 sentences produced by Sametran and 11 sentences produced by Google Translate were given the lowest fidelity rating “none”.

The performance of each machine translation program for micro evaluation (grammatical error analysis) can be seen in the below table:

| <b>Error Category</b>  | <b>Google Translate</b> | <b>Proçeviri</b> | <b>Sametran</b> |
|------------------------|-------------------------|------------------|-----------------|
| <b>Spelling</b>        | 0                       | 0                | 0               |
| <b>Not-found Word</b>  | 4                       | 1                | 3               |
| <b>Capitalization</b>  | 4                       | 5                | 4               |
| <b>Elision</b>         | 0                       | 0                | 0               |
| <b>Verb inflection</b> | 10                      | 3                | 9               |
| <b>Noun inflection</b> | 2                       | 3                | 13              |

|                         |    |    |    |
|-------------------------|----|----|----|
| <b>Other inflection</b> | 2  | 2  | 0  |
| <b>Rearrangement</b>    | 7  | 5  | 6  |
| <b>Category</b>         | 2  | 0  | 2  |
| <b>Pronoun</b>          | 0  | 1  | 0  |
| <b>Article</b>          | 0  | 0  | 0  |
| <b>Preposition</b>      | 1  | 0  | 3  |
| <b>Negative</b>         | 4  | 0  | 1  |
| <b>Conjunction</b>      | 0  | 0  | 1  |
| <b>Agreement</b>        | 4  | 0  | 4  |
| <b>Clause boundary</b>  | 0  | 0  | 1  |
| <b>Word Selection</b>   | 10 | 6  | 15 |
| <b>Expression</b>       | 2  | 3  | 2  |
| <b>Total</b>            | 52 | 29 | 64 |

Table 8: The micro evaluation of each machine translation program for form-focused text type.

Machine translation programs have displayed different performances while translating form-focused text type.

Google Translate made 52 errors, most of these errors were dictionary-related; 12 expression and word selection errors and 4 not-found words. In addition, verb inflection errors were dominant, and these inflection errors resulted in agreement and arrangement errors.

Proçeviri made the least errors among three MT programs. It made 29 errors, most of which was word and expression selection errors. In addition, the capitalization errors totaled up to 5, which was higher than that of other MT programs. Arrangement problems were also encountered by Proçeviri.



Sametran made 64 errors, more than other MT programs. Most of the errors stemmed from its dictionary; 17 word and expression selection errors, 3 not-found words and 2 category errors. In addition, Sametran made more noun inflection errors than other MT programs. 9 verb inflection errors and the resulting 4 agreement errors made the translation produced by Sametran received very low ratings in macro evaluation.

Form-focused text type, for which the machine translation programs were never considered suitable, has created different results for different programs. In the next section, the performances of MT programs for appeal-focused text type, another text type for which MT programs are considered to be unsuitable, are discussed.

### 6.3. APPEAL-FOCUSED TEXT TYPE

For appeal-focused text type, an internet advertisement of The Plaza, a luxurious New-York Hotel. ("The plaza new,") is chosen. Advertisement consists of 9 sentences and 255 words. The advertisement consists of very long sentences and abounds with adjectives.

The macroevaluation and micro evaluation of outputs of machine translation programs for each sentence are presented below:

| No.                    | Original  | Google Translate                                     | Proçeviri   | Sametran   |
|------------------------|---|--|---|--|
| 1                      | A century ago, The Plaza set the standard for luxury. | Bir yüzyıl önce, Plaza lüks için standart belirledi. | Bir yüzyıl önce, plaza lüks için standartı koydu. | Bir yüzyıl önce, lüks için standart kesin Plaza. |
| <b>Intelligibility</b> |   | Good (40%)   | Good (45%)  | Incomprehensible (60%)                           |
| <b>Fidelity</b>        |   | All (45%)  | Most (45%)  | None (60%)                                       |
| <b>Rank</b>            |   | Best (65%)   | Moderate (50%)                                    | Worst (100%)                                     |

| <b>Macro evaluation</b> | Annotators chose Google Translate’s translation as best among three translations. Google Translate’s translation also received high points in terms of intelligibility and fidelity. Proçeviri was ranked as moderate while all annotators chose Sametran’s translation as worst. Sametran’s translation also received the lowest points in intelligibility and fidelity ratings. |   |  |   |
|-------------------------|---|---|--|---|
| <b>Micro evaluation</b> | Google Translate made a noun inflection error in translating “standard”. The sentence was best translated by Proçeviri, although it was ranked as moderate.<br>Sametran has a rearrangement problem and a category error, mistaking verb “set” as adjective.  |   |  |   |
| <b>No.</b>              | <b>Original</b>   | <b>Google Translate</b>   | <b>Proçeviri</b>   | <b>Sametran</b>   |
| 2                       | Striking a balance between its storied past and limitless future, the passion and uncompromising service, which made the hotel a legend, has returned with a new and contemporary spirit.   | Onun katlı geçmiş ve sınırsız gelecek arasında bir denge, otel bir efsane yapılan tutku ve ödünsüz hizmet, yeni ve çağdaş bir ruhla geri döndü. | Onunkinin arasında bir dengeyi tarihsel geçmiş ve sınırsız geleceğe ayırmak, otele bir efsaneyi yapan tutku ve uzlaşmaz hizmet yeni ve çağdaş bir ruhla döndü. | Ve sınırsız gelecek ve ötesinde destansı arasında bir dengeye çarpmak, uzlaşmaz servis ve tutku, bir gösterge oteli yapan, yeni ve çağdaş bir ruh ile döndürdü. |
| <b>Intelligibility</b>  |   | Disfluent (40%)   | Disfluent (35%)  | Incomprehensibl   |

|                         |   |  |   |   |
|-------------------------|---|--|---|---|
|                         |   |  |   | e (50%)   |
| <b>Fidelity</b>         |   | Little (45%)   | Little (55%)  | None (50%)  |
| <b>Rank</b>             |   | Best (60%)   | Moderate (60%)  | Worst (90%)   |
| <b>Macro evaluation</b> | Translation produced by Google Translate and Proçeviri were both given the same intelligibility and fidelity ratings. However, Google Translate was ranked better than Proçeviri. Sametran was given the lowest scores in all three tests.  |  |   |   |
| <b>Micro evaluation</b> | <p>Google Translate made a word selection error for “storied”, and it made an error in verb inflection for “made” which was translated as passive instead of past tense. In addition, the word selection for “striking” is wrong. Proçeviri made a verb inflection error and a word selection error for “striking”, in addition there is a word selection error in translating “uncompromising”. The noun inflections of “hotel”, “past”, “future” and “legendary” are also wrong. The pronoun “its” was inflected incorrectly. The sentence needs to be rearranged.</p> <p>Sametran made the same word selection for “uncompromising”. “Striking” and “legend” were not translated with a proper word. There is a category error in translating “past”, it was mistaken as preposition instead of noun. In addition, there is an arrangement problem in the sentence. The inflections of the verbs “striking” and “returned” are also wrong. There are also a conjunction error and a preposition error in the sentence.</p> |  |   |   |
| <b>No.</b>              | <b>Original</b>   | <b>Google Translate</b>  | <b>Proçeviri</b>  | <b>Sametran</b>   |
| <b>3</b>                | Following a \$450 million renovation The Plaza Hotel offers 282 distinctive   | Bir 450.000.000 \$ yenileme sonrasında Plaza Hotel 102 suit olmak üzere 282 kendine özgü konuk odaları | Plaza otelinin, 102 takımını kapsayarak, 282 özgün misafir odasını sunar, bir 450 milyon \$ tamirini izleyerek. | Plaza oteline bir \$450 milyon yenileştirmesini izlemek, 102 paketi içermek, karakteristik guestrooms |

|                         |   |  |  |  |
|-------------------------|---|--|--|--|
|                         | guestrooms, including 102 suites.   | sunmaktadır.   |  | 282'yi önerir.   |
| <b>Intelligibility</b>  |   | Non-Native (30%)   | Disfluent (55%)  | Incomprehensible (60%)   |
| <b>Fidelity</b>         |   | Most (45%)   | None (45%)   | None (65%)   |
| <b>Rank</b>             |   | Best (95%)   | Moderate (70%)   | Worst (70%)  |
| <b>Macro evaluation</b> | <p>Translation produced by Google Translate received higher points in all three tests.</p> <p>Sametran was ranked the worst again, and it also received the lowest points for other tests.</p>  |  |  |  |
| <b>Micro evaluation</b> | <p>The sentence was best translated by Google. There is only an adjective inflection error for “\$”</p> <p>Proçeviri made two word selection errors in translating “suites” and “following”, and the sentence needs to be rearranged. The noun inflection of “renovation” and “guestroom” are also wrong.</p> <p>Sametran failed to recognize “guestrooms”, in addition to three word selection errors in translating “suites”, “following” and “offers”. The verb inflection of “following” is not correct. In addition, the adjective “282” was mistaken as noun.</p> |  |  |  |
| <b>No.</b>              | <b>Original</b>   | <b>Google Translate</b>  | <b>Proçeviri</b>   | <b>Sametran</b>  |
| <b>4</b>                | From the sumptuous decor to the impeccable white glove service, The Plaza Hotel returned to   | Kusursuz beyaz eldiven hizmeti görkemli dekor, Plaza Hotel, New York'un en ünlü adreste silinmez anılar oluşturmak için döndü. | Görkemli decordan kusursuz beyaz eldiven hizmetine, plaza oteli New York'un en bilinen adresinde | Kusursuz beyaz eldiven servisine masraflı dekordan, Plaza oteli Yeni York'un en çok kutlanan adresinde |

|                         |  |  |   |   |
|-------------------------|--|--|---|---|
|                         | create indelible memories at New York's most celebrated address.   |  | çıkılmaz anıları oluşturmak için döndü.   | silinmez bellekleri oluşturmaya döndürdü.                                       |
| <b>Intelligibility</b>  |  | Disfluent (45%)  | Non-Native (30%)  | Disfluent (35%)   |
| <b>Fidelity</b>         |  | Much (40%)   | Most (45%)  | None (45%)  |
| <b>Rank</b>             |  | Moderate (75%)   | Best (70%)  | Worst (95%)   |
| <b>Macro evaluation</b> | Proçeviri was ranked the best. Sametran was again given the lowest points for all tests.   |  |   |   |
| <b>Micro evaluation</b> | <p>Google Translate made two preposition mistakes in this sentence. In addition, there is a noun inflection error in “address”.</p> <p>In the translation produced by Proçeviri, there is one word selection error in translating “indelible”. The noun inflection of “memories” was also wrong.</p> <p>Sametran failed to recognize the proper name “New York”. The word selections of “celebrated”, “sumptuous” and “memories” are wrong. In addition, the preposition is not translated correctly. The verb inflection of “returned” is also wrong.</p> |  |   |   |
| <b>No.</b>              | <b>Original</b>  | <b>Google Translate</b>  | <b>Proçeviri</b>  | <b>Sametran</b>   |
| <b>5</b>                | While traditions remain, such as the opulent grandeur of the Beaux   | Gelenekler gibi Beaux Arts dekor zengin ihtişam ve muhteşem bir ambiyans olarak, devam ederken | Görenekler kalırken, mesela beaus sanatlarının zengin azameti efsanevi meşe odası ve meşe | Gelenekler kalırken özen gösterilmiş aydınlatma ve egzotik yeşilliği, havuzları |

|  |   |  |   |                               |
|--|---|--|---|-------------------------------|
| <p>Arts decor and the superb ambience of the legendary Oak Room and Oak Bar, exciting new offerings at The Plaza Hotel New York include the elegant Champagne Bar and stylish Rose Club, as well as an interior Tranquility Garden with reflecting pools, elaborate lighting and exotic foliage.</p> | <p>efsanevi Oak Room ve Mese Bar, Plaza Hotel New York'ta heyecan verici yeni teklifleri, şık Champagne Bar ve şık Gül Kulübü, yanı sıra içerir yansıtıcı havuzları, ayrıntılı aydınlatma ve egzotik bitki örtüsü ile iç Huzur Bahçesi.</p> | <p>barının decor ve şahane ortamı, plaza oteli New York'ta yeni teklifleri heyecanlandırmak zarif şampanya barı ve şık gül kulübünü içerir, iç bir sükunet bahçesine ek olarak havuzları yansıtmakla, ayrıntılı aydınlatma ve egzotik yapraklar.</p> | <p>yansıtmak ile bir iç Sakinlik Bahçe'si kadar iyi, şık gül rengi Kulüp ve zarif Şampanya Bar'ı Plaza otel Yeni York Ekle'sinde yeni teklifler heyecan-verici, efsanevi Meşe Oda'sı ve Meşe Bar'ının muhteşem ambiyansı ve Beaux Sanat'lar dekorunun zengin ihtişamı gibi.</p> |                               |
| <p><b>Intelligibility</b></p>  |   | <p>Incomprehensible (55%)</p>  | <p>Incomprehensible (50%)</p>   | <p>Incomprehensible (70%)</p> |
| <p><b>Fidelity</b></p>   |   | <p>None (45%)</p>  | <p>Little (40%)</p>   | <p>None (65%)</p>             |
| <p><b>Rank</b></p>   |   | <p>Moderate (40%)</p>  | <p>Best (40%)</p>   | <p>Worst (50%)</p>            |

|                         |  |   |   |   |
|-------------------------|--|---|---|---|
| <b>Macro evaluation</b> | This long sentence, which abounds in adjectives and conjunctions, received very low scores. None of the translations was comprehensible, and the fidelity ratings were also very low. Proçeşviri received a higher fidelity rating and it was ranked the best among three translations. Sametran was again given the lowest points.  |   |   |   |
| <b>Micro evaluation</b> | The long sentence created a lot of problems for machine translation programs.<br>Google Translate made an arrangement error. In addition, translation contains five incorrectly inflected nouns; “grandeur”, “ambience”, “decor”, “Club” and “Garden”.<br>Proçeşviri made two category errors in translating “exciting” and “reflecting”, in addition to conjunction and arrangement errors. The word selection for “traditions” is also wrong. There is also an article error. Sametran also had arrangement and conjunction errors. It was unable to recognize “New York” as proper name. In addition, Sametran made a category error in “include”, mistaking it for noun. The inflection and category of “reflecting” are also wrong. |   |   |   |
| <b>No.</b>              | <b>Original</b>  | <b>Google Translate</b>   | <b>Proçeşviri</b>   | <b>Sametran</b>   |
| <b>6</b>                | Old world luxury and elegance are complemented by discreet placement of the latest technology thoughtfully integrated throughout   | Eski dünya lüks ve zarafet düünceli bir şekilde misafirler yüksek çözünürlüklü televizyon, kablosuz yüksek hızda internet erişimi ve bir dokunmatik panel | Eski Dünya lüksü ve zarafet tamamlanır, en son teknolojinin ihtiyatli yerleştirmesi ile dikkatle konuklara yüksek-tanım televizyonunu getiren her oda | Eski Dünya lüks ve zarafet, oda ortamını özelleştirmek için ve bir dokunma panel ekranı her room getiren konuklar yüksek-tanım televizyonu, kablosuzu |

|                         |  |   |   |  |
|-------------------------|--|---|---|--|
|                         | each room bringing guests high-definition television, wireless high-speed internet access, and a touch panel screen to customize the room environment.   | ekran oda ortamı özelleştirmek için getiren her odada boyunca entegre son teknoloji gizli yerleştirme ile tamamlanmaktadır. | boyunca bütünleştirdi, kablosuz yüksek hızlı internet erişimi ve oda çevresini özelleştirmek için bir temas panel ekranı. | boyunca düşünceli bütünleşmiş en son teknolojinin ağız sıkı yerleştirmesi tarafından, son sürat Internet erişimi tamamlanır. |
| <b>Intelligibility</b>  |  | Incomprehensible (35%)  | Incomprehensible (55%)  | Incomprehensible (60%)   |
| <b>Fidelity</b>         |  | Little (45%)  | Little (70%)  | None (40%)   |
| <b>Rank</b>             |  | Best (75%)  | Moderate (50%)  | Worst (45%)  |
| <b>Macro evaluation</b> | Again, all translations of this long sentence received the lowest intelligibility points. Translations produced by Google Translate and Proçeviri were given the same fidelity rating, but Google Translate was ranked higher than Proçeviri.  |   |   |  |
| <b>Micro evaluation</b> | Google Translate had three noun inflection errors; “elegance”, “guests”, and “environment”. There is an unnecessary preposition “in”. In addition, the sentence needs a rearrangement.<br>Proçeviri had also a noun inflection error in translating “elegance” and a verb inflection error in translating “integrated”. The word selection for “environment” is another error. The sentence needs to be rearranged.<br>Sametran failed to recognize the word “room”, and there are two word selection errors in; “discreet”, and “wireless”. In addition, the noun |   |   |  |



| inflections of “elegance” and “guests” are wrong. |  |  |   |   |
|---|--|--|---|---|
| No.   | Original   | Google Translate   | Proçeviri   | Sametran  |
| 7   | Amenities at The Plaza hotel include the world-class Shops at The Plaza, featuring exclusive boutiques and purveyors of fine food, as well as health and wellness facilities including a Caudalie Vinotherapie Spa and Warren Tricomi Salon. | Plaza Hotel'de tesisinde Caudalie Vinotherapie Spa ve Warren Tricomi Salon dahil olmak üzere özel butikler ve kaliteli gıda satıcıları sunan Plaza'da dünya standartlarında Dükkanları, yanı sıra sağlık ve sağlıklı yaşam olanakları bulunmaktadır. | Plaza otelinde konforlar plazada dünya çapında dükkanlar, iyi yiyeceğin özel butikler ve tedarikçilerine yer vermek kapsar, bir Caudalie Vinotherapie kaplıcası ve tavşan kolonisi Tricomi salonunu kapsıyor olan sağlık ve iyilik tesislerine ek olarak. | Dünya-sınıf Plaza otel Ekle'sinde güzellikler, bir Caudalie Vinotherapie Kaplıca'sı ve Tavşanı çok olan yer Tricomi Salon'unu içeren kuyuluk olanakları ve sağlık kadar iyi, güzel yemenin özel kullanım butikler ve erzak müteahhidlerini belirtmek, Plaza'da alışveriş yapar. |
| <b>Intelligibility</b>                            |  | Non-Native (30%)   | Disfluent (45%)   | Incomprehensible (65%)  |
| <b>Fidelity</b>                                   |  | Most (35%)   | None (40%)  | None (65%)  |
| <b>Rank</b>                                       |  | Best (95%)   | Moderate (55%)  | Worst (70%)   |
| <b>Macro evaluation</b>                           | Translation produced by Google Translate received high points in terms of intelligibility and fidelity. Most of the annotators ranked it best.   |  |   |   |

|                         |   |   |   |   |
|-------------------------|---|---|---|---|
|                         | Proçeviri and Sametran received the same fidelity ratings, however, as Sametran's intelligibility rating was the lowest, it was ranked the worst.   |   |   |   |
| <b>Micro evaluation</b> | <p>The long sentence abounding in conjunctions couldn't be translated correctly by any of the machine translation programs.</p> <p>The best of three, translation produced by Google Translate, needs a rearrangement due to the errors in conjunction translation.</p> <p>Proçeviri failed to recognize proper name "Warren Tricomi". In addition, there are two verb inflection errors in translating "featuring" and "including". There is a conjunction error, which resulted in an arrangement error.</p> <p>Sametran also failed to recognize proper name "Warren Tricomi". In addition, there is a category error in "include", it was translated as noun instead of verb. Another category error occurred in translating "shops", which was rendered as a verb instead of noun. Furthermore, the word "wellness", which is not in the dictionary of Sametran, was translated literally as "kuyuluk", which can be regarded as a not-recognized word error. The word selections for "amenities", "purveyors", "featuring", and "food" are wrong. The sentence needs a rearrangement.</p> |   |   |   |
| <b>No.</b>              | <b>Original</b>   | <b>Google Translate</b>   | <b>Proçeviri</b>  | <b>Sametran</b>   |
| <b>8</b>                | As part of the lobby to roof renovation, the opulent Grand Ballroom and Terrace Room have been restored to their original   | Çatı yenileme lobi parçası olarak, zengin Büyük Balo Salonu ve Teras Odası orijinal ihtişamına restore edilmiş. | Çatı tamirine lobinin parçasının olduğu gibi, zengin büyük balo salonu ve teras odası onların orijinal azametine eski haline getirildi. | Yenileştirmeyi üstünü kaplamak için bekleme odasının bölümü olarak, zengin Bin Balo salonu ve Teras Oda'sı orijinal ihtişamlarını |

|                         |   |   |   |  |
|-------------------------|---|---|---|--|
|                         | grandeur.   |   |   | geri getirildiler.   |
| <b>Intelligibility</b>  |   | Non-Native<br>(30%)   | Non-Native<br>(30%)   | Incomprehensible<br>(55%)  |
| <b>Fidelity</b>         |   | Most (35%)  | Little (40%)  | None (50%)   |
| <b>Rank</b>             |   | Best (60%)  | Moderate (60%)  | Worst (75%)  |
| <b>Macro evaluation</b> | <p>Translation produced by Google Translate received high points, it was ranked the best.</p> <p>Sametran's translation again received the lowest points in all three tests.</p>  |   |   |  |
| <b>Micro evaluation</b> | <p>The "lobby to roof renovation" phrase was not translated accurately by any of the machine translation programs.</p> <p>Proçeviri had a word selection error in translating "as".</p> <p>Sametran made two word selection errors in translating "Grand" and "roof". "Roof" was rendered as a verb, thus scoring a category error.</p> <p>There is also a noun inflection error in translating "grandeur", which is the result of the absence of the preposition "to".</p> |   |   |  |
| <b>No.</b>              | <b>Original</b>   | <b>Google Translate</b>   | <b>Proçeviri</b>  | <b>Sametran</b>  |
| <b>9</b>                | Lavish social and business events will continue to be held in these historic landmarks, which are at the heart of The Plaza Hotel's legacy.   | Cömert, sosyal ve iş etkinlikleri Plaza Hotel'in mirasının kalbi olan bu tarihi yerlerde, düzenlenecek devam edecektir. | Bol sosyal ve iş olayları plaza otelinin mirasının kalbinde olan bu tarihi dönüm noktalarında yapılmaya devam edecek. | Bol sohbetli toplantısı ve iş olayları, Plaza otelin eskisinin kalbinde olan, bu tarihi dönüm noktalarında tutulacak devam edecek. |

|                         |   |                 |            |                        |
|-------------------------|---|-----------------|------------|------------------------|
| <b>Intelligibility</b>  |   | Disfluent (35%) | Good (25%) | Incomprehensible (40%) |
| <b>Fidelity</b>         |   | Much (45%)      | Most (30%) | Little (40%)           |
| <b>Rank</b>             |   | Moderate (45%)  | Best (70%) | Worst (70%)            |
| <b>Macro evaluation</b> | <p>Proçeviri's translation received high points in intelligibility and fidelity, and it was ranked the best among three.</p> <p>Translation produced by Sametran received lowest point for intelligibility, and it was again ranked the worst.</p>  |                 |            |                        |
| <b>Micro evaluation</b> | <p>Google Translate made an error in attributing the conjunction "and". In addition, it has a verb inflection error in translating "will continue to be held".</p> <p>The sentence was best translated by Proçeviri. There are only two word selection errors in "lavish" and "events".</p> <p>Sametran also made a verb inflection error as well as a word selection error for "will continue to be held". In addition, the word selections for "historic", "lavish" and "events" are wrong.</p> |                 |            |                        |

Table 9: The macro evaluation of each machine translation program for appeal-focused text type.

Below is presented the table which contains the times a sentence was ranked the best, worst or moderate, the intelligibility and fidelity ratings:

|                        | Google Translate | Proçeviri | Sametran |
|------------------------|------------------|-----------|----------|
| <b>Ranking</b>         |                  |           |          |
| <b># Best</b>          | 6                | 3         | 0        |
| <b># Moderate</b>      | 3                | 6         | 0        |
| <b># Worst</b>         | 0                | 0         | 9        |
| <b>Intelligibility</b> |                  |           |          |
| <b>Flawless</b>        | 0                | 0         | 0        |

|                         |   |   |   |
|-------------------------|---|---|---|
| <b>Good</b>             | 1 | 2 | 0 |
| <b>Non-Native</b>       | 3 | 2 | 0 |
| <b>Disfluent</b>        | 3 | 3 | 1 |
| <b>Incomprehensible</b> | 2 | 2 | 8 |
| <b>Fidelity</b>         |   |   |   |
| <b>All</b>              | 1 | 0 | 0 |
| <b>Most</b>             | 3 | 3 | 0 |
| <b>Much</b>             | 2 | 0 | 0 |
| <b>Little</b>           | 2 | 4 | 1 |
| <b>None</b>             | 1 | 2 | 8 |

Table 10: The macro evaluation of each machine translation program for appeal-focused text type.

Appeal-focused text, which was exemplified with an online hotel advertisement, was generally best translated by Google Translate. Out of 9 sentences, 6 sentences produced by Google Translate and 3 sentences produced by Proçeviri were ranked the best. 3 sentences produced by Google Translate and 6 sentences produced by Proçeviri were ranked as moderate. All of the sentences translated by Sametran were ranked the worst.

None of the sentences in the appeal-focused text was given the highest intelligibility rating. Only 1 sentence by Google Translate and 2 sentences by Proçeviri were rated as “good”. 3 sentences by Google Translate and 2 sentences by Proçeviri were rated as “non-native”. 3 sentences by Google Translate and Proçeviri and 1 sentence by Sametran were rated as “disfluent”. 8 sentences produced by Sametran were given the lowest intelligibility rating. This rating was also given to 2 sentences produced by Google Translate and Proçeviri.

In terms of fidelity, 1 sentence produced by Google Translate was given the highest rating “all”. 3 sentences produced by Google Translate and Proçeviri were rated as “most”. 2 sentences translated by Google Translate was rated as “much”. 2 sentences by Google Translate, 4 sentences by Proçeviri and 1 sentence by Sametran were rated as “little”. The

lowest fidelity rating “none” was given to 8 sentences by Sametran, 2 sentences by Proçeviri and 1 sentence by Google Translate.

The performance of each machine translation program for micro evaluation (grammatical error analysis) can be seen in the below table:

| <b>Error Category</b>   | <b>Google Translate</b> | <b>Proçeviri</b> | <b>Sametran</b> |
|-------------------------|-------------------------|------------------|-----------------|
| <b>Spelling</b>         | 0                       | 0                | 0               |
| <b>Not-found Word</b>   | 0                       | 1                | 6               |
| <b>Capitalization</b>   | 0                       | 0                | 0               |
| <b>Elision</b>          | 0                       | 0                | 0               |
| <b>Verb inflection</b>  | 3                       | 4                | 5               |
| <b>Noun inflection</b>  | 7                       | 8                | 2               |
| <b>Other inflection</b> | 4                       | 0                | 2               |
| <b>Rearrangement</b>    | 4                       | 5                | 3               |
| <b>Category</b>         | 0                       | 2                | 9               |
| <b>Pronoun</b>          | 0                       | 1                | 0               |
| <b>Article</b>          | 0                       | 1                | 0               |
| <b>Preposition</b>      | 4                       | 1                | 4               |
| <b>Negative</b>         | 0                       | 0                | 0               |
| <b>Conjunction</b>      | 2                       | 2                | 2               |
| <b>Agreement</b>        | 0                       | 0                | 0               |
| <b>Clause boundary</b>  | 0                       | 0                | 0               |
| <b>Word Selection</b>   | 1                       | 10               | 21              |
| <b>Expression</b>       | 0                       | 0                | 0               |
| <b>Total</b>            | 25                      | 35               | 54              |

Table 11: The micro evaluation of each machine translation program for appeal-focused text type.

Micro evaluation results have many parallels to the macro evaluation results.

Google Translate made the least errors, and most of which were inflection errors. Apart from these, conjunction and preposition errors were dominant.

Proçeviri made 35 errors. Dictionary-related errors were in majority; while there were also inflection errors.

Sametran made 54 errors, 36 of these errors stemmed from its dictionary. There were 9 category errors, and 9 inflection errors.

Appeal-focused text type, which was exemplified with an online hotel advertisement, has been translated very differently by different machine translation programs. In the next section, the performances of MT programs for audio-medial text type, which is exemplified with subtitles, are discussed.

#### **6.4. AUDIO-MEDIAL TEXT TYPE**

For audio-medial text type, an extract from the subtitles of “The Big Bang Theory” Season 8 Episode 1 is chosen. The extract consists of 28 sentences and 259 words. Subtitles include proper names, colloquial language, and idiomatic expressions.

The season premier starts with a scene where two characters, Leonard and Penny, see off Sheldon in a terminal, who decides to leave his friends and life after he witnesses lots of changes in a short time. In the second scene, Sheldon, without any pants, asks people for help by explaining his situation.

The macroevaluation and micro evaluation of outputs of machine translation programs for each sentence are presented below:

| <b>No.</b> | <b>Original</b> | <b>Google Translate</b>  | <b>Proçeviri</b>   | <b>Sametran</b>     |
|------------|-----------------|--------------------------|--------------------|---------------------|
| <b>1</b>   | We were worried | Senin hakkında endişeli. | Biz senin hakkında | Biz, senin hakkında |

|                         |  |                         |                      |                           |
|-------------------------|--|-------------------------|----------------------|---------------------------|
|                         | about you.   |                         | endişeliydik.        | endişelenildik.           |
| <b>Intelligibility</b>  |  | Incomprehensible (35%)  | Good (55%)           | Disfluent (35%)           |
| <b>Fidelity</b>         |  | Little (45%)            | All (50%)            | None (35%)                |
| <b>Rank</b>             |  | Moderate (70%)          | Best (100%)          | Worst (75%)               |
| <b>Macro evaluation</b> | All of the annotators ranked the translation produced by Proçeviri as best of them. The translation produced by Google Translate was ranked as moderate, as its intelligibility rating was the lowest. Sametran's translation ranked the worst, and it was rated with the lowest fidelity point. |                         |                      |                           |
| <b>Micro evaluation</b> | The sentence was best translated by Proçeviri.<br>Google Translate made a verb inflection error, which resulted in an agreement error.<br>Sametran also made the same verb inflection error, thus also scoring an agreement error.   |                         |                      |                           |
| <b>No.</b>              | <b>Original</b>  | <b>Google Translate</b> | <b>Proçeviri</b>     | <b>Sametran</b>           |
| 2                       | Don't be melodramatic  | Melodramatik olmayın.   | Heyecan verici olma. | Aşırı duygusal olmayınız. |
| <b>Intelligibility</b>  |  | Flawless (45%)          | Non-Native (30%)     | Flawless (50%)            |
| <b>Fidelity</b>         |  | Most (45%)              | None (60%)           | Most (42%)                |
| <b>Rank</b>             |  | Moderate (35%)          | Worst (80%)          | Best (50%)                |
| <b>Macro evaluation</b> | Translations produced by Google Translate and Sametran were both given the same intelligibility and fidelity ratings. However, Sametran was ranked the best. Translation produced by Proçeviri was given the lowest fidelity point and it was ranked the worst.                                  |                         |                      |                           |
| <b>Micro evaluation</b> | The sentence was translated accurately by both Google Translate and Sametran.  |                         |                      |                           |



| Proçeviri made a word selection error. |   |   |  |  |
|--|---|---|--|--|
| No.                                    | Original  | Google  | Proçeviri  | Sametran   |
| 3                                      | I'm just getting on a train and leaving forever.  | Ben sadece bir trene biniyor ve sonsuza kadar gidiyorum.                                | Ben sadece bir trene binyorum ve sonsuza dek terk ediyorum.                                    | Şimdi ben, bir treniyorum ve sonsuza dek bırakıyorum.  |
| <b>Intelligibility</b>                 |   | Good (30%)  | Non-Native (35%)   | Incomprehensible (65%)   |
| <b>Fidelity</b>                        |   | Most (35%)  | Much (30%)   | None (55%)   |
| <b>Rank</b>                            |   | Moderate (50%)  | Best (65%)   | Worst (95%)  |
| <b>Macro evaluation</b>                | Translation produced by Sametran was given the lowest points and it was ranked the worst by the majority of annotators. Google Translate's translation was given higher ratings in terms of intelligibility and fidelity than Proçeviri's translation. However, Proçeviri was ranked the best among three translations. |   |  |  |
| <b>Micro evaluation</b>                | The sentence was accurately translated by both Google Translate and Proçeviri.<br>Sametran made a category error and verb inflection error in translating "getting on a train"  |   |  |  |
| No.                                    | Original  | Google Translate  | Proçeviri  | Sametran   |
| 4                                      | So a few things don't go your way. And your best decision is to ride the rails like a   | Yani bir kaç şey yolunuzu gitmez. Ve en iyi kararın bir serseri gibi raylar sürmektir ? | Yani birkaç sey senin yoluna gitmez. Ve senin en iyi kararın bir aylak gibi rayları sürmektir? | Öyleyse az şey yolunuza gitmezler. Parmaklıklara binmek için en iyi kararın ve, bir aylaktan hoşlanıyor? |

|                         |  |                              |                        |                            |
|-------------------------|--|------------------------------|------------------------|----------------------------|
|                         | hobo?  |                              |                        |                            |
| <b>Intelligibility</b>  |  | Incomprehensible (50%)       | Disfluent (35%)        | Incomprehensible (75%)     |
| <b>Fidelity</b>         |  | None (45%)                   | None (30%)             | None (80%)                 |
| <b>Rank</b>             |  | Best (45%)                   | Moderate (50%)         | Worst (100%)               |
| <b>Macro evaluation</b> | The translations of this sentence were given the lowest fidelity ratings. Google Translate's and Sametran's translations were also given the lowest intelligibility ratings. Although Proçeviri received a higher intelligibility result than Google Translate, Google Translate was ranked the best while Proçeviri was ranked as moderate.   |                              |                        |                            |
| <b>Micro evaluation</b> | The sentence created many problems for machine translation programs. Google Translate made an expression selection error for "your way". In addition, it translated the expression "ride the rails" literally, thus scoring another expression selection error. Proçeviri made the same expression selection errors. Apart from above expression selection errors, Sametran made another error in translating "like", thus scoring a word selection error. |                              |                        |                            |
| <b>No.</b>              | <b>Original</b>  | <b>Google Translate</b>      | <b>Proçeviri</b>       | <b>Sametran</b>            |
| <b>5</b>                | Leonard, I am overwhelmed.   | Leonard, ben bunalmışımdır . | Leonard, ben ezileniz. | Leonard, ben alt edilirim. |
| <b>Intelligibility</b>  |  | Non-Native (55%)             | Incomprehensible (65%) | Non-Native (35%)           |
| <b>Fidelity</b>         |  | Most (35%)                   | None (70%)             | None (65%)                 |
| <b>Rank</b>             |  | Best (80%)                   | Worst (85%)            | Moderate (60%)             |
| <b>Macro evaluation</b> | The translation produced by Google Translate was ranked the best by the most of the annotators.  |                              |                        |                            |
| <b>Micro evaluation</b> | The sentence was best translated by Google Translate. It made only a verb inflection error.  |                              |                        |                            |

|                         |  |   |   |   |
|-------------------------|--|---|---|---|
|                         | Proçeviri made a verb inflection and a concomitant agreement error.<br>Sametran made a word selection error.   |   |   |   |
| <b>No.</b>              | <b>Original</b>  | <b>Google Translate</b>                 | <b>Proçeviri</b>                                | <b>Sametran</b>                                     |
| <b>6</b>                | Everything is changing, and it is simply too much.   | Her şey deęiřiyor, ve sadece çok fazla. | Her şey deęiřtiriyor ve o basitçe çok fazladır. | Her şey, deęiřtiriyor ve çok yalın řekilde o.       |
| <b>Intelligibility</b>  |  | Disfluent (40%)                         | Disfluent (40%)                                 | Incomprehensible (80%)                              |
| <b>Fidelity</b>         |  | Much (31%)                              | None (35%)                                      | None (75%)  |
| <b>Rank</b>             |  | Best (80%)                              | Moderate (90%)                                  | Worst (100%)  |
| <b>Macro evaluation</b> | Translation produced by Sametran was ranked the worst by all of the annotators. It also received the lowest intelligibility and fidelity rating. Google Translate's and Proçeviri's translations were given the same intelligibility ratings. As Google Translate's translation was found more faithful, it was ranked the best. |   |   |   |
| <b>Micro evaluation</b> | Google Translate has an absent pronoun.<br>Proçeviri made a verb inflection error.<br>Translation produced by Sametran needs a rearrangement.  |   |   |   |
| <b>No.</b>              | <b>Original</b>  | <b>Google Translate</b>                 | <b>Proçeviri</b>                                | <b>Sametran</b>                                     |
| <b>7</b>                | I need to get away and think.  | Ben uzaklařmak ve dūřünmek gerekir.     | Ben uzaklařmak ve dūřünmeye ihtiyaç duyarım.    | Ben, deplasman ve dūřünmeyi almaya ihtiyaç duyarım. |
| <b>Intelligibility</b>  |  | Disfluent (55%)                         | Non-Native (35%)                                | Incomprehensible (70%)                              |
| <b>Fidelity</b>         |  | Little (45%)                            | Most (40%)                                      | None (78%)  |
| <b>Rank</b>             |  | Moderate (95%)                          | Best (100%)                                     | Worst (95%)   |

|                         |  |                         |                       |                              |
|-------------------------|--|-------------------------|-----------------------|------------------------------|
| <b>Macro evaluation</b> | Translation produced by Proçeviri was ranked the best by all of the annotators. Most of the annotators ranked Google Translate as moderate and Sametran as worst. Sametran also received the lowest ratings.   |                         |                       |                              |
| <b>Micro evaluation</b> | Google Translate made an agreement error.<br>Proçeviri made a conjunction error, which resulted in a verb inflection error.<br>Sametran made an expression selection error, not recognizing “get away” as a phrasal verb. In addition, there is a conjunction error. |                         |                       |                              |
| <b>No.</b>              | <b>Original</b>  | <b>Google Translate</b> | <b>Proçeviri</b>      | <b>Sametran</b>              |
| <b>8</b>                | Sheldon...   | Sheldon ...             | Sheldon...            | Sheldon...                   |
| <b>9</b>                | Yes?   | Evet?                   | Evet?                 | Evet?                        |
| <b>Macro evaluation</b> | These sentences were excluded from macro evaluation as they were translated identically by MT programs.  |                         |                       |                              |
| <b>Micro evaluation</b> | Each program recognized the proper name successfully.  |                         |                       |                              |
| <b>No.</b>              | <b>Original</b>  | <b>Google Translate</b> | <b>Proçeviri</b>      | <b>Sametran</b>              |
| <b>10</b>               | I am gonna miss you.   | Seni ben seni özlerim.  | Ben seni özleyeceğim. | Ben, siz gonna bayanlarıyım. |
| <b>Intelligibility</b>  |  | Non-Native (45%)        | Flawless (70%)        | Incomprehensible (90%)       |
| <b>Fidelity</b>         |  | Much (30%)              | All (90%)             | None (95%)                   |
| <b>Rank</b>             |  | Moderate (95%)          | Best (100%)           | Worst (100%)                 |
| <b>Macro evaluation</b> | All of the annotators ranked Proçeviri as best and Sametran as worst. Translation produced by Proçeviri received highest ratings.  |                         |                       |                              |
| <b>Micro evaluation</b> | The sentence was best translated by Proçeviri.<br>Google Translate made pronoun error.<br>Sametran has a non recognized word “gonna”, which resulted in a category error for “miss”. The result of these errors is a verb inflection error.                          |                         |                       |                              |

| No.                         | Original   | Google Translate        | Proçeviri                            | Sametran                           |
|-----------------------------|--|-------------------------|--------------------------------------|------------------------------------|
| 11                          | Of course<br>you are.  | Sen tabii ki .          | Sen tabiisin.                        | Kursunsun.                         |
| <b>Intelligibility</b>      |  | Flawless (35%)          | Incomprehensible<br>(55%)            | Incomprehensible<br>(95%)          |
| <b>Fidelity</b>             |  | Little (55%)            | Most (30%)                           | None (95%)                         |
| <b>Rank</b>                 |  | Best (100%)             | Moderate (85%)                       | Worst (100%)                       |
| <b>Macro<br/>evaluation</b> | The sentence, which contains an ellipsis to the previous sentence, was best translated by Google Translate and worst translated by Sametran.   |                         |                                      |                                    |
| <b>Micro<br/>evaluation</b> | The inability of machine translation programs to recognize ellipsis between current and previous sentence resulted in failure to translate this short sentence. Google Translate and Proçeviri made verb inflection error, while Sametran made an expression selection error, failing to recognize “of course” as an expression. |                         |                                      |                                    |
| No.                         | Original   | Google Translate        | Proçeviri                            | Sametran                           |
| 12                          | You just<br>made that<br>easier.   | Sadece kolay<br>yaptın. | Sen sadece onu<br>daha kolay yaptın. | Sen şimdi, o daha<br>kolay yaptın. |
| <b>Intelligibility</b>      |  | Disfluent (40%)         | Good (25%)                           | Disfluent (50%)                    |
| <b>Fidelity</b>             |  | Little (55%)            | Most (30%)                           | None (50%)                         |
| <b>Rank</b>                 |  | Moderate (75%)          | Best (85%)                           | Worst (80%)                        |
| <b>Macro<br/>evaluation</b> | Proçeviri’s translation was ranked the best. Translations produced by Google Translate and Sametran received the same intelligibility rating, but as Sametran received the lowest fidelity rating, it was rated worst.   |                         |                                      |                                    |
| <b>Micro<br/>evaluation</b> | The sentence was translated best by Proçeviri.<br>Google Translate made an error in inflecting the adverb and there is an absent pronoun.<br>Sametran has made a pronoun mistake as well as a noun inflection error.   |                         |                                      |                                    |
| No.                         | Original   | Google Translate        | Proçeviri                            | Sametran                           |

|                         |  |  |  |   |
|-------------------------|--|--|--|---|
| <b>13</b>               | Excuse me.   | Afedersiniz.                                   | Beni mazur gör.  | Beni bağışlayın.  |
| <b>Intelligibility</b>  |  | Flawless (85%)                                 | Flawless (55%)   | Flawless (65%)  |
| <b>Fidelity</b>         |  | All (70%)                                      | All (60%)  | All (55%)   |
| <b>Rank</b>             |  | Best (100%)                                    | Moderate (40%)<br>Worst (35%)  | Moderate (40%)<br>Worst (35%)                             |
| <b>Macro evaluation</b> | This short sentence received the highest ratings in terms of intelligibility and fidelity. All annotators chose Google Translate's translation as best among the three. Proçeviri and Sametran were both ranked as moderate and worst. |  |  |   |
| <b>Micro evaluation</b> | The sentence was translated accurately by all programs.  |  |  |   |
| <b>No.</b>              | <b>Original</b>  | <b>Google Translate</b>                        | <b>Proçeviri</b>   | <b>Sametran</b>   |
| <b>14</b>               | Is it at all possible that you are knitting a pair of pants?   | Eğer bir pantolon örme olduğunu tüm mümkün mü? | O senin bir pantolonu örüyor olduğun mümkün herhangi bir şekilde mi? | Bir çift pantolonu örüyor olduğun o bütün olasıdır mıdır? |
| <b>Intelligibility</b>  |  | Incomprehensible (60%)                         | Disfluent (50%)  | Disfluent (55%)   |
| <b>Fidelity</b>         |  | None (65%)                                     | None (45%)   | Little (45%)  |
| <b>Rank</b>             |  | Worst (65%)                                    | Moderate (40%)   | Best (40%)  |
| <b>Macro evaluation</b> | Translation produced by Google Translate received the lowest ratings. Proçeviri's and Sametran's translation were given the same intelligibility ratings, but as Sametran was perceived as more faithful, it was ranked the best.      |  |  |   |
| <b>Micro evaluation</b> | Proçeviri made a noun inflection error for "pants" and the sentence needs to be rearranged.<br>Google Translate made an expression selection error for "at all". In  |  |  |   |

|                         |  |   |  |  |
|-------------------------|--|---|--|--|
|                         | <p>addition, the sentence needs to be rearranged.</p> <p>Sametran made a preposition error, attributing “at” to “possible”. In addition, the sentence needs to be rearranged.</p>  |   |  |  |
| <b>No.</b>              | <b>Original</b>  | <b>Google Translate</b>                             | <b>Proçeviri</b>   | <b>Sametran</b>  |
| <b>15</b>               | Oh, well, no, you are understandably terrified.  | Oh, hayır, sen anlaşılır korkmuş durumdasın.        | Oh, şey, hayır, sen anlaşılır şekilde dehşete düşmüştün. | Öyle mi? ,iyi, yok, anlaşılır şekilde sen korkutulursun. |
| <b>Intelligibility</b>  |  | Disfluent (40%)                                     | Good (35%)   | Incomprehensible (63%)                                   |
| <b>Fidelity</b>         |  | Little (35%)  | All (40%)  | None (80%)   |
| <b>Rank</b>             |  | Moderate (90%)                                      | Best (90%)   | Worst (100%)   |
| <b>Macro evaluation</b> | Translation produced by Sametran was again given the lowest ratings. All annotators ranked its translation worst.  |   |  |  |
| <b>Micro evaluation</b> | <p>The sentence was best translated by Proçeviri.</p> <p>Google Translate made a mistake in inflecting the adverb “understandably”.</p> <p>Sametran made a word selection error for interjection “well”. In addition, it failed to recognize the expression “be terrified”, translating it as passive voice, which can be regarded as a verb inflection error.</p> |   |  |  |
| <b>No.</b>              | <b>Original</b>  | <b>Google Translate</b>                             | <b>Proçeviri</b>   | <b>Sametran</b>  |
| <b>16</b>               | But, you know, allow me to explain.  | Ama, sen biliyorsun, ben açıklamak için izin verir. | Ama, sen biliyorsun, açıklamam için bana izin ver.       | Fakat, sen bilirsin, bana açıklamak için müsaade ediniz. |
| <b>Intelligibility</b>  |  | Disfluent (55%)                                     | Good (40%)   | Good (45%)   |
| <b>Fidelity</b>         |  | Little (45%)  | All (35%)  | Most (35%)   |
| <b>Rank</b>             |  | Worst (90%)   | Best (75%)   | Moderate (60%)   |
| <b>Macro</b>            | Translation produced by Google Translate was ranked the worst.   |   |  |  |

|                         |  |  |  |   |
|-------------------------|--|--|--|---|
| <b>evaluation</b>       | Proçeviri's and Sametran's translation were given the same intelligibility ratings, however, as Proçeviri was given the highest fidelity ratings, it was ranked the best among three translations.   |  |  |   |
| <b>Micro evaluation</b> | The expression “you know” was literally translated by all of the programs, which can be regarded as an expression selection error.<br>Google Translate made a pronoun error as well as a verb inflection error, which resulted in an agreement error.<br>Sametran made a verb inflection error for “explain”.  |  |  |   |
| <b>No.</b>              | <b>Original</b>  | <b>Google Translate</b>  | <b>Proçeviri</b>   | <b>Sametran</b>   |
| <b>17</b>               | 45 days ago, um, I embarked on a railroad journey of healing.  | 45 gün önce, um, ben bir iyileşme demiryolu yolculuğuna başladı. | 45 gün önce, um, ben bir iyileşme demiryolu seyahatine girdik. | 45 gün önce, um, ben iyileştirmenin bir demiryolu seyahatinde yükledim. |
| <b>Intelligibility</b>  |  | Disfluent (45%)  | Disfluent (40%)  | Incomprehensible (75%)  |
| <b>Fidelity</b>         |  | Little (35%)   | None (50%)   | None (70%)  |
| <b>Rank</b>             |  | Best (45%)   | Moderate (60%)   | Worst (85%)   |
| <b>Macro evaluation</b> | Translations of this sentence received very low ratings. Sametran was given the lowest ratings in all three tests.   |  |  |   |
| <b>Micro evaluation</b> | None of the machine translation programs could translate the sentence accurately.<br>Google Translate made a verb inflection error for “embarked”, and a concomitant agreement error.<br>Proçeviri made also the same verb inflection and agreement error.<br>Sametran made a noun inflection error for “healing”, in addition to expression selection error for “embark on”, and a preposition error. |  |  |   |
| <b>No.</b>              | <b>Original</b>  | <b>Google Translate</b>  | <b>Proçeviri</b>   | <b>Sametran</b>   |



|                         |   |  |   |   |
|-------------------------|---|--|---|---|
| <b>18</b>               | Because my university was making me do string theory, and my favorite comic book store burned down.   | Benim üniversite yapma Çünkü bana sicim teorisini yapmak, ve benim en sevdiğim çizgi roman mağaza yandı. | Benim üniversitem sicim teorisini yapmam için beni yapıyor olduğu için ve benim favori çizgi roman depomun yanıp kül olduğu için. | Çünkü üniversitem teoriyi gerdiğini yaptığımı beni yapıyordu, ve favori komik kitapçımdan aşağısında yazdı. |
| <b>Intelligibility</b>  |   | Disfluent (50%)  | Incomprehensible (45%)  | Incomprehensible (75%)  |
| <b>Fidelity</b>         |   | None (50%)   | Little (45%)  | None (80%)  |
| <b>Rank</b>             |   | Moderate (45%)   | Best (50%)  | Worst (85%)   |
| <b>Macro evaluation</b> | Although Proçeviri received a worse intelligibility rating than Google Translate, it was ranked the best among three as it was rated higher in terms of fidelity. Sametran was again given the lowest points in each test.  |  |   |   |
| <b>Micro evaluation</b> | The machine translation programs couldn't translate the expression "making me do", which can be regarded as expression selection error. Google Translate made a noun inflection error for "store". Proçeviri failed to inflect the verb "burned down". Sametran made three expression selection errors for "string theory", "comic books store" and "burned down". It has also a preposition error, not recognizing "down" as a part of phrasal verb "burned down". |  |   |   |
| <b>No.</b>              | <b>Original</b>   | <b>Google Translate</b>  | <b>Proçeviri</b>  | <b>Sametran</b>   |
| <b>19</b>               | And when my roommate got engaged, my  | Oda arkadaşım nişanlandı zaman, benim kız arkadaşım bir manevra sadece                                   | Ve benim oda arkadaşımın nişanlandığı zaman, benim kız arkadaşım benle  | Ve roommatemin, nişanlı aldığında, kız arkadaşım benim görmek için şimdi bir                                |

|                         |   |   |   |  |
|-------------------------|---|---|---|--|
|                         | girlfriend wanted to move in with me, which was no doubt a ploy just to see my, well, excuse my language, but my bathing suit parts.  | benim, iyi, benim dil bahane görmek için hiç şüphesiz benim, ama benim mayo parçaları ile taşımak istedi. | aynı eve taşınmayı istiyordu ki bu sadece görmek için şüphesiz bir hileydi benim, iyi, mazeret benim lisanım ama benim mayo kısımlarım. | manevra şüphe olmayanı, benimlede hareket etmeyi istedi, iyi, fakat yıkama dilek bölümlerim dilimi bağışlayın. |
| <b>Intelligibility</b>  |   | Incomprehensible (75%)  | Disfluent (40%)   | Incomprehensible (40%)   |
| <b>Fidelity</b>         |   | None (80%)  | Little (55%)  | None (80%)   |
| <b>Rank</b>             |   | Moderate (50%)  | Best (50%)  | Worst (60%)  |
| <b>Macro evaluation</b> | Google Translate and Sametran were both given the same intelligibility and fidelity ratings, but Google Translate was ranked as moderate and Sametran was again ranked the worst among three.   |   |   |  |
| <b>Micro evaluation</b> | <p>The long sentence, with an embedded sentence, created lots of problems for machine translation programs.</p> <p>Google Translate made a verb inflection error for “got engaged”. In addition, the expression selections for “excuse my language” and “move in with” are wrong. There is a noun inflection error in “parts”. The sentence needs to be rearranged.</p> <p>Proçeviri also made the same expression selection error for “excuse my language”. There are two noun inflection errors in “roommate” and</p> |   |   |  |

|                         |  |   |   |                                      |
|-------------------------|--|---|---|--------------------------------------|
|                         | <p>“parts”. The sentence needs to be rearranged.</p> <p>Sametran has an untranslated word “roommate”. It made two expression selection errors “got engaged” and “move in with”. Word selection for adverb “just” is also wrong. In addition, there is a preposition error in “with me”. The sentence needs to be rearranged.</p> |   |   |                                      |
| <b>No.</b>              | <b>Original</b>  | <b>Google Translate</b>                               | <b>Proçeviri</b>                                      | <b>Sametran</b>                      |
| <b>20</b>               | Uh, sir, may I use your phone?   | Ah, efendim, ben senin telefonunu kullanabilir miyim? | Uh, efendim, ben senin telefonunu kullanabilir miyim? | Uh sör?, telefonunu kullanabilirsin? |
| <b>Intelligibility</b>  |  | Non-Native (40%)                                      | Good (40%)  | Incomprehensible (73%)               |
| <b>Fidelity</b>         |  | All (35%)   | Most (40%)  | None (90%)                           |
| <b>Rank</b>             |  | Best (85%)  | Moderate (80%)  | Worst (100%)                         |
| <b>Macro evaluation</b> | Sametran received the lowest points in all three tests. In addition, while Proçeviri was given a better intelligibility rating than Google Translate, Google Translate was ranked the best.  |   |   |                                      |
| <b>Micro evaluation</b> | The sentence was accurately translated by both Google Translate and Proçeviri. Sametran made a subject-verb agreement error as well as word selection error in translating “sir”.  |   |   |                                      |
| <b>No.</b>              | <b>Original</b>  | <b>Google Translate</b>                               | <b>Proçeviri</b>                                      | <b>Sametran</b>                      |
| <b>21</b>               | I don't think so.  | Ben öyle düşünmüyorum.                                | Ben düşünmüyorum yani.                                | Ben, öyleyseyi düşünmem.             |
| <b>Intelligibility</b>  |  | Flawless (75%)  | Non-Native (35%)                                      | Incomprehensible (65%)               |
| <b>Fidelity</b>         |  | All (70%)   | None (50%)  | None (90%)                           |
| <b>Rank</b>             |  | Best (100%)   | Moderate (85%)  | Worst (90%)                          |
| <b>Macro</b>            | The sentence was best translated by Google Translate, it received the  |   |   |                                      |

|                         |  |   |  |   |
|-------------------------|--|---|--|---|
| <b>evaluation</b>       | highest ratings in all three tests. Sametran was again given the lowest points.  |   |  |   |
| <b>Micro evaluation</b> | The expression “I don’t think so” was translated literally by machine translation programs, thus each scored another expression selection error.   |   |  |   |
| <b>No.</b>              | <b>Original</b>  | <b>Google Translate</b>   | <b>Proçeviri</b>   | <b>Sametran</b>   |
| <b>22</b>               | Yeah, well, I understand that I'm half naked, but there is a reasonable explanation.   | Evet, ben yarı çıplak olduğumu anlıyorum, ama makul bir açıklaması var. | Evet, şey, ben benim kısmen çıplak olduğumu anlıyorum ama makul bir açıklama vardır. | Evet, iyi, ben yarı ispatsız olduğumu anlarım, fakat bir makul izah vardır. |
| <b>Intelligibility</b>  |  | Good (40%)  | Non-Native (40%)   | Incomprehensible (50%)  |
| <b>Fidelity</b>         |  | Most (35%)  | Most (45%)   | None (60%)  |
| <b>Rank</b>             |  | Best (75%)  | Moderate (80%)   | Worst (95%)   |
| <b>Macro evaluation</b> | Translations produced by Google Translate and Proçeviri received the same fidelity ratings. Google Translate was ranked the best as its intelligibility rating was higher than Proçeviri's.                                  |   |  |   |
| <b>Micro evaluation</b> | The sentence was best translated by Google Translate. Proçeviri made a noun inflection error in “explanation”. Sametran made a word selection error for “naked”. In addition, it made the same noun inflection as Proçeviri. |   |  |   |
| <b>No.</b>              | <b>Original</b>  | <b>Google Translate</b>   | <b>Proçeviri</b>   | <b>Sametran</b>   |
| <b>23</b>               | While I slept in my sleeper car, all my possessions  | Benim uyuyan arabada yattım iken, tüm eşyalarını çalındı.               | Ben benim uyuyan kişi vagonumda uyurken, benim bütün mallarım                        | Uykucu arabamda uyuyurken bütün iyeliklerim, çalındı.                       |

|                         |   |   |   |   |
|-------------------------|---|---|---|---|
|                         | were stolen.  |   | çalındı.  |   |
| <b>Intelligibility</b>  |   | Incomprehensible (40%)  | Non-Native (35%)  | Incomprehensible (60%)  |
| <b>Fidelity</b>         |   | Little (35%)  | Little (40%)  | None (70%)  |
| <b>Rank</b>             |   | Moderate (50%)  | Best (57%)  | Worst (90%)   |
| <b>Macro evaluation</b> | The sentence was best translated by Proçeviri. Google Translate was ranked as moderate thanks to its higher fidelity rating.  |   |   |   |
| <b>Micro evaluation</b> | The “sleeper car” was not translated accurately by machine translation programs, which can be regarded as expression selection error. Apart from this error, Google Translate made a verb inflection error for “slept”. In addition, the noun inflection for “possessions” is wrong. Proçeviri has an unnecessary pronoun. Sametran also made a word selection error in translating “possessions”. In addition, the verb inflection for “slept” is wrong. |   |   |   |
| <b>No.</b>              | <b>Original</b>   | <b>Google Translate</b>   | <b>Proçeviri</b>  | <b>Sametran</b>   |
| <b>24</b>               | Now, typically, I wear pajamas, but I recently adopted a hobo lifestyle and pajamas are the sleep-pants of the Man.   | Şimdi, tipik, ben pijama giyer, ama son zamanlarda bir berduş yaşam tarzı benimsemiş ve pijama insanın uyku-pantolon. | Şimdi, tipik olarak, ben pijamayı giyerim ama ben geçenlerde bir aylak yaşam tarzını benimsedim ve pijama adamın uyku-pantolonudur. | Şimdi, tipik olarak, ben, pajaması giyerim fakat Man Adas'ının uyku-pantolonu bir aylak yaşam biçimi ve pajamas are'si son zamanlarda evlat edinilen ben. |
| <b>Intelligibility</b>  |   | Incomprehensible (70%)  | Disfluent (30%)   | Incomprehensible (89%)  |

|                         |  |   |  |   |
|-------------------------|--|---|--|---|
| <b>Fidelity</b>         |  | Little (45%)  | Most (35%)   | None (90%)  |
| <b>Rank</b>             |  | Moderate (84%)  | Best (85%)   | Worst (100%)  |
| <b>Macro evaluation</b> | The sentence was best translated by Proçeviri. All of the annotators ranked translation produced by Sametran as the worst among the three.   |   |  |   |
| <b>Micro evaluation</b> | <p>The sentence was best translated by Proçeviri. There is only an extra pronoun.</p> <p>Google Translate made an error in inflecting the adverb “typically”, and three verb inflection errors and concomitant agreement errors “wear”, “adopt” and “are”.</p> <p>Sametran failed to recognize the word “pajamas”, “are” and “the man”, rendering “the man” as “Man Adas”. In addition, the word selection and verb inflection for “adopt” is wrong. Verb inflection for “pajamas are” is also wrong. The sentence needs to be rearranged.</p> |   |  |   |
| <b>No.</b>              | <b>Original</b>  | <b>Google Translate</b>   | <b>Proçeviri</b>   | <b>Sametran</b>   |
| 25                      | I'll have you know, Mahatma Gandhi wore no pants and a nation rallied behind him!  | Ben biliyorum gerekir, Mahatma Gandhi hiçbir pantolon giydi ve bir ulus arkasında yürüdü! | Ben yapacağım, sen biliyor muydun, mahatma Gandhi hiçbir pantolon giymedi ve bir ulus onun arkasında toplandı! | Ben senin bildiğine sahip olacağım Mahatma Gandhi'si, hiç pantolonu giymedi ve bir millet onun arkasında bir araya geldi! |
| <b>Intelligibility</b>  |  | Disfluent (42%)   | Disfluent (45%)  | Incomprehensible (50%)  |
| <b>Fidelity</b>         |  | None (40%)  | Little (35%)   | None (50%)  |
| <b>Rank</b>             |  | Worst (50%)   | Best (55%)   | Moderate (50%)  |
| <b>Macro</b>            | Google Translate and Sametran received the lowest ratings in terms of  |   |  |   |

|                         |  |                         |                              |                                     |
|-------------------------|--|-------------------------|------------------------------|-------------------------------------|
| <b>evaluation</b>       | fidelity. Sametran was also rated with the lowest point in terms of intelligibility. However, Sametran was ranked as moderate while Google Translate was ranked the worst.   |                         |                              |                                     |
| <b>Micro evaluation</b> | <p>Expression “I’ll have you know” couldn’t be translated accurately by machine translation programs.</p> <p>Apart from this error, Google Translate also made a negation error in “wore no pants”. In addition the pronoun “him” is absent in the translation.</p> <p>Proçeviri made a capitalization error for “Mahatma”.</p> <p>Sametran made two noun inflection errors in “Mahatma Gandhi” and “pants”.</p> |                         |                              |                                     |
| <b>No.</b>              | <b>Original</b>  | <b>Google Translate</b> | <b>Proçeviri</b>             | <b>Sametran</b>                     |
| <b>26</b>               | My good man...   | Benim iyi bir adam ...  | Benim iyi adamım...          | Iyi adamım...                       |
| <b>Intelligibility</b>  |  | Incomprehensible (45%)  | Flawless (35%)               | Good (50%)                          |
| <b>Fidelity</b>         |  | None (57%)              | All (30%)                    | Most (30%)                          |
| <b>Rank</b>             |  | Worst (100%)            | Best (60%)                   | Moderate (50%)                      |
| <b>Macro evaluation</b> | Google Translate was given the lowest ratings in terms of intelligibility and fidelity. All annotators chose its translation as worst among three.   |                         |                              |                                     |
| <b>Micro evaluation</b> | The expression was literally translated by all machine translation programs, which can be recognized as expression selection error.  |                         |                              |                                     |
| <b>No.</b>              | <b>Original</b>  | <b>Google Translate</b> | <b>Proçeviri</b>             | <b>Sametran</b>                     |
| <b>27</b>               | Now, before you walk away...   | Şimdi, önce yürüyüp ... | Şimdi, sen yürümeden önce... | Deplasmanı yürümeden önce, şimdi... |
| <b>Intelligibility</b>  |  | Incomprehensible (40%)  | Good (40%)                   | Incomprehensible (65%)              |
| <b>Fidelity</b>         |  | None (65%)              | All (42%)                    | None (90%)                          |

| <b>Rank</b>             |  | Moderate (85%)  | Best (100%)  | Worst (100%)   |
|-------------------------|--|---|--|--|
| <b>Macro evaluation</b> | All annotators ranked Proçeviri as best and Sametran as worst translator. Google Translate was ranked as moderate despite receiving the same intelligibility and fidelity ratings with Sametran.   |   |  |  |
| <b>Micro evaluation</b> | All machine translation programs failed to translate this short sentence. Each program made an expression selection error for “walk away”. Apart from this, translation produced by Sametran also needs a rearrangement.   |   |  |  |
| <b>No.</b>              | <b>Original</b>  | <b>Google Translate</b>   | <b>Proçeviri</b>   | <b>Sametran</b>  |
| <b>28</b>               | I know that I may appear deranged, but I am, in fact, a world-renowned physicist.  | Ben dengesiz görünebilir biliyorum, ama ben aslında, bir dünyaca ünlü fizikçi, duyuyorum. | Ben benim bozulan gözükebildiğimi biliyorum ama ben, aslında, dünya-ünlü bir fizikçiyim. | Ben düzenini bozulan sanabileceğimi bilirim, fakat benim bir dünya-ünlü fizikçisi, gerçekte. |
| <b>Intelligibility</b>  |  | Incomprehensible (45%)  | Disfluent (35%)  | Incomprehensible (60%)   |
| <b>Fidelity</b>         |  | Little (35%)  | Little (35%)   | None (70%)   |
| <b>Rank</b>             |  | Moderate (65%)  | Best (90%)   | Worst (75%)  |
| <b>Macro evaluation</b> | Translations of this sentence were given very low intelligibility and fidelity ratings. Sametran again received the lowest points in all three tests.  |   |  |  |
| <b>Micro evaluation</b> | Google Translate made a word selection error for “I am”, rendering it as “duyuyorum”. Proçeviri made a word selection error for “deranged”. In addition, the expression selection for “world-renowned” is also wrong. Sametran made the same word selection and expression selection errors as |   |  |  |



|   |
|---|
| Proçeviri. In addition, there is an agreement error and a pronoun error in its translation. |
|---|

Table 12: The macro evaluation and micro evaluation of MT programs for audio-medial text type.

Below is presented the table which contains the times a sentence was ranked the best, worst or moderate, the intelligibility and fidelity ratings:

|                         | <b>Google Translate</b> | <b>Proçeviri</b> | <b>Sametran</b> |
|-------------------------|-------------------------|------------------|-----------------|
| <b>Ranking</b>          |                         |                  |                 |
| <b># Best</b>           | 9                       | 15               | 2               |
| <b># Moderate</b>       | 13                      | 9                | 5               |
| <b># Worst</b>          | 4                       | 3                | 20              |
| <b>Intelligibility</b>  |                         |                  |                 |
| <b>Flawless</b>         | 4                       | 3                | 2               |
| <b>Good</b>             | 2                       | 6                | 2               |
| <b>Non-Native</b>       | 3                       | 6                | 1               |
| <b>Disfluent</b>        | 8                       | 8                | 3               |
| <b>Incomprehensible</b> | 9                       | 3                | 18              |
| <b>Fidelity</b>         |                         |                  |                 |
| <b>All</b>              | 3                       | 7                | 1               |
| <b>Most</b>             | 4                       | 6                | 3               |
| <b>Much</b>             | 2                       | 1                | 0               |
| <b>Little</b>           | 10                      | 5                | 1               |
| <b>None</b>             | 7                       | 7                | 21              |

Table 13: The macro evaluation of each machine translation program for audio-medial text type.

Audio-medial text type, which was exemplified with subtitles, was generally best translated by Proçeviri. 15 sentences by Proçeviri, 9 sentences by Google Translate and 2 sentences by Sametran were ranked the best. 13 sentences by Google Translate, 9 sentences by Proçeviri and 5 sentences by Sametran were ranked as moderate. The majority of sentences translated by Sametran was ranked the worst.

In terms of intelligibility, 9 sentences were given the highest rating , 4 sentences by Google Translate, 3 sentences by Sametran and 2 sentences by Sametran. 2 sentences by Google Translate and Sametran, and 6 sentences by Proçeviri were rated as “good”. 3 sentences by Google Translate, 6 sentences by Proçeviri and 1 sentence by Sametran were rated as “non-native”. 8 sentences by Google Translate and Proçeviri, and 3 sentences by Sametran were ranked as “diffluent”. The majority of sentences translated by Sametran was given the lowest intelligibility rating, 18 sentences were ranked as “incomprehensible”. 9 sentences by Google Translate and 3 sentences by Proçeviri were also given the lowest intelligibility rating.

There were 11 sentences rated with the highest fidelity rating. 7 of these sentences were produced by Sametran. 4 sentences by Google Translate, 6 sentences by Proçeviri and 3 sentences by Sametran were rated as “most”. 2 sentences by Google Translate and 1 sentence by Proçeviri were rated as “much”. 10 sentences by Google Translate, 5 sentences by Proçeviri and 1 sentence by Sametran were rated as “little”. The majority of sentences produced by Sametran was given the lowest fidelity rating. 7 sentences by Google Translate and Sametran were also given the lowest fidelity rating “none”.

The performance of each machine translation program for micro evaluation (grammatical error analysis) can be seen in the below table:

| <b>Error Category</b> | <b>Google Translate</b> | <b>Proçeviri</b> | <b>Sametran</b> |
|-----------------------|-------------------------|------------------|-----------------|
| <b>Spelling</b>       | 0                       | 0                | 0               |

|                         |    |    |    |
|-------------------------|----|----|----|
| <b>Not-found Word</b>   | 0  | 0  | 5  |
| <b>Capitalization</b>   | 0  | 1  | 0  |
| <b>Elision</b>          | 0  | 0  | 0  |
| <b>Verb inflection</b>  | 10 | 6  | 8  |
| <b>Noun inflection</b>  | 2  | 5  | 5  |
| <b>Other inflection</b> | 3  | 0  | 0  |
| <b>Rearrangement</b>    | 2  | 2  | 5  |
| <b>Category</b>         | 0  | 0  | 1  |
| <b>Pronoun</b>          | 5  | 2  | 2  |
| <b>Article</b>          | 0  | 0  | 0  |
| <b>Preposition</b>      | 0  | 0  | 4  |
| <b>Negative</b>         | 1  | 0  | 0  |
| <b>Conjunction</b>      | 0  | 1  | 1  |
| <b>Agreement</b>        | 7  | 2  | 3  |
| <b>Clause boundary</b>  | 0  | 0  | 0  |
| <b>Word Selection</b>   | 2  | 2  | 9  |
| <b>Expression</b>       | 12 | 11 | 19 |
| <b>Total</b>            | 44 | 30 | 62 |

Table 14: The micro evaluation of each machine translation program for audio-medial text type.

Machine translation programs displayed different performances in translating subtitles. However, their common weakness was their inability to select the right word and expression.

Google Translate made 44 errors, most of which was expression and word selection errors. In addition, there were 10 verb inflection errors, which resulted in 7 agreement errors. Google Translate also made more pronoun errors than other MT systems.

Proçeviri made the least errors. Out of 30 errors, 13 errors were expression or word selection errors. In addition, Proçeviri made 11 inflection errors.

Sametran made 62 errors, more than other MT systems. Most of these errors were dictionary-related. There were 19 expression selection and 9 word selection errors, and 5 not-found words. In addition, Sametran made more preposition error than others. Furthermore, there were 13 inflection errors in the sentences translated by Sametran.

## **6.5. DISCUSSIONS AND COMMENTS ON MACRO EVALUATION**

A comparative and contrastive macro evaluation of three different machine translation programs was conducted. Text samples which were chosen in accordance with the Reiss' typology (2000) were translated by three different machine translation programs. Then, the translations were rated and ranked in terms of fidelity, intelligibility and their general quality by 20 annotators. The ratings and rankings were sentence-based.

Content-focused text type was exemplified with an academic abstract. The text contained many abbreviations, financial terminology and long sentences.

- Google Translate's translations were ranked generally as "moderate" or "worst". Most of its translations were given two lowest intelligibility ratings. In terms of fidelity, its translation received all ratings, from the highest to the lowest. However, most of its translations received two lowest fidelity ratings.
- In terms of general quality, Proçeviri was ranked the best more than others. The intelligibility ratings of its translations were generally "non-native" and "disfluent". In terms of fidelity, it generally received "much" or "none" ratings. Its translations were never given the lowest fidelity rating.
- Sametran's translations were generally ranked as "moderate" or "worst". In terms of intelligibility, most of its translations were rated as "incomprehensible", which was the lowest rating. Its translations were mostly given the lowest ratings in terms of fidelity.

Form-focused text type was exemplified with an extract from a short story. The text included long sentences, short questions and a word-play.

- In terms of general quality, Google Translate's translations were rated as best 8 times, as moderate 7 times and as worst 10 times. The intelligibility and fidelity ratings of its translations ranked from highest to the lowest, but majority of them were given the lowest ratings.
- Proçeviri was generally rated best; none of its translations was ranked the worst. In terms of intelligibility and fidelity, its translations were usually rated with the highest two ratings, but they were given the lowest ratings; 7 times in intelligibility and 4 times in fidelity.
- In terms of general quality, the translations produced by Sametran were rated as moderate 10 times and as worst 15 times; none of its translations was ranked the best. In terms of intelligibility and fidelity, most of its translations were given the lowest ratings.

Appeal-focused text type was exemplified with an online hotel advertisement. The text included many long sentences and adjectives describing the hotel and its facilities.

- Translations produced by Google Translate were mostly ranked the best, they were never ranked the worst. In terms of intelligibility, the sentences produced by Google Translate received mostly "non-native" and "disfluent" ratings. The fidelity ratings differed greatly; its translation received the highest and lowest ratings, but the majority of them were rated as "most", "much" and "little".
- Proçeviri's general quality was mostly considered as moderate. The translation received similar fidelity and intelligibility ratings; they didn't get the highest ratings but they were given the lowest ratings.
- All of the sentences translated by Sametran were ranked the worst. In terms of intelligibility and fidelity, they were mostly given the lowest rating.

Audio-medial text type was exemplified with subtitles of a comedy serial. The text included many colloquial expressions and proper names.

- In terms of general quality, Google Translate was mostly ranked as “moderate”. Its translations were also given the worst and best ranks. Intelligibility ratings of Google Translate differed greatly, but most of the time, its translations were given two lowest ratings. In terms of fidelity, they were also mostly given two lowest ratings.
- Translations produced by Proçeviri were mostly ranked the best. In terms of intelligibility, the translations mostly received “good”, “non-native” and “disfluent” ratings. Its translations were rated as “all” 7 times, as “most” 6 times and as “none” 7 times in fidelity rating.
- Sametran was mostly rated worst. In terms of intelligibility and fidelity, its translations were generally given the lowest ratings.

It can be easily said that, the performances of MT programs remained same while translating different texts.

## **6.6. DISCUSSIONS AND COMMENTS ON MICRO EVALUATION**

A comparative and contrastive micro evaluation of three different machine translation programs was conducted. Text samples which were chosen in accordance with the Reiss’ typology (2000) were translated by three different machine translation programs. Then, the translations were analyzed and evaluated in accordance with Flanagan’s Grammatical Error Analysis categories (1994). The analysis was sentence-based.

Content-focused text type was exemplified with an academic abstract. The text contained many abbreviations, financial terminology and long sentences. The most common errors for machine translation programs were not-found words and arrangement.

- Translation produced by Google Translate had 11 verb inflection errors which resulted in mostly unintelligible sentences. Apart from these, many sentences needed to be rearranged to be understood.

- Proçeviri made the least errors among three machine translation programs. However, the dictionary capacity of Proçeviri when it came to abbreviations was very limited. In addition, the dictionary continued to be a problem in choosing right words and expressions.
- Sametran made the most errors among three machine translation programs. There were 11 not-found words, 12 word selection errors and 5 expression selection errors in 265 words long abstract.

Form-focused text type was exemplified with an extract from a short story. The first part of the extract was description while the second part of it was a dialogue. The text included long sentences, short questions and a word-play. The most common errors for machine translation programs were verb inflection and word selection.

- Google Translate made 10 verb inflection errors. Besides, there were 4 not-found word errors and 10 word selection errors.
- Proçeviri produced the translation with the least errors. There were 6 word selection errors. In addition, there were 5 capitalization errors and 5 arrangement errors.
- Translation produced by Sametran had the most errors. There were 15 word selection errors, 13 noun inflection errors and 9 verb inflection errors.

Appeal-focused text type was exemplified with an online hotel advertisement. The text included many long sentences describing the hotel and its facilities. The performances and errors of machine translations differed greatly for this text type.

- Translation produced by Google Translate had the least errors. Noun inflection and arrangement were its main problem areas.
- Proçeviri had 5 arrangement and 10 word selection errors.

- Translation produced by Sametran contained the most errors. Its errors were mainly due to dictionary; there were 6 not-found words, 9 category errors, and 21 word selection errors.

Audio-medial text type was exemplified with subtitles of a comedy serial. The text included many colloquial expressions and proper names. The most common error to three machine translation program was expression selection.

- Google Translate made 12 expression selection errors. In addition, there were 10 incorrect verb inflections.
- Translation produced by Proçeviri had the least errors. There were 11 expression selection errors.
- Sametran continued to suffer from its dictionary; there were 19 expression selection and 9 word selection errors. In addition, again due to the dictionary, there were 5 not-found words.

In conclusion, although the text types changed, MT systems made the same errors persistently and displayed similar performances.

|                                 | <b>Google Translate</b> | <b>Proçeviri</b> | <b>Sametran</b> | <b>Total error for text type</b> |
|---------------------------------|-------------------------|------------------|-----------------|----------------------------------|
| <b>Content-Focused</b>          | 39                      | 34               | 47              | 120                              |
| <b>Form-Focused</b>             | 52                      | 29               | 64              | 145                              |
| <b>Appeal-Focused</b>           | 25                      | 35               | 54              | 114                              |
| <b>Audio-Medial</b>             | 44                      | 32               | 62              | 138                              |
| <b>Total error of MT system</b> | 160                     | 130              | 227             |                                  |

Table 15: Total errors of MT systems for each text type.



In content-focused, form-focused and audio-medial text type, Proçeviri made the least errors. In appeal-focused text type, Google Translate made the least errors. Sametran produced more errors than other two MT systems in each text type.

The most errors were made in form-focused text type and the least errors were made in appeal-focused text type.

This chapter has dwelled on the case studies, which comprise the analysis part of the study. In the next chapter, these case studies are discussed within the light of research questions, and answers to these questions will be given.

## CHAPTER 7

### CONCLUSIONS AND RECOMMENDATIONS

This thesis has provided an evaluation of machine translation quality, with its multifarious dimensions including history, state-of-art, Turkish MT systems and different kinds of evaluation. Two different evaluations were conducted on four different text types in order to understand the quality of three different MT systems working from English into Turkish. The performance of these three different MT systems were compared and contrasted in the previous chapter with tables and charts.

This chapter focuses on the conclusions drawn from the analysis of MT systems; and it presents some recommendations regarding the MT and further research. Responses to main and sub research questions are given within the framework of analysis which has been presented in the previous chapter.

#### 7.1. CONCLUSIONS REGARDING THE RESEARCH QUESTIONS

The main and sub research questions are provided with answers, successively. The conclusions are drawn based on these answers.

##### 7.1.1. Main Research Question

**(RQ) Is it possible to evaluate MT quality consistently with different evaluation methods for different text types?**

Yes, it is possible. Quality of MT programs can be evaluated with different methods for different text types, and the results will be same. Findings of different evaluation methods

conducted in this study reveal that the quality of MT programs doesn't differ greatly from one text type to another, or from one evaluation method to another.

Among the three machine translation programs, Sametran made more errors than other programs in all text types, according to the results of micro evaluation. The sentences translated by Sametran were also ranked as worst. Among 74 sentences translated by Sametran, 50 were ranked the worst compared to the translations produced by Google Translate and Proçeviri.

Among the three machine translation programs, Proçeviri made less error than other programs in three text types, except for appeal-focused text type. Human annotators also ranked its translations best for most of the sentences. Among 74 sentences translated by Proçeviri, 41 were ranked the best and 26 were ranked as moderate.

Only one deviation was encountered in this study, which was the micro evaluation of MT performance for appeal-focused text type. Although the translations contained the least errors among other text types, the sentences received low ratings from human annotators. When the number of sentences which received two highest ratings in terms of intelligibility and fidelity and the times sentences were chosen as best are added up, it is seen that the translations of appeal-focused text bring up the rear among four text types:

|                                   | <b>Google Translate</b>                        | <b>Proçeviri</b> | <b>Sametran Sametech</b> |
|-----------------------------------|--|------------------|--------------------------|
|                                   | <b>Intelligibility (# of 2 Highest Points)</b> |                  |                          |
| <b>Content-Focused (12 sent.)</b> | 1  | 1                | 1                        |
| <b>Form-Focused (25 sent.)</b>    | 7  | 11               | 2                        |
| <b>Appeal-Focused (9 sent.)</b>   | 1  | 2                | 0                        |
| <b>Audio-Medial (28 sent.)</b>    | 6  | 9                | 4                        |

|                                   | <b>Fidelity (# of 2 Highest Points)</b> |    |   |
|-----------------------------------|---|----|---|
| <b>Content-Focused (12 sent.)</b> | 2                                       | 2  | 1 |
| <b>Form-Focused (25 sent.)</b>    | 9                                       | 16 | 2 |
| <b>Appeal-Focused (9 sent.)</b>   | 4                                       | 3  | 0 |
| <b>Audio-Medial (28 sent.)</b>    | 7                                       | 13 | 4 |
|                                   | <b>Ranking (# of "Best" Rating)</b>     |    |   |
| <b>Content-Focused (12 sent.)</b> | 2                                       | 7  | 1 |
| <b>Form-Focused (25 sent.)</b>    | 8                                       | 16 | 0 |
| <b>Appeal-Focused (9 sent.)</b>   | 6                                       | 3  | 0 |
| <b>Audio-Medial (28 sent.)</b>    | 9                                       | 15 | 2 |

Table 16: The added up numbers of sentences which received two highest ratings from human annotators in terms of fidelity and intelligibility and the times of translations chosen best in ranking.

The findings reveal that although it is possible to consistently evaluate the machine translation quality with different methods, human assessment remains to be the gold standard of quality. The number of errors encountered in a translation may not be parallel with the translation quality perceptions of humans.

### 7.1.2. Sub Research Questions

**(SRQ1) Is there a difference between micro evaluation and macro evaluation of MT performance for different texts in Reiss' typology exemplified with abstract, short story extract, online advertisement and subtitle?**

No, the results of micro evaluation and macro evaluation are parallel to each other in each text type.

Content-focused text type, exemplified with an abstract, was best translated by Proçeviri. In micro evaluation, Proçeviri made 34 errors, Google Translate made 39 errors and Sametran made 47 errors. In macro evaluation, Proçeviri was ranked the best 7 times, Google Translate was ranked the best 2 times and Sametran was ranked the best only 1 time. In terms of intelligibility, Proçeviri was generally considered to be “non-native” and “disfluent”, while Google Translate was considered to be “disfluent” and “incomprehensible”. Sametran was generally considered as “incomprehensible”. In terms of fidelity, Proçeviri was generally considered to preserve “much” and “little” of the source text meaning. While Google Translate was considered to preserve “little or “none”, Sametran was generally considered to preserve “none” of the meaning.

Form-focused text type, exemplified with an extract from a short story, was best translated by Proçeviri. In micro evaluation Proçeviri made 29 errors, Google Translate made 52 errors and Sametran made 64 errors. In macroevaluation, Proçeviri was ranked the best 16 times, Google Translate was ranked the best 8 times and Sametran was never ranked the best. In terms of intelligibility, Proçeviri was generally considered to be “flawless” and “good”, but it also received “incomprehensible” rating. Google Translate and Sametran were generally considered to be “incomprehensible”. In terms of fidelity, Proçeviri was generally considered to preserve “most” of the source text meaning. While Google Translate was considered to preserve “little or “none”, Sametran was generally considered to preserve “none” of the meaning.

Appeal-focused text type, exemplified with an online advertisement, was best translated by Google Translate. In micro evaluation Google Translate made 25 errors, Proçeviri made 35 errors and Sametran made 54 errors. In macro evaluation, Google Translate was ranked the best 6 times, Proçeviri was ranked 3 times while all of the sentences translated by Sametran were ranked the worst. In terms of intelligibility, Google Translate was generally considered to be “non-native” and “disfluent”, Proçeviri was considered to be “disfluent” while Sametran was mostly considered to be “incomprehensible”. In terms of fidelity, Google Translate was considered to preserve “most” of the source text meaning. Proçeviri was considered to preserve “most” and “little” of the source text meaning while Sametran was mostly considered to preserve “none” of the meaning.

Audio-medial text type, exemplified with subtitles, was best translated by Proçeviri. In micro evaluation, Proçeviri made 32 errors, Google Translate made 44 errors and Sametran made 63 errors. In macro evaluation, Proçeviri was ranked the best 15 times, Google Translate was ranked the best 9 times and Sametran was ranked the best 2 times. In terms of intelligibility, Proçeviri was generally considered to be “good”, “non-native” and “disfluent”. Google Translate was considered to be “disfluent” and “incomprehensible” while Sametran was mostly considered to be “incomprehensible”. In terms of fidelity, Proçeviri was considered to preserve “all”, “most” but sometimes also “none” of the meaning. Google Translate was considered to preserve “little” or “none” of the meaning while Sametran was mostly considered to preserve “none” of the source text meaning.

**(SRQ2) To what extent is the output of an MT system fluent (intelligibility) and faithful (fidelity) for the human annotators?**

It mainly depends on the length of the input sentence. The shorter the sentence, the better the translation, thus more intelligible and faithful. The ability of machine translation programs to resolve ambiguities and accurately inflect the verbs is still severely limited. The findings of the macro evaluation conducted in this study revealed that human annotators gave the sentences in question the lowest and highest ratings in terms of

intelligibility and fidelity. The highest scoring sentences in terms of intelligibility and fidelity in all text types and their translations are given below:

**Content-focused text type:**

**Sentence number 3:**

But the financial crisis and the Great Recession which began in Spring 2008 have dealt this optimistic picture a devastating blow.

Google Translate: Ancak mali kriz ve 2008 baharında başlayan Büyük Resesyon bu iyimser fotoğrafa bir yıkıcı bir darbe vurmuştur. (Intelligibility: Good, Fidelity: All)

**Form-focused text type:**

**Sentence number 4:**

"There must be more money!"

Google Translate: "Daha fazla para olmalı!" (Intelligibility: Flawless, Fidelity: All)

**Sentence number 6:**

The whisper was everywhere, and therefore no one spoke it.

Proçeviri: Fısıltı her yerde idi ve bu yüzden hiç kimse onu konuşmadı. (Intelligibility: Flawless, Fidelity: All)

**Sentence number 7:**

Just as no one ever says:

Proçeviri: Tam da hiç kimsenin şimdiye kadar demediği gibi: (Intelligibility: Flawless, Fidelity: All)

**Sentence number 10:**

"Because we're the poor members of the family," said the mother.

Proçeviri: "Biz ailenin yoksul üyeleri olduğumuz için", anne dedi. (Intelligibility: Good, Fidelity: All)

**Sentence number 13:**

The boy was silent for some time.

Google Translate: Oğlan bir süre sessiz kaldı. (Intelligibility: Flawless, Fidelity: All)

**Sentence number 15:**

"No, Paul.

Google Translate: "Hayır, Paul. (Intelligibility: Flawless, Fidelity: All)

Proçeviri: "Hayır, Paul. (Intelligibility: Flawless, Fidelity: All)

**Sentence number 16:**

Not quite.

Proçeviri: Tamamen değil. (Intelligibility: Flawless, Fidelity: All)

**Sentence number 21:**

"But it's lucre, not luck."

Proçeviri: "Ama o, şans değil, paradır". (Intelligibility: Flawless, Fidelity: Most)

**Sentence number 22:**

"Oh!" said the boy.

Google Translate: "Ah!" dedi oğlan. (Intelligibility: Flawless, Fidelity: All)

**Sentence number 25:**

If you're lucky you have money.

Google Translate: Eğer şanslıysanız para var. (Intelligibility: Non-native, Fidelity: All)



Proçeviri: Eđer sen řanslıysan, sen paraya sahipsın. (Intelligibility: Non-native, Fidelity: All)

**Appeal-focused text type:**

**Sentence number 1:**

A century ago, The Plaza set the standard for luxury.

Google Translate: Bir yüzyıl önce, Plaza lüks için standart belirledi. (Intelligibility: Good, Fidelity: All)

**Audio-medial text type:**

**Sentence number 1:**

We were worried about you.

Proçeviri: Biz senin hakkında endişeliydik. (Intelligibility: Good, Fidelity: All)

**Sentence number 2:**

Don't be melodramatic.

Google Translate: Melodramatik olmayın. (Intelligibility: Flawless, Fidelity: All)

Sametran: Aşırı duygusal olmayınız. (Intelligibility: Flawless, Fidelity: Most)

**Sentence number 10:**

I am gonna miss you.

Proçeviri: Ben seni özleyeceğim. (Intelligibility: Flawless, Fidelity: All)

**Sentence number 11:**

Of course you are.

Google Translate: Sen tabii ki. (Intelligibility: Flawless, Fidelity: Little)

**Sentence number 13:**

Excuse me.

Google Translate: Afedersiniz. (Intelligibility: Flawless, Fidelity: All)

Proçeviri: Beni mazur gör. (Intelligibility: Flawless, Fidelity: All)

Sametran: Beni bağışlayın. (Intelligibility: Flawless, Fidelity: All)

**Sentence number 15:**

Oh, well, no, you are understandably terrified.

Proçeviri: Oh, şey, hayır, sen anlaşılır şekilde dehşete düşmüsun. (Intelligibility: Good, Fidelity: All)

**Sentence number 16:**

But, you know, allow me to explain.

Proçeviri: Ama, sen biliyorsun, açıklamam için bana izin ver. (Intelligibility: Good, Fidelity: All)

**Sentence number 20:**

Uh, sir, may I use your phone?

Google Translate: Ah, efendim, ben senin telefonunu kullanabilir miyim? (Intelligibility: Non-native, Fidelity: All)

**Sentence number 21:**

I don't think so.

Google Translate: Ben öyle düşünmüyorum. (Intelligibility: Flawless, Fidelity: All)

**Sentence number 26:**

My good man...

Proçeviri: Benim iyi adamım... (Intelligibility: Flawless, Fidelity: All)

**Sentence number 27:**

Now, before you walk away...

Proçeviri: Şimdi, sen yürümeden önce.. (Intelligibility: Good, Fidelity: All)

The common thread to these highest scoring sentences is their length. It is also interesting to note that the highest two points in terms of fidelity and intelligibility were mostly given to the sentences in audio-medial text type. This can be again explained by their short length.

**(SRQ3) Which types of of MT errors have the highest impact on the human perception of translation quality?**

The micro evaluation was carried out in accordance with Flaganan's machine translation error classification. In this classification, there are 18 error categories. To understand which of these categories have the highest impact on the human perception of translation quality, the sentences which were ranked the worst by more than 85% of the human annotators are chosen and analyzed.

**Content-focused text type:**

**Sentence number 3:**

But the financial crisis and the Great Recession which began in Spring 2008 have dealt this optimistic picture a devastating blow.

Sametran: Fakat hangi büyük Gerileme ve mali kriz Spring'in 2008'ininde, bir yıkıcı darbe bu iyimser resimle ilgilendiğine başladı. (Intelligibility: Incomprehensible, Fidelity: None, Ranking: Worst (90%))

**Form-focused text type:**

**Sentence number 5:**

Yet nobody ever said it aloud.

Google Translate: Ancak kimse yüksek sesle söyledi. (Intelligibility: Non-native, Fidelity: None, Ranking: Worst (90%))

**Sentence number 6:**

The whisper was everywhere, and therefore no one spoke it.

Google Translate: Fısıltı her yerde, ve bu nedenle hiç kimse onu konuştu. (Intelligibility: Disfluent, Fidelity: Little, Ranking: Worst (90%))

**Sentence number 18:**

"Oh!" said Paul vaguely.

Sametran: Belirsizce Paul'u söyledi. "öyle mi?". (Intelligibility: Incomprehensible, Fidelity: None, Ranking: Worst (95%))

**Sentence number 21:**

"But it's lucre, not luck."

Google Translate: "Ama lucre, değil şans." (Intelligibility: Incomprehensible, Fidelity: None, Ranking: Worst (100%))

**Sentence number 22:**

"Oh!" said the boy.

Sametran: Delikanlıyı söyledi. "öyle mi?". (Intelligibility: Incomprehensible, Fidelity: None, Ranking: Worst (95%))

**Sentence number 25:**

If you're lucky you have money.

Sametran: Eđer sen paraya sahip olduėun řanslıysan. (Intelligibility: Incomprehensible, Fidelity: None, Ranking: Worst (100%))

**Appeal-focused text type:**

**Sentence number 1:**

A century ago, The Plaza set the standard for luxury.

Sametran: Bir yzyıl nce, lks iin standart kesin Plaza. (Intelligibility: Incomprehensible, Fidelity: None, Ranking: Worst (100%))

**Sentence number 2:**

Striking a balance between its storied past and limitless future, the passion and uncompromising service, which made the hotel a legend, has returned with a new and contemporary spirit.

Sametran: Ve sınırsız gelecek ve tesinde destansı arasında bir dengeye arpmak,uzlařmaz servis ve tutku, bir gsterge oteli yapan, yeni ve aėdař bir ruh ile dndrd. (Intelligibility: Incomprehensible, Fidelity: None, Ranking: Worst (90%))

**Sentence number 4:**

From the sumptuous decor to the impeccable white glove service, The Plaza Hotel returned to create indelible memories at New York's most celebrated address.

Sametran: Kusursuz beyaz eldiven servisine masraflı dekordan, Plaza oteli Yeni York'un en ok kutlanan adresinde silinmez bellekleri oluřturmaya dndrd. (Intelligibility: Disfluent, Fidelity: None, Ranking: Worst (95%))

**Audio-medial text type:**

**Sentence number 3:**

I'm just getting on a train and leaving forever.

Sametran: Şimdi ben, bir treniyorum ve sonsuza dek bırakıyorum. (Intelligibility: Incomprehensible, Fidelity: None, Ranking: Worst (95%))

**Sentence number 4:**

So a few things don't go your way. And your best decision is to ride the rails like a hobo?

Sametran: Öyleyse az şey yolunuza gitmezler. Parmaklıklara binmek için en iyi kararın ve, bir aylaktan hoşlanıyor? (Intelligibility: Incomprehensible, Fidelity: None, Ranking: Worst (100%))

**Sentence number 6:**

Everything is changing, and it is simply too much.

Sametran: Her şey, değiştiriyor ve çok yalın şekilde o. (Intelligibility: Incomprehensible, Fidelity: None, Ranking: Worst (100%))

**Sentence number 7:**

I need to get away and think.

Sametran: Ben, deplasman ve düşünmeyi almaya ihtiyaç duyarım. (Intelligibility: Incomprehensible, Fidelity: None, Ranking: Worst (95%))

**Sentence number 10:**

I am gonna miss you.

Sametran: Ben, siz gonna bayanlarıyım. (Intelligibility: Incomprehensible, Fidelity: None, Ranking: Worst (100%))

**Sentence number 11:**

Of course you are.

Sametran: Kursunsun. (Intelligibility: Incomprehensible, Fidelity: None, Ranking: Worst (100%))

**Sentence number 15:**

Oh, well, no, you are understandably terrified.

Sametran: Öyle mi? ,iyi, yok, anlaşılır şekilde sen korkutulursun. (Intelligibility: Incomprehensible, Fidelity: None, Ranking: Worst (100%))

**Sentence number 16:**

But, you know, allow me to explain.

Google Translate: Ama, sen biliyorsun, ben açıklamak için izin verir. (Intelligibility: Disfluent, Fidelity: Little, Ranking: Worst (90%))

**Sentence number 20:**

Uh, sir, may I use your phone?

Sametran: Uh söy?, telefonunu kullanabilirsin? (Intelligibility: Incomprehensible, Fidelity: None, Ranking: Worst (100%))

**Sentence number 21:**

I don't think so.

Sametran: Ben, öyleseyi düşünmem. (Intelligibility: Incomprehensible, Fidelity: None, Ranking: Worst (90%))

**Sentence number 22:**

Yeah, well, I understand that I'm half naked, but there is a reasonable explanation.

Sametran: Evet, iyi, ben yarı ispatsız olduğumu anlarım, fakat bir makul izah vardır. (Intelligibility: Incomprehensible, Fidelity: None, Ranking: Worst (95%))

**Sentence number 23:**

While I slept in my sleeper car, all my possessions were stolen.

Sametran: Uykucu arabamda uyuyurken bütün iyeliklerim, çalındı. (Intelligibility: Incomprehensible, Fidelity: None, Ranking: Worst (90%))

**Sentence number 24:**

Now, typically, I wear pajamas, but I recently adopted a hobo lifestyle and pajamas are the sleep-pants of the Man.

Sametran: Şimdi, tipik olarak, ben, pajaması giyerim fakat Man Adas'ının uyku-pantolonu bir aylak yaşam biçimi ve pajamas are'si son zamanlarda evlat edinilen ben. (Intelligibility: Incomprehensible, Fidelity: None, Ranking: Worst (100%))

**Sentence number 26:**

My good man...

Google Translate: Benim iyi bir adam ... (Intelligibility: Incomprehensible, Fidelity: None, Ranking: Worst (100%))

**Sentence number 27:**

Now, before you walk away...

Sametran: Deplasmanı yürümeden önce, şimdi... (Intelligibility: Incomprehensible, Fidelity: None, Ranking: Worst (100%))

Machine translation programs made a total of 85 errors while translating these sentences. Out of these 85 errors, 16 were verb inflection errors. Word selection and expression selection errors were 11 and 9, respectively. Apart from these, there were 7 not-found word errors.

The findings reveal that these errors, namely, verb inflection error, word and expression selection errors and not-found words have the highest impact on machine quality perception.

**(SRQ4) Does MT output quality change for different text types?**



Yes, it does change. The total number of errors made by three different machine translation programs for each text type is counted and presented in figures in the below table.

|                                 | <b>Google Translate</b> | <b>Proçeviri</b> | <b>Sametran</b> | <b>Total error for text type</b> |
|---------------------------------|-------------------------|------------------|-----------------|----------------------------------|
| <b>Content-Focused</b>          | 39                      | 34               | 47              | 120                              |
| <b>Form-Focused</b>             | 52                      | 29               | 64              | 145                              |
| <b>Appeal-Focused</b>           | 25                      | 35               | 54              | 114                              |
| <b>Audio-Medial</b>             | 44                      | 32               | 62              | 138                              |
| <b>Total error of MT system</b> | 160                     | 130              | 227             |                                  |

Table 17: Total errors of MT systems for each text type.

Machine translation programs made most of the errors in translating form-focused text type. Most of these errors were word selection errors. The least errors were made in translating appeal-focused text type, and most of these errors were word selection, noun and verb inflection errors.

Google Translate made most errors in form-focused text type; it was unable to select right words or expressions and it was unable to inflect the verbs accurately. Proçeviri made most errors in appeal-focused text type; its dictionary was limited. Sametran made most errors in form-focused text type; its dictionary and inflection capacity were limited.

## **7.2. ADDITIONAL CONCLUSIONS, OBSERVATIONS, RECOMMENDATIONS**

Other than the answers to main and sub research questions, this study has also provided additional conclusions, observations and recommendations pertaining to future research, which are presented in this section.

### 7.2.1. Additional Conclusions And Observations

Additional and noteworthy observation and conclusions drawn from the study can be enlisted as follows:

- a) There exists a higher level of consensus in ranking the worst translations than in ranking the best translations.
- b) The statistical machine translation system, Google Translate, produced mostly worse results than rule-based machine translation system, Proçe viri.
- c) It was expected for Google Translate to produce best results in audio-medial text type, as the subtitles were translated many times by different translators and they were available online. However, Google Translate produced best results in appeal-focused text type.
- d) The occurrence of the times when human annotators gave highest two intelligibility and fidelity ratings was in form-focused text type and then in audio-medial text type. This occurrence can be explained by the high number of short sentences in these texts.
- e) The rank of a sentence is proportionate to its fidelity and intelligibility rating, except for a small number of sentences. Thus, it can be said that ranking can replace rating in human evaluation.
- f) Content-focused text type, for whose translation MT programs were originally developed and which is said to produce better results, were actually not given the highest ratings in terms of fidelity and intelligibility. Dictionary-related errors; that is, not-found words, word and expression selection errors were frequent in the translation of content-focused text type.
- g) Statistical machine translation system, Google Translate made more verb inflection than other machine translation systems.
- h) Spelling and elision errors were not encountered in the micro evaluation of machine translation systems.

### 7.2.2. Recommendations

Machine translation and human translation show great differences from each other. However, just like the fact that no two human translators translate in the same way, no machine translation programs translate in the same way. This can be the biggest similarity between human translation and machine translation.

However, dissimilarities need to be reconciled and optimized in a way to benefit human translators in the most efficient way. The research for optimizing human-machine interaction in translation has been in progress. For the last two decades, a great deal of machine translation research has oriented towards either statistical machine translation or machine translation with post-editing.

Statistical machine translation has been proven to be effective between morphologically-poor languages like English and French, and although not very successful, SMT systems for Turkish have already been developed, like Google Translate and Bing Translator. The main requirement of statistical machine translation, bilingual or multilingual corpora, can be built to improve statistical machine translation, and to improve the understanding related to the human translation. The Translational English Corpus is such an initiative to develop understanding related to the translation universals ("The translational english,"). Compiled at the Centre for Translation & Intercultural Studies in Manchester University under the management of Mona Baker, the corpus aims to understand stylistic variations between individual translators and differences between non-translated and translated texts (translation universals). Apart from these, a multilingual corpus can be developed to understand differences between experienced and inexperienced translators for different languages. In addition, a spoken corpus can be compiled for understanding the development of trainee interpreters' performance over years.

Machine translation with post-editing has been the most-effective way for using machine translation output. The post-editing has also been one of the most important ways to evaluate the machine translation quality. There exist two main types post-editing for machine translation: light post-editing and full post-editing. Light post-editing denotes the

minimal intervention for the translation of information-only documents while full post-editing means a greater intervention. Another new area of research is regarding post-editing as a feedback mechanism and feed MT systems with the post-edited versions to increase quality. There are many post-editing guidelines which are usually company or institution-specific, such as KANT and European Commissions, inter alia. These post-editing guidelines can be integrated into the curricula of translation departments, and especially to the syllabus of MT and/or CAT lectures for developing students' post-editing skills.

Another recommendation can be on the issue of MTranslatability, the degree of translatability of a text for MT programs (Bernth & Gdaniec, 2001). There are various linguistic and cultural features which affect the MTranslatability. The most obvious one can be said to be the length of a sentence; very long and very short sentences are not translated fluently. The sentences which include many embedded sentences or having an ellipsis to the previous sentence create ambiguity for MT systems which are very hard to resolve. In addition, the verb inflection is a very complex issue for MT systems; gerunds and passive verbs are usually not inflected correctly. Idiomatic expressions, slangs and metaphors are mostly translated literally. Controlling the input sentence can be a very useful and effective way to improve machine translation quality.

### **7.2.3. Future Research**

Machine translation research has recently been oriented towards statistical machine translation and evaluation methods. In addition, spoken language translation and crowd-translation has been prominent research areas. The increasing number of new language-pairs in machine translation can be regarded as an indicator of the expanding coverage of MT.

This thesis has shed some light to machine translation evaluation, which is one of the promising research areas in MT. Future research can be conducted to answer the following questions:

- What is the quality of machine translation programs working from Turkish into English?
- What are the possible reasons for different success rates of machine translation programs working from Turkish into English and from English into Turkish?
- Is it possible to use post-editing effort as an evaluation method of MT performance?
- What do translators think about machine translation programs?
- What is the place of MT in the workflow of a professional translator?
- Is it possible to create a corpus for automatic subtitle translation from the available multilingual subtitles?
- Does the use of controlled language increase the practicality of MT in terms of the pre- and post-editing efforts?
- Is it possible to find a schema for reusing the human assessments in one MT evaluation schema for another MT evaluation schema?
- What are the differences between monolingual and bilingual human annotators evaluating the same texts?

## BIBLIOGRAPHY

- Allen, J. (2001). *Post-editing or no post-editing?* Regular column article in the International Journal for Language and Documentation, Issue 8, December 2000/January 2001. pp. 41-42.
- ALPAC. National Academy of Sciences, National Research Council, Division of Behavioral Sciences. (1966). *Language and machines: computers in translation and linguistics*. Washington: A report by the Automatic Language Processing Advisory Committee.
- Amaç*. (2015). Retrieved from <http://www.tnc.org.tr/index.php/tr/amac>
- Aref, M., Al-Mulhem, M., & Al-Muhtaseb, H. (1995, January). *English to arabic machine translation: A critical review and suggestions for development*. The fourth Saudi engineering conference. Retrieved from [http://coe.kau.edu.sa/Files/320/Researches/47269\\_18766.pdf](http://coe.kau.edu.sa/Files/320/Researches/47269_18766.pdf)
- Arnold, D. (2003). Why translation is difficult for computers. In H. Somers (Ed.), *Computers and Translation* Amsterdam: John Benjamins B.V.
- Arnold, D., Balkan, L., Meijer, S., Humphreys, R. L., & Sadler, L. (1994). *Machine translation an introductory guide*. London: Blackwell Publishers.
- Aziz, W., C. M. de Sousa, S., & Specia, L. (2012, May). *Pet: a tool for post-editing and assessing machine translation*. 8th international conference on language resources and evaluation (Irec), İstanbul, Turkey.
- Baker, A., & Hengeveld, K. (2012). *Linguistics*. West-Sussex: Wiley.
- Bar-Hillel, Y. (1959). *The present status of automatic translation of languages*. In S. Nirenburg, J. Hutchins & H. Somers (Eds.), *Readings in Machine Translation* London: The MIT Press.

- Berka, J., Černý, M., & Bojar, O. (2011). Quiz-based evaluation of machine translation. *The Prague Bulletin of Mathematical Linguistics*, (95), 77-86.
- Berth, A., & Gdaniec, C. (2001). Mtranslatability. *Machine Translation*, 16, 175-218. Retrieved from [http://download-v2.springer.com/static/pdf/121/art%3A10.1023%2FA%3A1019867030786.pdf?token2=exp=1429523733~acl=/static/pdf/121/art%253A10.1023%252FA%253A1019867030786.pdf\\*~hmac=5b227ea3d652005bb03be1b31e17568a961c1602741283b6abfaf8e4fbbb7346](http://download-v2.springer.com/static/pdf/121/art%3A10.1023%2FA%3A1019867030786.pdf?token2=exp=1429523733~acl=/static/pdf/121/art%253A10.1023%252FA%253A1019867030786.pdf*~hmac=5b227ea3d652005bb03be1b31e17568a961c1602741283b6abfaf8e4fbbb7346)
- Bird, S. (2003). Phonology. In R. Mitkov (Ed.), *The Oxford Handbook of Computational Linguistics*. Oxford: Oxford University Press.
- Bowker, L. (2003). Terminology tools for translators. In H. Somers (Ed.), *Computers and Translation*. Amsterdam: John Benjamins B.V.
- Bozşahin, C. (2002). The combinatory morphemic lexicon. “*Computational Linguistics*”.
- Brown, P., S. Della Pietra, V. Della Pietra, and R. Mercer. (1991). ‘Word-sense disambiguation using statistical methods’. *Proceedings of the 29<sup>th</sup> Meeting of the Association for Computational Linguistics (Berkeley)*.
- Buchmann, B. (1987). Early history of machine translation. In M. King (Ed.), *Machine Translation Today: The State of The Art*. Edinburgh: Edinburgh University Press.
- Cancedda, N., Dymetman, M., Foster, G., & Goutte, C. (2009). A statistical machine translation primer. In N. Cancedda, M. Dymetman, G. Foster & C. Goutte (Eds.), *Learning Machine Translation*. Cambridge, Massachusetts: The MIT Press.
- Çeviri yaparken nelere dikkat etmeliyim?. (n.d.). Retrieved from <http://www.sametran.com/index.php?content=dikkatedilmesi>
- Chris Callison-Burch, Cameron Fordyce, Philipp Koehn, Christof Monz, and Josh Schroeder. (2008). Further meta-evaluation of machine translation. In Proceedings of the Third Workshop on Statistical Machine Translation (WMT08), Columbus, Ohio.

- Chris Callison-Burch, Philipp Koehn, Christof Monz, Matt Post, Radu Soricut, and Lucia Specia. (2012). Findings of the 2012 workshop on statistical machine translation. In Proceedings of the Seventh Workshop on Statistical Machine Translation, pages 10–51, Montreal, Canada, June. Association for Computational Linguistics.
- Corporate profile*. (n.d.). Retrieved from <http://www.systransoft.com/translation-products/desktop/systran-7-premium-translator/>
- De Roeck, A. (1987). Linguistic theory and early machine translation. In M. King (Ed.), *Machine Translation Today: The State of The Art*. Edinburgh: Edinburgh University Press.
- Delavenay, E. (1960). *An introduction to machine translation*. London: Thames and Hudson.
- Dostert, L. E. (1955). The georgetown-ibm experiment. In W. N. Locke & A. D. Booth (Eds.), *Machine translation of languages* Cambridge: The MIT Press.
- Elliott, D., Hartley, A., & Atwell, E. (2003). *Rationale for a multilingual corpus for machine translation evaluation*. In Dawn Archer, Paul Rayson, Andrew Wilson and Tony McEnery (Eds.), *UCREL Technical Paper number 16. Special issue*.
- Erten, A. (1994). Çeviride bilgisayar kullanımı Çağı yakalamanın ve İletişimin sürat boyutu. *Gazi Üniversitesi İletişim Fakültesi Dergisi*, 1-2, 35-42
- Etchegoyhen, T., Bywood, L., Fishel, M., Georgakopoulou, P., Jiang, J., Van Loenhout, G., del Pozo, A., & Sepesy Maucec, M. (2014, May). *Machine translation for subtitling: A large-scale evaluation*. The 9th edition of the language resources and evaluation conference, Reykjavik, Iceland.
- Flanagan, M. (1994). *Error classification for mt evaluation*. Technology partnerships for crossing the language barrier: proceedings of the first conference of the association for machine translation, Columbia.



- Flaganan, M. (2009). *Recycling texts: Human evaluation of example-based machine translation subtitles for dvd*. (Doctoral dissertation).
- Fraser, S. (2013, December 2). *Sumat - an online service for subtitling by machine translation*. Retrieved from [http://cordis.europa.eu/fp7/ict/language-technologies/project-sumat\\_en.html](http://cordis.europa.eu/fp7/ict/language-technologies/project-sumat_en.html)
- Frequently asked questions about dg translation*. (27, May 2014). Retrieved from [http://ec.europa.eu/dgs/translation/faq/index\\_en.htm](http://ec.europa.eu/dgs/translation/faq/index_en.htm)
- Genzel, D., Uszkoreit, J., & Och, F. (2010). "poetic" statistical machine translation: Rhyme and meter. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing* (pp. 158-166). Association for Computational Linguistics.
- Gotti, F., Langlais, P. & Lapalme, G. (2014). Designing a machine translation system for Canadian weather warnings: A case study . *Natural Language Engineering*, 20, pp 399-433 doi:10.1017/S135132491300003X
- Güngördü, Z., & Oflazer, K. (1995). Parsing turkish using the lexical functional grammar formalism. *Machine Translation*, 10(4), 293-319.
- Hutchins, J., & Somers, H. L. (1992). *An introduction to machine translation*. London: Academic Press Limited.
- Hutchins, W. J. (1986). Machine translation: Past, present, future. *Computers and their applications*, ed B. L. Meek. New York: Halsted Press.
- Jiang, L and Zhou M. (2008). *Generating Chinese couplets using a statistical MT approach*. In *Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008)*, pages 377– 384, Manchester, UK, August. Coling 2008 Organizing Committee.
- Jurafsky, D., & Martin, J. H. (2006). *Speech and language processing*. New Jersey: Prentice Hall.

- King, M. & Perschke, S. (1987). Eurotra. In M. King (Ed.), *Machine Translation Today: The State of The Art*. Edinburgh: Edinburgh University Press.
- Kittredge, R. I. (2003). Sublanguages and controlled languages. In R. Mitkov (Ed.), *The Oxford Handbook of Computational Linguistics* New York: Oxford University Press.
- Knowles, Francis. (1982). The Pivotal Role of the Dictionaries in a Machine Translation System. In Lawson, Veronica, ed. *Practical Experience of Machine Translation*. North-Holland.
- Köksal, A. (2012). *A first approach to a computerized model for the automatic morphological analysis of Turkish and two related articles in Turkish*. (Doctoral dissertation).
- Kornfilt, J. (1997). *Turkish*. London and New York: Routledge.
- Küçük, D. & Turhan-Yöndem, M. (2007). A knowledge-poor pronoun resolution system for Turkish. *Proceedings of the 6th Discourse Anaphora and Anaphor Resolution Colloquium*.
- Lesk, M. (1986). 'Automatic sense disambiguation using machine readable dictionaries: how to tell a pine cone from an ice cream cone'. *Proceedings of ACM SIGDOC Conference (Toronto)*.
- Llitjos, F., Aranovich, R., & Levin, L. (2005). *Building machine translation systems for indigenous languages of Latin America*. Second conference on the indigenous languages of Latin America (CILLA II).
- Machine translation service*. (2015, April 27). Retrieved from [http://ec.europa.eu/isa/actions/02-interoperability-architecture/2-8action\\_en.htm](http://ec.europa.eu/isa/actions/02-interoperability-architecture/2-8action_en.htm)
- Maegaard, B. (1995). EUROTRA, History and results. In: *Proceedings of MT Summit V*, Luxembourg.

- Mausser, A. (2014, December 10). Google translate - now in 80 languages [Web log message]. Retrieved from <http://googletranslate.blogspot.com.tr/>
- McEnery, T. (2003). Corpus linguistics. In R. Mitkov (Ed.), *The Oxford Handbook of Computational Linguistics*. New York: Oxford University Press.
- Melby, A. (1992). The translator workstation. In J. Newton (Ed.), *Computers in translation: a practical appraisal*. New York: Routledge.
- Melby, A. K., & Warner, T. (1995). *The possibility of language*. Amsterdam / Philadelphia: John Benjamins Publishing Company.
- Metu turkish corpus*. (n.d.). Retrieved from <http://ii.metu.edu.tr/tr/corpus>
- Mikheev, A. (2003). Text segmentation. In R. Mitkov (Ed.), *The Oxford Handbook of Computational Linguistics*. New York: Oxford University Press.
- Mitkov, R. (2003). Anaphora Resolution. In R. Mitkov (Ed.), *The Oxford Handbook of Computational Linguistics*. New York: Oxford University Press.
- Munday, Jeremy (ed.) *The Routledge Companion to Translation Studies*. New York: Routledge, 20097
- Nirenburg, S. (1987). Knowledge and choices in machine translation. In S. Nirenburg (Ed.), *Machine Translation*. Cambridge: Cambridge University Press.
- Nyberg, E., Mitamura, T. & Huijsen, W. O. (2003). Controlled language for authoring and translation. In H. Somers (Ed.), *Computers and Translation*. Amsterdam: John Benjamins B.V.
- Och, Franz. "Breaking down the language barrier." *Google Official Blog*. Google Inc., 26 Apr 2012. Web. 12 Jun. 2013. <<http://googleblog.blogspot.com/2012/04/breaking-down-language-barriersix-years.html>>.

- Oflazer, K. and Kuruöz, İ. (1994). Tagging and morphological disambiguation of Turkish text. In *Proceedings of the 4th Applied Natural Language Processing Conference*, p. 144–149. ACL.
- Orhan, Z. and Altan, Z. 2006. *Impact of Feature Selection for Corpus-Based WSD in Turkish*, LNAI, Springer-Verlag, Vol. 4293: 868-878.
- Oulton, N. (2013). Medium and long run prospects for uk growth in the aftermath of the financial crisis. *CEP Occasional Papers*, Retrieved from [http://cep.lse.ac.uk/\\_new/publications/abstract.asp?index=4323](http://cep.lse.ac.uk/_new/publications/abstract.asp?index=4323)
- Papineni, K., Roukos, S., Ward, T., and Zhu, W. J. (2002). "BLEU: a method for automatic evaluation of machine translation" in *ACL-2002: 40th Annual meeting of the Association for Computational Linguistics* pp. 311–318
- Petitpierre, D. (1987). Software background for machine translation: a glossary. In M. King (Ed.), *Machine Translation Today: The State of The Art*. Edinburgh: Edinburgh University Press.
- Pugh, J. (1992). The story so far: an evaluation of machine translation in the world today. In J. Newton (Ed.), *Computers in translation: a practical appraisal*. New York: Routledge.
- Reiss, K. (2000). *Translation Criticism: The Potentials and Limitations. Categories and Criteria for Translation Quality Assessment*, tr. Erroll. F. Rhodes. Manchester: St. Jerome Publishing Ltd.
- Reiss, Katharina. "Text Types, Translation Types and Translation Assessment," In Andrew Chesterman (Ed.) *Readings in Translation Theory*. Helsinki, Oy Finn Lectura Ab, 1989, 105-115.
- Shann, P. (1987). Machine translation: a problem of linguistic engineering or of cognitive modelling. In M. King (Ed.), *Machine Translation Today: The State of The Art*. Edinburgh: Edinburgh University Press.




- Snover, M., Dorr, B., Schwartz, R., Micciulla, L., & Makhoul, J. (2006, August). *A study of translation edit rate with targeted human annotation*. 7th conference of the association for machine translation in the americas, Cambridge.
- Somers, H. (1992). Current research in machine translation. In J. Newton (Ed.), *Computers in translation: a practical appraisal* New York: Routledge.
- Somers, H. (2003). Sublanguage. In H. Somers (Ed.), *Computers and Translation* Amsterdam: John Benjamins B.V.
- Systran 7 premium translator*. (n.d.). Retrieved from <http://www.systransoft.com/translation-products/desktop/systran-7-premium-translator/>
- Tanner, A. (2007, March 28). *google seeks world of instant translations*. Retrieved from <http://www.reuters.com/article/2007/03/28/us-google-translate-idUSN1921881520070328>
- Tantuĝ , A. C., Oflazer, K., & Durgar El-Kahlout, I. (2008). *Bleu : A fine grained tool for bleu computation*. Language resources and evaluation conference lrec, Morocco.
- Taube, M. (1961). *Computers and common sense: the myth of thinking machines*. New York: Columbia University Press.
- The European Community, (1990). *Eurotra*. OFFICE FOR OFFICIAL PUBLICATIONS.
- The plaza new york*. (n.d.). Retrieved from <http://www.fivestaralliance.com/luxury-hotels/new-york-ny/the-plaza-new-york>
- The translational english corpus (tec)*. (n.d.). Retrieved from <http://www.monabaker.com/tsresources/TranslationalEnglishCorpus.htm>
- Trost, H. (2003). Morphology. In R. Mitkov (Ed.), *The Oxford Handbook of Computational Linguistics*. New York: Oxford University Press.

- Tucker, A. B. (1987). Current strategies in machine translation research and development. In S. Nirenburg (Ed.), *Machine Translation*. Cambridge: Cambridge University Press.
- U.S. House of Representatives (1960): Research on mechanical translation. Report of the Committee on Science and Astronautics, U.S. House of Representatives, Eighty-sixth Congress, Second Session, House report 2021, June 28, 1960. Washington: U.S. Government Printing Office.
- Van Slype, G. (1979). Critical Study of Methods for Evaluating the Quality of Machine Translation. Final Report, Bureau Marcel van Dijk / European Commission, Brussels.
- Vauquois, B. (2003). Automatic translation - a survey of different approaches. In S. Nirenburg, H. Somers & Y. Wilks (Eds.), *Readings in Machine Translation*. Cambridge: The MIT Press.
- Vilar, David, Jia Xu, Luis Fernando D'Haro, and Hermann Ney. (2006, May). Error Analysis of Statistical Machine Translation Output. In Proceedings of the 5<sup>th</sup> International Conference on Language Resources and Evaluation (LREC 06), pages 697–702, Genoa, Italy.
- Warwick, S. (1987). An overview of post-alpac developments. In M. King (Ed.), *Machine Translation Today: The State of Art* (pp. 22-37). Edinburgh: Edinburgh University Press.
- Weaver, W. (1949). Translation. In S. Nirenburg, H. Somers & Y. Willks (Eds.), *Readings in Machine Translation*. London: The MIT Press.
- Wheeler, P. (1987). Systran. In M. King (Ed.), *Machine Translation Today: The State of The Art*. Edinburgh: Edinburgh University Press.
- White, J. (2003). How to evaluate machine translation. In H. Somers (Ed.), *Computers and Translation*. Amsterdam: John Benjamins B.V.
- Wilks, Y. (1972). *Grammar, meaning and the machine analysis of language*. London: Routledge & Kegan Paul.

Wu, C. (2012, March 28). Using labels in translator toolkit [Web log message]. Retrieved from <http://googletranslate.blogspot.com.tr/>

Yule, G. (2010). *The study of language*. Cambridge: Cambridge University Press.

# Appendix 1: TEZ ÇALIŞMASI ETİK KURUL İZİN MUAFİYETİ FORMU

|  |
|--|
|  <p><b>HACETTEPE ÜNİVERSİTESİ</b><br/><b>SOSYAL BİLİMLER ENSTİTÜSÜ</b><br/><b>TEZ ÇALIŞMASI ETİK KURUL İZİN MUAFİYETİ FORMU</b></p>   |
| <p><b>HACETTEPE ÜNİVERSİTESİ</b><br/><b>SOSYAL BİLİMLER ENSTİTÜSÜ</b><br/><b>MÜTERCİM TERCÜMANLIK ANABİLİM DALI BAŞKANLIĞI'NA</b></p> <p style="text-align: right;">Tarih:26.05.2015</p>   |
| <p>Tez Başlığı / Konusu: Bilgisayar Çevirisi Kalitesinin Değerlendirmesi Yöntemlerinde Tutarlılık</p> <p>Yukarıda başlığı/konusu gösterilen tez çalışmam:</p> <ol style="list-style-type: none"> <li>1. İnsan ve hayvan üzerinde deney niteliği taşımamaktadır,</li> <li>2. Biyolojik materyal (kan, idrar vb. biyolojik sıvılar ve numuneler) kullanılmasını gerektirmemektedir.</li> <li>3. Beden bütünlüğüne müdahale içermemektedir.</li> <li>4. Gözlemsel ve betimsel araştırma (anket, ölçek/skala çalışmaları, dosya taramaları, veri kaynakları taraması, sistem-model geliştirme çalışmaları) niteliğinde değildir.</li> </ol> <p>Hacettepe Üniversitesi Etik Kurulları ve Komisyonlarının Yönergelerini inceledim ve bunlara göre tez çalışmamın yürütülebilmesi için herhangi bir Etik Kuruldan izin alınmasına gerek olmadığını; aksi durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.</p> <p>Gereğini saygılarımla arz ederim.</p> <p style="text-align: right;">26.05.2015<br/></p> <p><b>Adı Soyadı:</b> Özden Şahin<br/><b>Öğrenci No:</b> N12128381<br/><b>Anabilim Dalı:</b> Mütercim Tercümanlık Anabilim Dalı<br/><b>Programı:</b> İngilizce Mütercim Tercümanlık<br/><b>Statüsü:</b> <input checked="" type="checkbox"/> Y.Lisans <input type="checkbox"/> Doktora <input type="checkbox"/> Bütünleşik Dr.</p> |
| <p><b>DANIŞMAN GÖRÜŞÜ VE ONAYI</b></p> <p style="text-align: center;">Uygundur</p> <p style="text-align: center;"><br/>Prof. Dr. Aymil Doğan</p> <p><b>Detaylı Bilgi:</b> <a href="http://www.sosyalbilimler.hacettepe.edu.tr">http://www.sosyalbilimler.hacettepe.edu.tr</a><br/><b>Telefon:</b> 0-312-2976860 <b>Faks:</b> 0-3122992147 <b>E-posta:</b> <a href="mailto:sosyalbilimler@hacettepe.edu.tr">sosyalbilimler@hacettepe.edu.tr</a></p>   |





HACETTEPE UNIVERSITY  
GRADUATE SCHOOL OF SOCIAL SCIENCES  
ETHICS BOARD WAIVER FORM FOR THESIS WORK

HACETTEPE UNIVERSITY  
GRADUATE SCHOOL OF SOCIAL SCIENCES  
TO THE DEPARTMENT PRESIDENCY OF TRANSLATION AND INTERPRETATION

Date: 26.05.2015

Thesis Title / Topic: Consistency in the Evaluation Methods of Machine Translation Quality

My thesis work related to the title/topic above:

1. Does not perform experimentation on animals or people.
2. Does not necessitate the use of biological material (blood, urine, biological fluids and samples, etc.).
3. Does not involve any interference of the body's integrity.
4. Is not based on observational and descriptive research (survey, measures/scales, data scanning, system-model development).

I declare, I have carefully read Hacettepe University's Ethics Regulations and the Commission's Guidelines, and in order to proceed with my thesis according to these regulations I do not have to get permission from the Ethics Board for anything; in any infringement of the regulations I accept all legal responsibility and I declare that all the information I have provided is true.

I respectfully submit this for approval.

26.05.2015  
[Signature]

**Name Surname:** Ozden Şahin  
**Student No:** N12128381  
**Department:** Translation and Interpretation Department  
**Program:** English Translation and Interpretation  
**Status:**  Masters  Ph.D.  Integrated Ph.D.

**ADVISER COMMENTS AND APPROVAL**

Approved -

Prof. Dr. Aymil Doğan

## Appendix 2: ORJİNALLİK RAPORU

|  |
|--|
|  <p><b>HACETTEPE ÜNİVERSİTESİ</b><br/><b>SOSYAL BİLİMLER ENSTİTÜSÜ</b><br/><b>YÜKSEK LİSANS/DOKTORA TEZ ÇALIŞMASI ORJİNALLİK RAPORU</b></p>   |
| <p><b>HACETTEPE ÜNİVERSİTESİ</b><br/><b>SOSYAL BİLİMLER ENSTİTÜSÜ</b><br/><b>MÜTERCİM TERCÜMANLIK ANABİLİM DALI BAŞKANLIĞI'NA</b></p> <p style="text-align: right;">Tarih: 16.06.2015</p> <p>Tez Başlığı / Konusu: Bilgisayar Çevirisi Kalitesinin Değerlendirmesi Yöntemlerinde Tutarlılık.</p> <p>Yukarıda başlığı/konusu gösterilen tez çalışmamın a) Kapak sayfası, b) Giriş, c) Ana bölümler ve d) Sonuç kısımlarından oluşan toplam 158 sayfalık kısmına ilişkin, 14/06/2015 tarihinde şahsım/ tarafından Turnitin adlı intihal tespit programından aşağıda belirtilen filtrelemeler uygulanarak alınmış olan orijinallik raporuna göre, tezimin benzerlik oranı %4'tür.</p> <p>Uygulanan filtrelemeler:</p> <ol style="list-style-type: none"> <li>1- Kabul/Onay ve Bildirim sayfaları hariç,</li> <li>2- Kaynakça hariç</li> <li>3- Alıntılar hariç/dâhil</li> <li>4- 5 kelimedenden daha az örtüşme içeren metin kısımları hariç</li> </ol> <p>Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü Tez Çalışması Orijinallik Raporu Alınması ve Kullanılması Uygulama Esasları'nı inceledim ve bu Uygulama Esasları'nda belirtilen azami benzerlik oranlarına göre tez çalışmamın herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.</p> <p>Gereğini saygılarımla arz ederim.</p> <div style="text-align: right;">  </div> <p><b>Adı Soyadı:</b> Özden ŞAHİN</p> <p><b>Öğrenci No:</b> N12128381</p> <p><b>Anabilim Dalı:</b> Mütercim Tercümanlık Anabilim Dalı</p> <p><b>Programı:</b> İngilizce Mütercim Tercümanlık</p> <p><b>Statüsü:</b> <input checked="" type="checkbox"/> Y.Lisans <input type="checkbox"/> Doktora <input type="checkbox"/> Bütünleşik Dr.</p> |
| <p><b><u>DANIŞMAN ONAYI</u></b></p> <p>UYGUNDUR.</p> <div style="text-align: center;">  </div> <p>Prof. Dr. Aymil Doğan</p>  |



**HACETTEPE UNIVERSITY  
GRADUATE SCHOOL OF SOCIAL SCIENCES  
THESIS/DISSERTATION ORIGINALITY REPORT**

**HACETTEPE UNIVERSITY  
GRADUATE SCHOOL OF SOCIAL SCIENCES  
TO THE DEPARTMENT OF TRANSLATION AND INTERPRETATION**

Date: 16.06.2015

Thesis Title / Topic: Consistency in the Evaluation Methods of Machine Translation Quality.

According to the originality report obtained by myself by using the Turnitin plagiarism detection software and by applying the filtering options stated below on 14/06/2015 for the total of 158 pages including the a) Title Page, b) Introduction, c) Main Chapters, and d) Conclusion sections of my thesis entitled as above, the similarity index of my thesis is 4 %.

Filtering options applied:

1. Approval and Declaration sections excluded
2. Bibliography/Works Cited excluded
3. Quotes excluded
4. Match size up to 5 words excluded

I declare that I have carefully read Hacettepe University Graduate School of Social Sciences Guidelines for Obtaining and Using Thesis Originality Reports; that according to the maximum similarity index values specified in the Guidelines, my thesis does not include any form of plagiarism; that in any future detection of possible infringement of the regulations I accept all legal responsibility; and that all the information I have provided is correct to the best of my knowledge.

I respectfully submit this for approval.

**Name Surname:** Özden Şahin  
**Student No:** N12128381  
**Department:** Translation and Interpretation Department  
**Program:** English Translation and Interpretation  
**Status:**  Masters  Ph.D.  Integrated Ph.D.

**ADVISOR APPROVAL**

APPROVED.

Prof. Dr. Aymil Doğan

**Appendix 3: CD**

**THIS CD INCLUDES THE SURVEY AND ITS RESULT**