Universitat Autònoma de Barcelona

Terminological Quality Evaluation in Turkish to English Corpus-Based Machine Translation in Medical Domain

Gökhan Doğru

Terminological Quality Evaluation in Turkish to English Corpus-Based Machine Translation in Medical Domain

Gökhan Doğru

Doctorat en Traducció i Estudis Interculturals
Departament de Traducció i d'Interpretació i d'Estudis d'Àsia Oriental

Gökhan Doğru

Terminological Quality Evaluation in Turkish to English Corpus-Based Machine Translation in Medical Domain

Doctoral Thesis Supervised by:

Anna Aguilar-Amat

&

Adrià Martín-Mor

Bellaterra, February 2021

Table of Contents

Acknowledgments

This dissertation has been possible thanks to the great support and assistance from many great people. I would firstly like to thank my co-supervisors Anna Aguilar-Amat and Adrià Martín-Mor for their inspiring wisdom, never-ending patience, and insightful feedback about my work. They have been inspiring role models for me in the long way to not only being a good researcher but also being a good person. I also acknowledge all members of Tradumàtica Research Group for being so kind to share their knowledge and experience in our regular meetings. I am thankful to the Agency for Management of University and Research Grants (AGAUR) in Catalonia for funding my research with AGAUR-FI-2017 grant.

I am really grateful to Joss Moorkens for his eternal kindness and letting me have a great research visit experience in Dublin City University. I also appreciate the willingness of Rejwanul Haque from ADAPT Centre to help me organize my thoughts in my dissertation. I am candidly grateful to all the scientific community for creating the knowledge that makes new research possible. I deeply feel indebted to healthcare professionals who are fighting against COVID-19 disease relentlessly. Their superhuman efforts will never be forgotten.

I am also grateful to my family members who, despite being many kilometers away, were always by my side to support me: my father Ganim Doğru, my mother Nuray Doğru, my siblings Aylin, Okan, Onur and Çağdaş. A huge thanks goes to my friends in Barcelona who have always made me feel lucky and surrounded by eternal compassion and support. Ultimately, I would like to thank and dedicate my thesis to my dear wife Esra Çinar for her timeless support, care, and encouragement in every moment of the tough writing process. So glad that I have you by side!

List of Tables and Figures

List of tables

List of figures

List of Acronyms

BLEU	Bilingual Evaluation Understudy
CAT	Computer Assisted Translation
DQF	Dynamic Quality Framework
EU	European Union
GENCOR	Generic Corpus
LQR	Linguistic Quality Review
MQM	Multidimensional Quality Metrics
MT	Machine Translation
MWU	Multi Word Unit
NMT	Neural Machine Translation
PBMT	Phrase-based Machine Translation
QA	Quality Assurance
SMT	Statistical Machine Translation
TER	Translation Error Rate
TRENCARD	Turkish English Cardiology Corpus
UN	United Nations

Abstract

General quality aspects of machine translation (MT) such as adequacy and fluency are studied extensively, more fine-grained aspects such as the terminology translation quality have not received much attention especially in the context of translation studies. The objective of this study is to analyze the types and frequencies of terminology errors in custom statistical machine translation (SMT) and neural machine translation (NMT) with the goal of understanding how MT system type, corpus type and corpus size affect the terminology translation quality.

A Turkish – English parallel corpus obtained from cardiology journal abstracts was built from scratch for training domain-specific SMT and NMT engines. Then, this domain-specific corpus is combined with a mixed domain corpus and two more engines were trained. After conducting automatic evaluation and human evaluation on these 4 engines, terminology errors were annotated based on a custom terminology error typology. It was found that the types and frequencies of terminology errors are significantly different in SMT and NMT systems, and that changes in corpus size and corpus type had more drastic impact on NMT compared to SMT.

A key contribution of the dissertation to the MT research is the crafted language-agnostic terminology error typology which can be used for evaluating the relative strengths and weakness of different MT systems in terms of terminology. Besides, the finding that NMT systems exhibit different types of term errors with different frequencies implies that postediting guidelines conceived specifically for SMT systems could require changes to accommodate the behavior pattern of NMT..

Resum

Els aspectes generals de qualitat de la traducció automàtica (TA), com l'adequació i la fluïdesa, s'han estudiat àmpliament, però els aspectes més detallats, com la qualitat de la traducció de la terminologia, s'han subestimat, especialment en el context dels estudis de traducció. L'objectiu d'aquest estudi és analitzar els tipus i freqüències d'errors terminològics en la traducció automàtica estadística (TAE) i la traducció automàtica neuronal (TAN) personalitzades amb l'objectiu final de comprendre com el tipus de sistema de TA, el tipus de corpus i la mida del corpus afecten la qualitat de la traducció de terminologia.

Un corpus paral·lel turc-anglès obtingut a partir de resums de revistes de cardiologia va ser creat des de zero per entrenar motors de TAE i TAN de dominis específics. Després, aquest corpus es va combinar amb un corpus de domini mixt i es van entrenar dos motors més. Després de realitzar una avaluació automàtica i una avaluació humana en aquests 4 motors, els errors de terminologia es van anotar segons una tipologia d'errors de terminologia personalitzada. S’ha trobat que els tipus i freqüències dels errors terminològics són significativament diferents en els sistemes TAE i TAN, i que els canvis en la mida i tipus de corpus han tingut un impacte més dràstic en el TAN en comparació amb la TAE.

Una contribució clau de la dissertació a la investigació sobre TA és la tipologia d'error de terminologia independent del llenguatge per avaluar les fortaleses i debilitats relatives de diferents sistemes de TA en termes de terminologia. A més, la troballa que els sistemes TAN exhibeixen diferents tipus d'errors de terme amb diferents freqüències implica que les guies de postedició concebudes específicament per a sistemes TAE podrien requerir canvis per tal d’adaptar-se al nou patró de comportament de TAN.

Resumen

Los aspectos generales de calidad de la traducción automática (TA), como la adecuación y la fluidez, se han estudiado ampliamente, pero los aspectos más detallados, como la calidad de la traducción de la terminología, se han subestimado, especialmente en el contexto de los estudios de traducción. El objetivo de este estudio es analizar los tipos y frecuencias de errores terminológicos en la traducción automática estadística (TAE) y la traducción automática neuronal (TAN) con el objetivo final de comprender cómo el tipo de sistema de TA, el tipo de corpus y el tamaño del corpus afectan la calidad de la traducción de terminología.

Un corpus paralelo turco-inglés obtenido a partir de resúmenes de revistas de cardiología se creó desde cero para entrenar motores de TAE y TAN de dominios específicos. Luego, este corpus se combina con un corpus de dominio mixto y se entrenaron dos motores más. Después de realizar una evaluación automática y una evaluación humana en estos 4 motores, los errores de terminología se anotaron según una tipología de errores de terminología personalizada. Se ha encontrado que los tipos y frecuencias de los errores terminológicos son significativamente diferentes en los sistemas TAE y TAN, y que los cambios en el tamaño y tipo de corpus tienen un impacto más drástico en el TAN en comparación con el TAE.

Una contribución clave de la disertación es la tipología de error de terminología que se puede utilizar para evaluar las fortalezas y debilidades relativas de diferentes sistemas de TA en términos de terminología. Además, el hallazgo de que los sistemas TAN exhiben diferentes tipos de errores en los términos con diferentes frecuencias implica que las pautas de posedición que se prepararon para los textos resultantes de TAE deben actualizarse para adaptarse al nuevo patrón de comportamiento de TAN.

Introduction

Machine translation (MT) has become ubiquitous in the last two decades thanks to the advances in research and industrial implementations. In line with their investments in developing machine translation systems, major technology companies such as Google, Microsoft etc. provide generic MT platforms for translating any kind of content on the Web. Google provides MT in more than 100 languages translating an average of 143 billion words a day (Way, 2019). Besides, language service providers and independent translators benefit from MT either by connecting their computer-assisted translation (CAT) tools to these and other platforms or by developing their custom MT technologies. This growing interest in MT has been enabled by two corpus-based paradigms: statistical machine translation (SMT) and neural machine translation (NMT).

While the research on SMT begun by the end of the 1980s and slowly grew in the 1990s, the introduction of free and open source Moses Toolkit in 2006 accelerated its adoption (Koehn, 2020, p. 37). Coupled with that, increasing amount of free and open parallel corpus (for example, parliamentary proceedings of the European Parliament: Koehn, 2005) and faster computers allowed companies to create their own, custom MT systems, and researchers to experiment with different SMT research setups. The translation quality of these SMT systems varied based, among other things, on language pair, parallel corpora size and corpus type. Languages such English, French and Spanish which are official languages of international organizations such as the EU and the UN had significantly more free and open parallel corpora and hence were the first to benefit from the advances in SMT. Nevertheless, SMT systems trained with low-resource languages such as Turkish, Kazakh etc. did not yield satisfactory results both due to the scarcity of training corpora and their grammatical distance to their usual MT pair, English, since SMT's algorithm did not provide adequate results with morphologically rich languages. However, in time, best practices were developed in the translation industry to allow for some degree of standardization for the widespread use of SMT properly; for example, Translation Automation User Society (TAUS) developed postediting guidelines[1] in 2010 in collaboration with scholars and industry professionals.

After 2015, a new approach to corpus-based MT, NMT, began to compete with the state-of-the-art SMT in terms of translation quality (Koehn, 2020, p. 33). Especially after Google's transition to NMT (Wu et al., 2016), the interest in the topic increased more and many comparative studies (Bentivogli, 2016; Castilho et al., 2017; Toral&Sanchez-Cartagena, 2017, Shterionov et al., 2018) were published comparing NMT with SMT. Going beyond comparison, Hassan et al. (2018) even claimed that the quality of their Chinese to English NMT system is "at human parity". Toral et al. (2018) warned against such claims due to the methodological concerns in human quality evalution where crowd-sourced, non-expert translators are included. Besides, Castilho et al. (2018) stated, in their fine-grained comparison study including MT in English to four languages (German, Russian, Greek and Portuguese), that in some language pairs and domains, SMT still provided better quality while NMT is usually claimed to perform overwhelmingly. It should be noted that the above-mentioned studies are usually conducted using language pairs with high amounts of general domain and specific domain, open parallel corpora. Similar, comparative studies with Turkish to

[1] MT Post-editing Guidelines. https://www.taus.net/academy/best-practices/postedit-best-practices/machine-translation-post-editing-guidelines (last access: 26.01.2021)

English SMT and NMT require large amounts of high quality, specific domain, open, parallel corpora. The currently available open corpora for Turkish to English MT in Opus Corpus (Tiedemann, 2012) consist of volunteer translations from Turkish to English in limited number of domains including news, subtitling and IT while, for instance, there are high quality corpora from the UN, European Medicines Agency etc. for Spanish to English language pair to be used for MT training. This scarcity of parallel corpora implies that there is a growing need for effective corpus preparation for low resource languages. And in connection with this scarcity, the comparative translation qualities as well as the weaknesses and strengths of Turkish – English SMT and NMT systems are understudied.

Another understudied aspect of MT that is of interest in our study is the role of terminology in MT quality evaluation. While terminology is a mature discipline with its own theories (including translation theories of terminology) and its principles for studying concepts, its role in MT use, training and evaluation has not been studied comprehensively. Especially its role in MT evaluation has been understudied (Haque et al., 2019). MT quality is evaluated through automatic and human evaluation methods. Automatic evaluations use reference human translations for comparison with MT outputs while human evaluations require collaboration with human reviewers for assessing the MT quality usually based on fluency and adequacy (Castilho et al., 2018). More fine-grained evaluations are possible through error typologies such as the MQM error typology (Lommel, 2018), which specifies 8 main error categories. Terminology is one of the most vital parts of specialized translation and although terminology error is one of the error categories, error typologies usually lack a more fine-grained categorization of terminology error types. Besides, few MT specialists differentiate between specialized terms and other lexical items in their automatic and human evaluations (an exception is Haque et al., 2019). Such a fine-grained terminology error typology plays an important role in systematically understanding how SMT and NMT behave differently considering specialized terminology and allows i). MT specialists to make informed decisions about the preprocessing and/or postprocessing phases of MT development, and ii). translators and posteditors to know what to expect from an MT output based on the system type. Besides, considering the ubiquitous use of MT by general public (see the Google example above), it is crucial to measure the influence of MT on terminology in target cultures/languages. Annotated terminology sets such as the work of Scansani et al. (2019) and terminology error typologies applicable for all MT language pairs such as Haque et al. (2019) are needed for a comprehensive understanding of the impact of MT on terminology. Particularly a terminology error typology prepared from the translation studies perspective will be helpful for providing insights to translators and translation industry about MT.

We observe that thanks to the user-friendly interfaces and permissive licenses for both free (such as MTradumatica[2], OpenNMT[3]) and commercial MT platforms (such as KantanMT[4]), MT customization which involves the use of custom translation resources such as translation memories, and procedures to create personalized MT engines is becoming more accessible for translators, small and mid-size companies and researchers from translation studies departments. Hence, a comprehensive research on the overall phases of MT customization and evaluation may provide beneficial insights for translation stakeholders.

---

[2] https://github.com/tradumatica/mtradumatica (last access: 27.01.2021)

[3] https://opennmt.net/ (last access: 27.01.2021)

[4] https://kantanmt.com/ (last access: 27.01.2021)

Objectives of the research

The main objective of the thesis is to explore the types and frequencies of terminology errors in Turkish to English, custom SMT and NMT engines trained with specific domain (medical, and specifically cardiology) and mixed domain parallel corpora. There are 3 research questions that we aim to answer to reach this main objective:

Q1: Can large amount of high quality, specific domain, Turkish to English parallel corpora be created in semi-automatic procedures that can be used by translators?

Q2: Are the performances of SMT and NMT engines different when they are trained with the same specific domain and mixed domain corpora?

Q3: Are there qualitative and quantitative differences between SMT and NMT engines with regard to the types and frequencies of terminology errors?

By controlling all the main processes and components of MT customization (unlike blackbox systems such as Google Translate or Microsoft Translator), we can draw insights about how each parameter contributes to the overall MT quality and terminology translation quality.

While scholars usually study MT in English to their native language direction, this thesis concentrate on the Turkish to English MT for a few reasons. Firstly, a growing number of doctors in Turkey needs to publish in international journals and custom, Turkish to English MT can be helpful both for them and for researchers from other countries willing to understand the scientific work produced by Turkish medical scholars; the global COVID-19 disease pandemia has shown us that research on coronavirus coming rapidly from all countries to World Health Organization is crucial to halt the progress of the pandemia. Our bilingual cardiology corpora, when published openly, can be used by MT specialists to improve their medical, custom MT systems.

For answering Q1, parallel corpora preparation techniques are explored with the goal of creating a highly specific domain, Turkish to English cardiology corpora from scratch. Bilingual, academic cardiology journals are utilized as resources for building the parallel corpora.

For handling Q2, the obtained parallel corpora are trained in SMT and NMT systems in KantanMT platform; then these specific corpora are combined with a mixed domain corpora and 2 more engines (one SMT and one NMT) are trained; automatic and human evaluation are performed with the objective of comparing the relative translation qualities.

And for the Q3, firstly, a manual term annotation is performed on a sample set of 100 sentences. 11 term error categories are specified and a manual error annotation is conducted on the 100 sentences translated by each of the 4 engines.

The Structure of the Thesis

The main body of the thesis includes a general introduction, three parts containing six chapters and a final chapter on conclusions. Ultimately, a bibliography and annexes are attached.

Part I covers the theoretical and conceptual framework and is organized in two chapters. Chapter 1 surveys the history of machine translation, defines the state-of-the-art machine translation paradigms with a major focus on corpus-based MT systems. Besides, it reports the MT studies with the Turkish language. The chapter also concentrate on corpus and its preparation for machine translation training. Chapter 2 covers the theoretical and conceptual aspects of terminology and its relevance and significance for machine translation. It starts with definitions of key concepts in terminology, and then reports how each MT paradigm handles terminology. Finally, a comprehensive section on Turkish medical terminology describes how medical terminology is used and translated in the context of Turkish.

Part II covers the methodological aspects of the thesis and describes the qualitative and quantitative methods employed in the study starting from the corpus preparation methodology to the terminology evaluation strategies in MT. Chapter 3 consists of two main sections. The first section specifies the tools and methods that we used to create (i) a specific domain Turkish to English parallel corpus from cardiology journals and (ii) a test corpus from the same source; and to compile (iii) a mixed domain parallel corpora from available free and open online corpora. The second section of this chapter is related to the tools and methods used for training SMT and NMT engines. Using the cardiology corpus we trained one SMT and one NMT engines. Then using the mixed domain corpus and specific domain corpus together, we made another round of training to have 2 more engines (one SMT, one NMT). After describing the training process, we explain how automatic and human evaluations of these 4 engines were conducted. Chapter 4 describes the methodology adopted for terminology annotation and terminology error categorization. Custom terminology categories are defined and the methods of manual terminology evaluation in 4 engines are specified.

Part III reports and discusses the quality evaluation results for MT engines in general and terminology translation in particular. Chapter 5 exposes the results of human and automatic evaluation in 4 engines and compares the results based on corpus size, corpus type and MT system type. Chapter 6 reports the results of terminology translation qualities of these engines. The types and frequencies of terminology errors are reported for each engine and then these results are compared between engines based on engine type, term n-gram length and source sentence length.

Once all results are given and discussed, the conclusions will revisit the research questions of the thesis, provide the overall conclusions, explain the limitations of the study and project about the possible future research lines.

Part I. Literature Review

Chapter 1. Machine Translation

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.

-Warren Weaver, 1955

If a lion could speak, we could not understand him.

-L. Wittgenstein, 1953

Machine translation is “the attempt to automate all, or part of the process of translating from one human language to another (Arnold et al., 1994, p. 1)”. When this endeavor began, the ultimate objective was fully automatic high-quality translation (FAHQT). However, after it was understood that this ultimate goal is far from a reality, machine translation (MT) practitioners determined more modest goals such as aiding the translators in translation process and gisting (Fiederer and O’Brien, 2009) where “[t]ranslation does not have to be perfect to bring across the meaning” (Koehn, 2009, p. 21.). Machine translation research began in the 1940s (Arnold et al., 1994, p. 12). Since that time, different methods of automatic translation have been developed with varying degrees of success. Way&Hearne (2011) gather these studies under three titles: rule-based machine translation (RBMT), example-based machine translation (EBMT) and statistical machine translation (SMT). However, since the advent of neural machine translation (NMT), there is a tendency to simplify the categorization in two main categories: rule-based machine translation and corpus-based machine translation. In the broadest sense, corpus-based machine translation includes three main approaches: EBMT, SMT and NMT. However, considering the popularity of their adoptions, and despite the presence of hybrid version, RBMT, SMT and NMT have been the most widely studied and deployed approaches. After a brief outline of the history of machine translation, these approaches will be explained in detail.

1.1. Short History of Machine Translation

The first studies on machine translation dates back to 1940s when the first computers began to be used in the United States (Koehn, 2009). These studies were firstly made for defense purposes against Russia with objective of decoding (translating) Russian communications in a rapid and reliable way. However, the primary trials did not meet the expected quality thresholds. And finally, the notorious ALPAC (Automatic Language Processing Advisory Committee) report in 1964 assessed the performance of MT research and concluded that the output quality of MT was not at a desirable level to be worth the funding expenditure and suggested that there were no advantages of using MT (Koehn, 2009, pp. 15-16). And this caused a domino effect in the field and decreased the number of studies on MT dramatically (Arnold et al., 1994, p. 13). However, though in a smaller scale, MT research continued to be carried on in the 1960s, 1970s and 1980s. As Way&Hearne (2011) states, these studies were collectively referred to as rule-based machine translation, under which Transfer and Interlingual systems were dominant (p. 231). Then, in the 1980s, two new approaches came: EBMT and SMT approaches, both of which, to some extent, are corpus-based strategies. However, at the end, SMT (in fact, a type of it, phrase-based SMT) has become dominant because of its success.

1.1.1. Rule-Based Machine Translation

Rule-Based Machine Translation is one of the oldest approaches to machine translation. It depends on thorough understanding and representation of linguistic information about source language and target language (Way&Hearne, 2011, p. 232). In this approach, the semantic, morphological, and syntactic rules of both languages are represented, and bilingual dictionaries are prepared for the system (Hutchins, 1986; Arnold et al., 1994). When an input sentence (source language) is given, it is firstly analyzed in terms of semantic, morphological, and syntactic structure and a target sentence is generated based on the available resources in the system.

The linguistic analysis and representation in RBMT may have different levels of complexity and as Way&Hearne (2011) observes:

> Generally, the more abstract the representation assigned to the input sentence, the more complex the task of mapping this into an intermediate representation and the more complex the task of generating a final translation. (p. 222)

Together with the necessity of handcrafted bilingual glossaries, this increased complexity makes RBMT preparation a time-consuming and costly task. Besides, even with detailed syntactic analyzers, ambiguities in source sentences cause problems; adapting the system to a particular domain is usually expensive and require the creation of glossaries in the relevant domain. Despite these shortcomings, RBMT had practical applications and continued to be the dominant approach until the 1990s. Early commercial RBMT efforts include Systran[5]'s system in 1970s, Météo MT system used in Canada, Logos and METAL (Koehn, 2020, p. 35).

Example-based machine translation (EBMT) was firstly developed in Japan by Nagao (1984, cited by Way&Hearne, 2011, p. 233). From a historical perspective, it stands between RBMT and SMT since it utilizes both RBMT procedures and SMT procedures to some extent. EBMT is based on an analogy principle. It uses a corpus-based strategy, namely it uses parallel corpora to translate sentences. When an input sentence is given, it is firstly compared by the source language in the corpora and then similar sentences are found. And finally, parts that can be used in the translation of the sentence are extracted, and they are recombined to generate the translation (Way&Hearne, 2011, p. 233). According to Way&Hearne (2011) this approach lacked a well-formed probability model which harmed its scalability to larger applications. Coinciding with the advent of SMT, EBMT did not find much industrial implementation as compared to SMT, and other corpus-based MT systems.

[5] https://www.systransoft.com/ (last access: 10.02.2021)

1.2. Corpus-based Machine Translation

The time-consuming, labor-intensive, and expensive nature of RBMT and EBMT has led scientists to investigate more efficient ways for creating machine translation systems (Way&Hearn, 2011; Koehn, 2009; Kenny&Doherty, 2014). These efforts have encouraged the birth of corpus-based approaches, in which instead of manually writing the grammatical rules and preparing lexical dictionaries for source and target languages from scratch, the scientists created algorithms that allow computers to learn automatically from a bilingual[6] corpus how to translate based on statistical probabilities. The two most common corpus-based machine translation systems are SMT and NMT. Both systems have a step called “training” in which the system is given at least one large bilingual parallel corpus, and the most probable translations for words, phrases or sentences are calculated (a process which may take long time depending on the computational power of the computer used for training). In the following sections, we will explain how each system works, how they are used and how they are different from each other in detail. Nevertheless, before entering into details of these systems, it will be beneficial to explain what corpora are, how they are prepared and used in machine translation since both corpus quality and quantity are very important factors in machine translation. Besides, any corpus-based machine translation system training needs to begin by the preparation of a training corpus or corpora.

The preparation process broadly requires the selection, collection, alignment, and cleaning steps, each of which require an automatic and/or manual processing. The following section will describe corpora and their preparation for machine translation.

[6] Depending on the approach, monolingual corpora or multilingual corpora are also used in the training process.

1.2. Corpus-based Machine Translation

1.2.1. Understanding Corpora

In this section, the definition of corpus, corpus types, corpus creation methods and tools are explained. Although a general overview regarding corpus will be given, the focus will be on corpus creation and machine translation. Both translation practice and translation theories have been deeply influenced by corpus linguistics (Zanettin, 2013, p. 20). Translation memory systems, machine translation systems and product-based translation studies are largely based on the use and research of corpora. Hence, a comprehensive understanding of corpora is a crucial starting step. Kenny (2009, p. 59) defines a corpus as “a collection of texts that are the object of literary or linguistic study”. However, as observed by many scholars, this collection of texts is generally in electronic format and have some other specific characteristics (Saldanha&O’Brien, 2013, p. 55; Zanettin, 2012, p. 7; Bowker&Pearson, 2002, p. 9). Bowker&Pearson (2002, p. 9) highlight four main characteristics: a corpus should be “electronic” (in a machine-readable digital format), “authentic” (should not be artificially created but be a natural representation of the language concerned), “large” (big enough to be representative for the purpose of study or use) and have “specific criteria” of selection (some criteria limiting which texts can enter into the corpus). In their view, these are the characteristics that differentiate corpus from other bodies of text.

There are different categorizations of corpus types. While Zanettin (2012) begins categorizing corpus according to how many languages involved (“monolingual”, “bilingual” or “multilingual”) while Bowker&Pearson (2002) have a more comprehensive typology depending on purpose, modality, language, time frame, size flexibility and didactic value, respectively. It includes “General reference corpus vs special purpose corpus”, “Written vs spoken corpus”, “Monolingual vs multilingual corpus", "Synchronic vs diachronic corpus", "Open vs closed corpus" and "Learner corpus". Starting by this general level categorization, corpus design decisions can be made according to the needs of the researcher or practitioner.

Each discipline using corpus tend to concentrate on some specific types of corpus. In the case of machine translation, the type of text used in the corpus determines the scope and capability of the machine translation system. Hence, there is a tendency to differentiate between "limited domain" (also referred to as "specific domain") and "open domain" (also referred to as "general domain") corpora (Koehn, 2009). In the cases of both SMT and NMT, it is argued that the quality of machine translation output from a specific domain may be better when the engine is trained on the same specific domain corpus, e.g. translating a medical domain text using an MT engine trained on medical corpus (Lumeras&Way, 2017; Wołk&Marasek, 2015). To illustrate a specific domain corpus, the EMEA Corpus[7] (European Medicines Agency) on OPUS Corpus Project (Tiedemann, 2012) is a multilingual, special purpose, specific domain (medical), written corpus. Figure 1 shows a Spanish – English sample from this corpus. A parallel corpus contains text from a language and its translation into another

[7] EMEA Corpus. http://opus.nlpl.eu/EMEA.php (last accesed: 02.12.2019)

(src)="s4.1"> ABILIFY
(trg)="s4.1"> ABILIFY

(src)="s5.1"> EPAR summary for the public
(trg)="s5.1"> Resumen del EPAR para el público general

(src)="s6.1"> This document is a summary of the European Public Assessment Report (EPAR) .
(trg)="s6.1"> En el presente documento se resume el Informe Público Europeo de Evaluación (EPAR) .

(src)="s6.2"> It explains how the Committee for Medicinal Products for Human Use (CHMP) assessed the studies performed , to reach their recommendations on how to use the medicine .
(trg)="s6.2"> En él se explica cómo el Comité de Medicamentos de Uso Humano (CHMP) ha evaluado los estudios realizados con el medicamento a fin de emitir unas recomendaciones sobre su uso .

(src)="s6.3"> If you need more information about your medical condition or your treatment , read the Package Leaflet (also part of the EPAR) or contact your doctor or pharmacist .
(trg)="s6.3"> Si desea más información sobre su enfermedad o el tratamiento de la misma , le aconsejamos que lea el prospecto (incluido en el EPAR) o pregunte a su médico o su farmacéutico .

(src)="s6.4"> If you want more information on the basis of the CHMP recommendations , read the Scientific Discussion (also part of the EPAR) .
(trg)="s6.4"> Si desea más información sobre el fundamento para las recomendaciones del CHMP , le aconsejamos que lea el Debate Científico (incluido en el EPAR) .

(src)="s7.1"> What is Abilify ?
(trg)="s7.1"> ¿Qué es Abilify ?

(src)="s7.2"> Abilify is a medicine containing the active substance aripiprazole .
(trg)="s7.2"> Abilify es un medicamento que contiene el principio activo aripiprazol .

(src)="s7.3"> It is available as 5 mg , 10 mg , 15 mg and 30 mg tablets , as 10 mg , 15 mg and 30 mg orodispersible tablets (tablets that dissolve in the mouth) , as an oral solution (1 mg/ ml) and as a solution for injection (7.5 mg/ ml) .
(trg)="s7.3"> Se presenta en comprimidos de 5 mg , 10 mg , 15 mg y 30 mg , como comprimidos bucodispersables (que se disuelven en la boca) , de 10 mg , 15 mg y 30 mg , en solución oral (1 mg/ ml) y en solución inyectable (7,5 mg/ ml) .

(src)="s8.1"> What is Abilify used for ?
(trg)="s8.1"> ¿Para qué se utiliza Abilify ?

Figure 1. A Sample from the Spanish English EMEA Parallel Corpus

language, in which usually the sentences are aligned side by side. On the other hand, a comparable corpus does not contain the translation of the source text, but it is "on the same subject, all the same type of text (e.g. instruction manual, technical report, etc.), all from the same time frame, etc." (Bowker&Pearson, 2002). These are the two subcategories of multilingual corpora. Hence, it shall be stated that EMEA corpus above is a parallel corpus, not a comparable corpus.

Parallel corpora are not called "translation corpora" since one language is not necessarily the source language and the other is the target language. In this design, the directionality is not strictly important. The texts in two languages can be equally "authoritative" or "created equally" (Zanettin, 2012). Translation memory created in computer-assisted translation tools are common parallel corpora examples. And in the last decades, translation memories have been increasingly used for machine translation training. While bilingual parallel corpora are the most common type of corpus used in machine translation, comparable corpora are used to train machine translation engines when parallel corpora are not available especially in the cases of low-resource languages. Wikipedia[8] or multilingual news websites are common resources for comparable corpus creation. There is a parallel corpora version of Wikipedia hosted in OPUS Corpus created by Wołk&Marasek (2014) as well. Monolingual general domain and specific domain corpora are also used in machine translation. In the following section, parallel and monolingual corpus design methods and tools will be described.

1.2.2. Corpus Design Methods & Tools

Finding free and open monolingual or bilingual corpus for research or commercial use is a challenging task for MT practioners. With the exceptions of corpora made available by the European Union or United Nations and some other independent sources such as open corpora by Softcatalà in the case of Catalan[9], there are not many reliable sources of corpus for all languages. The case is even worse for languages that are not official languages of international organizations, and do not have enough digital content. For these reasons, most of the time, researchers or practitioners need to design and prepare their own corpora, often by harvesting the Web for digital content. Zanettin (2012, p. 154) makes a similar observation about parallel corpus:

> Given the scarcity of easily available sources of parallel corpora beyond a few domains and text types, researchers have turned their attention to the Web in an attempt to devise automatic or semi-automatic procedures to harvest parallel text.

As it can be inferred from this reference, there are both automatic and semi-automatic methods for corpus preparation. Although many tools have been developed to automatize the corpus preparation process, there are still some steps in which human intervention is needed to assure the quality of the corpus. This is especially relevant in the cases of corpus preparation for machine translation systems where corpus quality

[8] Wikipedia. https://www.wikipedia.org/ (last accessed: 02.12.2019)

[9] https://github.com/Softcatala/ca-text-corpus (last access: 10.02.2021)

conditions the machine translation output quality. Generally, the purpose of the researcher or practitioner determines which the corpus preparation steps will be included or excluded. For example, Koehn (2005) suggests five steps for corpus preparation for statistical machine translation:

- obtain the raw data (e.g., by crawling the web)
- extract and map parallel chunks of text (document alignment)
- break the text into sentences (sentence splitting)
- prepare the corpus for SMT systems (normalisation, tokenisation)
- map sentences in one language sentences in the other language (sentence alignment)

The overall steps of corpus preparation for machine translation are presented in Figure 2 based on the classifications of Koehn (2005) and Esplà-Gomis et al. (2019); these steps can also be applied to any corpus creation effort with minor manual modifications in the process (This model will be used for Turkish-English parallel corpus preparation in our study in Chapter 3). Parallel corpus creation from local files does not need the web crawling step, and monolingual corpus creation may skip the alignment step but both will mostly follow a similar process. In summary, the first step is a decision-making step in which firstly resources for corpus content (such as news websites, governmental websites, user-generated content websites etc.) are selected. Once the selection step is completed, using a tool, this content shall be crawled from the web. Unnecessary files shall be eliminated and only the files with relevant content shall be included in the corpus and aligned in document level. Since these files will be in the form of web pages, they will be formatted. The text content shall be filtered out from the formatting such as HTML tags through a parsing tool. Later, the source files are aligned sentence by sentence with the target files using an alignment tool. In the next step, these files are compiled to create one corpus file[10], if possible, for ease of processing. Then, depending on the quality expectation of the corpus, the compiled corpus can be cleaned from inline tags, foreign characters, misaligned sentence pairs and other "noisy" content that potentially harms the quality of machine translation. Finally, since machine translation systems expect certain file types, the resulting corpus shall be saved in a file format permitted by the MT system. It shall be highlighted that many machine translation platforms today support a variety of file types and, therefore, in some cases, this step may be unnecessary. Below we explain in detail each step together with the corpus preparation tools in the context of corpus creation for machine translation. Lastly, it shall be noted that although there are a variety of tools (e.g. applications, software, toolkit etc.) for each preparation step with some manual interventions, there are some others aimed at automatizing all this process and yield a ready-to-use corpus such as Sketch Engine (Kilgarriff et al., 2014) and Bitextor[11]. We will elaborate on them in the following sections.

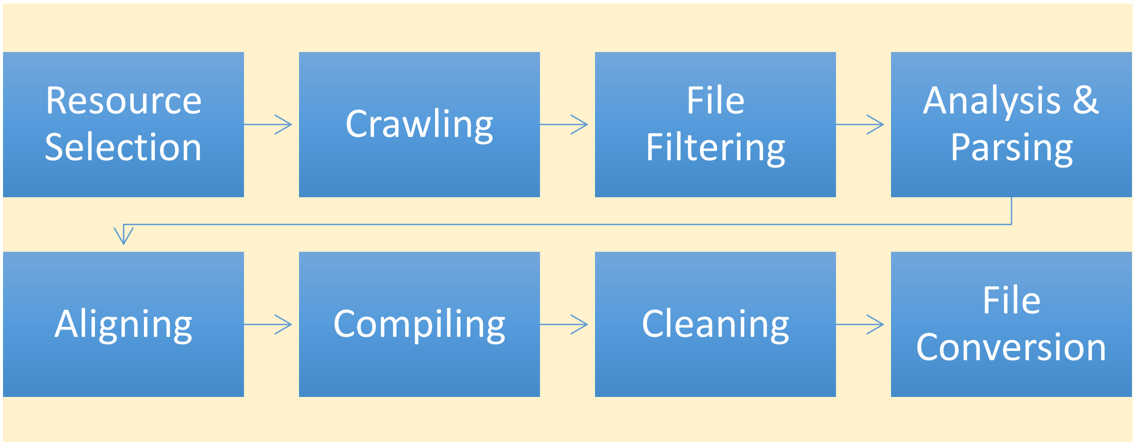


Figure 2. Main Corpus Preparation Steps

[10] More than one files can be created but having only one file makes the cleaning process agile and cleaning steps can be conducted only one time this way.

[11] https://bitextor.readthedocs.io/en/latest/ (last access: 09.12.2019)

1.2. Corpus-based Machine Translation

i. Resource Selection

Corpus-based machine translation require large amounts of bilingual parallel corpus to give high quality translation results. Besides, as it is explained above, the use of specific domain corpora for a special purpose MT engine may provide better translation results for the domain involved. A Catalan – English medical MT engine trained on Catalan – English medical parallel corpora will provide better medical translations compared to a general ("open") domain engine with the same language pair. It can be assumed that the narrower the domain is, the better the translation quality will be. Hence, deciding the scope of the corpus and selecting the type of texts that will be involved is the first step of corpus preparation. Yet, as pointed out in Doğru et al. (2018), this creates a paradox for the MT specialist:

> "although we will be more likely to create a better engine when we have a specific domain, the more our domain is specific, the less amount of text we will likely have". (p. 14)

The data-hungry nature of machine translation systems leads translation companies and researchers to search for more and more parallel corpus in the corpus data selection phase. Translation companies may begin by their legacy translation memories and use them as training corpus either by classifying them according to client or subject matter. Depending on the need, they may also resort to free and open source parallel corpora resources such as OPUS. Researchers may also start with these corpora since they constitute ready datasets in the right file format. However, depending on the language pair, domain and/or data quality, they may need to search for more data. And generally, they browse the Web to select and obtain the linguistic data they need. "The World Wide Web is the largest content distribution system and the most extensive and accessible repository of textual data" (Zanettin, 2012). In this vast space of content, some criteria can guide selecting the correct resources. These criteria include relevance of content (based on the specific domain sought for), the amount of textual data, copyright restrictions and the ability to handle the file types. A researcher or practitioner may create a bilingual legal corpus, search for relevant content using keywords in search engines, find open, public data sources[12] and download them using a web crawler.

ii. Crawling

The method used for downloading and storing web content is called web crawling. A web crawling tool can download a whole website, a number of URLs from different websites, different URLs from different websites based on user-defined keywords, some filetypes (for example, only HTML or only PDFs) from a website depending on the parameters decided by the user. The textual content of downloaded website(s) can then be used for creating a monolingual or bilingual corpus. What Koehn (2005) calls "the raw data" is obtained in this step. There are many tools for web crawling with different parameters and/or platforms. The crawler programs are sometimes called "web spiders" too (Koehn, 2005). In the parallel corpora preparation literature, software such as HTTrack Website Copier[13], Wget[14], and WebBootCaT (Baroni et al, 2006) are commonly utilized. HTTrack is a multiplatform, "free (GPL, libre/free software)" desktop software for downloading websites. It is possible to download a whole website or a part of it by configuring requested file types. Wget is a command-line software for retrieving files from websites on GNU operating system. It is also highly configurable,

[12] When sufficient public data is not available, researchers and practitioners may resort to general data compilation and then data anonymization by removing sensitive or private data either manually or automatically.

[13] https://www.httrack.com/ (last access: 12.09.2019)

[14] https://www.gnu.org/software/wget/ (last access: 12.09.2019)

which means that the user can configure which type of data to download from the relevant website. An example command in Ubuntu Terminal[15] for Wget is given below:

```
wget --mirror -p --convert-links --content-disposition --trust-server-names -P corpus http://khd.tkd.org.tr/
```

This Wget command will download the content of the website "http://khd.tkd.org.tr/" into the computer and save all the files in a folder called "corpus". Note that images, style sheets and other files will also be downloaded, and further filtering shall be required either before running the command by adding file filter or after the download, by manually eliminating the unnecessary files. There are some corpus-specific web crawlers as well. WebBootCaT is a good example of such crawlers. It is the web server version of BootCat[16] and it can crawl content, filter out unnecessary files and compile a corpus-based on seed words, URLs, or websites. This tool is particularly useful for monolingual specific domain corpus creation. Or a user can firstly download the first language option of a website and the second language option of it and then try to align these two corpora. However, for a good alignment, website structure shall be the same for the two language versions of the website. Aside from these tools, it is possible to crawl data from the Web through programming languages such Python with advanced functionalities. However, these techniques require advanced programming skills.

iii. File Filtering

The objective of file filtering is to prepare the files to be aligned sentence by sentence. Hence, firstly, the files shall be sorted and aligned correctly so that source_file_1 can be aligned with target_file_1 and source_file_2 with target_file_2 etc. One can end up

[15] A terminal is where a specific code is executed in an operating system.

[16] https://bootcat.dipintra.it/?section=home (last access: 09.12.2019)

having two folders with equal number of files both for source and target languages. However, some crawled files may not be suitable for corpus creation. Clearly, image files, videos, style sheets or other code files are irrelevant and can be deleted (if they are not filtered out already by the crawler program). And among the textual files, if the objective is to create a bilingual corpus, files that do not have translated counterparts may be useless and need to be removed. In these scenarios, consistency in file names is crucial and files shall be paired in an easy-to-process manner. Generally, file-level alignment is a manual task, however, once unnecessary files are removed, some batch file renaming operations can be conducted using operating system features such as Windows PowerShell if the names for source and target files are not easily recognizable. While HTTrack may change downloaded file names by default, Wget yields more consistent file names. Once, files are organized in an ordered fashion, they can be analyzed for filtering purposes. Since in each step of corpus preparation the objective is to create a corpus as large as possible, file selection shall be conducted in a way allowing for using maximum number of crawled files.

iv. Analysis & Parsing

Web files are downloaded with their source codes where textual content co-exist with formatting codes such as HTML or XML, and/or other kinds of codes such as Javascript, PHP etc. The textual content shall be separated from these codes to be used in machine translation or corpus analysis tools. Generally, the objective is to obtain plain texts from source files, which later can be used for sentence alignment. While computational linguists and computer scientists may have advanced tools for file parsing, in the context of translation, most computer-assisted translation tools such as OmegaT, Memoq etc. have file filtering components which allow importing different file formats and extracting the translatable strings from source code and then export the target strings properly together with the source code once the translation is completed. Okapi Framework[17] includes powerful tools for filtering files. For example, the Okapi Filters Plugin for OmegaT have filter options for HTML, JSON, InDesign IDML etc. These tools also allow for writing custom filtering rules when a standard filter is not enough. Table 1 shows the filtering of an XML file.

```
<?xml version="1.0" encoding="utf-8"?>
<myDoc>
 <prolog>
  <author>Zebulon Fairfield</author>
  <version>version 12, revision 2 - 2006-08-14</version>
  <keywords><kw>horse</kw><kw>appaloosa</kw></keywords>
  <storageKey>articles-6D272BA9-3B89CAD8</storageKey>
 </prolog>
 <body>
  <title>Appaloosa</title>
  <p>The Appaloosas are rugged horses originally breed by
the <kw>Nez-Perce</kw> tribe in the US Northwest.</p>
  <p>They are often characterized by their spotted coats.</p>
 </body>
</myDoc>
```

Table 1. Translatable content in an XML file.

It is also possible to apply more than one filter to a particular file if, for example, it has HTML and XML code together, or some other programming languages together. These kinds of filters are called cascading filters. Regular expressions are used for creating these custom filters. Considering the variety of programming languages, either a manual or automatic strategy shall be implemented to detect the code structure of each file and decide what part of this code is relevant or is desired to be extracted. In these

[17] https://okapiframework.org/ (last access: 09.12.2019)

scenarios, it is important that the all the files involved have similar code structures. Otherwise, the filtering step is going to be a very cumbersome.

v. Aligning

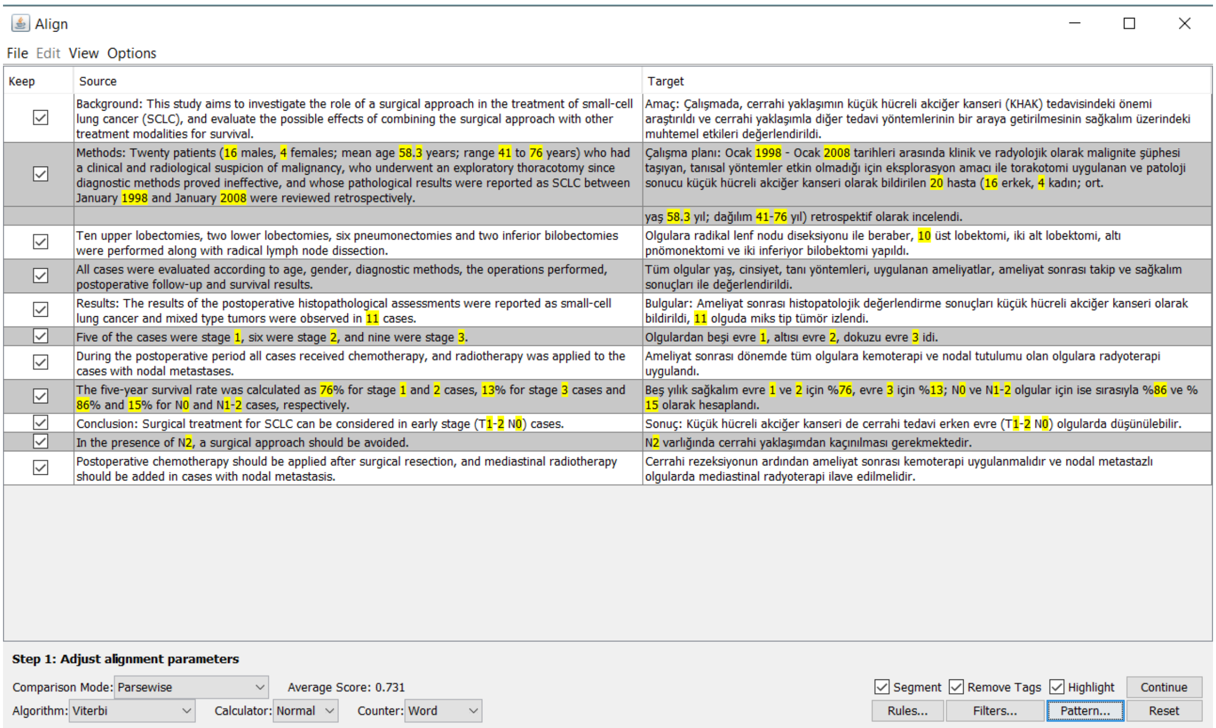
Sentence level alignment is a crucial step for corpus-based machine translation since systems are trained to detect the possible patterns in sentence level. Researchers working on parallel corpus analysis also benefit from sentence alignment. Koehn (2009) highlights that work on (automatic) sentence alignment has begun as early as 1990s with scientists such as William A. Gale, Kenneth W. Church and P. Brown. Sentence alignment requires mapping source sentences into target sentences. Kraif (2002) provides a comprehensive definition:

> Aligning consists in finding correspondences, in bilingual parallel corpora, between textual segments that are translation equivalents. (p. 275)

While correspondences are understood as sentences, as Zanettin (2012) observes, the concept of "sentence" can be controversial and is a problem that alignment researchers, among other researchers, need to deal with. A title, list items, table items or other string pieces are not sentences strictly speaking, yet they need to be aligned. Some CAT tools with alignment feature may prefer to align segment by segment. In the context of CAT tools, a segment is a piece of text delimited by punctuation marks and/or line breaks. Sentence length, numbers, mark-up, tags, document structure or anchor terms can be used to find equivalence between sentences. Some source sentences can be divided into two sentences when translated, or a sentence maybe omitted in the translation. Alignment tools such as HunAlign[18] can calculate a confidence score for aligned sentences and allow the user to remove sentence pairs below a score threshold. In other

[18] Hunalign sentence aligner. URL: https://github.com/danielvarga/hunalign (last access: 11.12.2019)

tools such Memoq's LiveDocs alignment feature, after the automatic sentence alignment, the user can edit the alignments before further processing. In CAT tool context, paired segments are then converted into a translation memory in the form of a TMX file.



Keep	Source	Target
☑	Background: This study aims to investigate the role of a surgical approach in the treatment of small-cell lung cancer (SCLC), and evaluate the possible effects of combining the surgical approach with other treatment modalities for survival.	Amaç: Çalışmada, cerrahi yaklaşımın küçük hücreli akciğer kanseri (KHAK) tedavisindeki önemi araştırıldı ve cerrahi yaklaşımla diğer tedavi yöntemlerinin bir araya getirilmesinin sağkalım üzerindeki muhtemel etkileri değerlendirildi.
☑	Methods: Twenty patients (16 males, 4 females; mean age 58.3 years; range 41 to 76 years) who had a clinical and radiological suspicion of malignancy, who underwent an exploratory thoracotomy since diagnostic methods proved ineffective, and whose pathological results were reported as SCLC between January 1998 and January 2008 were reviewed retrospectively.	Çalışma planı: Ocak 1998 - Ocak 2008 tarihleri arasında klinik ve radyolojik olarak malignite şüphesi taşıyan, tanısal yöntemler etkin olmadığı için eksplorasyon amacı ile torakotomi uygulanan ve patoloji sonucu küçük hücreli akciğer kanseri olarak bildirilen 20 hasta (16 erkek, 4 kadın; ort.
		yaş 58.3 yıl; dağılım 41-76 yıl) retrospektif olarak incelendi.
☑	Ten upper lobectomies, two lower lobectomies, six pneumonectomies and two inferior bilobectomies were performed along with radical lymph node dissection.	Olgulara radikal lenf nodu diseksiyonu ile beraber, 10 üst lobektomi, iki alt lobektomi, altı pnömonektomi ve iki inferiyor bilobektomi yapıldı.
☑	All cases were evaluated according to age, gender, diagnostic methods, the operations performed, postoperative follow-up and survival results.	Tüm olgular yaş, cinsiyet, tanı yöntemleri, uygulanan ameliyatlar, ameliyat sonrası takip ve sağkalım sonuçları ile değerlendirildi.
☑	Results: The results of the postoperative histopathological assessments were reported as small-cell lung cancer and mixed type tumors were observed in 11 cases.	Bulgular: Ameliyat sonrası histopatolojik değerlendirme sonuçları küçük hücreli akciğer kanseri olarak bildirildi, 11 olguda miks tip tümör izlendi.
☑	Five of the cases were stage 1, six were stage 2, and nine were stage 3.	Olgulardan beşi evre 1, altısı evre 2, dokuzu evre 3 idi.
☑	During the postoperative period all cases received chemotherapy, and radiotherapy was applied to the cases with nodal metastases.	Ameliyat sonrası dönemde tüm olgulara kemoterapi ve nodal tutulumu olan olgulara radyoterapi uygulandı.
☑	The five-year survival rate was calculated as 76% for stage 1 and 2 cases, 13% for stage 3 cases and 86% and 15% for N0 and N1-2 cases, respectively.	Beş yılık sağkalım evre 1 ve 2 için %76, evre 3 için %13; N0 ve N1-2 olgular için ise sırasıyla %86 ve % 15 olarak hesaplandı.
☑	Conclusion: Surgical treatment for SCLC can be considered in early stage (T1-2 N0) cases.	Sonuç: Küçük hücreli akciğer kanseri de cerrahi tedavi erken evre (T1-2 N0) olgularda düşünülebilir.
☑	In the presence of N2, a surgical approach should be avoided.	N2 varlığında cerrahi yaklaşımdan kaçınılması gerekmektedir.
☑	Postoperative chemotherapy should be applied after surgical resection, and mediastinal radiotherapy should be added in cases with nodal metastasis.	Cerrahi rezeksiyonun ardından ameliyat sonrası kemoterapi uygulanmalıdır ve nodal metastazlı olgularda mediastinal radyoterapi ilave edilmelidir.

Table 2. Sentence alignment in OmegaT. Among other things, numbers (shown in yellow) are used for finding equivalence.

vi. Compiling

Corpus compilation generally includes annotation, however, depending on the purpose of the corpus, especially in the case of machine translation, annotation may not be necessary. Plain-text sentence alignment in the format of TXT may be enough for machine translation system. Other than aligned plain texts, the most common corpus type is formatted and encoded translation memory, as Zanettin (2012, p. 76-77) underlines:

> Encoding standards such as the XML Corpus Encoding Standard (XCES) for use in the encoding of language corpora, including multilingual and parallel aligned ones, and the Translation Memory Exchange (TMX) format, an industrial standard for the

> exchange of translation memories between different TM systems, have been developed for use by research and professional communities.

Since translation memories are already aligned are produced after translation process, they can be easily deployed in machine translation systems in the format of TMX. Many MT systems natively support TMX import. Therefore, corpus shall be compiled either in the format of TMX or aligned plain text in the format of TXT with the proper encoding (usually UTF-8). CAT Tools such OmegaT or Memoq allows for importing aligned sentences pairs into a single translation memory, which then can be exported as a TMX file. Compiling a large number of aligned files into one single compiled file can facilitate the next steps considerably. For example, all the text-level cleaning steps can be processed on one single file only one time.

vii. Cleaning

Parsing and automatic alignment do not provide bilingual sentence pairs that are ready to be used in machine translation. Sentence pairs coming from formatted files can have HTML inline tags, wrongly encoded characters; untranslated sentences where the source and the target can be present in the compiled file. The quality of the corpus data plays a significant role in the quality of machine translation. "Garbage in, garbage out" motto was used in SMT community to define the impact of the quality of training data on the MT output. Today, in the case neural machine translation, it is even more relevant since "NMT […] has been shown to be more sensitive to noisy data" (Gupta et al., 2019). For this reason, having high-quality clean parallel corpus is crucial and a cleaning step is usually needed. Translation memory maintenance tools such as Heartsome TMX Editor[19] and Goldpan TMX/TBX Editor, and Okapi Framework tools

[19] https://github.com/heartsome/tmxeditor8 (last accessed: 16.12.2019)

such as CheckMate and Olifant can used to remove these noisy contents. Besides this manual cleaning process, machine translation systems may also have a preprocessing step in which sentence pairs with certain criteria can be rejected and not used in the MT training. Very long sentences, sentence pairs with significantly different length of words (a source sentence may have 5 words while the target have 25, which deems this pair unusable for the training), duplicate sentences etc. can be eliminated. When dealing with text coming from Web content, the ability to remove tags becomes important. Goldpan TMX/TBX Editor and Olifant are particularly efficient in removing tags (or in general, codes).

viii. File Conversion

The last step of corpus preparation is to have the corpus files in a file type that is acceptable by the machine translation system. While most machine translation platforms such as MTradumàtica, KantanMT, ModernMT etc. have the possibility to import TMX files, some may require two separate plain text files aligned sentence by sentence. If the users have a TMX files, they can convert into two corpus files using Rainbow. In the case of SMT, monolingual files can also be saved as plain text files with the required encoding.

ix. Complete Corpus Creation Tools & Available Corpus Resources

Thanks to advances in Natural Language Processing and corpus technologies, currently there are some tools which aim to streamline all the corpus preparation and creation process in a single pipeline. Bitextor is a good example of such a pipeline. This tool starts by crawling a website and ends up with a clean corpus. Esplà-Gomis (2019, p. 119) outlines the corpus creation steps of Bitextor as follows:

> (1) downloading HTML documents from the Internet; (2) preprocessing, normalizing and augmenting information from these documents; (3) aligning documents that are parallel; (4) aligning the segments in each of the document pairs identified; (5) filtering noisy data, deduplicating and formatting the output.

It can be observed that the steps are similar to the ones outlined in Figure 2. Bitextor is used in the ParaCrawl[20] project to create bilingual corpus resources for the languages in the European Union. Sketch Engine is also able to directly crawl, clean and compile a corpus. However, in this case, the corpus will be monolingual. The user will need to crawl two language versions of a website, have two corpora and then align them in an external tool. However, sentence by sentence alignment may not be guaranteed if the downloaded and cleaned files in the compiled corpus are not aligned or the same. It seems that this tendency to integrate different steps of corpus preparation will continue considering that most, if not all, of the components of these tools are free and open source. The above-mentioned Sketch Engine also allows for extensive corpus analysis. It hosts a large collection of corpora from different languages and subjects. Finally, AntConc[21] is also a popular free desktop toolkit for corpus analysis. However, it is mainly used for analyzing already compiled corpora.

MT researchers use free and open parallel corpora for MT training and evaluation to allow for the replicability of their results and avoid any infringement of copyright. One of the biggest repository of open parallel corpus is OPUS Project. This project hosts large-scale open corpora from different resources including the above-mentioned ParaCrawl project, the United Nations, and the European Union institutions. The corpus files are presented in XML, raw, TMX and some other formats to be used in different scenarios. It is also possible to query some of the corpora and make analysis.

---

[20] https://paracrawl.eu/index.php/about (last access: 17.12.2019)

[21] https://www.laurenceanthony.net/software/antconc/ (last access: 17.12.2019)

1.2. Corpus-based Machine Translation

In the following section, we will see how monolingual and parallel corpus are used in two different approaches: SMT and neural machine translation.

1.2.3. Statistical Machine Translation

SMT had been the dominant research area in machine translation and had been widely adopted in the industry over the last twenty years (Lumeras&Way, 2017) until the paradigm shift towards neural machine translation accelerated by the paper of Wu et al. (2016) from Google in late 2016. In this section, we will firstly explain SMT, and then concentrate on its components, types, the available tools for creating SMT systems as well as the limitations it has.

SMT is a corpus-based machine translation paradigm that uses parallel corpora to infer statistically probable translation patterns, and automatically translates a previously unseen sentence based on these statistical probabilities (Brown et al., 1988; Koehn, 2009; Hearne&Way, 2011; Way&Hearne, 2011). Unlike RBMT, the grammatical rules for source language and target language or the bilingual glossaries are not readily provided. In principle, the SMT system is expected to learn them from the provided corpora (Kenny&Doherty, 2014, p. 278).

SMT has two different processes namely training and decoding (Hearne&Way, 2011, p. 205). The training phase requires a parallel corpus for computing a statistical translation model and a monolingual corpus for computing a language model. The resulting translation model is a probabilistic bilingual dictionary with probabilities assigned to every possible translation of words or phrases[22]. Usually the product of the

[22] The concept of "phrase" is used flexibly in SMT and may not necessarily be a grammatical phrase, in other words, chunks such "called a", "will not come to", "internally a" are phrases in the context of SMT.

translation model is also referred to as a phrase table (Koehn, 2009, pp. 148-149). A language model is a probabilistic grammar of the target language computed through a monolingual corpus. Hearne&Way (2011, p. 205) highlight that the size of monolingual corpus is typically larger than the parallel corpora. However, it is also possible to use only the target side of the parallel corpus for training a language model. While the objective of the translation model is to produce an accurate translation for a given source sentence in terms of content, the objective of the language model is to make the sentence as fluent as possible. Hence, the translation model aims to solve the accuracy problem while language model aims for the fluency problem. Once the training phase is completed, the trained machine translation engine (translation model + language model) can be used for translating sentences new to the training corpus. This second phase of SMT is called decoding. When a source sentence is introduced to the engine as an input, the mathematically most probable target sentence is produced as an output. The formulation of this operation by Hearne&Way (2011) is provided below:

> "The decoding process essentially treats translation as a search problem: given the sentence to be translated, search over all possible translations permitted by the translation model, and all possible reorderings thereof, for the one which is assigned the highest overall probability according to the translation and language models." (p. 206)

In Figure 3, we illustrate the process of creating an SMT engine. For the decoding to produce high quality translations in terms of adequacy and fluency, (Way&Hearne (2011, p. 235), the training process shall be conducted with enough parallel corpora and monolingual corpora.

1.2. Corpus-based Machine Translation

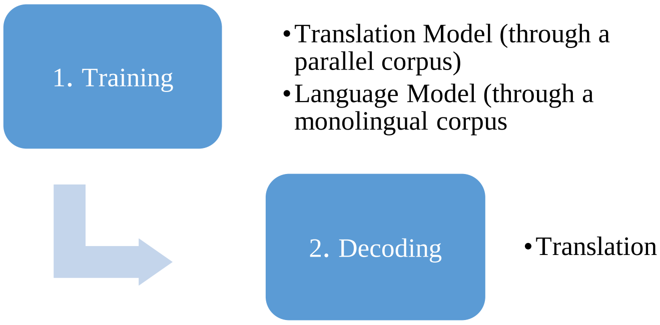


Figure 3. 2 Phases of SMT.

At this point, it should be highlighted that the version of SMT referred above depends on a noisy channel model which is a model for decoding the translation results. According to Lumeras&Way (2017), this model is "surpassed by the log-linear model" (p. 23) which allows for the inclusion of components other than translation model and language model such as a reordering model for yielding higher quality translations. The mathematical formulations of these models are provided in Fig. 4.

$$\text{Noisy-channel model: } \mathit{Translation} = \mathit{argmax}_T \; P(S|T) \cdot P(T)$$

$$\text{Log-linear model: } \mathit{Translation} = \mathit{argmax}_T \sum_{m=1}^{M} \lambda_m \cdot h_m(T, S)$$

Figure 4. Noisy channel model and log-linear model as formulated in Hearne&Way (2011, p. 206)

The first model includes a decoding algorithm (*argmax*$_T$), the translation model expressed as *P*(S|T) and the language model expressed as *P*(T). On the other hand, log-linear model expresses features (e.g. reordering model etc.) more generally and instead of having only translation model and language model, it can include *m* number of features (Hearne&Way, 2011, p. 207). While mathematical explanation of these models is beyond the limits of this study, it should be noted that in log-linear model, there is more flexibility in terms of adding more components.

Language model. Let us now concentrate on what Language Model is and how it is trained. According to Koehn (2009, p. 181) the language model “measures how likely it is that a sequence language model of words would be uttered by a [target language] speaker” in one way to support the translation model to yield fluent, target language sentences. The language model can be calculated based on unigram, bigram or n-gram models, namely one-word, two-word, or multi-word models.

(1a). Bir kadın mavi bir araba sürüyor. (English: A woman is driving a blue car.)

We can illustrate this calculation starting with a unigram model. The Turkish sentence (1a) above has six unigrams. While the probability of “kadın”, “mavi”, “araba” and “sürüyor” occurring in a target sentence is 1/6, the probability for “bir” is 2/6, which is higher. The unigram level of occurrences in each sentence in the corpus are calculated similarly and a probability is assigned to each of them; then, this probability is used in the decoding process. However, as Kenny&Doherty (2014, p. 279) highlight, this method presents two issues: it gives higher score to shorter sentences and does not consider word order. To solve these problems, bigrams or n-grams are calculated. In a bigram scenario for sentence (1a), the bigrams will be “Bir kadın”, “kadın mavi”, “mavi bir”, “bir araba”, “araba sürüyor” and the occurrence probability of each bigram will be 1/5. Nevertheless, the most common models use trigrams (Koehn, 2009, p. 183). It is possible to use higher number of n-grams and usually the limit is 7-grams because of computational restrictions (Kenny&Doherty, 2014, p. 280).

Translation model. The translation model calculates alignments between source sentence n-grams and target sentence n-grams. Similar to the language model, the first models have aimed for one-word alignments in the 1980s and 1990s by IBM's Candide project (Koehn, 2009, p.81). In its simplest form, the translation model aligns sentences word by word:

Yo	tengo	un	coche.
↓	↓	↓	↓
I	have	a	car.

Table 3. One-gram alignment (1b).

However, the alignment of source words and target words is not always direct since word-order is not the same in all languages and some words in some languages may not have counterparts in the corresponding language. Hence a reordering strategy (Koehn, 2009, p. 129) is implemented as illustrated below:

Yo	tengo	un	coche	azul.
↓	↓	↓		
I	have	a	blue	car.

Table 4. Word alignment with reordering (1c).

After, word alignments are completed, larger n-grams (phrases) can be aligned as well to create phrase tables (Kenny&Doherty, 2014, p. 283).

He will visit	us	at the hospital.
Bizi	hastanede	ziyaret edecek.

Table 5. N-gram alignments (1d).

Note that in the English – Turkish sentence pair in (1d), one 3-gram ("He will visit") is aligned with a 2-grams in Turkish ("ziyaret edecek"); a 3-gram ("at the hospital") is aligned with a unigram ("hastanede"). These one-to-many and many-to-many alignments are essential for completely capturing the meaning of a sentence. In the translation model, usually phrases of up to 7-grams are allowed. However, intuitively, when the number of words in a phrase increase, its probability of occurrence decreases; and larger phrases are computationally more demanding, hence an upper limit of 7-grams may be established. Once all phrase alignments ("phrase table") are completed and their probabilities are assigned in the training phase, the translation model is ready to be used for translation.

Moses Toolkit. The launch of the free and open-source SMT toolkit Moses (Koehn et al., 2007) had an important role in accelerating the industrial and academic implementation of SMT (Lumeras&Way, 2017). This toolkit includes a training pipeline and a decoder[23]. In the training pipeline, there are preprocessing tools such as GIZA++ for word-alignment, HunAlign for sentence alignment, other tools and scripts for tokenization, truecasing and cleaning, and probabilistic translation and language model preparation (See the section below for definitions of the concepts). Lastly, Moses has a tuning algorithm "where the different statistical models are weighted against each other to produce the best possible translations"[24]. The convenience of Moses has motivated companies to create products based on it. Custom MT service providers such

[23] http://www.statmt.org/moses/?n=Moses.Overview (last access: 17.02.2020)

[24] *Ibid.*

as KantanMT[25], Systran[26], Slate[27] created platforms based on Moses for custom SMT training. Other open-source projects for wider adoption of SMT such as MTradumàtica[28] (Martín-Mor, 2017) and ModernMT[29] are also Moses-based.

Preprocessing in SMT. In the section on corpus preparation (section 1.2.2.), we describe how a corpus is prepared. However, before the SMT training there are three preprocessing steps that it needs pass through: tokenization, truecasing and cleaning (Koehn, 2016). In the tokenization process, sentences are separated into words and words are separated from punctuation marks. For example, the sentence "I'm playing football." becomes "I 'm playing football ." (in a treebank tokenizer scenario). In truecasing step, initial words in each sentence in the corpus are converted to their most probable case to decrease data sparsity. The sentence "My house is bigger than my car" becomes "my house is bigger than my car" and hence, "My" and "my" are not considered as two different tokens. Proper names may create issues, for example truecasing "Apple" as a company and as a fruit may require additional annotation steps. In the cleaning step, very long sentences, misaligned sentences, or empty sentences are removed; the types of sentences to be removed can be configured by the user.

Types of SMT. The state-of-the-art SMT is based on phrases, not words (Forcada, 2010, p. 220), which is also the reason why it is called phrase-based SMT or PBMT. While this approach remained dominant until the second half of 2010s, possible ways to increase performance in some language pairs and domains are still being studied by

[25] https://kantanmt.com/ (last access: 17.02.2020)

[26] https://www.systransoft.com/ (last access: 17.02.2020)

[27] https://www.slate.rocks/ (last access: 17.02.2020)

[28] https://mtradumatica.uab.cat/ (last access: 17.02.2020)

[29] https://www.modernmt.com/ (last access: 17.02.2020) Currently, it has an NMT version as well.

researchers. Above, we have mentioned that SMT does not utilize linguistic knowledge in the training phase unlike RBMT. In the last decade, there have been studies to create linguistically informed, hybrid SMT systems; these approaches that try to integrate morphology, syntax or semantics in the preprocessing or postprocessing steps are collectively called factored PBMT systems.[30] For example, Hassan et al. (2006) investigate a syntax-based PBMT and integrate syntactic information to the Arabic-English engine while Koehn&Hoang (2007) concentrate on word-level annotation in the form of linguistic markup or automatically generated word categories. These approaches try to find possible solutions to improve the performance by handling the limitations of SMT. The next section will look at these limitations

Limitations of SMT. SMT has provided some advances in terms of performance, time and cost compared to RBMT, even though its approach also brings some challenges. Firstly, it relies on large parallel corpora and monolingual corpora (Way, 2009, p. 31), which may not be immediately available for some language pairs. Hence, low resource languages cannot make use of SMT until they have sufficient amount of training data. In such a scenario, RBMT may be more useful since it does not depend on corpora. SMT provides higher translation quality when trained with specific domain corpora, and sentences from the same specific domain are translated. For this reason, even if there may be sufficient general parallel corpora for a particular language pair such as the datasets in OPUS Corpus, specific domain training data may not be available. For instance, a large subtitling corpus[31] is available for Turkish – English language pair in OPUS Corpus. Nevertheless, this does not mean that an engine trained with this corpus

[30] http://www.statmt.org/moses/?n=Moses.FactoredModels (last access: 17.02.2020)

[31] http://opus.nlpl.eu/OpenSubtitles.php (last access: 17.02.2020)

will provide high quality translation for a medical text full of medical terminology. There are some other limitations of SMT which cannot be solved with right type or amount of corpora and are more related to the SMT algorithm itself. Koehn (2004, p. 35) highlights that SMT is "bad at large-scale reordering; add, drop, change of function words for non-local reasons, and correct syntax on sentence level". Reordering problems are particularly relevant in longer sentences, and language pairs from very distant language families (such as Turkish and English). Way, (2019, p.316) also point out that SMT does not perform well with morphologically complex languages. Languages such as Turkish, German, Finnish and Hungarian are within this category. Since these languages express meaning in fewer compound and/or highly inflected long words, word or phrase alignment in translation model and language model training may hinder the training process. The approaches mentioned in the previous section try to find solutions to some aspects of these problems. However, in its core, SMT translates in sub-sentence level (e.g. with phrases, words or even morphemes represented by n-grams), and one of the key difference of the next paradigm (Neural Machine Translation) that will be covered is that it considers the whole sentence while translating. Below we will explain the neural machine translation (NMT). At the end of the NMT section, we will compare these two paradigms for known differences.

1.2.4. Neural Machine Translation

Neural machine translation (NMT) is "a new breed of corpus-based machine translation" (Forcada, 2017, p. 292), which means that just like SMT, it uses a corpus of previously translated strings to train an engine and then translate new strings based on a encoder-decoder architecture. The main difference from SMT is that NMT "uses a completely different computational approach: *neural networks*" (Forcada, 2017, p. 292). According to Koehn (2020), "[a] neural network is a machine learning technique that takes a number of inputs and predicts outputs". The resemblance to biological neurons is only an analogy, and neurons in artificial neural networks are represented with numbers. The activation and output of these artificial neurons "depends on the stimuli they receive from other neurons and the strength of the connections along which these stimuli are passed" (Forcada, 2017). This machine learning technique had deeply influenced how machine translation systems are trained and deployed.

Way (2019, p. 316) notes that while SMT has been the dominant paradigm for the nearly last 25 years, this has started to change in favor of NMT after the results obtained in International Workshop on Spoken Language Translation in 2015. And after Google Translate[32] transitioned to NMT in 2016, the success of NMT has become more publicly known, and popular MT providers such as Microsoft Translator[33], Amazon Translate, DeepL[34] have begun to provide NMT based engines. Early academic studies such as Bentivogli et al. (2016) and Luong&Manning (2015) have reported clear improvements in favor of NMT in quality both in human and automatic evaluations. On the other hand, in their study on 3 different text domains, Castilho et al. (2017) and (Castilho et al., 2018) show that there are still cases in which PBMT provides better quality, and there is space for improvement. It should also be highlighted that these studies have been made only on particular language pairs and some specific domains, which implies that the case may be different in other scenarios. We will focus on the differences between SMT and NMT after we talk about the training and encoding-decoding phases of NMT.

[32] "A Neural Network for Machine Translation, at Production Scale". https://ai.googleblog.com/2016/09/a-neural-network-for-machine.html (last access: 25.02.2020)

[33] https://www.microsoft.com/en-us/translator/ (last access: 25.02.2020)

[34] https://www.deepl.com/en/translator (last access: 25.02.2020)

NMT Training. NMT is trained on parallel corpora to create "a single ('monolithic') large neural network whose connection weights are all jointly trained." (Forcada, 2017, p. 301). In the training, firstly, each word both in source and target language is converted to a unique encoding vector for representing them numerically (Lanners, 2019). In other words, they are converted to "neurons", and then the semantic network between these neurons is computed "to determine weight or strength of each of the connections between neurons so that the desired results are obtained" (Forcada, 2017, p. 295). These connections can be calculated for one single layer of network, or multiple layers of networks ("deep"). Another distinctive process in NMT training is that the system "assigns the maximum likelihood to the whole reference translation for all of the source – target pairs in the training set." (Forcada, 2017, p. 295). Besides, the distributed representations are trained in a way to be aware of both the context of the source and target sentences. Once the training is completed, source sentences can be introduced for translation.

Encoding and Decoding in NMT. Once a source sentence is introduced into an NMT engine, it is firstly encoded into a machine-readable format; namely, "vector embeddings of individual words" are formed (Forcada, 2017, p. 297). After encoding is completed, the decoding process begins by translating one word at a time. Forcada (2017) dubs this process as "machine translation as predicting the next word" based on the trained neural network. In an English-Turkish NMT engine trained with enough parallel corpora, a sentence like "I play football", the decoder would probably produce in the first loop: "I play football: Ben" since given such an English sentence, the engine will infer that "Ben" (which means "I" in Turkish) is the most probable first word. Then, in the second loop, "I play football: Ben futbol" will be produced. And in the last loop, "I play football: Ben futbol oynuyorum" will be yielded.

Another key concept in NMT is attention mechanism. According to Forcada (2017), this mechanism:

> "pays attention (responds) not only to the last representation built by the encoder […] but also to the whole sequence of representations built during encoding […] through an appropriate additional set of neural connections and layers." (p. 299)

Figure 5 shows how decoding works in OpenNMT's decoder[35] with an attention mechanism. Thanks to this mechanism, long distance relations in sentences are captured better and more natural translations are created (though not necessary more accurate ones). Other than attention-based models, there are convolutional models (Gehring et

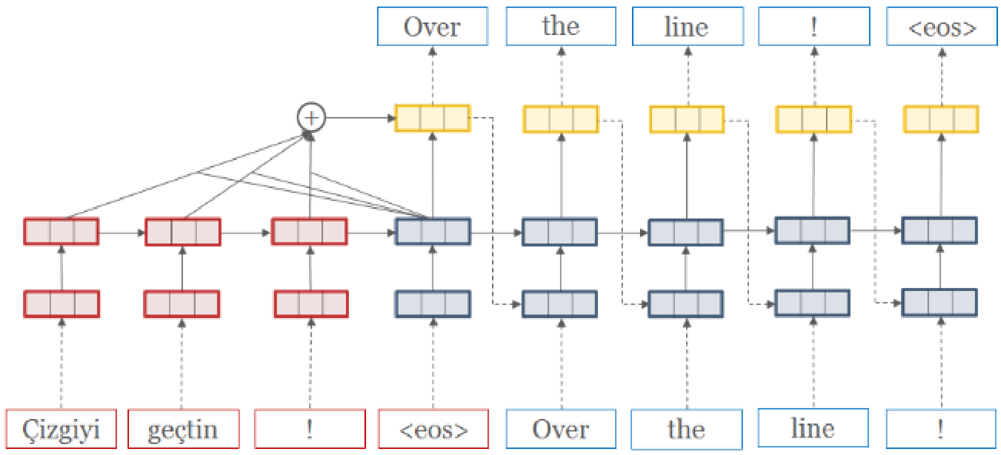


Figure 5. OpenNMT's decoder illustrated. <eos> represents "end of sentence".

al., 2017) which take into a few words to the left and to the right of a word instead of whole sentence during decoding (Forcada, 2017), and transformer neural networks modes which do not use the encoder-decoder architecture and only use the attention mechanism[36]. One of the novel features brought by NMT is the use of byte pair encoding for solving the problem of rare words in the training corpus. According to Koehn (2020, p. 225), byte pair encoding is "[a] popular method for creating an inventory of subword units and legitimate words". Words in a corpus are separated

---

[35] https://opennmt.net/ (last access: 25.02.2020)

[36] https://kantanmtblog.com/2018/12/18/which-neural-mt-model-should-i-choose/ (last access: 25.02.2020)

character by character and in each iteration, most frequent characters are merged. Koehn (2020, p. 226) state that "[a]t the end of this process, the most frequent words will emerge as single tokens, while rare words consist of still unmerged subwords." The idea behind using using subwords is that unknown words mostly consist of known parts such as morphemes. Sennrich et al. (2016) highlights that named entities, cognates and loan words and morphologically complex words are candidates that can be divided into subwords. The use of this byte pair encoding strategy has the potential to improve the performance of MT especially in morphologically rich languages such as Turkish.

Differences between SMT vs NMT. Both SMT and NMT need a huge amount of parallel data for training (Way, 2019). However, it has been shown that NMT is more sensitive to noisy data (Koehn&Knowles, 2017; Rikters, 2018) and will learn to replicate the errors in the corpus. Hence, the quality of corpus becomes even more significant.

One of the most challenging aspect of NMT is the very long training times. NMT is computational demanding (Way, 2019; Forcada, 2017; Shterionov et al., 2018). While it is possible to train an SMT engine in one day using a normal computer, an NMT engine requires days or months (Way, 2019).

When it comes to the difference in terms of output, we should firstly consider the differences between the decoders of these two paradigms. In SMT, when a source sentence is translated, it is divided into phrases and the most probable translation counterparts are found for them, and finally a reordering mechanism is applied. On the other hand, in NMT, translation is produced word by word by considering the totality of the source segment (Forcada, 2017). For this reason, the way a target sentence is generated is quite different. This leads to translations to be less fluent in SMT and follow as much as possible the grammar of the source sentence.

Turkish Source	**Google PBSMT**
Bu çalışmada koroner anjiyografi yapılan hastanın ikinci bir işleme girmeksizin koroner anjiyografiyi takiben elle şekillendirilen kateterle karotis anjiyografi yapılabilirliğini ve konvansiyonel katetere göre pratiklik, güvenirlik ve görüntüleme başarısını karşılaştırdık.	In this study, patients who underwent coronary angiography, coronary angiography followed by a second processing to enter shaped by hand carotid angiography catheter feasibility and practicality compared to conventional catheters, imaging, we compared the reliability and success.
Google NMT	**Human Translation**
In this study, we compared the ability of the patient undergoing coronary angiography to carotid angiography with a hand-shaped catheter following coronary angiography and practicality, reliability and imaging success compared to conventional catheter.	We have evaluated the feasibility of concomitant carotid angiography after coronary angiography with the same catheter, and the practicality, safety and success of image acquisition with respect to conventional catheters.

Table 6. The translation of a Turkish sentence into English by Google PBSMT, Google NMT and Human Translation.

For example, the Google PBSMT translation of the Turkish sentence in Table 6 has a grammatical structure very close to the source while the Google NMT version is comparatively fluent. However, this does not mean that the translation is better since some words may be omitted in favor of fluency in NMT. For example, the phrase "ikinci bir işleme" is omitted in Google NMT while it is present in Google PBSMT as "a second processing" which can be considered a correct translation. NMT models are known to produce such translation issues.

Koehn&Knowles (2017) report 6 big challenges for NMT. Firstly, they attribute the above-mentioned problem to the inability of NMT to perform well in out-of-domain scenarios. They argue that in out-of-domain situations, NMT sacrifices adequacy for fluency. Secondly, and in line with the first challenge, they state that NMT does not perform well in low training data scenarios. Thirdly, low-frequency words are still problematic since their probability of occurrence is low, though byte pair encoding is used (cf. section 1.2.4.). The fourth challenge is related to long sentence translation, despite the fact that according to the authors, NMT still performs comparatively better than PBSMT in very long sentences. The fifth and sixth challenges are related to attention mechanism and word alignment model, and the way decoding works in long search spaces.

Although there are still many challenges for NMT, it is being widely adopted today, especially thanks to open projects such as OpenNMT[37], Marian[38] and others.

1.3. Quality Evaluation in Machine Translation

MT studies bring with them the issue of evaluating the translation quality. The quality of the MT systems shall be evaluated for determining their fitness for purpose and understanding if there is an improvement compared to other previous systems. MT quality evaluation[39] is commonly performed with two different methods: automatic evaluation and human evaluation (Castilho et al., 2018). In this chapter we review these methods, both of which have certain advantages and limitations in translation quality assessment. Human evaluation of MT output can provide lots of information about the quality of the MT system while it is slow, subjective, costly, and inconsistent (Castilho et al., 2018). Automatic MT evaluation can be fast, objective, cheap (most of the time free) and consistent over time (Way, 2018). However, it is usually dependent on a "golden-standard" human translation reference, which may not always be present; besides, as it will be shown below, a source sentence can be translated in various

[37] https://opennmt.net/ (last access: 25.02.2020)

[38] https://marian-nmt.github.io/ (last access: 25.02.2020)

[39] We use "quality evaluation" and "quality assessment" interchangeably following the definitions of Multidimensional Quality Metrics (MQM) framework: http://www.qt21.eu/mqm-definition/definition-2015-12-30.html

different ways, which may give lower scores to accurate translations differing stylistically from the reference human translation (Way, 2018).

Before focusing on human and automatic evaluation of MT, it should be noted that a translation can be produced by a human or machine[40], and the quality assessment of this translation can be conducted by human evaluators or automatically. Table 7 shows possible translation & evaluation scenarios. While quality assurance (QA) is distinct from automatic MT evaluation, they are grouped under "automatic assessment" since they are both performed automatically on translation outputs.

	Human Assessment	**Automatic Assessment**
Human Translation	Human Reviewers with/without DQF/MQM based models	QA Tools (ApSIC, XBench, Okapi CheckMate etc.)
Machine Translation	Human Reviewers with DQF/MQM based models	Automatic Evaluation Metrics (BLEU, TER, F-Measure)

Table 7. Quality assessment types for human and machine translation.

Human evaluation of translation has been done human reviewers in the industry for long a time. These evaluations were usually vague and subjective and leading to disagreement between clients and language service providers (LSPs) and/or between LSPs and translators (Lommel, 2018). In 1990s, first efforts were made to create a systematic and standard quality evaluation model, and the most outstanding one has

[40] It should also be noted that today thanks to computer-assisted translation tools, human translators are interacting with MT output in different ways and the translated segments usually include MT-translated segments, human-translated segments as well as translated segments coming from the translation memories as exact matches and fuzzy matches, which implies that line between human and machine translation is blurring. This fact is highlighted in Castilho et al. (2018, p. 27).

been the LISA[41] QA Model (Lommel, 2018). This model and future models have aimed to identify error categories, assign severities and weights to each category, and yield an overall score for the translation quality (Castilho et al., 2018). Among these models, the extensive error typology developed within the scope of QTLaunchPad project[42] by German Research Centre for Artificial intelligence (DFKI) is called Multidimensional Quality Metrics (MQM)[43]. This model defines over 100 issue types under 8 top level dimensions: accuracy, design, fluency, local convention, terminology, verity, style, and internationalization. This model also ties metrics to client specification, in other words it gives a flexibility of selecting the error/issue types to be evaluated (Lommel, 2018). The error typology presented in MQM will be discussed more deeply in the Methodology chapters.

Another important error typology is TAUS'[44] Dynamic Quality Framework (DQF) (O'Brien, 2012). This model has 6 error categories: accuracy, linguistic, terminology, style, country standards and layout. Later, as of 2014, DQF and MQM have been integrated and harmonized[45] in order to avoid any market confusion and accelerate improvement in quality practices (Lommel, 2018, p. 124). One of the key changes after this harmonization is the upgrade of "terminology" and "style" as stand-alone categories other than being subcategories of accuracy and fluency. In total, there are 50 issue types in this new typology.

---

[41] Localization Industry Standards Association. Although this association was closed, its model has survived in the industry.

[42] http://www.qt21.eu/launchpad/index.html (last accessed 28.02.2020)

[43] http://www.qt21.eu/mqm-definition/definition-2015-12-30.html (last accessed: 28.02.2020)

[44] Translation Automation User Society. https://www.taus.net/dqf (last accessed: 28.02.2019)

[45] https://www.taus.net/academy/news/press-release/dqf-and-mqm-harmonized-to-create-an-industry-wide-quality-standard (last accessed: 11.04.2020)

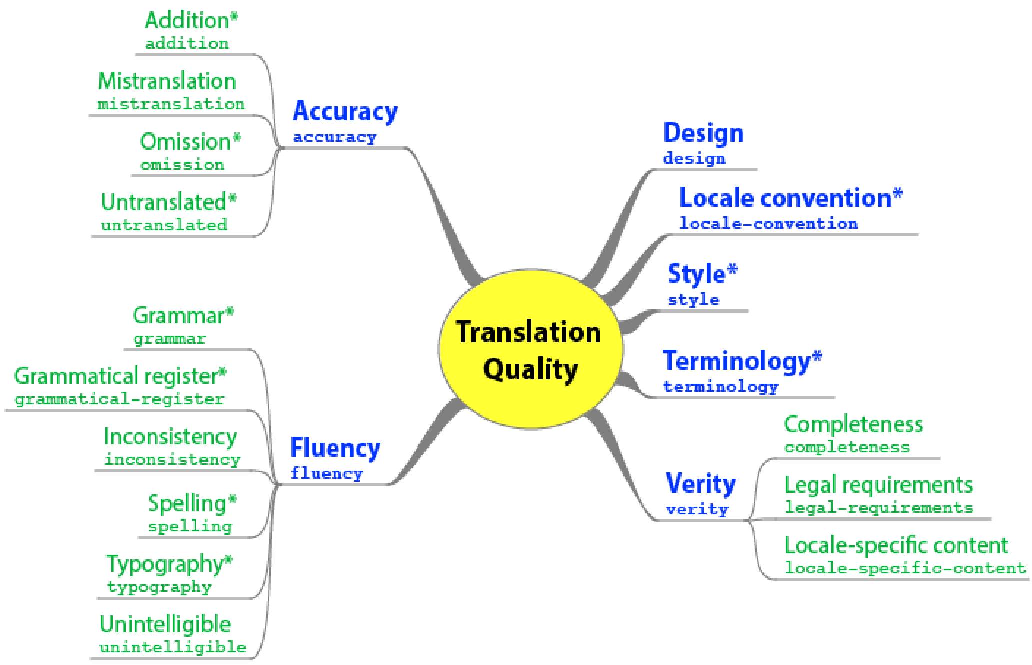


Figure 6. DQF/MQM integrated error typology available online and in Lommel (2018, p. 125).

As Castilho et al. (2018) highlights, the abovementioned two models and their harmonized version are designed to be used for evaluating both human translation and machine translation output independent of their process of production. They are being adopted in both academic studies and in the industry. Companies specialized in translation quality evaluation such as ContentQuo[46] use the framework of MQM and provide the possibility to utilize harmonized DQF/MQM model. KantanMT[47] has a feature called KantanLQR which includes different evaluation models including the harmonized version of DQF/MQM.

Error annotation plays a key role in this framework. It should be noted that the types and frequencies of these error categories are usually annotated by human evaluators no matter by which agent the translation has been done (human or machine). For this reason, in machine translation, this type of quality evaluation is usually referred to as

[46] https://contentquo.com/ (last accessed: 11.04.2020)

[47] https://kantanmt.com/ (last accessed: 11.04.2020)

human evaluation of MT. Human evaluations in MT usually measure fluency and accuracy by using scales such as Likert scale (Castilho et al., 2018). Aspects such readability, comprehension, acceptability, and usability may also be measured, and rankings are performed among different MT systems (Castilho et al., 2018). However, the presence of a human in the loop poses a problem for these models in terms of subjectivity, though arguably it is minimized by the presence of standard metrics. Furthermore, the slow and expensive nature of the application of these methods leads researchers to look for automatic methods.

1.3.1. Automatic MT Evaluation

In this section we review the automatic evaluation metrics (EAMs) used for measuring the quality of machine translation output. The focus of the section is not on quality assurance tools such Okapi CheckMate, ApSIC XBench etc. which are used for automatically checking possible issues (number error, inconsistent tag use etc.) in human translations and are performed in CAT tool environment. Instead, our focus will be on specific EAMs which usually make n-gram comparison between an MT output and a human reference translation (Way, 2018). The ultimate objective of these metrics is to yield results consistent with human judgments (Castilho et al., 2018; Koehn, 2010). Since immediate, repeatable feedback about an MT engine's quality is desirable for constant improvement, these automatic methods have been an inseparable part of MT workflow. Customized MT providers tend to have the option to see the automatic quality scores for the trained engine according to different metrics.

Word Error Rate (WER) is one of the earliest efforts to automatically calculated the quality of an MT output by comparing it to a human translated reference translation and computing insertions, deletions, substitutions and taking into consideration the relative length of the sentences (Nießen et al. 2000; Castilho et al. 2018). The most commonly used quality evaluation metric is BiLingual Evaluation Understudy (BLEU) score (Shterionov et al. 2018). This metric has been proposed by Papineni et al. (2002) as an automatic score which correlates with human judgments. A score between 0 and 1 is assigned to an MT-translated sentence based on its weighted similarity to a human reference translation (some companies such as KantanMT use a score between 0% and 100%). The higher the score, the better the translation quality. This score is computed according to n-gram matches between the MT-translated sentence and reference sentence as well as the sentence length and word order (Shterionov et al., 2018).

No	Source	Reference/Target	KantanMT Output	BLEU
1	Çalışma Dizisi Yükle	Working Set	Load Worksheet	47%
2	Tamamlanmamış ya da Geçersiz Yapılandırma	Incomplete or Invalid Configuration	Incomplete or Invalid Configuration	100%
3	Biçim alanı	Style area	Format space	47%
4	Öznitelik durumu	Attribute case	Attribute state	51%
5	. Ağaç Dosyalarında üst seviye klasörleri kaldırabilir ve yeni yerel klasörler ekleyebildiğiniz. RMB Yeni Üst Klasörleri. Kullanın ve bir yerel klasör seçin ya da uzak klasör seçin be ftp: // kullanıcı@ sunucu nun formu içinde uzak sunucu adı girin.	. that you can add new local and remote top level folders to the Files Tree? Use RMB New Top Folder. and select a local folder or enter the remote server name in form of ftp: / /user@server and select the remote folder.	. that you can remove top-level folders and add new local folders. You can use RMB New Top Folder. Use a local folder and select a local folder. Select a local folder and choose the remote server name inside ftp:/user@server.	34%
6	dakika	min	minutes	47%
7	Renklendirilmiş nesne, değişik dizinlerde değişik dosya türüne sahip. Ne yapılacağını seçiniz.	The highlighted item has a different type in the different directories. Select what to do.	The Colored object contains different file types in different folders. Select this.	7%
8	Bellekten çağır	Recall from memory	Memory from Memory	59%

Figure 7. A section of the BLEU score calculation in KantanMT platform. 8 Turkish sentences are translated by the custom TR-EN engine and these sentences are than compared to the human reference translations. Factors such as n-gram match, sentence length and word order are considered.

Throughout the development of SMT, BLEU score has been used for its ease of use and applicability for any languages. BLEU is commonly used in machine translation research. Companies such as Microsoft and Google report the improvement of their MT systems in reference to the BLEU score as well as human evaluations. However, studies such as Way (2018) and Shterionov et al. (2018) find shortcomings in use of BLEU score and other n-gram based automatic metrics in neural machine translation. Due to these shortcomings especially affecting languages such as Turkish, Tantuğ et al. (2008) created a custom BLEU metric called BLEU+. Other commonly used n-gram based metrics include F-Measure and Translation Error Rate (TER). According to Shterionov et al. (2018) describe F-Measure as:

> [T]he harmonic mean of the precision and the recall of a system. Similar to BLEU, it is concerned with the comparison of candidate translations to a set of reference translations at the *n*-gram level. Precision is the fraction of the number of correctly translated *n*-grams to the total number of translated *n*-grams; recall is the fraction of the number of correctly translated *n*-grams to the total number of reference *n*-grams.

Usually, the higher the F-Measure, the better the translation performance is. Castilho et al. (2018), Way (2018) and Shterionov et al. (2018) also seem to agree that the character-based version of F-Measure called chrF and developed by Popovic (2015) may be more reliable because it works at character level and can capture matches sub-sentence features. Lastly, TER aims to calculate the number of edits needed to match the MT-translated sentence to the human reference sentence. Usually, the lower the number of edits needed, the better the translation quality is.

Automatic evaluation metrics alone are not enough to determine the quality of a machine translation system or engine. When possible, they are coupled with different human evaluation methods. Besides, there is still a search for better automatic metrics for the evaluation of neural machine translation systems since the currently available ones are thought to be biased towards SMT systems (Way, 2018).

1.4. Turkish and Machine Translation Paradigms

Turkish is an agglutinative language with a productive inflectional and derivational morphology (El-Kahlout&Oflazer, 2006). Turkish words generally consist of root morphemes and then derivational or inflectional suffixes are combined with them. According to Oflazer&El-Kahlout (2007), “Turkish employs about 30,000 root words and about 150 distinct suffixes.” This makes Turkish a challenging language for corpus-based MT systems because a single Turkish word[48] may need to be mapped to a long phrase in English in the translation model of the SMT or NMT. For example, the Turkish word *yapabilemeyeceğimizden* can be translated as "since we cannot do [it]" and it can be divided into its morphemes as such: yap+a+bil+e+me+y+eceğ+imiz+den. This may require the use of morpheme analyzers (El-Kahlout&Oflazer, 2010) in the training phase or special evaluation metrics (Tantuğ et al., 2008) to evaluate the engine performance with Turkish. Finnish is one of the closest languages to Turkish in the EU in terms of its feature of agglutination. El-Kahlout&Oflazer (2006, p. 7) report that the BLEU score for Finnish to English SMT is 21.8 and for English to Finnish, it is 13.0 which is one of the lowest scores among European languages. This may provide a hint about the possible performance of Turkish↔English engines.

The sentence structure of Turkish also provide a challenge for MT as well. El-Kahlout (2009) states that although Turkish is a Subject+Object+Verb language, the word order seems to be flexible. This fact affects the calculation of probabilities in the translation model and may decrease the quality of the output. For example, "Hasta iyileştikten sonra taburcu edildi."[49] (SOV), "İyileştikten sonra taburcu edildi hasta." (OVS), "Taburcu edildi hasta iyileştikten sonra." (VSO) are all permissible and grammatical in Turkish. All these forms are used in domain-general texts and, as Koehn (2009) observes, when sentences are longer, finding patterns may be more difficult. However, when the scope of the domain is narrowed, for example to medical texts, the first option (SOV) will probably be more common, which may affect positively the output of the

---

[48] As an extreme example to illustrate this fact, the single word "muvaffakiyetsizleştiricileştiriveremeyebileceklerimizdenmişsinizcesine" literally means "As though you happen to have been from among those whom we will not be able to easily/quickly make a maker of unsuccessful ones". Though it is highly improbable to come across this word in a Turkish text, most Turkish words are long due to their combination with inflectional and derivational suffixes.

[49] "The patient was discharged after recovering."

SMT system. This implies the positive affect of using a domain-specific strategy. Since specific text types already constrain selection of terminology and flexibility of grammar, and generally follow certain patterns, they tend to provide higher scores in quality metrics.

Efforts have been made to develop MT systems with Turkish since the times of RBMT. For example Apertium Turkic[50] aimed to provide Turkic linguistic resources for RBMT. While most of the studies focus on English – Turkish language pair, there are also some studies that aim to benefit from the similarities between Turkish and other Turkic languages. In this section we survey the Turkish MT studies and how they have evolved.

The Turkish Thesis Center[51] include 13 PhD theses and 27 master's degree thesis on machine translation in its database[52]. More than half of these studies investigate English – Turkish language pairs (24 studies) while others explore MT between Turkish and other Turkic languages such as Crimean Tatar, Kirghiz, Turkmen, Kazakh, Azeri. In fact, the earliest study is a master's thesis on RBMT from Turkish to Azeri (Hamzaoglu, 1993). While there are 6 EBMT, 9 SMT and 1 NMT studies, most of the remaining studies are based on RBMT. This high number of RBMT studies in comparison to other paradigms can be explained by two main reasons among others: i). lack of access to the

[50] https://wiki.apertium.org/wiki/Apertium_Turkic (last access: 21.02.2021)

[51] https://tez.yok.gov.tr/UlusalTezMerkezi/giris.jsp. All master and PhD theses written in Turkish Universities are uploaded into the Turkish National Thesis Center.

[52] Interestingly, only 3 theses are written by translation studies students while 36 theses are written by computer science students and one is written by an electrical engineering student. This shows that the interest in machine translation in translation studies departments is still low in Turkey.

high-quality parallel corpora for MT training, and ii). the struggle for exploiting the similarities between Turkish and other Turkic languages through MT rules.

Corpus-based MT studies on Turkish require the availability of large parallel corpora sets. Turkish – English corpus collection of Opus includes 52 million sentence pairs from 19 corpora as of January 2020. Another comprehensive Turkish corpus project is the TS Corpus Project which "is a free and independent project that aims building Turkish corpora, NLP tools and linguistic data sets" (Sezer, 2017). It hosts Turkish monolingual corpora and English – Turkish parallel corpora for analysis purposes. Currently, it includes 14 corpora, 1.3 billion tokens. Another large-scale project is Turkish National Corpus which is a monolingual corpus with more than 50 million words (Aksan&Aksan, 2014). The contents of TS Corpus and Turkish National Corpus are only available for analysis and corpus download for machine translation is not possible. However, we found no publicly available medical or legal parallel corpora for Turkish ↔ English, which impedes the study of custom MT in these domains.

In Chapter 3, we will explain how we created a Turkish – English parallel corpora out of cardiology abstracts from online journals using the steps described above.

1.4.1. Turkish Rule-based Machine Translation

A comprehensive study on Turkish RBMT is published recently by Alkım&Çebi (2019) and uses a combined approach with interlingual and transfer systems for Turkic languages. The product of the study called MT-Turk[53] includes Turkish, Kirghiz and Kazan Tatar. The evaluation results are reported in terms of BLEU score and oscillate between 7.2 (Turkish to Kazan Tatar) and 21.71 (Kirghiz to Turkish). Oflazer&Saraclar

---

[53] http://nlpapps.cs.deu.edu.tr/MTTurk/MTTurk.aspx (last access: 16.04.2020)

(2018) covers other Turkish RBMT works especially on morphological disambiguation.

1.4.2. Turkish Statistical Machine Translation

An early study on English to Turkish SMT with 20k sentences reports a BLEU score of 0.0913 (El-Kahlout&Oflazer, 2006). Tyers&Alperen (2010) conduct an English to Turkish SMT study with 208k news domain sentences (SETIMES corpus) and achieve a BLEU score of 20.90. Bektaş et al. (2016) train English to Turkish and Turkish to English SMT engines with the same corpus (again SETIMES), and different configurations using Moses Toolkit and Turkish morphological analyzer, and they achieve a maximum BLEU score of 15.06 in Turkish-English pair, and a maximum BLEU score of 8.59 in English-Turkish pair. El-Kahlout&Oflazer (2010) conducts a similar study on English to Turkish SMT with 56K sentences from mixed domains (news texts and documents from NATO, EU, and foreign ministry sources). They achieve a BLEU score of 25.17, among other things, because they use “a selectively segmented morphemic representation with various additional steps” including “[r]eranking the 1000-best outputs” and reordering English phrases to make them more like Turkish morpheme structure. Tantuğ et al. (2008) evaluate the SMT system of Oflazer&El-Kahlout (2007) with a custom version of BLEU which they call “BLEU+”. BLEU does not perform optimally with agglutinative languages since even a minor suffix addition to a word as compared to the reference word leads to penalization in BLEU score. For example, in the standard definition of BLEU score if the MT system outputs “kitapların” (“of the books”) and the reference word is “kitaplar” (the book), the translation will be considered inaccurate. Tantuğ et al. (2008) create a version of BLEU which takes into consideration the word roots in the process of word comparison, and hence they solve what they call “all-or-none nature of word comparison” in BLEU score. Their baseline BLEU score of 27.64 rises up to 33.12 when the word roots are considered.

1.4.3. Turkish Neural Machine Translation

The English to Turkish SMT quality in Google was considered very low (as illustrated by the BLEU scores above). In survey on Turkish translation students, Şahin (2015) reports that more than 50% of the students finds English to Turkish MT (in the context of the study, Google SMT) “inadequate” and “only useful for drafting”. The study concludes with an expectation that better approaches to Turkish MT would be developed in the future. One year later, NMT started to gain popularity as of September-November 2016 after Google’s announcement[54] that it had begun to transition to NMT in its translation platform because NMT “has been generating exciting research results for a few years”. One of its first pairs where NMT is implemented English-Turkish. Besides, EMNLP[55] 2018 included English-Turkish language pair for the NMT shared translation task in news domain, which lead to more research papers on EN-TR NMT. Burlot et al. (2018) compare the results from 6 shared translation tasks with BLEU scores varying between 24.84 and 48.42. Ataman (2018) conducts a study on English-Turkish NMT using both SETIMES corpus and a custom corpus of 35K sentences, from which she obtains a BLEU score of 13.77. She also trains a multilingual engine with English, Turkish and Kurdish (adding approximately 14K sentences in English-

[54] Found in translation: More accurate, fluent sentences in Google Translate. https://www.blog.google/products/translate/found-translation-more-accurate-fluent-sentences-google-translate/ (last accessed: 20.04.2020)

[55] https://www.statmt.org/wmt18/translation-task.html (last accessed: 20.04.2020)

Kurdish and Kurdish-Turkish language pairs) and achieves a slightly higher score of 13.97.

It can be observed that BLEU scores for English-Turkish language pair fluctuate considerably (as it does for EN-TR SMT). This fluctuation may be due to the underlying algorithms, corpus quality, corpus size and corpus type. Furthermore, it implies the necessity to compare the results using different automatic evaluation metrics and human evaluation strategies.

Chapter 2. State-of-the-Art Terminology from Machine Translation Perspective

It is the task of terminologists to find a happy medium between authenticity and internationalization.

-Maria Teresa Cabré, 1998

In order to understand both the effect of terminology on machine translation and the effect of machine translation on terminology, a solid understanding of terminology theories and principles is necessary. In this chapter, we firstly begin with a historical overview of the terminology discipline, its principles, key concepts, and methodology (2.1.2). Then, we continue with the challenging aspects of terminology for translators. After that, we explain how different machine translation paradigms handle terminology (2.3) and how terminology is affected by machine translation (2.4). In the fifth section, we elaborate on terminology in translation evaluation and finally, we dedicate a section on Turkish medical texts and terminology.

The studies that have opened the way to create an autonomous discipline of terminology began as early as 1930s. Many terminology scholars see the works of Eugen Wüster in the 1930s as the works founding the principles of terminology. Protopopescu (2013) states that as an engineer, Wüster aimed to systematize and standardize terminology to avoiding ambiguities and solve practical linguistic problems. One of Wüster's basic principles was to begin terminology work with concepts and then pass to naming (an onomasiological approach). In what later is called General Theory of Terminology, he defended univocity, only one term referring to one concept, which deliberately excluded polysemy and synonymy. This was inevitable in a terminology work aiming for standardization. However, in the extended version of this theory, Wüster and his followers allowed synonymy to some extent, spoken forms of terms are admitted and phraseology is added to the study of terminological units (Protopopescu, 2013, p. 3). This approach has been influential until the end of 1980s when M. T. Cabré has developed a communication theory of terminology. Focusing on the theoretical aspects of terminology from communication perspective against a Wüsterian willingness for homogeneity, Cabré (2003) starts with two assumptions:

> The first assumption is that terminology is simultaneously: a set of needs, a set of practises to resolve these needs, and a unified field of knowledge. The second assumption is that the elements of terminology are the terminological units. (p. 182)

Her first assumption shows the pragmatic nature of her theory while the second assumption focuses on the object of the study of terminology. According to this second assumption, terminological units are units of knowledge (concepts), units of language (terms) and units of communication (situation), hence they are multidimensional and this multidimensionality distinguishes them from ordinary words.

Closely, another recent theory that has similar foundations to communication theory of terminology is the frame-based terminology theory (Faber et al., 2005; Faber et al., 2007) which has a cognitive approach to terminology and argues that the best way to study specialized knowledge units is through texts. This approach concentrates on three efforts: "(1) conceptual organization; (2) the multidimensional nature of terminological units; and (3) the extraction of semantic and syntactic information through the use of multilingual corpora."[56] This theory differentiates from the communication theory of terminology by considering redundant and unnecessary the need for making a distinction between a word and a term. Finally, there has been a tendency to focus on multi word units (MWU) arising from computational efforts for those specified words or 'words with n-modifiers'.[57]

Aside from the above-mentioned approaches the interest in terminology discipline is still increasing and new theories with multidisciplinary approaches are born. Other major theoretical approaches to terminology as recounted by Campo (2013) include sociocognitive terminology (Temmerman, 2000), cultural approach to terminology

---

[56] http://lexicon.ugr.es/fbt (08.12.2020)

[57] Cf. Aguilar-Amat, 1990.

(Diki-Kidiri, 2007), textual approaches with a technological component (Bourigault&Slodzian, 1998). One clear trend in these works is that the importance of the concepts as the units of our thought (not just as lexical labels) is being increasingly understood. And the second trend is related to the use of technology: today with the developments in information technologies, the study and practice of terminology are being computerized and advanced. We agree with Campo (2012) that “[g]reater understanding of methodological issues through basic research will result in more coherent and more precise practices” (p. 135) and that development in terminology will continue to be influenced by closely related disciplines.

Terminology theories have also influenced both the practice and study of translation. Conceptualizations such as special(ized) language, special(ized) text or specialized translation have shaped our understanding of the role and use of terminology in these contexts.

2.1. Theoretical Approaches to the Study of Terminology

Terminology goes hand in hand with special languages. For this reason, it is useful to start this section by exposing the meaning of special and general languages and their differences, if there are. Many terminology scholars such Cabré (1999; Aguilar-Amat et al. 2006; Kocourek, 1992; Hoffmann, 1979; Sager et al., 1980; Kageura, 2015) and others have investigated this dichotomy for properly understanding their relationship to terminology. After stating that “[t]he set of rules, units and restrictions that form part of the knowledge of most speakers of a language constitutes the common or general language” (p. 59), Cabré (1999) surveys different definitions of special language by some of the abovementioned scholars and comes to the conclusion that special languages “refer to the subsets of language that are pragmatically characterized by three variables: subject field, type of user, and type of situation in which communication takes place." (p. 65) According to this definition, she assumes that the subject fields expressed by special language are different from general knowledge, they are created or communicated formally by subject field experts under professional or scientific criteria. To sum up, she thinks that the major aim of special language is "allowing objective, precise, and unambiguous exchange of information particularly between subject field experts and professionals" (p. 77). However, she admits that it is difficult to draw a line between general language and special language. And in her referred work, she sees special language as a subset of general language which is, further, a subset of language. While Sager et al. (1980) draws a clear distinction between general language and special language, scholars such as Ciapuscio (2003) agree with Cabré (1999) and see a continuum between the two.

When we pass to levels of specializations in texts, (Hoffmann, 1998, pp. 56-69) cited in (Pérez, 2009), considers two types of variations: horizontal and vertical. The horizontal variation is related to areas of specialties or themes (medical, legal etc.) and vertical variation considers levels of abstraction, linguistic forms, scopes, and participants. With this framework in mind, Pérez (2009) argues that the distinctive features of specialized texts with higher abstraction levels manifest themselves in conciseness, precision, depersonalization, and their systematical approach. In the morpho-syntactic level, high usage of present tense, passive voice and nominalization are common with the purpose of being objective and universal. With the same aim, in the lexical-semantic level, abstract, mono-referential, specific concepts/terms are common and are used consistently. However, this does not necessarily mean that specific terms/concepts cannot be used in less specialized texts. From a translation point of view, Aguilar-Amat et al. (2005) highlights the importance of specificity of the used terms (the relationship of the denomination with the referent) in any text and considers that the degree of specialization is not an inherent question for terminology or thematic fields, because unclassified texts still deal with the same terminology problems in translation. In her view, texts become specialized with the quantitative predominance of specified terms of the same domain.

Below we will concentrate on different meanings of terminology, its subject matter, and its applications in translation. Finally, we would like to analyze the implications of the terminology use in machine translation. We focus on terminology from a translational point of view, our exposition also has implications for knowledge engineering, language planning, and cultural contrastivity. Terminology is a multidisciplinary field with its own methods and principles studying concepts, conceptual systems, and their labels —terms (Loening&Sonneveld, 1994; Cabré, 1999). The core subject matter of terminology efforts, concepts, are defined aptly by Loening&Sonneveld (1994):

> Concepts are the results of our segmentation and classification of the results of cognition, observation and experimentation; therefore, they also form a definite system which predetermines the systemic character of the terms expressing them. (p. 4)

While the meanings of the concepts in technical and scientific fields keeps changing —dynamicity, see Faber (2011), Kageura (2015), terminologists aim to keep track of this change, facilitate communication in an adaptable and flexible way by developing, formulating and/or classifying conceptual systems. By organizing conceptual knowledge in such a way, terminology also helps multilingual communication (i.e. translation and interpreting) (Loening&Sonneveld, 1994, p. 5). And being a multidisciplinary field, it acquires its knowledge from different fields such as linguistics, translation studies, philosophy as well as information technologies.

However, since it uses lexical items to express concepts in language, usually there is a confusion about its subject matter especially in comparison to subfields of linguistics (such as lexicology or semantics). Therefore, it is crucial to highlight particularly the difference between terminology and lexicology. Both Cabré (1999, p. 38) and Aguilar-Amat (1998) agree that the objects of study and methodologies of these two disciplines are different. Lexicology studies all words of a language, including both general and specific ones, their change and evolution in time, while terminology concentrates on concepts related to specific meanings in specialized fields, as mentioned above. The approach of lexicology is semasiological; namely from the word to the meaning or different meanings such as the case of polysemy while the approach of terminology is onomasiological, namely from the meaning/concept to the word/sign/term. What semasiology, in lexicography through dictionaries, will describe as polysemy, terminology's onomasiological method describes as homonymy (same item, different concepts, i.e. entries). For instance, considering a lexical item like *virus*, a lexicological (semasiological) approach would describe at the same entry both the medical (1) and the computational science meanings (2), while terminology will establish two different entries as they are two different concepts: 1) virus and 2) computer virus.

Hence, the distinction between lexicography and terminography is particularly significant. Cabré (1999, pp. 30-31, 110) points out that "lexicography deals with the principles and methods of writing dictionaries" and "[t]erminography involves gathering, systematizing, and presenting terms from a specific branch of knowledge or human activity". While the product of both of these fields is the same (dictionaries), the scope and purpose of these dictionaries are different. It should be highlighted that terminographic work does not only require collection of the terms in a particular subject field, but also the organization/classification of the conceptual relations (cause-effect, part-whole etc.; sometimes in form of tree diagrams) that these terms might have. Besides, unlike a purely descriptive methodology for describing the vocabulary of a given language in lexicography, one of the objectives of terminography is to standardize and/or clarify the concepts of a certain scientific or professional field. Considering the distinctions above, we can now continue with a relevant distinction between words and terms but we should note that the debate on this distinction is still continues and different theories of terminology either blur or reinforce this distinction that remains from the ancient and new world after scientific revolution. According to the communication theory of terminology of Cabré (1999), both words and terms share systematic linguistic features and refer to elements within reality (either they can be concrete, abstract, or imagined such as ‘alien’). However, terms refer to very specific concepts in specialized scientific or professional domains. In terminology, concepts come before and independent of words that represent them. Regarding grammatical categories, terms are predominantly in noun phrase forms while there are no strict rules banning the use of other grammatical categories such as adjectives, verbs etc. (Cabré, 2010, p. 358). Cabré also states that words and terms differ pragmatically, they are mostly used by professionals or used for communicating specialized knowledge effectively. In summary, “[a] term has a linguistic form and a content representing the concept” (Cabré, 1999, p. 95). Besides, the author also adds the condition of being associated with a field of expertise as a prerequisite for being a term. Developing a new Translation Theory of Terminology (as a part of Communication Theory of Terminology), Aguilar-Amat et al. (2005) approaches differently to the criteria of identifying a term. She argues that thanks to the technological developments (i.e. Internet) the possibility of communicating through specialized terms has become easier beyond specialized domain boundaries and hence, she prioritizes the correct/complete specificity of words to identify them as terms: high specificity and high understandability, plus high degree of equivalence and disambiguation, become the basic criteria. This theory is also more permissive to term variations for concepts compared to the perspective of standardizing the denominations of concepts to one, single denomination. Binomial way of building terms (a generic name and a limiting modifier; for example, "cellular ischemia") until a maximum level of specificity is achieved plays a foundational role in this theory based on the big number of bigram terms. Here, the objective is not obtaining univocity, i.e., a single name for a single concept. A concept may be denoted by different specific terms as long as the denomination is specific enough. We can observe that collocations occupy a significant space in Aguilar-Amat's theory (1993). Words such as "operation" or "terminal" may be ambiguous and not specified enough (low specificity, high ambiguity). Only when they are collocated with limiting modifiers such "surgical" (surgical operation), or nouns like "bus" (bus terminal) can they be understood correctly by translators (both human and machines) and hence be called terms and have an equivalence. This position is useful to consider that a standard translation for a word/term can change when adding modifier for specification: for example, the standard equivalence for (es) *ablación* would be (en) ablation, but:

(es) *ablación endometrial* – (en) endometrial ablation
(es) *ablación genital* – (en) genital mutilation
(es) *ablación del tabique nasal* – (en) excision of the nasal septum
Having these approaches in mind, we will look at the relationship between translation and terminology below and explain some key issues in terminology that create problems for translators and finally see the implications of these issues for machine translation.

2.1. Theoretical Approaches to the Study of Terminology

In translation practice, translation of specialized texts requires dealing with the terms used in the particular field of expertise because experts use terms in their texts (Cabré, 2010, p.357). This makes translators one of the most common users of terminology work. Translators firstly need to find the target language equivalent terms for source language terms. Note that this first step is not a creation process but a discovery process entailing research. They benefit from monolingual, bilingual and/or multilingual terminology dictionaries and other resources, monolingual textual documentation, and knowledge bases as well as bilingual or multilingual text resources in the forms of corpora (Cabré, 2010, p. 361). This search assumes that an equivalent term in the target language exists and is identifiable. (Achkasov, 2014, 218) states the following: "'to translate a term' implies, first and foremost, identification of a related (identical) concept in the conceptual system of the target language." The problems begin to arise when such a target term is not readily available in the target language (there is a gap) or there are multiple target terms for the single source concept (variations) or analogies. Not having a target term inevitably leads to coining new terms (neologism) or borrowing (loan translation) while having multiple options require criteria for selecting a particular term candidate (selection of a term based on its frequency of use, its being a neologism or the preference by the commissioner of the translation work etc.), a strategy for dealing with synonyms. To illustrate, the two authority terminology dictionaries on Turkish – English cardiology have 3 equivalents of en. "cardiac arrest": tr. "kardiyak arest", "kalp durması" and "yürek durması". Both "yürek" and "kalp" mean "heart" in English yet, "kalp" is an Arabic word while "yürek" is a Turkish word. Denominative variation is common in Turkish medical language since some experts prefer "pure" Turkish terms and others rely on the use of borrowed Latin and Greek rooted terms. Usually, Turkish medical translators will need to decide which term to use (see section 2.6). Other than these absolute equivalents, Cabré (1999, p. 110) counts five more synonym cases for terms:

i). initialisms and their full forms (for example: “COPD” = “Chronic obstructive pulmonary disease”),

ii). abbreviations and their full forms (Gln = Glutamine),

iii). scientific names and popular names (heart attack = myocardial infarction),

iv). standard form and dialectal form,

v). two synonymous terms having orthographic variants of the same term (Turkish example: “kardiyak arest” is sometimes written as “kardiak arrest” or “kardiyak arrest”).

Considering the objective of terminology to allow for clear, concise, and specific communication, proliferation of synonyms normally is not a desired phenomenon (while there are objections for such a view: see (Freixa, 2006; Aguilar-Amat et al., 2005). But, from the perspective of translation (and machine translation), each situation above, creates certain problems. In Section 4.3., we categorize correct and incorrect term translations in MT by also considering the list above.

Another relevant issue, referred in the description of onomasiology and semasiology, is related to polysemy and homonyms. Polysemy is the property of a lexical unit to have multiple meanings and is common in the general language. Since a single lexical denomination for a single concept in specific subject field is desired (according to the pragmatic situation), polysemy is treated as homonymy in terminology. In other words, at least in theory, it is expected that a concept should have one single designation in one particular subject field (Cabré, 1999, p. 40). In other words, the meanings of homonymous concepts are delimited by their domain/subject field. In terminography, the concept of “memory” in psychiatrics and in computer science will be defined quite differently. Note that homonymy usually affects the multi-word collocations compared to the unigram terms (“human memory” vs. “computer memory” or “CPU memory” or

“RAM memory”). Their co-occurrence with other words makes them more specific and in a translation scenario, it becomes easier to find an equivalent term, if there is one in the target language. But it is also where the strategy of translating a term by calquing leads to greatest problems. Let us consider the example of the superconcept “bank” and its collocations:

(en) riverbank – (tr) *nehir kıyısı*

(en) investment bank – (tr) *yatırım bankası*

(en) bank of clouds – (tr) *bulut yığını*

In Turkish, the superconcept of bank is translated differently in all three cases due to the specification contribution of the modifier. As it might be expected, this phenomenon has implications for machine translation. Literal translation of such concepts according to their most frequent occurrences by MT system may break the conceptual link between the concept and its label.

2.2. Terminological Challenges in Translation Practice

Cabré (2010) highlight that translators tend to firstly consult already available terminology resources to identify an equivalent concept and look for other strategies only when they are convinced that an equivalent is not available. Not finding an already existing equivalent may incline translators to come up with a new term candidate, i.e. creating target neologisms and/or neonism. Achkasov (2014, p. 219) states that translators should not develop terminology and use already existing resources. However, in reality, translators end up creating ad hoc terminologies (in comparison to benefiting from systematic terminology work) (Cabré, 2010, p. 359). Drifting apart from the systematic nature of terminology and adopting a semasiological approach in translating terminology concerns Achkasov (2014):

2.2. Terminological Challenges in Translation Practice

> Performative aspects of LSP translation contradict prescriptions, and, in turn, translators in many ways affect terminology. They often seem to be unaware and unconcerned about such rules and focus more on labels ('words') than concepts, use contextual and multiple designations or descriptive ways of concept representations within target texts, which confronts the nature on terminology. The latter issue has been the point of attention as well, predominantly as a matter-of-course though undesirable and non-systemic byproduct of terminology use. (pp. 210-211)

While such a view shows the clear gap between terminology theory and the practical approach in translation, translators follow systematic approaches (through their terminology education and training) to deal with terminological challenges of translation. In the next paragraph we list the terminological challenges encountered by translators according to Aguilar-Amat et al. (2009).

Translators come across many different terminology challenges including issues resulting from acronyms, analogies, synonyms (variations of denominations), collocations and neologisms. We provide examples are from medical domain in English to Turkish translation but these challenges are similar in other domains as well. It may be assumed that these challenges for translators are also relevant and cause problems of equivalence in machine translation.

In Aguilar-Amat's approach, the first challenge for translators is related to acronyms in the source language. Acronyms are used to economize either written or oral communication and may pose difficulties during translation depending on a variety of factors. In each scenario, the translator needs to know the unabbreviated form for finding the proper equivalent. In the first case, there may be standard equivalent acronyms in target language. They may be abbreviated either according to the initials of the source term (en. "AIDS" = tr. "AIDS" or en. "HIV" = tr. "HIV") or according to the initials of the target term (en. "COPD (chronic obstructive pulmonary disease)" = tr. "KOAH (kronik obstrüktif akciğer hastalığı)". Here, the translator needs to identify the standard equivalent in target language. However, as Erten (2016, p. 62, cited in

Koprulu, 2017), highlights the same abbreviation may be used for more than one terminological unit: “CT”: “Cellular therapy”, “Cerebral tumor”, “Clotting time”, “Corneal transplant”, “Cerebral thrombosis”. In such a case, the translator should be aware of the context and understand properly which meaning is intended and select the correct abbreviation or translation strategy accordingly. Acronyms may also have multiple equivalents in the target. (Koprulu, 2017, p. 258) gives the example of “chronic obstructive pulmonary disease (COPD)” which can also be abbreviated as COLD (chronic obstructive lung disease) or COAD (chronic obstructive airway disease). Koprulu (2017, p. 259) points out that there are also situations in which the acronyms used in the source text do not have standard equivalent acronyms in the target and may either be neologisms or particular to source text author. Depending on the domain expertise of the translator, giving the source and target unabbreviated terms together in the first occurrence in the text or a descriptive translation may solve the translation problem.

The next challenge for terminology is the use of synonyms (different denominations of terminological units referring to the same concept), also called variants. We have mentioned above that terminology scholars have been hesitant about the proliferation of synonyms in scientific and professional domains since they may decrease standardized, efficient communication between experts. The translators may come across synonyms in different forms and they should make a proper choice, considering communication, experts, and language coherence. One source term may have multiple synonyms in the target creating a challenge for the translator to select the most appropriate one (one-to-many). Conversely, different synonymous terms may be used in the source text and the translator needs to understand that they refer to the same term in the target (many-to-one). Lastly, multiple synonyms may be present for the same concept both in source and target languages. Koprulu (2017, p. 261) notes that understanding the type, purpose and audience of the text may give clue to the translator in selecting the proper target term. “Tuberculosis” in English can be translated as “tüberküloz” in professional medical communication (doctor-to-doctor context) and “verem” in a less formal situation (doctor-to-patient or patient-to-patient) (different contexts of medical communication: Moraes, 2014 cited in Guner (2020, p. 347). Besides, translators also benefit from frequency of use searches in specific domain bilingual corpora to take a decision about the term.

Koprulu (2017, p. 256) also warn about the challenges that homonymous terms may create and gives the example of “cardia” which is usually used in collocations related to heart. However, the term also means “the region following the “z-line” of the gastroesophageal junction”. In a gastroenterology context, misunderstanding the meaning of “cardia” may cause some serious translation and/or health problems. Observing the collocations and understanding their meaning within context may help increase the specificity and find the correct target term.

The use of analogical components to refer to concepts is common in all domains of expertise. According to Pena&Andrade-Filho (2010, p. 619), analogies are useful tools for “human reasoning and learning [and] for resolving problems and providing arguments”. For example, with the analogy to a fruit, the term “strawberry gallbladder” has a visual clue about the form of the gallbladder. Similarly, in “sickle cell anemia”, the shape of the affected red blood cells is resembled to a sickle which is an agricultural tool. While the Turkish translation of this term uses the same analogy, in some situations the analogies in the target language may be different (and maybe no analogy is present in the common equivalent term) and literal translation of the source analogy leads to inaccurate translation. In Turkish, “kulakçık” literally means a “little ear”. When collocated with “kalp”, it becomes “kalp kulakçığı” and its English equivalent is “heart atrium”. In Latin, “atrium” literally means an “entry hall”. It is obvious that the same analogy is not used in Turkish. Hence, a literal translation into English would cause a translation problem. Similarly, “kulak salyangozu” (the snail-shaped component of the human ear) literally means “ear snail” in Turkish and should not be translated literally since its equivalent in English is “cochlea” (snail in Latin).

Turkish term	**literal translation**	**English equivalent**
kalp kulakçığı	little heart ear	heart atrium
kulak salyangozu	ear snail	cochlea

Table 8. Analogical terms in Turkish and their translations.

These analogies especially create problems when they occur alone in a sentence and are not collocated with other modifiers. As Aguilar-Amat&Torres-Hostench (2021) highlight, the translator needs to compensate for the missing modifier and take a translation decision accordingly. For example, if TR → EN translators see “salyangoz” in an ear-related medical text, they should compensate it with “kulak” to achieve “kulak salyangozu” and find its equivalent. Pena&Andrade-Filho (2010, p. 617) include a list of analogies used in medicine including “cigar” (elongated nuclei with relatively blunt ends), “carrot” (elongated nuclei with blunt ends), “cartwheel” (radiating chromatin) for describing nuclei.

Considering that terms as multi-word units (n-grams) do not solve problems in the context of terminology for translators, we need to differentiate them from phraseologies in languages before we pass to phrasal terms or terminological collocations. According to Cabré (1999, p. 91), phrasal terms are “lexical structures with a terminological value (even though they are made up of words)”. On the other hand, phraseologies are structures that occur in discourse frequently and have quasi-standardized forms but they do not denote a specific concept, hence they do not have a terminological value. She gives examples of phraseology from computer science: “press the key”, “edit a text” etc. While there are strategies to deal with phraseologies in translation, our focus here is on terminological collocations and below we will look at them more closely.

Collocations differentiate from other co-occurrences by their high level of understandability and specificity. Benson (1985, p. 5; cited in Cigan, 2018), state that “two features that distinguish a collocation from a free phrase are its frequency and rather low possibility that semantic concept contained in a collocation can be expressed by other word connections”. Below we will see that connotations are also divided into further categories and among these categories, terminological collocations have the highest specificity and understandability. They are comprised of two or more concepts to specify a certain reality (Aguilar-Amat et al., 2005). They include a nucleus or head as well as a modifier. The importance of terminological collocations is derived from their specificity in denominating concepts. This specificity, in the context of translation, helps translators to identify more easily the meanings of source concepts and find their target equivalence. The Turkish word “kapak” can mean “cover”, “lid”, “cap” in Turkish when it is not specified with a modifier. However, when collocated with another word in medical domain and the term “kalp kapağı” is formed, the equivalent of this medical term in English becomes “heart valve”. If we were to translate this term as “heart cover”, the specificity would be lost and the reference to the concept would be ambiguous. It is also worth mentioning how the analogy of “cover” in Turkish is expressed with a different analogy in English, “valve” (see the section of analogy above). Detecting and finding the equivalents of terminological collocations plays a crucial role for avoiding these kinds of shifts of meaning.

Aguilar-Amat et al. (2005) note three broad types of collocations: attributive collocations, terminological collocations, and idiomatic collocations. Modifiers (noun or adjective) in attributive collocations do not increase specificity of the nucleus (example: “strong tea”) and can be quantified (“very strong tea”). In idiomatic collocations, at least one of the referents related to expressed concepts is usually opaque and figurative (“black market”). They can’t be quantified (*“very black market”) and this feature of syntactic rigidity also affects terminological collocations. Terminological collocations are the one of the key objects of terminology study. They are mostly binomial. According to Aguilar-Amat et al. (2005), “terminological collocations have a set of well-known morphosyntactic, semantic and conceptual properties that undergo few variations whether we talk about one language or another.” Syntactically, they consist of noun + noun (“heart transplantation”), adjective + noun (oxidative stress) and other combinations of adjective + noun phrases (“chronic obstructive pulmonary disease”). Understanding and mapping the grammatical categories of terminological collocations help extract them monolingually or bilingually from parallel or comparable corpora both for knowledge extraction and for translation purposes.

In translation, the correct identification of these collocations in source language and finding their equivalents in the target language are crucial for translation to transfer meaning of the concepts correctly. Frame-based terminology theory (as stated in the website[58] of the research group) also assigns a special role to terminological collocations:

> Specialized language units are mostly represented by compound nominal forms that are used within a scientific or technical field, and have meanings specific of this field as well as a syntactic valence or combinatory value. (cited from the website)

These compound nominal forms correspond to terminological collocations. And applying wrong translation techniques such as literal translation (as we have shown in the example above) while translating them may create target terms that are not used or understood by special field experts. In section 2.3. and 2.4., the relevance of this challenge for machine translation is highlighted.

Last challenge for translating terminology is related to neologisms in Aguilar-Amat's (2021) approach. Since scientific and professional disciplines keeps developing and changing, new concepts are being born every day. Cabré (1999) specifies four criteria for considering a unit as a neologism:

> a. diachrony: a unit is a neologism if it has arisen recently
> b. lexicography: a unit is a neologism if it is not in dictionaries
> c. systematic instability: a unit is a neologism if it exhibits signs of formal instability (e.g. morphological, graphic, phonetic) or semantic instability
> d. psychology: a unit is a neologism if speakers perceive it as a new unit (p. 205)

In a particular language, normally official language planning institutions such Consortium of the Centre for Terminology TERMCAT[59] in Catalonia and Turkish Language Association[60] coin the terms for new concepts based on standard guidelines in regular intervals. Translators follow these institutions and benefit from their monolingual, bilingual or multilingual resources (Cabré, 2010). However, sometimes, resources may not be available and the translators may need to coin the new terms

---

[58] http://lexicon.ugr.es/fbt (08.12.2020)

[59] https://www.termcat.cat/en (last access: 10.12.2020)

[60] https://www.tdk.gov.tr/ (last access: 10.12.2020)

(which according to some authors may not be their duty under normal conditions: see Achkasov (2014)). There are many prescriptive techniques for translation that can also be applied for the translation of terminology (see the paragraph below). The use of these techniques may carry the risk of not referring properly to the source concept. Nevertheless, since these translation techniques solve communication problems pragmatically, their application is common in translation practice.

The 18 translation techniques developed by Molina&Hurtado (2002), parallel to the Vinay&Darbelnet (1977), should be appliable to terminological issues (Aguilar-Amat et al. 2009). Orozco-Jutorán (2016: pp. 165-166) also proposes 12 translation techniques for the same purpose. We present these two sets of techniques in a single table. As it can be observed from the table, most of the techniques overlap and the use of equivalent terms is also included as a technique.

Hurtado&Molina (2002)	Orozco Jutorán (2016)
Adaptation	Total or regular equivalent
Amplification	Contextual equivalent
Borrowing (pure and neutralized)	Functional equivalent
Calque	Borrowing
Compensation	Calque or lexical translation
Description	Neutralization
Discursive creation	Neologism or lexical creation
Established equivalent	Periphrastic translation
Generalization	Translator’s note
Linguistic amplification	Omission or reduction
Linguistic compression	Addition or explicitation
Literal translation	Compensation
Modulation	
Particularization	
Reduction	

Substitution	
Transposition	
Variation	

Table 9. Translation techniques that can be used for translating terminology.

To illustrate the use of these techniques, a common technique used in English to Turkish medical translation is the neutralized borrowing (see the third row in left column above: "borrowing (pure and neutralized)"). Latin or Greek rooted terms in English are borrowed and neutralized considering Turkish pronunciation and grammar rules of Turkish: en. "Carotid artery stenting" = tr. "Karotis arter stentlemesi". Considering the large number of techniques that can be employed by a translator, it becomes interesting to observe how machine translation can deal with the translation of terms in a specialized field, which we describe below.

2.3. The Use of Terminology in Machine Translation

Considering the complex relationships between concepts and their representations through terminological units, terminology continues to be one of the biggest challenges for machine translation. Machine translation paradigms deal with terminology at a shallow level and usually do not differentiate terminological units from ordinary lexical units, which, in turn, causes a loss between the concept and the term denoting it (see Haque et al. (2019a). Each MT paradigm handles terminology differently at different steps of the preparation with varying degrees of complexity and success. In this section, we will cover the use of terminology in RBMT, SMT and NMT, results obtained by them, and different strategies for improving terminological quality.

2.3.1. Terminology in Rule-Based Machine Translation

RBMT uses monolingual and bilingual dictionaries. Source language dictionary, bilingual dictionary and target language dictionary need to be built for the architecture of the RBMT system (Reynolds, 2015). While carefully prepared and extensive term lists can provide accurate and consistent terminology, terminology creation takes time and is expensive. As Reynolds (2015) highlights very large amounts of terms are needed, and when they are introduced, all the dictionaries in the system shall be synchronized. In an article comparing RBMT, SMT and NMT, Senellart (2016) sees "the huge, contextual and always expanding volume of terminology units"[61] as one of the biggest limitations of RBMT. In principle, well-formed, dynamic bilingual dictionaries including terminological collocations and their equivalents may help improve the translation quality of RBMT systems. However, creating these dictionaries is costly and time-consuming, and for each domain a new dictionary shall be created since these systems are highly influenced by polysemy and can create incorrect translations since terminological units may have different meanings in different domains. To illustrate this point, Aguilar-Amat (2002) gives the structural representations of "ordenador de quinta generación" as a terminological collocation in a RBMT system:

```
%% 23
%% ordenador de quinta generación
%% N1_P1_A_N2
MWU= {/cat=np} [*{},
        {mwu_lu=MWU/cat=n,e_lu=N1},
         {imwu=yes/cat=pp}[{/e_lu=P1},
               {/cat=np}[ *{},
                       {/cat=ap,head=A},
                       {/cat=n,e_lu=N2},
                       *{imwu=no/cat=?}]],
         *{} ].
```

[61] Comparing Neural MT, SMT and RBMT – The SYSTRAN Perspective. http://kv-emptypages.blogspot.com/2016/09/comparing-neural-mt-smt-and-rbmt.html (last access: 28.04.2020)

In cases like these, all occurrence patterns of terminological collocations need to be known beforehand and these representations should be written from scratch.

2.3.2. Terminology in Statistical Machine Translation

SMT and NMT are trained on and learn from corpus. In principle, corpus-based MT systems reflect the content of their corpora. Hence, when corpora with in-domain terminology is introduced, the MT engine will provide better results for in-domain translation considering terminology. Since the SMT algorithms extract monolingual and bilingual multi-word phrases (n-grams starting from unigrams until 7-grams; Galbrun, 2009, pp. 13-14) from the corpus-based on the preferences of the developers, they can produce equivalent terminological collocations in the translation depending on the size and domain of the training corpus. For such a result, corpus should be both of a specific domain and have enough size because calculations are based on probabilities of co-occurrences (Lumeras&Way, 2017, p. 33). Since specificity in terminology increases with terminological collocations, ideally the frequency of the collocations should be high in the training corpus to provide terminologically accurate translations since in SMT, statistical probabilities usually determine the end result. In the decoding phase when a new source string is introduced for translation, if the terminological units in the source have a low frequency of occurrence in the training corpus, they may be inserted in the target sentence without translation. These units may have synonyms which are not related to the domain and their insertion in the target sentence may create inaccurate translation. In another scenario, the subcomponents of the terminological collocation may be in the training corpus and the complete terminological collocation may not be there, hence again an inaccurate translation may be created because of a literal translation. To illustrate this with the example used section 2.2., the translation of tr. “kulak” and tr. “salyangoz” as en. “ear” and en. “snail” may be present in the training corpus but tr. “kulak salyangozu” (as a bigram) may not be present. Hence instead of the accurate translation “cochlea”, en. “ear snail” could be produced as a translation. This shows the importance of large specific domain parallel corpora. We can assume that alignments of higher numbers of n-grams (for example: 4-grams in Turkish aligned with 4-grams in English: tr. “sağ posteriyor serebral arter”: en. “right posterior cerebral artery”) may help produce more accurate terminological units in the target.

Considering the translation of other terminological issues such as acronyms and synonyms, SMT again benefits from the probabilities provided by phrase tables. If the acronyms do not occur in the bilingual training corpus, they are inserted without any change (which may lead to accurate translation without a conscious conceptual link; in other cases, the transfer of the abbreviation into the target will be inaccurate). In the case of synonyms, instead of the conceptual link between the concept and the term, the frequency of occurrence in the corpus will determine which synonym will be used in the translation. In a mixed domain corpus, this may lead to inaccurate or out-of-domain terminological units to be used.

Aside from the content of the corpus, as Scansani (2020, pp. 23-24) observes, SMT architecture allows for implementing terminology in preprocessing and postprocessing steps, which may improve terminology accuracy.

Lumeras&Way (2017) explain two scenarios in which terminology can be integrated into SMT engine. Firstly, in preprocessing step, the desired terms[62] can be included into

[62] A client-provided bilingual term list or a standard bilingual term list.

the training corpus. In other words, terms in the target sentences can be replaced with desired terms with a function like find/replace. A term list can also be added for the training. Nevertheless, as Eisold (2017) pinpoints, these two types of interventions do not fully guarantee the accurate translation of the terms in the engine because the actual translation will depend on their probabilistic distribution within the bilingual corpus (translation model) and monolingual corpus (language model). Besides, Lumeras&Way (2017: 26) state that overall MT quality (e.g. in BLEU score or other scores) may slightly decrease due to these interventions. This drop in quality may be caused by the fact that usually term lists include terms in their nominal form but they may occur in inflectional forms in the training corpus. The second scenario uses a "run-time approach" (Eisold, 2017); namely, term list is integrated to the pipeline after the training of the phrase table (see section 1.2.3. for the significance of phrase table) is completed and runs together with MT decoder, phrase table and language models (if available, other parameters) (Lumeras&Way, 2017, p. 26). In the tuning phase, this term list is assigned a score by MERT, and if the quality of the sentence increases in regard to BLEU, the terms are inserted to the target sentence; if not, they are ignored. Lumeras&Way (2017) finally state that in the first scenario, the user has more control over terminology in comparison to the second scenario. In short, it is possible to intervene in terminology translation in different stages of SMT training and translation.

2.3.3. Terminology in Neural Machine Translation

NMT is also based on corpora and is more sensitive to the content of the corpus compared to SMT to a point that it sacrifices adequacy in favor of fluency when low quality out-of-domain corpus is introduced (Koehn&Knowles, 2017, p. 28). Therefore, the quality and type of the corpus play a significant role for accurate terminology translation in NMT. Besides, the NMT decoder considers the source sentence as a whole while producing the target sentence, which leads to improved grammatical distribution in the sentence, and hence more fluency (See subsection on Encoding and Decoding in NMT).

In NMT, forcing terminology into the pipeline is not readily supported (Lambert, 2018) because of the end-to-end architecture of NMT. Considering the importance of custom terminology, this hinders the practical NMT use. Different methods have been proposed to solve this problem. A straightforward solution is domain adaptation, in which initially a general domain system is trained and then training is continued for a few iterations by specific domain data (Koehn&Knowles, 2017, p. 29).

For other current methods on terminology integration in NMT, Lambert (2018, 2020) provide a comprehensive overview. Of the seven different articles that the author cites, the methods implemented can be grouped under two general strategies: i). enriching the training corpus with tagging and ii). restructuring the beam search (a common search component in MT to retrieve target words in reference to source words) to constrain the NMT decoding to yield the requested terminology. Each of these strategies comes with their limitations and costs. According to Lambert (2018), the methods used by Hokamp&Liu (2017), Post&Vilar (2018), Anderson et al. (2017) and Hasler, et al. (2018) use different beam search architectures to improve the recall of terms in the corpus (the strategy is referred to as “constrained decoding”). However, these methods may undermine translation speed of the system due to the computational requirements and cause reordering problems in the target sentences. The method of Crego et al (2016) as well as the three methods explained by Wang et al. (2019) are focused on tagging the terminology in the training corpus. First approach tags terms in source and target sentences with tags like <start> and <end>:

> "He has a <start> cardiovascular disease <end>|||Onun <start> kardiyovasküler hastalığı <end> var"[63]

This facilitates the detection of the terms in the training. The second approach which is called "mixed phrase" adds the target term near the source term:

> "He has a cardiovascular disease kardiyovasküler hastalığı |||Onun kardiyovasküler hastalığı var"

And the third approach combines these two approaches and both tags the terms and appends target terms in the source sentence (tagging + mixed phrase):

> "He has a <start> cardiovascular disease <middle> kardiyovasküler hastalığı <end>|||Onun <start>kardiyovasküler hastalığı<end> var"

There is also an extra embedding step to differentiate the source and target terms. With the third approach with embedding on a Chinese-English and German-English news domain NMT engines, Lambert (2020) reports that the percentage of accurately translated terms (named entities in this case) increases from 74% to 99.40% in the Chinese-English engine, and from 91.2% to 99.3% in the German-English engine. The author adds that these remarkable results are obtained without sacrificing the translation speed. However, the external term list needs to be ready to be implemented into the training data with the abovementioned methods. This external resource may not be readily available, and its creation may be labor intensive. To sum up, the approaches above demonstrate that exciting developments are happening for improving the terminological accuracy of the NMT systems. Evaluation studies with this approach on other language pairs including morphologically rich, free-form languages such as Turkish can provide a clue about the its effectiveness.

[63] Example sentences are adapted from Lambert (2018).

In the following section, we will look at the terminology evaluation strategies both for human translation and different MT systems.

2.4. The Effect of Machine Translation on Terminology

Freely available generic machine translation systems such as Google Translate[64], Microsoft Bing Translator[65], Linguee[66] are increasingly being used both by general public and professionals. Especially the use of these systems in professional domains may create impacts that are not desired from the point of view of terminology. Vieira et al. (2020) reviews this impact in medical and legal domains and warns about the risks of using MT in these domains. While corpus-based MT systems may avoid these impacts to some extent by the implementations of preprocessing and postprocessing methods to ensure the desired terminology is used consistently, the use of generic MT systems may not guarantee consistent or accurate terminology. In this section we investigate how terminology as a whole can be influenced by the MT systems.

In today's dominant MT paradigms (SMT and NMT) in most languages and workflows, the translation quality of an MT engine is dependent on, among other things, the type and quality of the bilingual training corpus (see section 1.2.1 on corpus). Hence, accurate translation of terminology in a subject field can only be ensured if the bilingual training corpus include the relevant terms both in the source and target sides. Besides, since both NMT and SMT are based on probabilistic distributions of the terms, the presence of these terms is necessary but not sufficient. Their frequency of occurrence should also be sufficient in the corpus so that once a new sentence with source terms is

[64] https://translate.google.com/ (last access: 21.12.2020)

[65] https://www.bing.com/translator (last access: 21.12.2020)

[66] https://www.linguee.es/ (last access: 21.12.2020)

to be translated, the accurate equivalents can be fetched by the decoder. If the frequency of occurrence of a particular term is not sufficient, it will not get the highest score as a candidate equivalent and it will not be translated accurately.

In generic MT systems trained with large mixed domain corpora, this phenomenon may be more problematic since most frequent term candidate and its probabilistic equivalent may be from a different domain. A professional or a non-professional may input a legal sentence in an MT system and the MT system may be trained on mostly with medical corpora. This may cause an inaccurate transfer of the terms into the target language. Besides, systems such as Google Translate train their engine between English and another language, and for language pairs that do not have English, they use English as a pivot language[67]. For example, translation from Spanish into Turkish is done in this order: Spanish → English, English → Turkish. Inevitably, this tripartite transfer may lead to loss of meaning and a break between a term and the concept that it designates.

The solution to the above-mentioned problem is generally to train customized MT systems with specific domain bilingual texts including specialized terminology. However, usually the corpus size is not large enough and significant probabilistic occurrences of terms may not be guaranteed for accurate translation. In cases of low frequency, terms are inserted into target sentence without translation. Term variation (synonyms) can still be a problem here if terms in the training corpus are not consistently used across different documents. For instance, an MT engine trained on both professional medical texts (doctor-to-doctor) and public healthcare texts (doctor-to-patient) may yield inconsistent term translations when a specialized text is translated.

[67] https://cloud.google.com/translate/docs/languages (last access: 21.12.2020)

2.4. The Effect of Machine Translation on Terminology

To illustrate with an example we previously used, a sentence with “myocardial infarction” may need to be translated to Turkish as “miyokard enfarktüsü” in a specialized register; however, since its common translation as “kalp krizi” is more frequent in the corpus, the MT will choose it as the equivalent term. This does not necessarily mean that Greco-Latin rooted medical terms are more correct than their Turkish rooted equivalents. As we have mentioned above, while there are struggles to suggest Turkish rooted medical terms, experts still tend to use borrowed terms.

The use of MT systems may lead to neologisms as well. Due to the high amount of MT consumption (due to a possible lack of human translation and high demand for information from a source language; for example, information from Chinese machinery field or medical research in China into Turkish), non-standard target equivalents may become prevalent in a scientific or technical field. If there is no language planning or standardization effort in the receiving culture, such a flow of neologisms may hinder effective communications. Besides, the translation of terminological collocations may be problematic. If a sentence is not properly disambiguated in the MT phase, the collocates of a terminological collocation may be translated separately. In the best scenario, they create an acceptable neologism. In other scenarios, their separate translations may break the link between the term and the concept it denominates, and the overall translation of the term may be meaningless. Similarly, as Aguilar-Amat (2005) notes, if the modifier in a binomial term is not present in a collocation (uncompensated), MT engine will recognize the term and, hence, produce an incorrect translation. The example below clearly shows how uncompensated terms may lead to incorrect translations in specialized texts.

2.4. The Effect of Machine Translation on Terminology

Turkish	**Google Translate**	**Translation with Correct Term**
Hasta kriz yüzünden acil servise alındı.	The patient was taken to the emergency room because of the crisis.	The patient was taken to the emergency room because of the heart attack.

Table 10. Translation example from Google Translate.

The author of the source text (like in the example above) may aim to economize the text size and assume that the shortened terms are understandable from the context. However, such a strategy is detrimental to MT since its context knowledge in the translating phase is limited to the words within the source words. In a larger scale of societal level where generic MT systems are used repeatedly by the general public and experts, the desired specification of terminological units run the risk of getting lost on the way, which, in turn, deteriorates specialized communication in the target culture. Hence, the influence of these systems on terminology should be a concern for language/term planning organizations.

We have seen in section 2.2 that acronyms are already challenging for translators. Different methods of forming acronyms and their non-standard uses also pose problems for MT. Guidelines[68] for preprocessing training corpora or texts to be translated usually suggest avoiding acronyms. The reason for this phenomenon is that MT training algorithms do not differentiate acronyms from other lexical items in a sentence and processed based on the same probabilistic rules. As a result, very common acronyms such NATO, UN, WHO are usually translated correctly by the MT engines. Yet, as it

[68] IBM Cloud Machine Translation Tips. https://cloud.ibm.com/docs/GlobalizationPipeline?topic=GlobalizationPipeline-globalizationpipeline_tips (last access: 22.12.2020)

may be expected, less common and/or non-standard acronyms in the source text may not be translated correctly. This challenge is particularly relevant in medical domain where the use of acronyms and abbreviations is very common. Curiously, there are not enough studies on the use of acronyms in machine translation. In practice, MT expert suggest preprocessing texts to be translated and type the extended synonyms of the acronyms. However, this can take time and be costly.

Lastly, terminological collocations formed by at least one analogical item may be translated literally by the MT engine and the obtained target collocation may not be the standard equivalent for the source collocation. The example in section 2.3.2 (tr. "kulak salyangozu", en. *"ear snail") shows how this equivalence problem may happen. In contexts of no translation quality check (for example multilingual communication between experts), this new terminological collocation equivalents may enter into the target culture and remain as synonyms, adding to the proliferation of synonyms in terminology systems.

The issues that we have listed above continue to receive criticisms from translation terminology scholars such as Aguilar-Amat (in press). MT practitioners suggest using preprocessing of training corpora by focusing on terminological issues in general or postprocessing the MT outputs based on bilingual glossaries. Nonetheless, due to the syntactical and semantic differences between languages, these strategies may not provide correct terminology translations. More research is needed to expose the terminological issues arising from machine translation. For this purpose, terminology-focused translation evaluation can show which terminological error types are frequent and, in turn, relevant measures can be taken both by translators and MT practitioners to provide high quality terminology in MT outputs. Below we explore translation quality evaluation with a focus on terminology translation in MT.

2.5. Terminology in Translation Evaluation

The accuracy of terminology plays a significant role in the overall accuracy of a translation. For this reason, terminology is listed as one of the 8 main dimensions of MQM/DQF error typology (Lommel, 2018). The "terminology" error dimension includes two subcategories: i). inconsistent with a termbase and ii). inconsistent use of terminology.

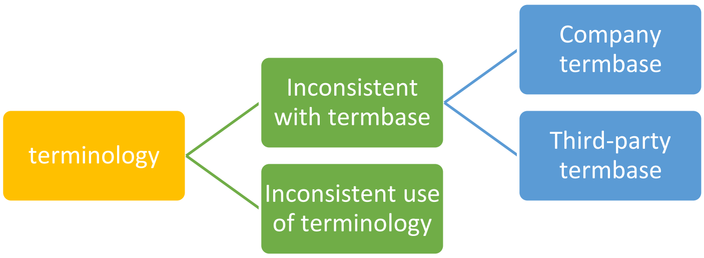


Figure 8. Terminology dimension in DQF/MQM Harmonized Error Typology. Figure adapted from the MQM website[69].

As it can be observed from Figure 8, the first subcategory is further broken down to inconsistency with a company termbase and third-party termbase. The second subcategory is based on the document level inconsistency of term use. In the guidelines of the DQF/MQM, it is stated that because of the existence of a separate category of accuracy, simple mistranslations of terms are to be considered accuracy errors. For this reason, the terminology category in this typology gives more weight to consistency of term usage. A reference company termbase or third-party termbase is required for determining if there is a terminology error in a translation. And these termbases are

[69] http://www.qt21.eu/mqm-definition/issues-list-2015-12-30.html#terminology (last access: 29.04.2020)

human made and can have some mistakes or from a terminological perspective, they can be obsolete. And the other subcategory (inconsistent use of terminology) requires access to a full text and the ability to compare term use across the text. In other words, this categorization assumes a document-level translation evaluation. QA features integrated in CAT tools, and the stand-alone QA tools such as XBench compare source and target strings with a given bilingual term list, and flag possible mismatches or missing terms in targets; however, they may produce false positives when inflectional languages are involved. When it comes to machine translation, quality evaluation is usually performed at sentence level. Besides, these subcategories do not provide enough insights about the types of terminology errors that an MT system yields, which may be important for improving the MT system and determining a postediting strategy. Below we will cover terminology error typology in MT which is a rarely studied research area.

MT comparative evaluation studies usually consider terminology simply as an issue of lexical choice (Scansani et al. 2019) without inspecting the details of terminology errors (Haque et al. 2019b). Scansani (2020: 22-23) presents a comprehensive overview of NMT vs. PBSMT evaluation studies on different language pairs including Bentivolgi et al. (2016), Toral&Sanchez-Cartagena (2017), Castilho et al. (2018) and Van Brussel et al. (2018). For the description of the differences between SMT and NMT, see the subsection “Differences between SMT vs NMT”. The mentioned studies investigate lexical choice (not only specific domain words but also other function words) errors such as addition, omission, mistranslation, and the evaluations are either performed automatically or by human translators. The author concludes that there are mixed results in terms of lexical choice and some results (such as Van Brussel et al. (2018)) question the assumed superiority of the NMT system (Scansani, 2020, p. 23). It should be highlighted that these studies do not differentiate between specialized terms and other lexical items in language. Not having a clear definition of what a term is in the above-mentioned studies does not give us sufficient information about how NMT or PBSMT handle terminology translation.

There are only a few studies that examine the types of terminology errors in MT systems (Haque et al. 2019a, 2019b; Scansani et al. 2019; Scansani, 2020). The evaluation strategy of Scansani et al. (2019) contains a term annotation step by two human annotators in which Italian-English single-word and multi-word terms are annotated in a corpus of 2055 parallel sentences. In this annotation, the encountered terms are labeled as “disciplinary”, “education” and “education equipment”. Later, this “gold standard” sample set is used for evaluating the terminology use in two different NMT systems. They use a metric called Term Hit Rate (THR) obtained from the study of Farajian et al. (2018). This metric “takes in a list of annotated terms in each reference sentence and looks for their occurrence in the MT output” (Scansani et al., 2019). It can compute exact or fuzzy term matches. While this approach can be beneficial to identify accurate, inaccurate or partially accurate terminology translation, it does not provide a fine-grained error analysis for MT systems; and the challenging aspects of terminology for MT such as acronyms, analogies, terminological variations (synonyms) and neologisms cannot be identified. Besides, what constitutes a term is determined by professional human annotators and rigorous criteria such as high specificity or low ambiguity (as might be expected by terminologists) are not applied, and hence single-word terms are common in their dataset which may lead to disambiguation problems for the MT system and, in turn, more inaccurate MT results.

Considering the need for understanding more thoroughly how MT handles terminology, Haque et al. (2019) implements a different annotation strategy and suggests a more terminology-focused error typology. This is the one of most detailed error typology concentrating solely on terminology in MT in the literature. In this typology, the translated terms by PBSMT and NMT are not only identified as correct/incorrect but the nature of their incorrectness is also determined and annotated in their approach. The application of this methodology is interesting to compare relative strength and weakness of PBSMT and NMT systems in view of terminology. The reference language pair in their case is Hindi-English. The authors firstly identify 5 categories for the terminology errors but later increase this number to 8 error categories. The 8th category is "other error" which may include any other possible kind of error not covered in the previous 7 categories. The table below gives the details of these categories.

Error category	Abbreviation	Description
reorder error	RE	the translation of a source term forms the wrong word order in the target
inflectional error	IE	the translation of a source term inflicts a morphological error
partial error	PE	the MT system correctly translates part of a source term into the target and commits an error for the remainder of the source term
incorrect lexical selection	ILS	the translation of a source term is an incorrect lexical choice
term drop	TD	the MT system omits the source term in translation
source term copied	STC	a source term or part of it is copied verbatim to target
disambiguation issue in target	DIT	although the MT system makes a potentially correct lexical choice for a source term, its translation-equivalent does not carry the meaning of the source term
other error	OE	there is an error in relation to the translation of a source term, whose category, however, is beyond all remaining error categories

Table 11. Terminology error types described in Haque et al. (2019, p. 439). Table is created according to the original definitions in the paper.

2.5. Terminology in Translation Evaluation

Haque et al. (2019) apply these categories on what they refer to as “single-word” and “multi-word” terms. In an extended version of the study Haque et al. (2019b) also makes a distinction between “superterms” (e.g., “bus terminal”, “computer terminal”) and “subterms” (e.g., “terminal”) and annotate both of these categories. We have already highlighted that single-word terms lack specificity and hence are problematic both for human translators and machine translation. Terminological collocations (“superterms”) are more specific and, when translated, may have all error categories (a subterm may not have the error category of reordering). The classification of Haque et al. (2019a) also lacks a focus on acronyms and analogies in terminology and only implicitly covers the concept of neologisms (see the classification of correct translation of terms below).

The authors also provide a classification for correct term translation in order to understand the diversity of their MT model in view of translating domain terms and how similar or different is the target term to the reference term in view of lexical and inflectional variation. Their correct term classification has 7 categories. These categories are given in the table below.

Correct Term (CT) Class	Abbreviation	Description
CT given the reference term	CTR	The translation of a source term is the reference term
CT given one of the LIVs	CTV	the translation of a source is one of the LIVs of the reference term
variation missing	VM	a source term is correctly translated into the target, but the translation is neither the reference term nor any of its LIVs
correct inflected form	CIF	A source term is correctly translated into the target, but the translation is neither the reference term nor any of its LIVs. However, the base form of the translation of the source term is identical

		to the base form of either the reference term or one of the LIVs of the reference term
correct reorder form	CRF	a source term is correctly translated into the target, and the translation includes those words that either the reference term or one of the LIVs has, but the word order of the translation is different to that of the reference term or one of the LIVs
correct reorder and inflected form	CRIF	this class is a combination of both CIF and CRF
other correct	OC	a source term is correctly translated into the target, whose category, however, is beyond the all remaining correct categories.

Table 12. Correct terminology categories described in Haque (2019, pp. 439-440). Table is created according to the original definitions in the paper.

Having established a firm classification for correct and incorrect terms, the authors proceed to compare English-Hindi and Hind-English PBSMT and NMT systems trained on judicial corpus. Their systems are evaluated and annotated by two human evaluators. They report that their "NMT systems commit fewer lexical, reordering and morphological errors than the PB-SMT systems" (p. 444). Since Hindi is a morphologically rich language like Turkish, a terminology error annotation task in Turkish-English language pair can utilize a similar typology.

The recent studies of both Scansani et al. (2019) and Haque et al. (2019a, 2019b) demonstrate that there is a growing interest in analyzing more deeply the terminological quality of machine translation systems since terminology plays a crucial role for accurate translation. Particularly, the comprehensive terminology error categorization and the terminology-focused MT quality evaluation by Haque (2019a, 2019b) provides a novel methodology for investigating terminology in machine translation especially from computer science perspective. With this in mind, we create an error categorization adapted from the above-mentioned methodology by considering the perspective of terminology discipline with the aim of reflecting the terminological translation challenges.

Finally, it should be noted that term error annotation is preceded by a source and reference target term annotation ideally to be conducted automatically. However, as Inkpen et al. (2016) notes the resulting term quality may not be always be high and a validation step may be needed. Especially morphologically rich languages such as Turkish pose problems for term annotation or extraction (Hakkani-Tür, 2018; Li et al., 2016). Considering these challenges, researchers such as Haque et al. (2019) resort to manuel annotation with the asssitance of extracted candidate terms. However, deciding on the implementation of manual or semi-automatic annotation strategy depends on the available resources such as bilingual glossaries (like in the case of Haque et al., 2019) and limitations such as time, cost and human annotator availability. Acronym annotation is also problematic and while there are studies such as Islamaj Doğan et al. (2014) to annotate acronyms and their long forms, a system to be used in the workflows of translators or translation scholars is not readily available.

2.6. Turkish Terminology and the Medical Domain

The earliest comprehensive terminology efforts in Turkey began in 1920s and 1930s in parallel to the efforts to modernize Turkish language as a whole (Cengizhan&Tanı, 2010) right after the foundation of the Turkish Republic in 1923. Language planning was one of the key agendas of the newly founded republic. Both the common language and the technical and scientific language of the time were dominated by Arabic, Persian, and French words. In 1932, Society for Research on the Turkish Language ("Türk Dili Tetkik Cemiyeti"), which was later called Turkish Linguistic Society ("Türk Dil Kurumu"), was established as the only responsible authority to replace these words with "pure" Turkish words both through the use of available Turkish words as well as neologisms based on the rules of Turkish (Eker, 2013; Cengizhan&Tanı, 2010). One of the commissions of this new society was called Dictionary and Term Branch ("Sözlük ve Terim Kolu") and conducted lexicographic and terminographic works. Considering the methodological differences of these two fields, later, in 1934, they were divided into 2 different commissions (Eker, 2013). Term commission examined many scientific disciplines including philosophy, physics, mathematics, biology, extracted Arabic, Persian, French and other foreign language terms and suggested new Turkish terms. The "(re)turkification" movement, as coined by (Karaman, 2014, p. 10), developed different methods to suggest new terms including using colloquialism (borrowing words from colloquial language and coining scientific terms with them), borrowing terms from Turkic dialects such as Uzbek, Turkmen; reviving archaic Turkish words from old texts; compounding; literal translation from Arabic and Persian and other methods (see the cited work).

Eker (2013, p. 76) notes that until 1960s, the objective of the commission was only to find new Turkish terms for foreign terms in the language. Namely, terminology was understood as a nomenclature creation work and the scientific approach to analyzing concepts and their relations was lacking. After 1960, the commission continued to coin new terms and publish systematic terminology works. During this "(re)turkification period", the influence of Arabic and Persian terms significantly decreased but French and later English words have kept entering into the language. Karaman (2014), on her historical corpus study on geography terms, finds that while 94% of the terms were Arabic terms and only 3% were in Turkish in 1937, Arabic terms decrease to 10% and Turkish terms increase to 70% in 1985. In the same period, she states that terms from western languages increase from 3% to 10% in geography. This trend seems to repeat in other disciplines as well. Today there is an increasing amount of terms entering into Turkish language from English and the commission (under different administrative names) and Turkish Linguistic Society as a whole still recommend new terms through different dictionaries. Besides, there are scientific and professional associations (such as engineering and medical associations), universities and other state bodies (such as Turkish Standards Institute) which are also publishing terminology dictionaries based on various term construction principles. While some adopt a purist, prescriptive Turkish term coining strategies, others prefer to be descriptive and document the current use of the terms without suggesting neologisms (see the two cardiology dictionaries, one purist and the other descriptive). Because of this situation, as Eker (2013) highlights, there is a lack of synchronization and standardization, which in turn, lead to the proliferation of term variations (synonyms).

2.6.1. Medical Texts and Medical Terminology

This section approaches the specifics of medical domain, medical terminology, and the characteristics of Turkish medical language. Medical texts convey knowledge about latest medical practices, procedures, diseases among other things. The intensive use of Greek and Latin rooted words, prefixes and suffixes differentiates medical texts from other texts. As Guner (2020, p. 342) states, the dominance of Greek and Latin in medical terminology is caused by the internationalization need of medicine:

> Medical texts are far from cultural aspects; because they particularly explain general facts, cases and processes which affect people throughout the World equally. Using an international language including Graeco-Latin terms facilitates the professional process for both writers and readers.

The density of Greek and Latin terms varies according to the subtypes of medical texts. Montalt (2011, p. 81) groups medical genres under four categories in his social function-oriented distinction: research, professional, educational, and commercial.

2.6. Turkish Terminology and the Medical Domain

Research-based medical texts are "those used by researchers to communicate their findings and arguments: original articles, case reports, doctoral theses, etc." (p. 81). Professional medical texts are used in hospitals and other medical contexts (for example, in doctor-to-patient communication). Educational medical texts can be aimed at other medical experts or general public and hence their terminology use may differ. Finally, commercial medical texts may also be targeting other medical professionals and/or general public; their registers will vary accordingly.

Another interesting categorization of technical texts (including medical text) is made by Newmark (1988). The author argues that a technical text can be professional, academic, or popular. Professional medical texts may include document types used in hospital contexts; academic medical context may contain journal articles and life science books and popular medical texts may cover documents prepared by healthcare professionals to inform the general public. The two classifications presented here are both helpful for analyzing the density of medical terms in different medical texts.

Montalt (2011) observes that medical terminology is highly dynamic and neologism is very common thanks to the new discoveries, diseases, and procedures. For example, the pandemic outbreak of 2019-2020 has brought "coronavirus disease (COVID-19)"[70] term. Another remarkable characteristic of medical terms is the widespread variation between the academic/professional terms and their popular counterparts. Montalt (2011) gives the example of "cephalalgia" in specialized use, and "headache" in popular use. Erten (2007, p. 68) also draws a distinction between "medical language" and "daily language". According to her, "heart attack" term is an example of daily use while

[70] Coronavirus. https://www.who.int/health-topics/coronavirus#tab=tab_1 (last access: 02.04.2020)

"myocardial infarction" is the more specific medical language example; the corresponding Turkish translation are "kalp krizi" and "miyokard enfarktüsü".

Turkish medical terminology also includes lots of Greek and Latin rooted terms. Erten (2007, p. 49) reports that until 1927, Persian, Arabic and Turkish rooted medical terms were used. After this date, instead of Arabic alphabet, Latin alphabet started to be used. This, coupled with the influence of French culture and translations from French, facilitated the adoption of Greek and Latin rooted international medical terminology. Paradoxically, since that time, there have been struggles to create "pure Turkish" medical terminology (see the previous section). For example, Turkish Cardiology Society (TCS) has a terminology guideline on cardiology[71] which recommends the use of Turkish terms instead of Latin or Greek rooted terms. For instance, the term "cardiomyopathy" is usually translated as "kardiyomiyopati" in Turkish; TCS suggests using "yürek kası hastalığı" (lit. "heart muscle disease"). From our observation, both types of Turkish terminology (pure Turkish and nonpure Turkish terminology) are simultaneously used in medical texts, which leads to increased variation in terminology. A certain level of variation may be considered necessary for preserving ethnological points of view and diversity of natural languages and their conceptual systems. However, since MT systems are mostly based on frequency, they have the potential to reflect the consensus of the majority (in case they are trained on sufficiently representative parallel corpora).

We have already studied the common terminology challenges for translators in the previous section 2.2. We will briefly mention how these challenges affect Turkish

[71] Kardiyoloji (Yürek Bilimi) Türkçe Terimler Kılavuzu

medical language and its translation from and into English. Firstly, English and Turkish acronyms of the same medical concepts are used simultaneously in Turkish medical texts. Aslan et al. (2004) lists some examples: tr. “büyüme hormonu” (en. “growth hormone”) can be seen as “BH” and “GH” in Turkish texts. The same phenomenon also occurs with the simultaneous use of the acronyms for Greek and Latin rooted terms and their pure Turkish synonyms in Turkish. The proliferation of synonyms and their acronyms decreases standardization and deteriorates effective medical communication. Aslan et al. (2004, pp. 18-19) states that common rules for abbreviating medical terms is not available and suggests 14 non-binding principles to guide the use of abbreviation in medical texts. Another challenge is related to analogies. There are many anatomical terms constructed through analogy.

Turkish medical term	**Analogical component**	**English term**
kürek kemiği	kürek = shovel	scapula
göğüs kafesi	kafes = cage	thorax
kaval kemiği	kaval = pipe	tibia
omurilik soğanı	soğan = onion, bulb	spinal bulb

Table 13. Turkish terms constructed by analogy.

Not being aware of the equivalents of these terms may lead to inaccurate translations. As mentioned in section 2.2, target terms may have different analogical components or no analogy at all. In case only the Turkish analogical component is used in a sentence (for example, “soğan” from the table above), translator may need to compensate it with the component of the full terminological collocation (“omurilik soğanı”) and prefer to use the equivalent term in full form.

There are various studies that concentrate on collocations in Turkish language (Eken, 2016; Ozkan, 2012; Metin&Karaoglan, 2010; Sarikas, 2006). However, these studies cover collocations in the general Turkish language and none of them focus on terminological collocations in medical texts. The general grammatical categorization of collocations by Sarikas (2006) can be adapted to specify medical collocations in Turkish. The formal structure of medical collocations consists of noun + noun or adjective + noun phrases. The number of adjectives or nouns may increase but adjective always comes before the noun. For example, it is possible to make adjective + noun + noun or adjective + adjective + noun collocations. Turkish language does not have determiners or gender. Syntactically, collocations come together in a sentence without any auxiliary connectors such as of "of" in English. The last noun in the collocation is the head of the phrase and may include suffixes depending on the sentence.

Structure	**Turkish Example 1**	**English Equivalent**	**Turkish Example 2**	**English Equivalent**
noun + noun	kalp yetersizliği	heart failure	sinüs ritmi	sinus rhythm
adjective + noun	derin bradikardi	profound bradycardia	yüksek lipoprotein	high lipoprotein
adjective + noun + noun	akut böbrek hasarı	acute kidney injury	sağ kalp fonksiyonu	right heart function
adjective + adjective + noun	ani kardiyak ölüm	sudden cardiac death	koroner yavaş akım	slow coronary flow

Table 14. Turkish collocations in different grammatical categories.

Based on the above-mentioned considerations, it can be argued that understanding the medical terminological collocations and their structures, analyzing their components for the presence of any culture specific analogies, and identifying their target equivalents may be key competences for translators.

As in other languages, neologisms are common in Turkish medical terminology and as Guner (2020) highlights in detail, the usual method of adopting a term is borrowing:

2.6. Turkish Terminology and the Medical Domain

> Medical terms in Turkish medical texts, especially in academic and professional ones, enter the Turkish language system by the method of transcription/borrowing. This transcription is mainly phonetic, and phonetic features of the source language are kept while transcribing the term into Turkish. As well as transferring phonological characteristics of a source term, the newly coined term is adapted to the orthographic features of the Turkish language. (p. 341)

When translators encounter a terminological collocation which is not available in medical terminology dictionaries, they also apply this method and contribute to the nomenclature of the medical domain. This method seems to be less risky than creating pure Turkish terms without any validation from medical professionals. Yet, it increases the amount of borrowed terms in the Turkish medical language, which contradicts the historical struggle for the turkification of the scientific language.

Yazıcı (2018) pinpoints an interesting dilemma about Turkish medical translation. There is a growing demand for Turkish to English translation for international journals. However, there are not enough medical translators who are native in English and translate from Turkish. Hence, translations are performed by Turkish natives, which is usually not a preferred strategy in professional settings. The author suggests that while this gap can be narrowed by increased specialization of translators, there is also a need for division of labor and inclusion of native English reviewers whose linguistic knowledge is essential for the translation quality. We believe that medical domain specific Turkish to English MT systems can also help to narrow this gap.

2.6.2 Cardiology Terminology

In this thesis, we create a cardiology-based parallel corpora for the purpose of MT training and terminology quality evaluation. As heart related disorders are the main causes of death all over the world, we considered rapid information exchange through machine translation about cardiology to be an important part of the struggle for the treatment of these disorders. There are different disorders related to heart functions.

2.6. Turkish Terminology and the Medical Domain

Having a map of cardiology as a domain can help to build a complete specific corpora both automatically and semiautomatically.

Cardiology is an academic discipline and the branch of medicine that focuses on the disorders of human heart. It deals with subjects such as anatomy of the heart, heart

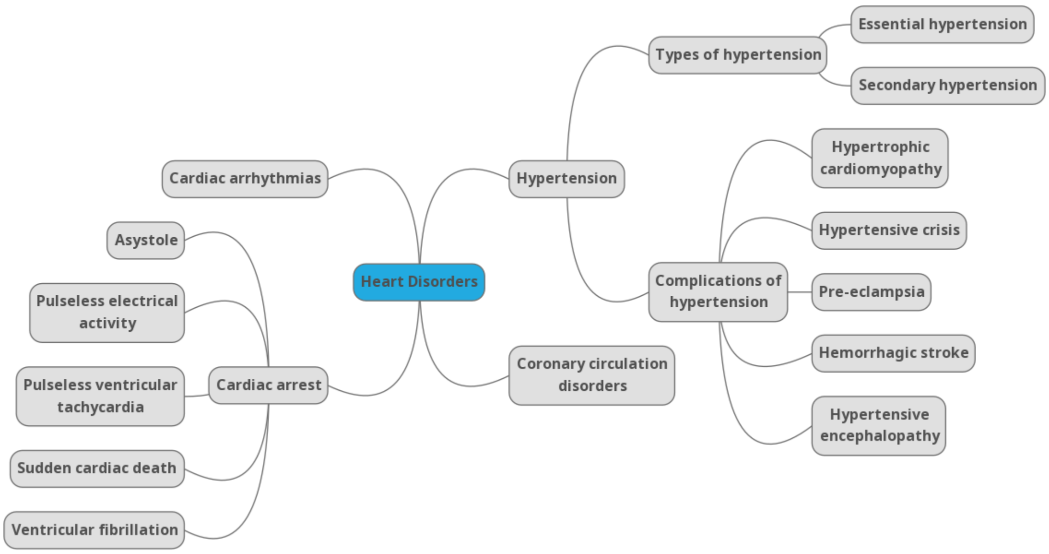


Table 15. A tree diagram showing the conceptual relationships between heart disorders

disorders, physical examination techniques of the heart, and procedures and medications to treat heart disorder. Each of these subjects has their own specific concepts with specific relationships between them.

The table above shows different levels of conceptual relationships between heart disorder related cardiology concepts. These concepts are denoted with specific terms in English and other languages. A more detailed bilingual version of the diagram is provided below.

1) **Hypertension**; **Hipertansiyon**
    a. Types of hypertension; Hipertansiyon türleri
        i. Essential hypertension; Esansiyel hipertansiyon
        ii. Secondary hypertension; Sekonder hipertansiyon
    b. Complications of hypertension; Hipertansiyon komplikasyonları

i. Hypertrophic cardiomyopathy; Hipertansiyon komplikasyonları
ii. Hypertensive crisis; Hipertansif kriz
iii. Pre-eclampsia, Eclampsia; Preeklampsi, Eklampsi
iv. Hemorrhagic stroke; Hemorajik inme
v. Hypertensive encephalopathy, hypertensive nephropathy, hypertensive retinopathy; Hipertansif ensefalopati, hipertansif nefropati, hipertansif retinopati

2) Cardiac arrhythmias; Kardiyak aritmiler

a. Atrial fibrillation; Atriyal fibrilasyon
b. Atrial flutter; Atriyal çarpıntı
c. Heart block; Kalp bloğu
d. Long QT syndrome; Uzun QT sendromu
e. Premature atrial contractions; Prematüre atriyal kasılmalar
f. Premature ventricular contractions; Prematüre ventriküler kasılmalar
g. Sick sinus syndrome, Bradycardia-tachycardia syndrome; Hasta sinüs sendromu, Bradikardi-taşikardi sendromu
h. Supraventricular tachycardia; Supraventriküler taşikardi
i. Torsades de pointes; Torsades de Pointes
j. Ventricular fibrillation; Ventriküler fibrilasyon
k. Ventricular tachycardia; Ventriküler taşikardi

3) Coronary circulation disorders; Koroner dolaşım bozuklukları

a. Atherosclerosis; Ateroskleroz
b. Coronary artery disease; Koroner arter hastalığı
   i. Acute coronary syndrome; Akut koroner sendrom
   ii. Angina pectoris; Angina pektoris
   iii. Myocardial infarction; Miyokardiyal enfarktüs
c. Restenosis; Restenoz

4) Cardiac arrest; Kardiyak arrest

a. Asystole ("flatline"); Asistol ("düz çizgi")
b. Pulseless electrical activity; Nabızsız elektriksel aktivite
c. Pulseless ventricular tachycardia; Nabızsız ventriküler taşikardi
d. Sudden cardiac death; Ani kardiyak ölüm
e. Ventricular fibrillation; Ventriküler fibrilasyon

Table 16. Some of the heart disorder types given in English and Turkish.

As it can observed from the table above (detailing some of the heart disorders) that just like in other medical domains, Greco-Latin rooted terms are common and are usually in terminological collocation form in cardiology. The translations of the terms into Turkish mostly maintain these Greco-Latin roots and are borrowed and either neutralized (for instance, en. “atherosclerosis” / tr. Ateroskleroz) or left the same (for instance, en. “angina pectoris” / tr. “angina pectoris”. While experts in Turkish Society of Cardiology advocate the use of Turkish term equivalents and have published

2.6. Turkish Terminology and the Medical Domain

Kardiyoloji (Yürek Bilimi) Türkçe Terimler Kılavuzu (2009) (Guide for Cardiology Turkish Terms), many cardiologists continue to use borrowed terms (as in the table above). Those selected terms allowed us to navigate and get our bilingual corpora, and they were also used to observe and measure term translation by machine translation engines.

Having established the theoretical and conceptual framework of our thesis, we will explain our research methodology of corpus preparation, machine translation and evaluation, and terminology annotation and evaluation in MT in the following 2 chapters. Chapter 3 explains the parallel corpora preparation process and machine translation training and evaluation phases. Chapter 4 focuses on the methodology that we developed for annotating and evaluating terminology errors.

Part II. Research methodology and design

Chapter 3. Corpus Design and Machine Translation Customization and Evaluation

"Language is a process of free creation; its laws and principles are fixed, but the manner in which the principles of generation are used is free and infinitely varied. Even the interpretation and use of words involves a process of free creation."

Noam Chomsky, 1968

"Every linguistic sign is located on two axes: the axis of simultaneity and that of succession."

Roman Jakobson

2.6. Turkish Terminology and the Medical Domain

In this chapter, we are going to explain the methodology that we used for training and evaluating 4 specific-domain Turkish → English MT systems: 2 SMT engines and 2 NMT engines. We conducted a multifaceted comparison between these 4 MT engines. A custom parallel corpora compiled from cardiology journal abstracts has been developed and used in the training tasks together with a corpus of mixed domain (PBMT Engine 1: cardiology corpus; PBMT Engine 2: cardiology corpus + mixed domain corpus; NMT Engine 1: cardiology corpus; NMT Engine 2: cardiology corpus + mixed domain corpus). Chapter 3 covers the process of corpus preparation and machine translation customization. The trainings were performed using a state-of-the-art SMT and NMT platform, KantanMT, and automatic evaluation scores for each engine were reported immediately after the trainings in the form of F-Measure, BLEU and TER scores. After this step, we passed to human evaluation, and professional translators conducted a side-by-side comparison of the four engines, ranked them according to their qualities and gave them adequacy and fluency scores in KantanMT LQR dashboard. This step has provided us the chance to check if there is a correlation between automatic evaluation and human evaluation.

Once this global evaluation is completed, we perform a comparative qualitative evaluation on terminology translation in SMT and NMT. For this evaluation, we benefited from and extended the methodology suggested by Haque et al (2019) and firstly annotated the terms in both source and target sides of our test corpus. Then these sentences are translated using 4 engines. The translation qualities of these terms are manually evaluated based on the categories suggested by Haque (2019), and the encountered errors were annotated separately for each engine. Once the annotation is completed, terminological qualities of each engine were reported quantitatively in relation to the error types and their frequency. Below we will explain our methodology in detail. Figure 9 summarizes the four main methodological steps followed in the study.

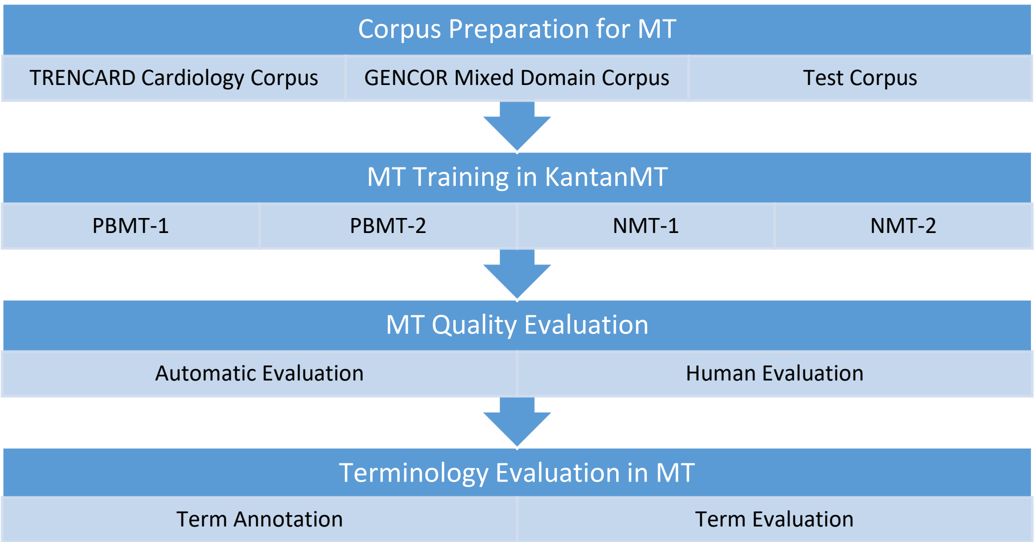


Figure 9. An overview of the four main methodological steps.

3.1. Corpus Preparation for Machine Translation

There are three different bilingual corpora complied for MT training and evaluation in this study: i). a custom specific domain (cardiology) corpora built by us following the corpus preparation strategies. ii). a mixed domain corpora acquired from a free and open parallel corpora repository and iii). a small, specific domain (cardiology) corpora for the purpose of general MT evaluation. The first and third corpora are compiled from Turkish and English abstracts included in four different academic Turkish cardiology journals and have highly specialized medical language. The cleaned total size of the first corpus is 49693 aligned sentences and 788046 source words. Hereinafter, this corpus will be referred to as TRENCARD (Turkish English Cardiology) corpus. The second main corpus is a collection of 8 corpora from different domains in Opus Project. This mixed domain corpus has 5,668,129 source words and 381,322 sentences. Hereinafter, this corpus will be referred to as GENCOR. Lastly, the third corpus which we call Test Corpus has 11015 source words and 677 source sentences. While the preparation of TRENCARD and Test Corpus involved all the corpus preparation phases, the preparation of GENCOR only included corpus selection/curation phase. Below we firstly explain how we prepared TRENCARD, and its features. Secondly, we briefly explain the features of GENCOR, and our design procedure. Finally, we provide details of our Test Corpus.

3.1.1. TRENCARD: Turkish - English Parallel Corpus from Cardiology Journals

There is no free and open Turkish English medical parallel corpus available for use in Opus Corpus or in the Web. While European Medicine Agency (EMEA) corpus[72] is available in official European languages in OPUS, it does not have Turkish. Hence, we created the cardiology domain corpus that we named TRENCARD. The design of TRENCARD has required the use of several tools and methods. A semi-automatic methodological procedure is followed as described in the second chapter. The objective has been to create a clean parallel corpus which is in a very specific domain (cardiology) and is large enough to be used in the machine translation training tasks. Corpus preparation starting from the selection decisions of the resources till the compilation of the corpus are explained below. There are four cardiology related journals with permissive licenses that we crawled: *Archives of the Turkish Society of Cardiology*[73], *Turkish Journal of Cardiovascular Nursing*[74], *Turkiye Klinikleri Journal of Cardiology*[75] and *Turkish Journal of Thoracic and Cardiovascular Surgery*[76]. We describe how we prepared the corpus from each journal. Although a similar procedure

---

[72] European Medicine Agency (EMEA) corpus: http://opus.nlpl.eu/EMEA.php (last access: 08.01.2020)

[73] https://www.archivestsc.com/about-the-journal (last access: 10.01.2020)

[74] http://khd.tkd.org.tr/EN/about (last access: 10.01.2020)

[75] https://turkiyeklinikleri.com/journal/kardiyoloji-dergisi/1300-0314/identity/en-index.html (last access: 10.01.2020)

[76] http://tgkdc.dergisi.org/static.php?id=2 (last access: 10.01.2020)

was implemented with each journal, there are some differences in parsing or cleaning phases. Below these differences are described in detail. In the following paragraph we explain why we selected these journals, and content type we extracted.

Since we aim to compare the terminological quality of machine translation systems, we chose to work with a very specific domain instead of a more general domain like medical domain. In a narrower domain, terminology tend to be used more consistently. The translation of a term used in cardiology may be translated differently in oncology. Besides, as Koehn (2009, p. 54) observes, "[r]estricting the domain simplifies the machine translation problem dramatically". However, when the domain is very specific, the biggest challenge is to build a parallel corpus big enough for the training. Hence, we crawl more than one journal website.

Abstracts in academic cardiology journals were chosen based on corpus size, reliability of translation, existence of structured data for crawling and terminological density. Another important aspect in compiling a corpus from the web is related to licenses. The journals that we crawled for the TRENCARD corpus had the license of Attribution-NonCommercial 4.0 International (CC BY-NC 4.0)[77]. This license allows for sharing and adapting content under the terms of appropriate credits and noncommercial use. Abstracts in Turkish medical journals are generally bilingual and are in Turkish and English. The remaining parts of the journal issues are usually monolingual. For this reason, we opted to build our TRENCARD corpus from abstracts. Nonetheless, abstracts have typically between 100-300 words and are short texts, which require the compilation of large number of abstracts. An advantage of using abstracts is that their

[77] Attribution-NonCommercial 4.0 International (CC BY-NC 4.0). https://creativecommons.org/licenses/by-nc/4.0/

content is very structured. A cardiology abstract starts with a section called "background" or "objective", then includes "methods", "results", "conclusion" and "keywords". This structure allows for the creation of a specific domain better performing machine translation engine. And most importantly, since word count is limited in abstracts, more terminology is used to express the content in a minimum number of words, which leads to terminological density. Considering that these four scientific journals are authoritative in their fields and are published openly, the translations of the abstracts are of high quality. In combination, these four cardiology journals reflect the last 30 years of cardiology study in Turkey. The oldest online issues are from 1990. And the content of English and Turkish abstracts represents the terminology used in this field.

Corpus Preparation from Archives of the Turkish Society of Cardiology. *Archives of the Turkish Society of Cardiology* "is a peer-reviewed journal that covers all aspects of cardiovascular medicine"[78], is published in Turkish and English, and "accepts papers on a wide range of topics, including coronary artery disease, valve diseases, arrhythmias, heart failure, hypertension, congenital heart diseases, cardiovascular surgery, basic science and imaging techniques"[79]. The journal is published by the Turkish Society of Cardiology, which "is the scientific, nonprofit, nongovernmental organization of Turkish cardiologists, established on May 21st, 1963. Its 2360 members cover almost all the academicians and practitioners of cardiology and the related specialists in Turkey."[80] The online archive of the journal includes the period between

---

[78] https://www.archivestsc.com/about-the-journal (last access: 13.01.2020)

[79] https://www.archivestsc.com/about-the-journal (last access: 13.01.2020)

[80] Turkish Society of Cardiology. https://tkd.org.tr/en/menu/10/history (last access: 13.01.2020)

1990-2020. As of 2005, each volume includes 8 issues and a varying number of supplements. Issues have editorial section, original articles, case reports, case images and a section called perspectives.

The website of the journal is in Turkish and English, and access to the contents of the articles is open. Besides, the website has a very well-structured HTML that allows the processing of the pages easily. Using the main URL of the website (https://www.archivestsc.com/), we downloaded all the website into our computer by the default workflow of HTTrack Website Copier. All the items inside the website are downloaded including HTML files, style sheets, images, PDFs etc. Since folder structure is maintained, all English abstracts and Turkish abstracts are included in separate folders. We deleted all the unnecessary files and only left the HTML pages which include each abstract (there is one abstract in each HTML page). For each language, there happen to be around 390 pages. However, we realized that file names for each page was changed in a way that makes it hard, if not impossible, to align files. The files go with the names such as “jvi0a30”, “jvi0aa5” etc. Normally, on the website, each HTML page is hosted under a URL structure such as “website name + /jvi.aspx?un=TKDA-72699”. Here the number after the “TKDA” attributes to the unique identity of each page. As we were not able to align the files properly in this setting, we looked for another website crawler and found WGET, the description of which is given in Chapter 1. WGet works on Ubuntu operating system. Since we work in a Windows OS, we downloaded Oracle VirtualBox to be able to work on Ubuntu. After opening the terminal on Ubuntu, we typed this code to download all the website again:

```
wget --mirror -p --convert-links --content-disposition --trust-server-names -P TurkishCard https://www.archivestsc.com/
```

3.1. Corpus Preparation for Machine Translation

The name “TurkishCard” is given for the folder that will include all the files and subfolders. Similar to HTTrack Website Copier all content including javascript files, style sheets, images and other elements are downloaded. And as before, folder structure is maintained. Yet, different from the previous download, file names are kept as they are in the website. After the completion of the download, we moved the folder back to Windows. Then, we removed all the unnecessary files again and only left the files with abstracts. Each page including Turkish abstract has the names such as “jvi.aspx_pdir=tkd&plng=tur&un=TKDA-00090” in which the section of “jvi.aspx_pdir=tkd&plng=tur&un=” is the same in all the Turkish abstracts. In the English abstracts, the names have structure of “jvi.aspx_pdir=tkd&plng=eng&un=TKDA-24582” where “jvi.aspx_pdir=tkd&plng=eng&un=” is standard in all abstracts. This consistent pattern has let us use batch tasks. Windows 10 has an advanced command-line terminal called Power Shell[81] through which it is possible to batch rename files. In order to simplify the handling of the files and match Turkish and English files, we used the Power Shell terminal for renaming by using the following command:

```
PS D:\Academia\2019 - 2020 Thesis Completion Phase\PhD\Chapter 7_Methodology\Ready Corpora 2019\1. TKDA Journal\v2_TKDA Journal\www.archivestsc.com\uzun en> Dir | Rename-Item -NewName {$_.name -replace "jvi.aspx_pdir=tkd&plng=tur&un=","tr-"}
```

Through this command, we replaced “jvi.aspx_pdir=tkd&plng=tur&un=” with “tr-”. Hence, the file names are abbreviated to short names such as “tr-TKDA-0900” which is easier to handle. We repeated the same process for the English abstracts. We realized that abstracts are repeated two times in both folders. We deleted the repeated ones. Therefore, we have around 3920 abstracts. There was a difference of 8 abstracts

[81] Windows Power Shell. https://docs.microsoft.com/en-us/powershell/scripting/getting-started/getting-started-with-windows-powershell?view=powershell-7 (last access: 13.01.2020)

between Turkish abstracts folder and English abstracts folder. Since it is a time-consuming task to find the non-matching files, we changed the names of Turkish abstracts temporarily from “tr-” to “en-” so that they became equal to the English ones in terms of name, and we used a software called AllDup 4.2 to identify and delete the non-matching files. After finding and deleting the non-matching files, we renamed the Turkish abstracts again. We ended up having 3918 files for each language, in other words, 3918 abstracts for Turkish and English ready. With this step, we complete the file alignment process and pass to the sentence by sentence alignment process. The side by side comparison of the folder with Turkish files and the folder with English files is displayed in Figure 10.

After the file selection and alignment steps, we passed to the sentence level alignment.

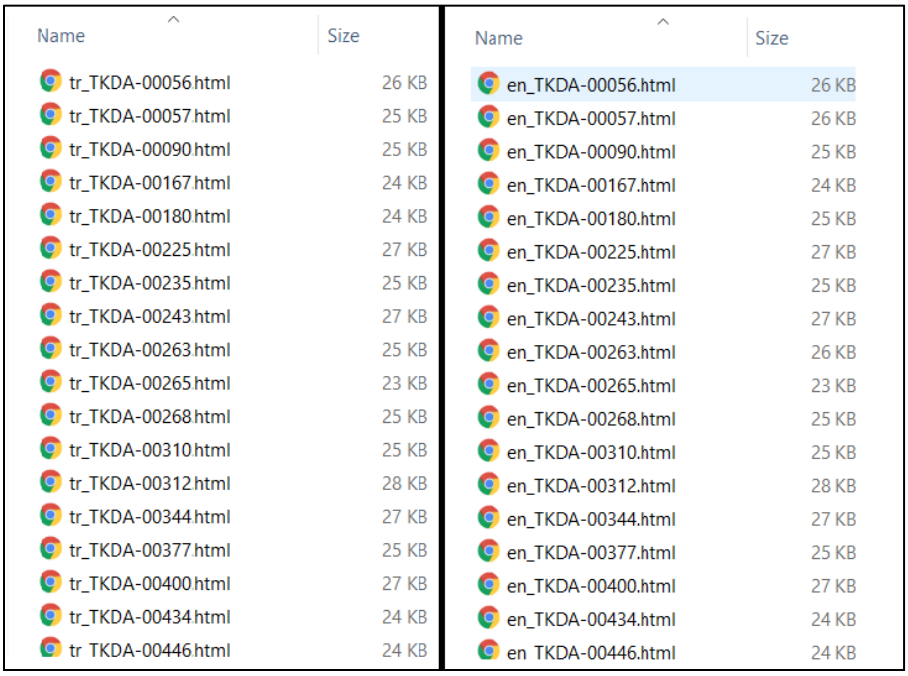


Figure 10. Folder containing Turkish abstract files (on the left side) and English abstract files (on the right side)

But before we need to analyze the files and decide on the most efficient strategy for parsing the files since we only need the abstracts in each page. Manual copy-paste of each abstract into a plain text is very time-consuming[82]; therefore, a strategy for parsing each file and then aligning the files sentence by sentence provides advantages. The free and open source CAT tool OmegaT has an alignment feature. However, it is not possible to batch process files and parse a certain part of a file. HunAlign has required advanced technical skills for achieving our goals. For these reasons, we opted for Memoq's LiveDocs feature in which it is both possible to customize the file parsing filter and include only a certain part of the file, and right after that, align these files directly. And when the automatic alignment is completed, an editor window is opened for editing the misaligned sentences. Finally, when all sentence alignments are confirmed, all the sentences can be imported into a translation memory. To sum up, our steps hereinafter are as follows:

1). Open Memoq's LiveDocs feature and Create a new corpus

2). Add alignment pairs to the new corpus (All files in Turkish folder into one side and all files in English folder into another side)

3). Create a parsing filter for Turkish side and one for the English side

4). Select the correct language encoding

5). Start the filtering and aligning process

6). Check alignment editor for any mismatch

7). Import all sentences into a translation memory

8). Export this translation memory in TMX format

[82] In fact, in the first year of the doctoral studies, we tried to manually copy each abstract from the website and paste it directly to a plain text file. Our speed was half or one volume each day for 3 days. Soon we realized that it is not an efficient way of corpus preparation, and have started to look for automatic methods.

When an HTML file is imported into Memoq, the program uses its default HTML filter[83].

i). Parsing the Turkish abstracts. All text content (menu items, website related general texts etc.) will be imported with this default filter. However, this can be changed, and either other default filters can be applied, or custom filters can be created using regular expressions to extract only a specific part of a file. Besides, more than one filter can be used to filter content from the code selectively. These consecutive filters are called cascading filters. Our cascading filters have a regex text filter (to extract only the relevant content) and a HTML text filter (to correctly visualize HTML tags in the file). In order to configure this filter for extracting the English and Turkish abstracts, we analyzed a few HTML files. Note that for this filter to work on 3918 files, the structural pattern shall be the same in all of them. Figure 11 shows the HTML structure of the file.

[83] We use the term “filter” with the same meaning as “parser”.

3.1. Corpus Preparation for Machine Translation

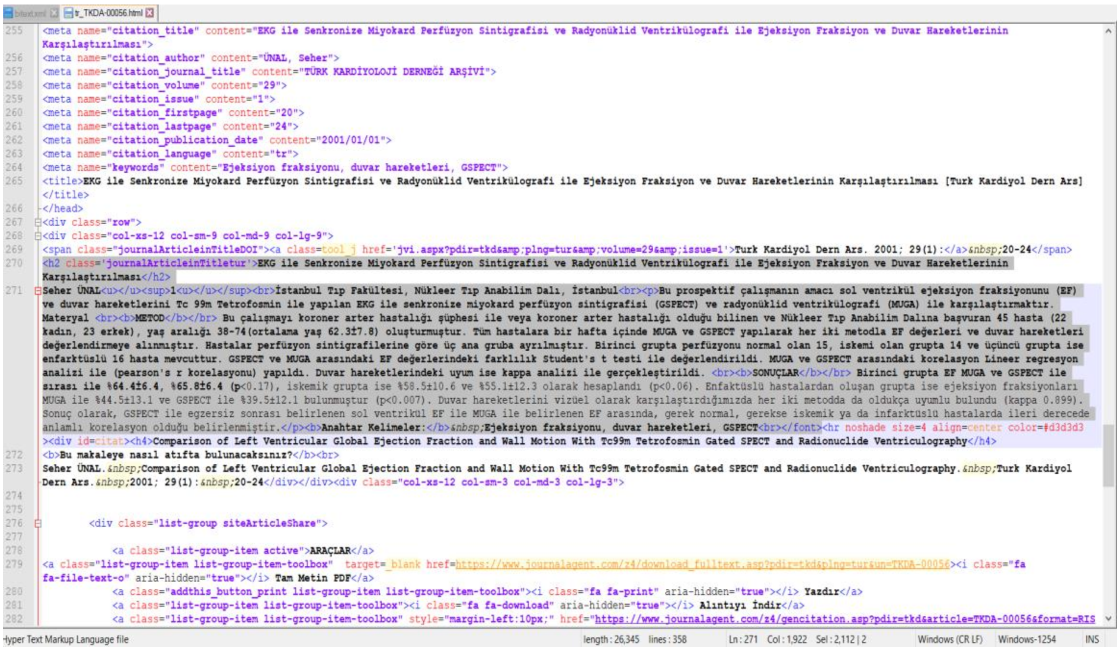


Figure 11. HTML structure of one of the files including Turkish abstract as displayed in Notepad++ Editor. Grey area is the intended area to be extracted and the rest shall not be imported.

Our regex filter was able to filter the relevant text from the file. All abstracts had the same title structure between “<h2 class='journalArticleinTitletur'>” and “</h2>” tags. Secondly, content between “
<p>” and “<hr noshade size=4 align=center color=#d3d3d3>” was desirable. In other words, we wanted to filter the content of the abstract till the end of the keywords. Consequently, we had 2 regex rules for our first filter, which are as follows:

Import only the following content:
1- <h2 class='journalArticleinTitletur'>.*</h2> (Import everything between these two tags)
2-
<p>.*<hr noshade size=4 align=center color=#d3d3d3> (Import everything between these two tags)

The “.*” regular expression symbols mean “every character”. As it can be observed from Figure 11, there are some inline tags that remain in between the extracted content which is perceived as plain text such as “
METOD</br>”. These characters both affect the sentence segmentation during alignment, and then machine translation quality during training. Hence, they need to be removed in the following steps. To facilitate this process, we applied a second filter (HTML filter) over the extracted content to recognize these characters as HTML tags. This allows for removing HTML tags automatically in a tool with such a feature. The last important consideration in this step is the selection of the correct encoding for the language. Windows encoding (Windows-1254) provided the correct character set. And after the application of the cascading filter (regex text filter + HTML filter), we ended up having the segmented, clean Turkish abstract. Figure 12 shows how the filtered Turkish abstract and the English abstract look in LiveDocs Alignment Window.

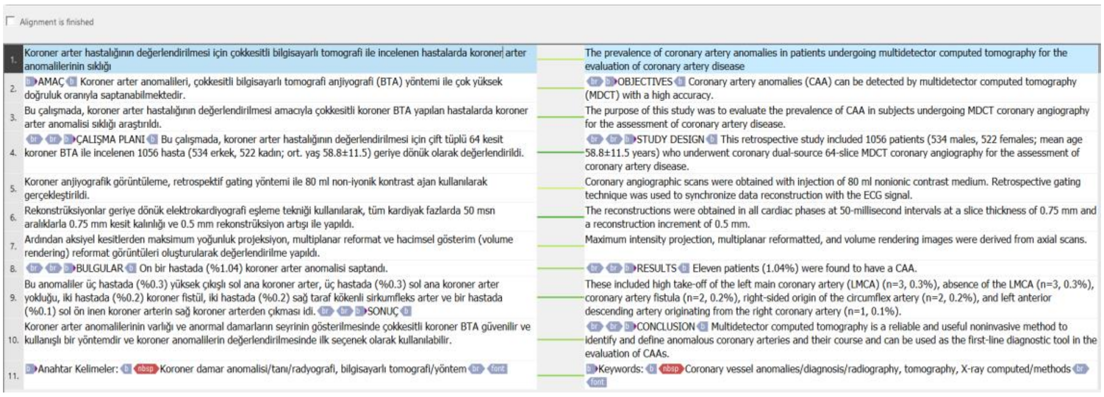


Figure 12. Alignment of source and target files in Memoq. Purple and red colored contents are HTML tags.

ii). Parsing the English abstracts. We followed a similar procedure for the English abstracts. Again, we used a cascading filter configuration for parsing. Only the first rule of the regex filter was slightly different:

Import only the following content:
1- <h2 class='journalArticleinTitleeng'>.*</h2> (Import everything between these two tags)
2-
<p>.*<hr noshade size=4 align=center color=#d3d3d3> (Import everything between these two tags)

Note that above only "<h2 class='journalArticleinTitleeng'>" tag is different in the filter. Another difference is with the encoding. We selected Western European (Windows) encoding for the English files. Once the parsing filters are set for both languages, we initiated the alignment step. Figure 12 shows how two files are aligned in Memoq. In this case, there are no misalignments. However, there may be some misalignments that need to be edited in some cases. After making sure with some tests that this parsing and alignment methodology can be used for all the files, we conducted a batch processing for aligning all the 3918 Turkish files with 3918 English files using the corresponding filters. The process is completed with 27279 aligned sentences (segments).

All these sentences are compiled in a translation memory and are exported as a TMX file. In order to remove any tag, special character or any other kind of noise, we further processed this corpus using the cleaning tools. Once we combine all the content of the 3918 Turkish abstracts and 3918 English abstracts within a single TMX file, we can start the cleaning step. As mentioned in the Chapter 1, translation memory maintenance tools such as Heartsome TMX Editor and Goldpan TMX/TBX Editor, and Okapi Framework tools such as CheckMate and Rainbow are useful for implementing batch cleaning operations such as removing inline HTML tags, inconsistent numbers, duplicate sentences etc. Essentially, this process is similar to translation quality assurance.

3.1. Corpus Preparation for Machine Translation

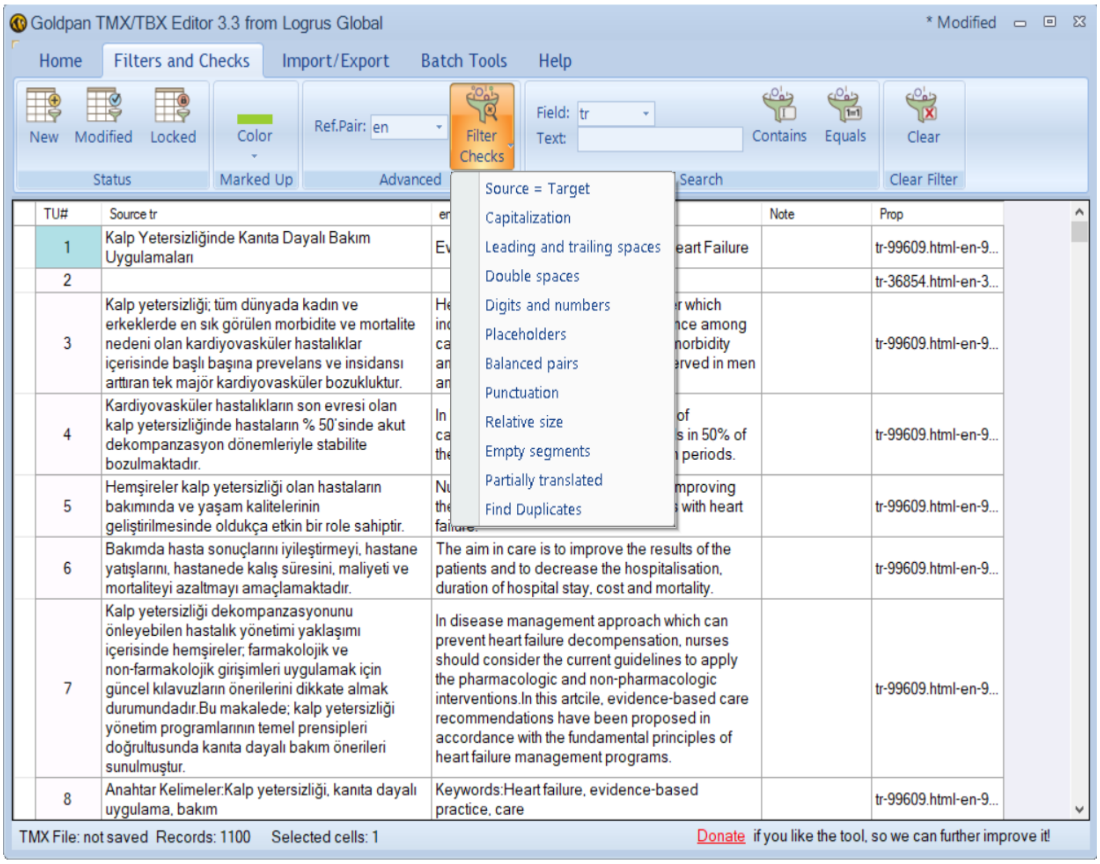


Figure 13. Goldpan TMX/TBX Editor has filter checks that allow 12 different checks to remove inconsistencies. Besides the Home tab includes a function for cleaning the tags in different formats. We used these options to optimize our corpus.

We have only used GoldPan to clean our corpus. When we complete cleaning, we ended up having a corpus in the form of a TMX file. Since KantanMT supports importing TMX files, we do not need to further process our file. In some machine translation tools, the files need to be prepared as two aligned plain text files. For example, older versions of MTradumàtica require two distinct corpus files. Rainbow tool can be used for this operation. It shall be noted that MT toolkits or systems such as KantanMT include internal operations for cleaning the corpus automatically similar to the cleaning procedures mentioned above. We will elaborate on this topic in section 3.3 when we handle the training process.

3.1. Corpus Preparation for Machine Translation

After importing the TMX file into Rainbow, going to Utilities > Conversion Utilities > File Format Conversion > Parallel Corpus Files and executing the command will create an aligned Turkish plain text file and an align English text file with the desired encoding. Although not needed for training purposes, we conducted this step to be able to analyze the files when required. In the table below, we summarize the steps and the tools we used for preparing our first journal corpus.

	Phase	Tools
1.	Crawling & Downloading	WGET
2.	Analysis	NotePad++
3.	Parsing	Memoq (Filters Features)
4.	Aligning	Memoq (LiveDocs Feature)
5.	Compiling	Memoq (Export to TMX feature)
6.	Cleaning	GoldPan TMX/TBX Editor
7.	File Conversion	Okapi Rainbow

Table 17. The tools and steps that we have followed for building the corpus.

Following these steps, we created a corpus of 27279 sentences, 496327 source words with a word/sentence rate of 18.19. Below we display the profile of our corpus together with some meta information about the it.

Journal Name	***Archives of the Turkish Society of Cardiology***
Domain	Cardiology
UNESCO Code	3205.01
Source Word Count	496327
Target Word Count	570082

Sentence Count	27279
Source Word / Sentence Rate	18.19
Target Word / Sentence Rate	20.89

Table 18. The corpus profile of the first journal as calculated by Memoq's Statistics feature.

Lastly, we would like to mention the challenges encountered while preparing this corpus. Our first trials of crawling resulted in randomly (or in a way that does not allow to grasp the naming pattern) named files in HTTrack; to solve this problem, we passed to WGET in Ubuntu. Secondly, while the encoding of the HTML pages is in UTF-8, using this encoding has yielded noisy characters. After trial and error, we found the correct encoding for the corpus. Some translations have not been made sentence by sentence into English, which resulted in misalignments. In order to minimize this, we used terms as anchors in Memoq so that alignment reliability has increased. We derive these terms from the "keywords" section of the abstracts.

We used the procedure explained in the Table 17 for all the 4 journals with some minor changes. Therefore, in the next sections, we will briefly explain preparation procedure by repeating the same steps one by one.

Corpus Preparation from Turkish Journal of Cardiovascular Nursing. *Turkish Journal of Cardiovascular Nursing* is also a journal published by Turkish Society of Cardiology, more specifically by the Cardiovascular Nursing Technicians Working Group. It is "an Open Access, peer-reviewed e-journal that considers scientific research, case reports, reviews, translations, letters to the editor, news and abstracts presented at the National Congress of Cardiology"[84]. The topics covered include "the

[84] http://khd.tkd.org.tr/EN/about (last access: 23.01.2020)

field of coronary artery disease, valvular heart disease, arrhythmias, heart efficacy, hypertension, congenital heart disease and all articles related to the coronary intensive care nursing."[85] The website of the journal, being of the same society, has a similar design to the previous; hence, we utilized a similar procedure for this journal. We crawled the website with WGET, selected the relevant files, analyzed them, and determined the parsing filter that will use and implemented it. We used a cascading filter which has regex text filters and HTML filter just like we did in the first journal. The regex for Turkish abstract includes only one rule:

Import only the following content:
1. <h2>.*

<hr noshade size=4 align=center color=#d3d3d3><h2>

The reason why we used only one rule is that the structure of the HTML allowed for easier parsing. And the above tags are only included once in the file. If they were to be included more than once it can be impossible to use these rules. The example of the fourth journal will illustrate this kind of situation. After the regex text filter, an HTML filter is applied too. The cascading filter for the English abstract include a regex filter and one HTML filter. Regex text includes one 1 rule:

Import only the following content:
1. <h2

<hr noshade size=4 align=center color=#d3d3d3><font color=#515151><h2>

As it can be observed the same parsing filter rule is used for both cases. The only difference between the English and Turkish files are the order of the abstracts are changed. In the first case, Turkish abstract is above while in the second case the English is above. Hence, using the same filter configuration, we were able to filter and align the journal abstracts. After the alignment, we imported the content into translation memory

[85] http://khd.tkd.org.tr/EN/about (last access: 23.01.2020)

and exported it as TMX file. In total, we have 1093 sentences, 17471 source words and 21019 target words.

Journal Name	***Turkish Journal of Cardiovascular Nursing***
Domain	Cardiology
UNESCO Code	3205.01
Source Word Count	17471
Target Word Count	21019
Sentence Count	1093
Source Word / Sentence Rate	15.98
Target Word / Sentence Rate	19.23

Table 19. Corpus profile of the second journal as calculated by Memoq's Statistics feature.

We also followed similar cleaning procedure and removed the tags from the TMX files as well as unnecessary content. This corpus preparation does not have any challenge. The only problem is that it is relatively small. However, its content is useful for the TRENCARD corpus.

Corpus Preparation from Turkiye Klinikleri Journal of Cardiology. *Turkiye Klinikleri Journal of Cardiology* is another cardiology journal focusing on research publication in Turkey. It was published between 1988 – 2005. Its archive is distributed under Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) Creative Commons license. After crawling its website by WGET, we developed a strategy to parse the files. We obtained 1018 files (abstracts). In this case, the files were downloaded with their original article title names without number reference. Hence, file alignment has not been possible. However, each file includes both the Turkish and English abstracts. Using only one file in parsing step, we were able to divide the source and target content and, and then align them sentence by sentence. In a similar setup, we used a cascading filter with regex text filter and HTML filter. The regex for Turkish abstracts includes only one rule:

Import only the following content:
1. <div class="summaryMain">ÖZET
.* </div>

And the regex for English abstracts also includes only one rule:

Import only the following content:
1. <div class="summarySub">ABSTRACT
.* </div>

The abstract content is between tags that are occurring only once in the file. Otherwise, this kind of extraction would not be possible. The fourth journal will illustrate this case.

Journal Name	***Turkiye Klinikleri Journal of Cardiology***
Domain	Cardiology
UNESCO Code	3205.01
Source Word Count	118314
Target Word Count	133997
Sentence Count	7384
Source Word / Sentence Rate	16.02
Target Word / Sentence Rate	18.14

Table 20. Corpus profile of the third journal as calculated by Memoq's Statistics feature.

After the parsing and alignment, we obtained our TMX file and cleaned it from HTML tags. The results yielded 7384 sentences, 118314 source words, 133997 target words as it is displayed in the table below.

Corpus Preparation from Turkish Journal of Thoracic and Cardiovascular Surgery. The Turkish Journal of Thoracic and Cardiovascular Surgery is the last journal that we crawled for corpus preparation. It is also a journal published in Turkish and English:

> "Turkish Journal of Thoracic and Cardiovascular Surgery is an international open access journal which publishes original articles on topics in generality of Cardiac, Thoracic, Arterial, Venous, Lymphatic Disorders and their managements. These encompass all relevant clinical, surgical and experimental studies, editorials, current and collective reviews, technical know-how papers, case reports, interesting images, How to Do It papers, correspondences, and commentaries."[86]

It is also licensed under Creative Commons Attribution-NonCommercial 4.0 International License. The website of the journal has a similar structure to the first two journals. However, in this case, we cannot extract the content using the cascading filter (regex text filter and HTML filter) that we used before since the content of the abstracts are not included between unique tags or texts. For example, "<div class="col-lg-12 col-md-12 col-sm-12 col-xs-12 makale-ozet">" tag is repeated several times; hence, when we try to parse the content between this tag and another tag, the tool cannot decide where to start the parsing. We tried different combinations of regex text filters; however, we were not able to extract the content with this method. Since a typical abstract file in this journal does not include too much noisy content, we decide to use a default HTML filter. The noisy content that had to be imported include menu items, journal description in English as well as some meta data about how many times an abstract is viewed, authors etc. Since these contents are going to be both in source segment and target segment without change, we can remove them by utilizing the "Remove Duplicate" function in GoldPan TMX editor. To sum up, we parsed the abstracts with HTML filter, aligned them and exported them into TMX. Then in GoldPan, we removed the tags, the duplicated and in the end only the abstract sentences remained. This led to spending more time in the cleaning phase. However, the result

[86] http://tgkdc.dergisi.org/static.php?id=4

was the same. In the end, we had 13937 sentences, 155934 source words, 182284 target words.

Journal Name	***Turkish Journal of Thoracic and Cardiovascular Surgery***
Domain	Cardiology
UNESCO Code	3205.01
Source Word Count	155934
Target Word Count	182284
Sentence Count	13937
Source Word / Sentence Rate	11.18
Target Word / Sentence Rate	13.07

Table 21. Corpus profile of the fourth journal as calculated by Memoq's Statistics feature.

It can be observed that it is possible to achieve the same corpus by concentrating on different steps of the corpus preparation procedure. What is important is to analyze the files in detail, know what each tool in the stack can do and implement the steps accordingly.

TRENCARD Corpus Compiled. Following the corpus preparation procedure, we obtained 4 corpora on cardiology in the format of TMX. Despite the fact that multiple files can be combined into a single TMX, keeping them as separate files will allow for a more fine-grained analysis of the results for each corpus. It should be noted that machine translation systems have their own cleaning procedure in which they reject some sentences based on predefined criteria. Hence, we will be able to see how many words/sentences are rejected from each corpus. These will be shown in the training step in section 3.3.

3.1. Corpus Preparation for Machine Translation

Table 22 shows the compiled profile of TRENCARD. We created a corpus of 49693 sentences, 788046 source words and 907382 target words. The source word / sentence count gives an idea about the average length of sentence in the corpus.

Name	**TRENCARD CORPUS**
Domain	Cardiology
UNESCO Code	3205.01
Source Word Count	788046
Target Word Count	907382
Sentence Count	49693
Source Word / Sentence Rate	15.85
Target Word / Sentence Rate	18.25

Table 22. TRENCARD Corpus words, sentence count and word/sentence counts as calculated by Memoq's Statistics feature.

3.1.2. GENCOR: Turkish - English General Domain Parallel Corpus

We will use a general domain corpus for training an SMT engine and an NMT engine in combination with TRENCARD corpus. We call this general domain corpus GENCOR and it is derived from 8 publicly available bilingual corpora from different domains. These corpora are available in OPUS Corpus. The total source word count of GENCOR is 5668129 words and the number of sentences is 381,322 with a source word to sentence ratio of 14.86.

Corpus Name	Domain	Source Word	Sentence	W/S

EUBookShop	Information	482,649	22070	21.87
PHP	IT	74,042	9057	8.18
Infopanniki	Information	164,693	13173	12.50
WMT2019 News	News	197,288	10007	19.71
Ubuntu	IT	29,290	7285	4.02
KDE	IT	650,294	130731	4.97
Bianet	News	713,504	31749	22.47
Wikipedia	Information	3,356,369	157250	21.34
Total	General	5,668,129	381322	14.86

Table 23. Components of GENCOR are broken up according to corpus name, domain, source word count, sentence count and word/sentence rate.

One limitation of GENCOR may be that most of it is comprised of volunteer translations. This does not necessarily mean that translations in these corpora are of low quality; it only implies that professional translation processes (translation by a professional translator, review, and proofreading) and quality standards may not be guaranteed.

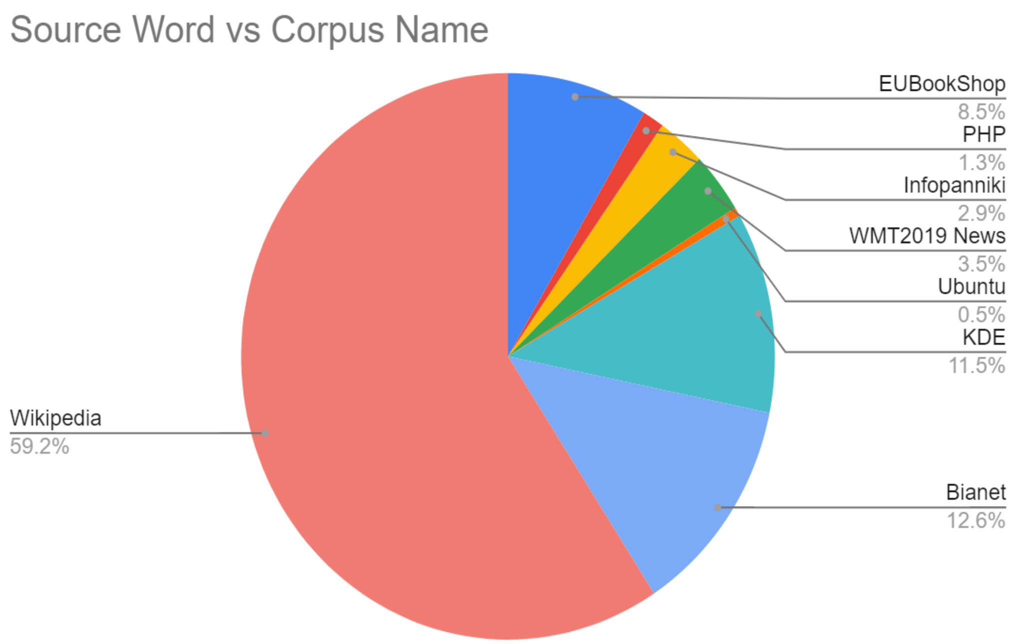


Figure 14. The share of each corpus within GENCOR.

The biggest component of GENCOR is the Wikipedia corpus[87] (Wołk&Marasek, 2014) which constitutes %59.2 of the corpus. This corpus includes sentences from Wikipedia's informative articles of a broad subject spectrum. For this reason, we considered it to be a proper subcorpus for our general domain corpus. PHP[88], Ubuntu[89] and KDE4[90] corpora are from IT localization projects translated by volunteer communities. As it can be observed from the Table 23, average sentence sizes are very small in these 3 corpora. EUBookShop[91] and Infopankki[92] have informative content and translations are conducted by professional translators in these two cases. Bianet

---

87 http://opus.nlpl.eu/Wikipedia-v1.0.php (last access: 27.01.2020)

88 http://opus.nlpl.eu/PHP-v1.php (last access: 27.01.2020)

89 http://opus.nlpl.eu/Ubuntu-v14.10.php (last access: 27.01.2020)

90 http://opus.nlpl.eu/KDE4-v2.php (last access: 27.01.2020)

91 http://opus.nlpl.eu/EUbookshop-v2.php (last access: 27.01.2020)

92 http://opus.nlpl.eu/infopankki-v1.php (last access: 27.01.2020)

corpus[93] (Ataman, 2018) is from a newspaper that publish news in Turkish, English and Kurdish. WMT2019 News[94] corpus also includes news articles from different subjects. It is the test set used in ACL 2019 Fourth Conference on Machine Translation. Considering these corpora, the terminology can be expected to be very varied. The objective is to measure how much terminological variation will occur when cardiology related texts are introduced for machine translation.

3.1.3 Bilingual Test Corpus for MT Quality Evaluation

We finally created a bilingual test corpus of 677 sentences using the methodology explained in section 3.1.1. We call this corpus Test Corpus. It is used both in human evaluation of the 4 MT engines, and their evaluation of terminological issues. The sentences are extracted from the bilingual abstracts published between January 2020 – October 2020 in the journal of *Archives of the Turkish Society of Cardiology*. Original raw test corpus included 725 sentence pairs. After removing duplicates and sentences which were left in English in both source and target side, we ended up with 677 segments. Since our training corpus of cardiology (TRENCARD) only includes the abstracts until the issue of May 2019, our test corpus is not included in the MT training. Below the details of the test corpus are given.

Name	**Test Corpus**
Domain	Cardiology
UNESCO Code	3205.01
Source Word Count	11015

[93] http://opus.nlpl.eu/Bianet-v1.php (last access: 27.01.2020)

[94] http://opus.nlpl.eu/WMT-News-v2019.php (last access: 27.01.2020)

3.1. Corpus Preparation for Machine Translation

Target Word Count	13293
Sentence Count	677
Source Word / Sentence Rate	15,40
Target Word / Sentence Rate	18,38

Table 24. Summary of the test corpus.

The average source sentence length in the corpus is 15.35 and the median is 14. We randomly selected a sample of 100 sentences for MT human evaluation as well as terminological quality evaluation based on sentence length, accuracy of the translations[95] and presence of cardiology terminology in the sentence. Besides, out of this sample of 100 sentences, we reserved the segments that include the keywords for abstracts and created a reference bilingual terminology list to be use for term annotation and terminology error categorization.

[95] Since alignment is made automatically and a minor revision is made after this operation, some sentences may still be misaligned. Besides, in some cases, the translations are observed to be freely extended (in terms of number of words) or summarized (probably to obey word limits in abstracts).

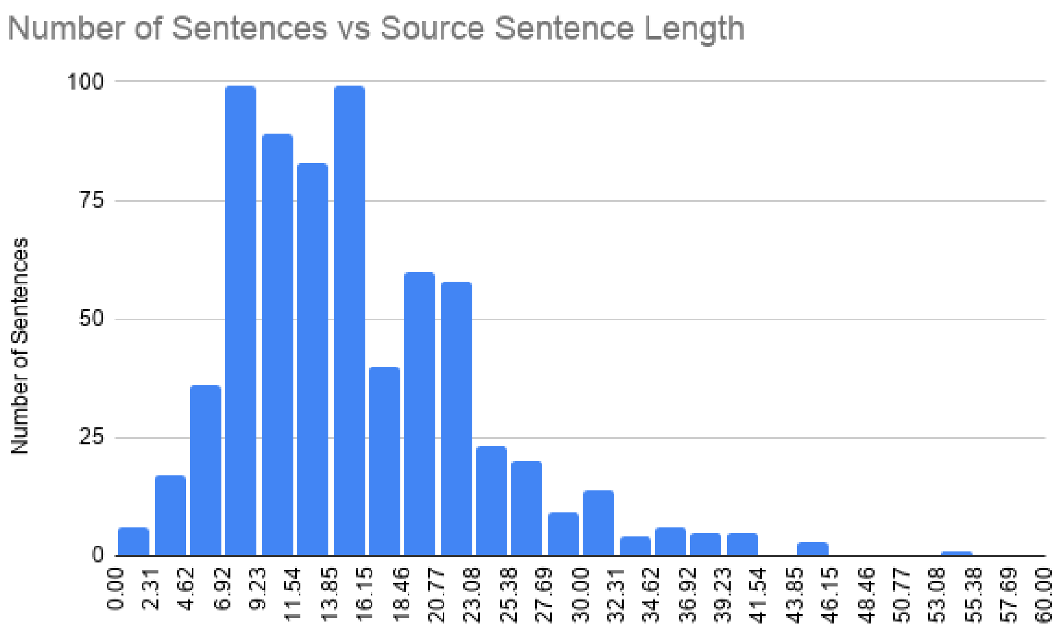


Figure 15. Histogram of sentence length in our test corpus.

3.1.4. Summary

Three corpora were created for TR → EN MT training and evaluation using corpus preparation tools and methods. TRENCARD corpus consists of bilingual abstracts from 4 cardiology journals, and is a highly specific domain corpus. GENCOR is a general domain corpus containing 8 publicly available bilingual corpora. Test Corpus is created from the bilingual abstracts in a cardiology journal and has 677 sentences. As a whole, this chapter shows that large amounts of corpora can be elaborated by using a synthesis of automatic and semi-automatic methods and tools.

3.2. Machine Translation Customization and Evaluation

In this section, we will describe the training process for our four MT engines in KantanMT using our custom corpus and a mixed domain corpora. A summary of the MT engine type, corpus type, source word count and sentence count are provided in the table below. Our SMT engines are called PBMT-1 and PBMT-2 to specify that they are phrase-based machine translation engines. Our NMT engines are called NMT-1 and NMT-2. We initially describe the overall characteristics of KantanMT, and why we selected it for the training steps.

Engines	Corpus Type	Word Count (Source)	Sentence Count
PBMT-1	Cardiology	788046	49693
PBMT-2	Cardiology + General Domain	6456175 (788046 + 5668129)	431015 (49693 + 381322)
NMT-1	Cardiology	788046	49693
NMT-2	Cardiology + General Domain	6456175 (788046 + 5668129)	431015
Test Corpus	Cardiology	11015	677

Table 25. Corpus type, source word count and sentence count for each engine.

KantanMT is “a commercial platform for custom MT” (Shterionov et al., 2018). The SMT side of the platform is based on Moses[96]. The NMT side of the platform is based on Transformer Neural Networks (TNN)[97]. It is possible to customize both NMT and SMT engines, measure their quality performance (both automatic and human evaluations), translate text directly in the platform, deploy and integrate into different devices and translation management systems[98]. The dashboard of the platform is divided as KantanMT and KantanLQR. KantanMT dashboard includes all customization features as well as other auxiliary features. KantanLQR dashboard provides linguistic quality review for measuring the quality of MT engines. It is possible to assign reviewers and let them check or compare qualities of the outputs from

[96] https://kantanmtblog.com/2015/01/21/moses-use-case-kantanmt-com/ (10.06.2020)

[97] https://kantanmtblog.com/2019/01/14/joint-study-confirms-kantan-tnn-delivers-remarkable-quality-scores/ In this blogpost, KantanMT researchers compare three models: Transformer Neural Networks (TNNs), Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) and find that TNN provides better fluency and adequacy compared to other models.

[98] https://kantanmt.com/aboutcompany.php (last access: 09.06.2020)

different MT engines based custom quality metrics or standard quality metrics such as DQF and MQM. These features make KantanMT a state-of-the-art MT customization platform both for SMT and NMT.

There are a number of reasons for selecting KantanMT as the main customization platform in our study. First of all, basic concepts of SMT are "difficult to understand", "challenging" and "intimidating" for translation professionals and translation students (Kenny&Doherty, 2014). Besides, commonly used systems such as Moses[99] require high technical knowledge for using Linux operating system, terminal, and scripting languages. Graphical user interfaces (GUI) have been developed for bridging this gap between SMT systems and translators. MTradumàtica[100] (Martín-Mor, 2017), which is developed by Tradumàtica Research Group in Universitat Autònoma de Barcelona, is a good example of this tendency. It is a free and open Moses-based cloud platform with an easy to use GUI. Currently, it runs in beta version. A desktop software example of this tendency is Slate Desktop[101] which is also based on Moses and runs locally on desktop. Its GUI allows translators to easily import their translation memories and create a custom MT engine. In the case of NMT, the above-mentioned technical difficulties are coupled with hardware limitations because NMT is computationally very demanding. Forcada (2017: 302) observes that while there are many free and open NMT toolkits "installing, configuring, and using them requires skills that are not usually possessed by professional translators, even if one has access to the kind of specialized hardware needed." Considering these limitations and the fact that the focus of our study is the evaluation of terminology translation in SMT and NMT, we decide

[99] http://www.statmt.org/moses/ (last access: 10.06.2020)

[100] https://github.com/tradumatica/mtradumatica (last access: 10.06.2020)

[101] https://www.slate.rocks/ (last access: 10.06.2020)

to have a single training platform. KantanMT has a GUI both for NMT and SMT and a nearly identical workflow. And since the trainings are performed in the cloud, there are no local computational requirements. Thanks to these features, we were able to concentrate on our corpus and the quality evaluation of our engines without spending too much time on the training process. While MTradumàtica could have been used as well since it has an easy-to-use customization workflow and the ability to evaluate engines, it is only possible to train SMT in it. We aim to keep the training procedures as standard as possible (in other words, as similar as possible) and KantanMT has provided this capability for us.

One limitation of using a closed, proprietary system such as KantanMT is that an advanced user cannot modify the underlying algorithm to suit their needs. Below we present how we trained each of our engines.

3.2.1. Training of the 2 SMT Engines

We firstly trained a phrase-based machine translation engine called PBMT-1 in KantanMT using TRENCARD cardiology corpus. As explained in section 3.2.2. in Chapter 3, TRENCARD corpus include 4 corpora compiled from 4 cardiology journals. After configuring the project by selecting system type (SMT, not NMT) and language pair (Turkish (tr) → English(en)), we uploaded the 4 corpora in TMX format. KantanMT gives the option of adding its own stock data in the language pair. However, we did not used its own training data in none of the trainings. The training corpora and procedure should be the same for allowing more reliable measuring of the quality differences between our SMT and NMT engines. The training phase has taken 16 minutes.

We followed the same steps for PBMT-2. The only difference is the training corpora used in this engine. Here we used both TRENCARD corpus and a GENCOR corpus which include 8 corpora from different domains. Again, monolingual corpus is not included. This training has taken a bit longer to complete since the corpus size is significantly larger. Training is completed in 44 minutes.

3.2.2. Training of the 2 NMT Engines

After completing the training of SMT engines, we trained the 2 NMT engines. We followed exactly the same steps. NMT-1 engine includes only TRENCARD cardiology corpus. After importing the translation memories comprising it, we started to build the engine. No stock data or monolingual data is included in the training. Since the training is performed on the servers of KantanMT and our parallel corpus is not very large, the training was completed in a reasonable time of 1 hours and 37 minutes. The training has had 9 training cycles. The second engine called NMT-2 has followed the same procedure. It was trained by the combination of TRENCARD + GENCOR corpora. It has taken 11 hours and 44 minutes to complete the training.

Once training is completed, KantanMT builds analytics both for corpus statistics and for automatic MT quality evaluation. Reported corpus statistics include total source word count, unique word count, number of rejected words etc.

3.2.3. Automatic and Human Evaluation of the Engines

Automatic MT Evaluation. Evaluation of the quality of the engines was performed in two levels: automatic evaluation and human evaluation. Below we will firstly describe how automatic evaluation metrics are deployed and then we will explain in detail the methodology that we followed for the human evaluation.

3.2. Machine Translation Customization and Evaluation

3 MT evaluation metrics are provided at the end of training in KantanMT for both SMT and NMT: F-Measure, BLEU and TER. All these metrics require a test set with human reference translations. KantanMT automatically allocates 500 sentences from the training corpus and keeps them as a separate test corpus and does not include them in the training phases of the model. These sentences are later translated by the SMT and NMT engines and compared against human reference translations using F-Measure, BLEU and TER metrics. Once evaluations are completed, a summary of evaluation results is provided in the dashboard with graphics. In other tabs, details of each metric are also provided. Besides, it is possible to check general word count and unique word count together with amount of rejected corpus content. Words that are not aligned with any words or phrases are also shown in gap analysis tab.

Aside from the abovementioned 3 metrics, NMT includes one more metric called "perplexity". In KantanMT's platform, perplexity is defined as follows:

> Perplexity is the measurement of how well a probability distribution or probability model predicts a sample. A low perplexity indicates the probability distribution is good at predicting the sample.[102]

We will report perplexity while comparing our NMT engines with each other. However, since SMT does not have this metric, it will not be used for comparing SMT and NMT engines.

Human Evaluation. Automatic evaluation alone, though useful and practical, may not be sufficient to conclude which MT engine performs better. For this reason, a human evaluation is also needed. For this task, we have machine translated 100 sentences with cardiology terms from our test corpus (see Chapter 3 – Sample Set section) using our 4 engines. Then, using KantanLQR, 5 professional translators has evaluated the quality

[102] https://kantanmtblog.com/2017/08/04/get-the-best-from-neural-mt-with-quality-data/ (last access: 15.06.2020)

of the translations in terms of adequacy and fluency and ranked them from the best to the worst. Below we explain the features of our source segments, translators who participated in the evaluation task, features of the evaluation platform and the method of conducting the evaluation task.

We selected 100 sentences from our test corpus-based on the presence of a cardiology related term both in source and target sides and having a source sentence length between 5 – 29 words. In our selection process, we have firstly removed the sentences that do not include any cardiology terms in the source side. We referred to two cardiological terminology dictionaries, *Kardiyoloji (Yürek Bilimi) Türkçe Terimler Kılavuzu* (2009) (Guide for Cardiology Turkish Terms) and *Pediyatrik Kardiyoloji Terimler Kılavuzu* (2017) (Guide for Pediatric Cardiology Terms), for verifying the existence of the terms. After these filtering steps, we ended up having 612 sentence pairs with terms. Out of these 612 sentences, we randomly selected 100 sentences based on their word lengths. There are 4 sentence pairs from each lengths: 4 source sentences having 4 words, 5 source sentences having 4 words etc. Thanks to this custom test data, it became possible to evaluate how NMT and PBMT behave according to source sentence length.

3.2. Machine Translation Customization and Evaluation

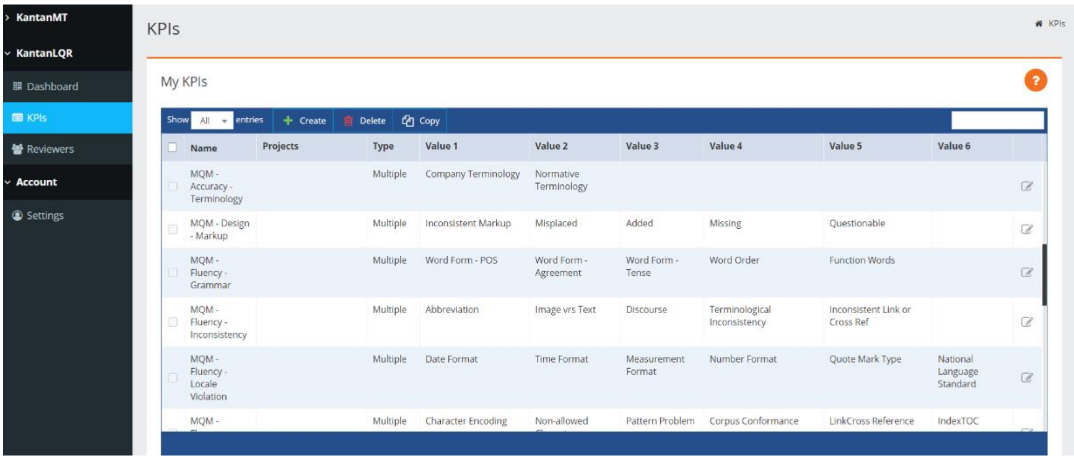


Figure 16. An overview of KantanLQR's key performance indicators. While it is possible to use MQM framework, it is also possible to add custom indicators.

5 professional Turkish translators participated in the quality evaluation task. The translators filled out an anonymous form to identify their professional competences (See Annex III for the form). Then they were instructed about how to access to KantanLQR and how they should perform the evaluation. Once the instructions are completed, the translators began and completed the evaluation.

All human evaluation tasks were performed online on KantanLQR dashboard. This dashboard is part of KantanMT and is used for conducting human review of (machine) translated content. It is possible to conduct quality evaluation, A/B test and postediting tasks on the platform. Reviewers can be invited for a certain task through adding their e-mails. Both industry standard key performance indicators such as MQM (Multidimensional Quality Metrics) as well as other widely used indicators such as adequacy, fluency, terminology etc. can be implemented in quality evaluation tasks. Lastly, once the evaluation task is completed, results are shown in graphic visualizations. In summary, KantanLQR’s capabilities include language quality review, distributed workflow, key performance indicators, customizing test plans, real time data visualization, productivity, and reliable industry standards[103].

Our human evaluation task consists of evaluating the quality of 4 MT engines comparatively and individually by conducting an A/B Test, and assigning adequacy and fluency scores to each sentence. A/B Test requires to sort 4 target sentences from the best to the worst in each round evaluation. KantanLQR has a scale of stars and the

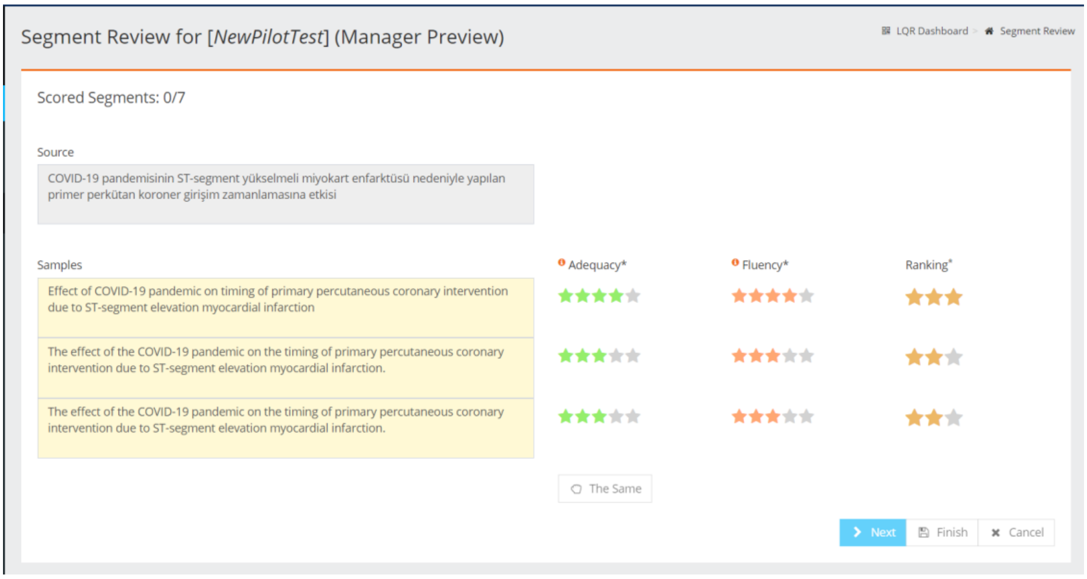


Figure 17. An example of a A/B Test and adequacy/fluency evaluation. There are 3 MT results to be evaluated. Not that assigning the same ranking is permissible.

more stars a segment has, the higher its ranking is. If the reviewers consider that two segments are equal, they can assign them the same number of stars. Reviewers see MT output from 4 engines in a randomized fashion. In other words, they will not know which target sentence comes from which engine.

The reviewers also need to assign adequacy and fluency scores. Adequacy measures how much of the meaning in the source sentence is expressed in the target sentences, namely it is a measure of accuracy. The reviewers need to select a number of stars from

---

[103] Automate language quality review for better translation outputs. https://www.kantanmt.com/overview-kantanlqr.php (26.10.2020)

a scale of 5 stars (more stars mean better accuracy). Fluency measures the grammaticality and readability of a sentence and focuses more on stylistic aspects. Again, reviewers need to select from a scale of 5 stars with more star meaning better fluency. As it can be observed from Figure 17, there are orange colored info icons which provide the reviewer additional information about the meaning of each star for each key performance metric. The reviewers are asked to complete the evaluation task in a single session. Once all reviewers complete their evaluation, KantanLQR visualizes global results in a dashboard with the overall scores each engine get. Besides, it is possible to view the results coming from each reviewer.

Chapter 4. Methodology for Terminology Evaluation in MT

The whole of language is a continuous process of metaphor, and the history of semantics is an aspect of the history of culture; language is at the same time a living thing and a museum of fossils of life and civilisations.

-Antoni Gramsci, 1971

4.1. Methodology for Terminology Annotation and Evaluation

The last step of the MT quality evaluation involves comparative subjective evaluation of terminology errors in our 4 engines with the purpose of exploring how each engine handles terminology depending on their MT system type and corpus type. In this chapter, we explain how we selected and annotated terms in our 100 sample sentences, which terminology error categories we adopted, and how we manually analyzed errors in the 4 MT engines.

4.2. Term Annotation in Sample Sentences

We used the same 100 Turkish - English sentences sampled from Test Corpus with the aim of evaluating terminology translation. As mentioned in Section 3.3.3, these 100 Turkish sentences and their English translations were selected based on the availability of at least one cardiology terms both in source and target side.

We developed a subjective, manual annotation strategy and did not resort to automatic annotation. We have already discussed in Chapter 2 that the difference between a term and a word is not always clear and is controversial. Besides, since our source language Turkish is an agglutinative language, any automatic term recognition tool may not recognize Turkish terms with lots of suffixes (cf. section 2.5 for a discussion of term extraction and annotation). Such an automatic strategy would also require a large ready-made bilingual terminology list in compatible file formats which was not available to us during the task. Lastly, unlike previous studies, we aim to evaluate acronyms as well and they cannot be annotated automatically in an efficient way (cf. section 2.5). Considering all these facts and with objective of creating a gold standard Turkish → English term-annotated bilingual corpora, we used a manual annotation method.

4.2. Term Annotation in Sample Sentences

Using a spreadsheet in which source sentences and target sentences are in separate columns, we firstly identified each term and then write them in separate cells in the same row (one term per cell) with the purpose of extracting them for later use in the terminology evaluation phase. Unigrams, bigram terms and n-gram terms up to 8-grams are annotated. While no term length limit is stipulated, the longest terms, though very less frequent, had 8 words; only 10 terms with more than 4 words were available in the test set. Unspecified, general terms such as "patient", "heart", "head" etc. are not annotated and not included in the evaluation. Only unigrams with specified meanings in cardiology such as "arrhythmia", "cardiomyopathy" are included. Since specificity is usually achieved in terminological collocations, all cardiological terminological collocations are annotated. Despite the fact that this might have an effect on the objectivity of this part of the research, and being this an essential step to terminology evaluation, this effect was compensated for through the following measures. i). Glossary check: We checked 2 cardiology terminology dictionaries, *Kardiyoloji (Yürek Bilimi) Türkçe Terimler Kılavuzu* (2009) (Guide for Cardiology Turkish Terms) and *Pediyatrik Kardiyoloji Terimler Kılavuzu* (2017) (Guide for Pediatric Cardiology Terms), in order to validate the termness of a lexical unit. ii). Reference keyword check: We also used as reference the keywords sections of the abstracts in test corpus[104]. Once we completed annotating the terms (see Annex VIII and Annex IX), we passed to the manual terminology error analysis step. Below we explain our methodology for error categorization and analysis.

[104] Test corpus is accessible through: https://github.com/gokhandogru/TurkishEnglishParallelCorporaandMTEvaluation (last acces: 12.02.2021)

4.3. Error Categorization & Analysis

Terminology error categories helps us understand what kinds of terminology errors each MT engine type produces. After specifying the terminology error types in each MT engine, it becomes possible to make a statistical analysis of the distributions and frequencies of error types. For this purpose, a comprehensive terminology error typology is required. We already mentioned in chapter 2 that there are only few studies concentrating on the variety of terminology errors that MT systems commits. While the error categories by Haque et al. (2019) paves the way for a comprehensive automatic terminology error analysis, it is complemented with other term error categories such as acronyms which are important from a translator's point of view. Below we explain our error categories.

Before conducting a deep manual error analysis in MT engine results, we firstly realized a binary analysis of the term translations in 4 engines simply checking if a particular term is translated correctly or incorrectly considering the reference human translation and the terminology dictionary resources. Hence, a target term translation is to be labeled as "correct" if it is the same as the one in human reference translation or if it is a synonym available in one of the above-mentioned terminology resources. And in all remaining cases (omission, insertion without translation etc.), it is deemed incorrect. After a first binary analysis (correct/incorrect), a subsequent detailed analysis assigned categories of terminology errors. When qualitative analysis is completed, we conducted a quantitative analysis of terminology errors, compare the rates of such errors in each engine and discuss the possible reasons for these error types. Throughout the analysis process, in order to minimize the effect of subjectivity, we compared MT results to human references and resorted to the cardiology dictionaries as well as the bilingual keywords in the test corpus.

The reason for errors in terminology translation in MT may be due to semantic, morphosyntactic, or related to MT system or training corpus. Extending the error categories of Haque et al. (2019), we identified 11 different error types. Acronym mistranslation, target term extension and shift of grammatical category are not included or nested under other categories in the mentioned study.

Error category	**Description**
Partial term translation	Source term is partially translated into target term
Incorrect term equivalent	Totally incorrect term translation. None of the components of the term are correct.
Morphosyntactic problem in target term	Incorrect suffixes or prefixes and/or misplaced syntactic operators leading to shifts / errors of meaning.
Reordering error	Components of a terminological collocation are separated from each other in a sentence.
Source term insertion	Source term is inserted without a translation.
Acronym mistranslation	An acronym is dropped or mistranslated.
Term drop	A source term is dropped and not translated in the target sentence.
Term extended	MT system adds a modifier to a term and changes/extends the meaning of the source term unnecessarily.
Term with incorrect grammatical category	A term in noun form is translated as an adjective, or as other grammatical categories.
Literal Translation	Source term is literally translated causing a loss between the concept and the term especially in cases on terminological analogies.
Other term errors	A terminology error that cannot be categorized within the categories above.

Table 26. 11 error categories used for terminology translation evaluation in MT.

The table above summarizes each error category. We assign “partial term translation” error category when at least one of the components of a target term is left without translation, or it is not translated. If none of the components of a term are translated correctly, we mark the error as an "incorrect term equivalent". When a term has suffixes or prefixes that cause changes in meaning, we consider that there is a "morphosyntactic problem in target term". Besides, in some languages, operators such as "of", "in", "as" within terms may be misused and can cause errors in terms. When the there are other words between the components of term or the components needs to change place (instead of "cardiovascular disease", "disease cardiovascular"*[105] is used), we specify the error as "reordering error". If the source term is conveyed directly to target sentence without translation, we consider "source term insertion error". If the source term is not translated or included in the target sentence, we consider it as "term drop". "Term extension" can be considered as the opposite of term drop and occurs when irrelevant or meaning changing modifiers are appended to the term translation. In some cases, a bi-gram term may be combined with a modifier that was not present in the source sentence simply because it is statistically more probable. For example, tr. "tedavi etkinliği" has the translation of "treatment effectiveness" but it is translated as "percutaneous treatment effectiveness" where "percutaneous" is not mentioned in the source. If a source term in noun form is translated in adjective form or other form, we assign "term with incorrect grammatical category" error. To illustrate, tr. "diyabet" is translated as "diabetic" while it should be "diabetes" in a target sentence. If a term is translated literally and has lost the conceptual link in a way that does not trigger a neologism, we consider it a "literal translation" error. For example, tr. "diyabetik sonlanım" is translated as "diyabetik endpoints"* where the word tr. "sonlanım" is literally translated and the actual terminological meaning is not conveyed. The correct translation in the reference is "diabetic outcome". Lastly, acronyms as terms require

[105] "*" here means that the expression is incorrect.

contextual judgment and subject matter expertise to decide if the they are correctly conveyed to target side. Especially in cases of direct transfer of source acronym into target without a change. In English to Turkish medical translation, usually conserving the source form of the acronym is accepted while in Turkish to English pair, a Turkish acronym will have to be translated into English, or else, the comprehension might be compromised (See the discussion of acronyms in Chapter 2). Hence, we consider case by case the acronyms in the sample set and unless an acronym is a correct English acronym in target, it is considered as an "acronym mistranslation". In other words, if acronym is dropped, it is also considered under the same error category since only a binary error annotation is applied to acronyms. "Other term errors" category include all the remaining terminology errors including ambiguous cases where more than one term assignation is possible.

Lastly, we also detail the categories of correct terms to be able to compare the behaviors of different MT systems. We define 4 correct term categories which are provided below.

Correct term category	**Description**
Correct term considering reference	Term translation is equivalent to the term in the reference human translation.
Correct synonym	Term is a denominational variation (synonym). It is available in reference terminology dictionaries and/or training corpus.
Possible equivalent	Not available in reference resources but a plausible term candidate to be considered a neologism.
Correct acronym	Acronym is conveyed or translated correctly.

Table 27. Correct term categories.

It should be noted that our typology deviates from the correct term typology of Haque et al. (2019) in that our typology considers neologisms as possible equivalents if they are plausible translations even though they are neither in the reference corpus nor in terminology dictionaries. In such cases, we follow the specificity principle of Aguilar-

Amat et al. (2005) to consider a term translation as a plausible candidate. It can be observed from the table that we considered correct acronyms as a separate class.

After assigning these terminology error categories and correct terminology categories to the 100 sentences translated by our 4 MT systems, we conclude our terminological evaluation by conducting a comparative statistical analysis to determine relative strengths and weakness of each engine. We analyzed the frequency of correct/incorrect term translations according to source term n-gram lengths, and source sentence lengths in each engine. In both of these analyses, acronym translation is not included since this analysis cannot be applied to acronyms since they are unigrams. We determined five n-gram categories: 1-gram, 2-gram, 3-gram, 4-gram and >4-gram source term categories. The objective of this categorization is to analyze the effect of source term n-gram length on MT systems. In other words, the objective is to observe if there is a correlation between term length and correct term translation in PBMT and NMT. The sentence length in the sample set varies between 5 words and 29 words. Sentences with 1-4 words are not included for a few reasons: firstly, it may not be possible to observe phenomena such as partial term translation error or term extension in very short sentences. Secondly, our term analysis is based on term translation in the context of a sentence. We divided source sentences into 3 categories according to their lengths: sentences with 5-10 words, sentences with 11-20 words and sentences 21-29 words. The objective in this categorization is again to measure the correlation between sentence length and term translation in different MT systems. The results of this terminology evaluation are provided in Chapter 6.

Part III. Study Results: analysis and reflections

Chapter 5. Automatic and Human Evaluation Results for MT Engines

Can the word ‘best’ mean anything at all, except to some particular person in some particular mood? Perhaps not — so if we allow the word to stand as an absolute, you, or you, or perhaps you, may be appalled at omissions or inclusions or, never having read me before, may even be impelled to cry out, ‘Good heavens, are those his best?.

-Isaac Asimov, 1973

In this chapter, we report automatic and human evaluation results for our 4 engines in a systematic and detailed way. The first half of the chapter presents the evaluation results based on automatic evaluations metrics as measured by KantanMT platform: F-Measure, BLEU and TER. Initially, specific domain cardiology TR → EN engines (PBMT-1 and NMT-1) are compared side by side. Afterwards, mixed domain TR → EN engines (PBMT-2 and NMT-2) are compared in the same way. An overall discussion of the automatic evaluation results is provided at the end. The second half of the chapter provides the results of the human evaluation of each engine based ranking, adequacy and fluency tasks conducted by 5 human reviewers. At the end of the chapter, we discuss our overall results and contrast them with other studies.

5. 1. Automatic Evaluation of TR → EN Specific and Mixed Domain MT Engines

Automatic evaluation metrics (AEM) help gain insights rapidly about translation quality of an MT engine and compare different engines by subjecting all of them to exactly the same criteria. These aspects are especially important in MT development phases which require lots of iterations of quality evaluation, in which case repetitive human evaluations might be slow, subjective, and costly. Although there is a growing number of criticism about the effectivity of these evaluation methods (Way, 2018), they are still widely used in MT studies and paired with human evaluation tasks for increased confidence about the overall quality. We used 3 different evaluation metrics, all of which make a sentence-level comparison between a reference sentence and a machine translated sentence. A summary of the automatic evaluation scores for our 4 engines is given below in Figure 18. A more detailed analysis is provided at the end of the section.

5. 1. Automatic Evaluation of TR → EN Specific and Mixed Domain MT Engines

500 sentences were reserved from the training corpora automatically and when the training was completed, these sentences were translated by each engine. Then machine translation outputs are compared to reference human translations and F-Measure, BLEU and TER scores are calculated based on these 500 sentences. Note that this randomly selected test corpus of 500 sentences is different for each engine. A sample of 50 sentence from each engine is provided in Annex II.

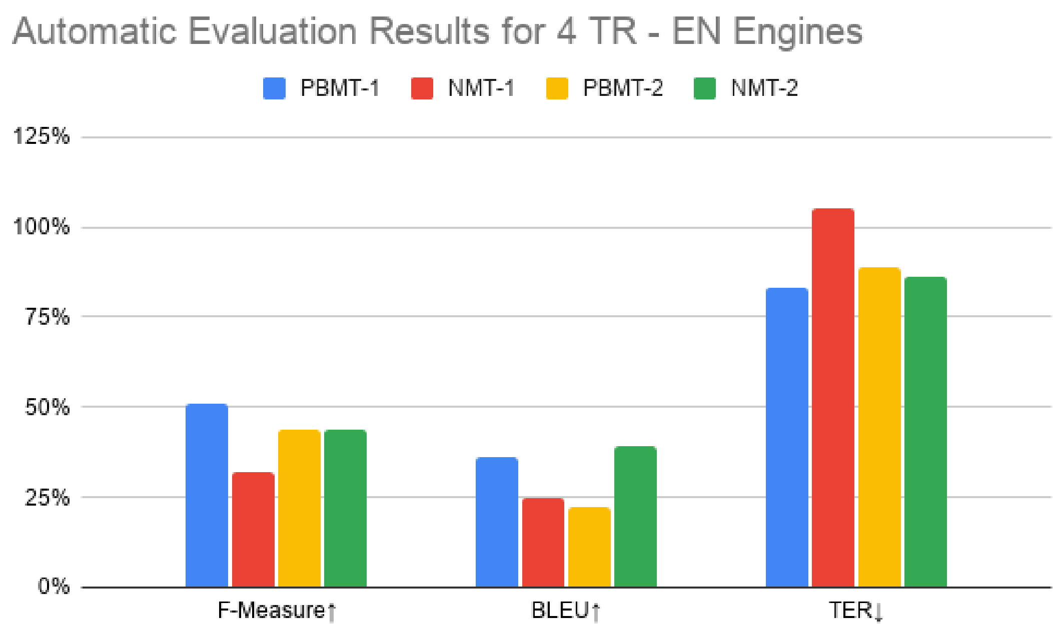


Figure 18. Summary of the automatic evaluation results for the 4 engines: PBMT-1, NMT-1, PBMT-2 and NMT-2.[106]

PBMT-1 vs. NMT-1. We trained both PBMT-1 and NMT-1 engines with TRENCARD corpus (For the details of TRENCARD corpus, see Chapter 3). Table 29 provides the details of the corpus statistics for PBMT-1 and NMT-1. The raw corpus includes

[106] Note that while higher score means better translation quality in F-Measure and BLEU, low score means better translation quality in TER.PBMT-1 has the highest F-Measure score, NMT-2 has the highest BLEU score and again PBMT-1 has the best TER score. TER is calculated based on the ratio between the number of edits (i.e., additions, deletions and substitutions) and the number of words in the reference translation. If the number of edits is more than the number of words in the reference translation, the TER ratio is going to be bigger than 1 and hence the percentage score will be above 100%. That's why NMT-1 has a percentage higher than 100%.

approximately 800,000 words. However, KantanMT applies a further preprocessing steps to remove sentences that decrease training performance. The rejected sentences are reported under "Reject" tab in BuildAnalytics[tm] page. It is possible to see the rejected number of words as well as their percentage. Reasons to be rejected include "significant difference in segment lengths", "Difference in placeholder counts" and "Segment too long to be meaningful during training". Examples of rejected sentences are given in Table 28. After rejected words are subtracted, 702,424 words entered into the training. The number of unique words (in other words, the vocabulary size) of this corpus is 85,877 in PBMT-1 and 85,647 in NMT-1. It is curious that different number of unique word is found although the same corpus is used for both trainings. This phenomenon might be explained by the possible change in the definition of word in KantanMT pipeline for PBMT and SMT. In other words, word pairs with a dash might be considered a single word in one case, and two words in the other case.

Turkish Sentence	**English Sentence**	**Rejection Reason**
[1] Olgularımızın mortalite oranı %2.6dır.	[0] The mortality rate of thrombolytic theraphy in our group was 2.6%.	[Error 104] Difference in placeholder counts.
[6] Sonuç:	[129] Conclusion: Incidence of ventricular dysrhytmia was low as expected because of early operation age and short follow-up period in our study population.	[Error 102] Significant difference in segment lengths.
[113] Bu çalışma, akut anterior miyokard infarktüsünde inferior derivasyonlardaki ST segment çökmesinin önemini belirlemek amacı ile yapılmıştır.Sol ön inen koroner arterdeki lezyonun yeri ile ST segment çökmesi arasındaki ilişki saptanmaya çalışılmış ve yine inferior derivasyonlardaki ST segment çökmesinin anterolateral derivasyonlardaki	[101] The project is planned to evaluate the clinical importance of ST segment depression(STD) in inferior leads in patients with acute anterior myocardial infarction(AAMI).Correlation between inferior STD, elevation in anterolateral regions and the level of obstruction in left anterior descending coronary artery(LAD) is observed. 28 patients with AAMI are grouped according to obstruction level being either before (G1, n=15) or after(G2, n=13) the first	[Error 105] Segment too long to be meaningful during training.

ST segment amplitüdlerinden ne kadar etkilendiği irdelenmiştir.Akut anterior miyokard infarktüslü 28 hastaya infarktüsten sonraki ilk 30 gün içerisinde koroner anjiyografi uygulanmış ve lezyon yerinin birinci diagonal dal öncesi(G1; n=15) ya da sonrasında(G2; n=13) olmasına göre vakalar ikiye ayrılmıştır.Gruplar arasında yaş, koroner anjiografi olana kadar geçen süre, Streptokinaz uygulanımı ve tutulan ek damar sayısı arasında herhangi bir fark kaydedilmemiştir. Proksimal lezyonlu grupta, D2,D3 ve aVF derivasyonlarındaki ST segmentinde distal lezyonlu gruba göre anlamlı çökme izlenmiştir.	diagonal branch according to coronary arteriography performed approximately in 30 days after AAMI.No difference was observed between the 2 groups according to thrombolytic therapy application, time till coronary arteriography performed, other coronary artery involvement or totally occluded LAD.

Table 28. 3 different reason for not including sentences in the training.[107]

Besides, a gap analysis report is provided for the words that are untranslated when sample machine translations are generated for making automatic evaluation. These words remain untranslated because the MT system cannot generate a translation candidate based on the given corpus. The possible reason for low number of unknown words in NMT-1 is the use of subword units. Note that the use of subword units may have a significant effect on the translation of terms. Table 29 outlines the corpus details used in both trainings.

Engine Name	**Total Word Count**	**Unique Word Count**	**Rejected Words**	**Unknown Words**

[107] "Difference in placeholder counts" was the most common rejection reason and just like the first row, many sentences can be edited quickly and later added to re-training.

5. 1. Automatic Evaluation of TR → EN Specific and Mixed Domain MT Engines

PBMT-1	702,424	85,877	91,053	238
NMT-1	702,424	85,647	91,053	38

Table 29. Corpus statistics for PBMT-1 and NMT-1.[108]

Once TRENCARD corpus is uploaded to KantanMT, the training is initiated without adding further monolingual corpora. Since the training corpus is relatively small, the trainings are completed in reasonably short times. PBMT-1 training has taken 16 minutes while NMT-1 training has taken 1 hour 37 minutes. Annex I illustrates the training steps comprehensively for each engine. Table 30 demonstrates the automatic evaluation results for PBMT-1 and NMT-1. The F-Measure, BLEU and TER scores for PBMT-1 are 51%, 36% and 83%, respectively. On the other hand, F-Measure, BLEU and TER scores for NMT-1 are 32%, 25% and 105%, respectively. Considering these scores, PBMT-1 performs significantly better in three of the metrics.

Engine Name	F-Measure↑	BLEU↑	TER↓
PBMT-1	**51%**	**36%**	**83%**
NMT-1	32%	25%	105%*

Table 30. F-Measure, BLEU and TER scores of PBMT-1 and NMT-1 Engines.

[108] Total monolingual source count, unique source word count, number of words rejected in preprocessing, and the number of untranslated words in the machine translated sentences used for automatic evaluations. Note that there are slight differences in the number of unique words. This change is likely to be due to the changing definition of what a word means in the KantanMT pipeline.

This significant difference between these two engines may be due to our relatively small corpus size. NMT engines are known to perform worse than PBMT engines in small size corpus scenarios (Koehn&Knowles, 2017).

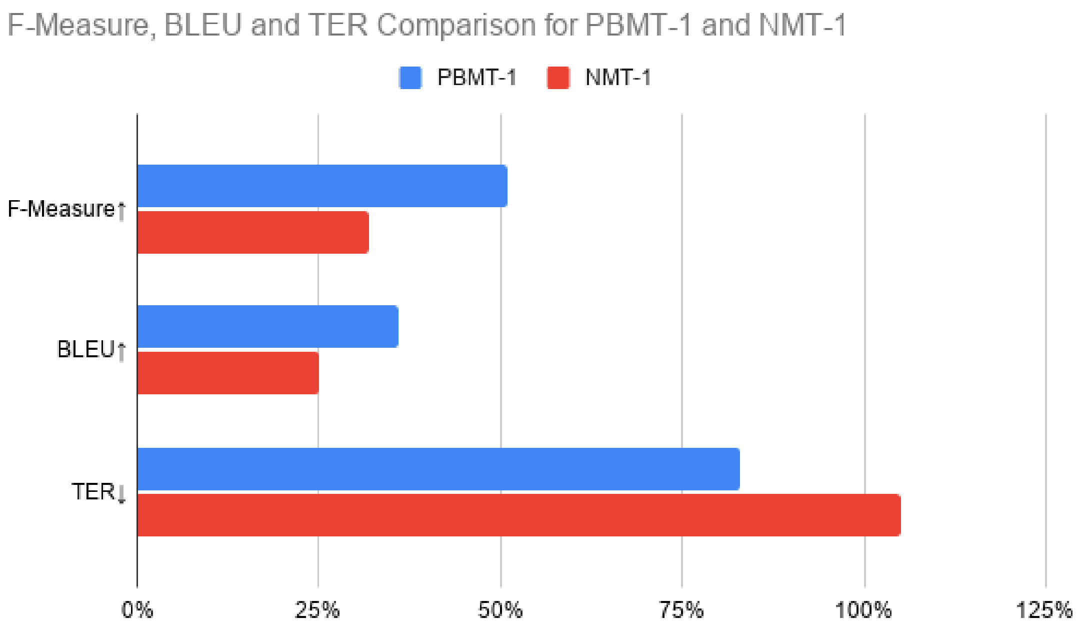


Figure 19. This horizontal bar chart shows how significantly PBMT-1 performs better than NMT-1.

The F-Measure score of 51% for PBMT-1, while still not sufficient for a well performing engine[109], implies an average knowledge of target domain and language. On the contrary, the F-Measure score of 32% for NMT-1 is low and implies below average knowledge of target domain and language according to KantanMT's quality thresholds. Considering BLEU score, a 50% score is a good score[110] and will require less postediting effort. Neither of the engines reaches that score. However, PBMT-1 is better than NMT-1 by 11%. Finally, TER score of maximum 30% is recommended by

[109]According to KantanMT, an engine with a score of 70% and over is a well performing engine that will need less postediting. F-Measure in BuildAnalytics. https://kantanmt.zendesk.com/hc/en-us/articles/204656689-F-Measure-in-BuildAnalytics (last access: 14.10.2020)

[110] BLEU in BuildAnalytics. https://kantanmt.zendesk.com/hc/en-us/articles/205355285-BLEU-in-BuildAnalytics (last access: 14.10.2020)

KantanMT[111] and our engines are very far from this score with 89% and 105%. This means that a high amount of postediting effort will be needed for both of the engines. But again PBMT-1 has a better score compared to NMT-1.

In conclusion, transitioning from PBMT to NMT using the same small size, specific domain TR → EN corpus has not increased the performance of the MT engine in context of the automatic evaluation metrics. The quality scores for F-Measure, BLEU and TER decreased by 19%, 11% and 22%. Below, we will look at PBMT-2 and NMT-2 scores and finally discuss the results all together.

PBMT-2 vs. NMT-2. After training PBMT-1 and NMT-1 engines, we trained PBMT-2 and NMT-2 engines in KantanMT with considerably larger corpora. The language pair is Turkish → English and the MT platform is KantanMT.

Engine Name	Total Word Count	Unique Word Count	Rejected Words	Unknown Words
PBMT-2	4.657.519	616,231	955,006	359
NMT-2	4.657.520	615,724	955,006	132

Table 31. Corpus statistics for PBMT-2 and NMT-2 engines. Total source word count, unique word count, rejected word count and unknown words during automatic quality evaluation are provided.

[111] https://kantanmtblog.com/2015/07/28/what-is-translation-error-rate-ter/ (last access: 14.10.2020)

5. 1. Automatic Evaluation of TR → EN Specific and Mixed Domain MT Engines

PBMT-2 and NMT-2 are trained on a dataset consisting of both TRENCARD cardiology corpus and GENCOR mixed domain corpus. Training corpus statistics are provided above in Table 31. Both source word count and unique word count are above the KantanMT's recommended thresholds (2 million words for source word count and 500,000 for unique word count). The number of rejected words was 955,006 for both engines while gap analysis has yielded 359 unknown word for PBMT-2 and 132 word for NMT-2. The training time for PBMT-2 was 44 minutes while it was 11 hours and 44 minutes. Training steps are provided in Annex I.

Engine Name	F-Measure↑	BLEU↑	TER↓
PBMT-2	44%	22%	89%
NMT-2	44%	39%	**86%**

Table 32. F-Measure, BLEU and TER scores for PBMT-2 and NMT-2. Higher scores are displayed in bold. Note that F-Measure scores are equal. NMT-2 has a significantly better BLEU score and there is 3% difference between TER scores in favor of NMT-2.

F-Measure, BLEU and TER scores for PBMT-2 and NMT-2 are given in Table 32. It can be observed that while F-Measure scores are equal (44%), NMT-2 has a significantly better BLEU score (39%) in comparison with PBMT-2 (22%).

5. 1. Automatic Evaluation of TR → EN Specific and Mixed Domain MT Engines

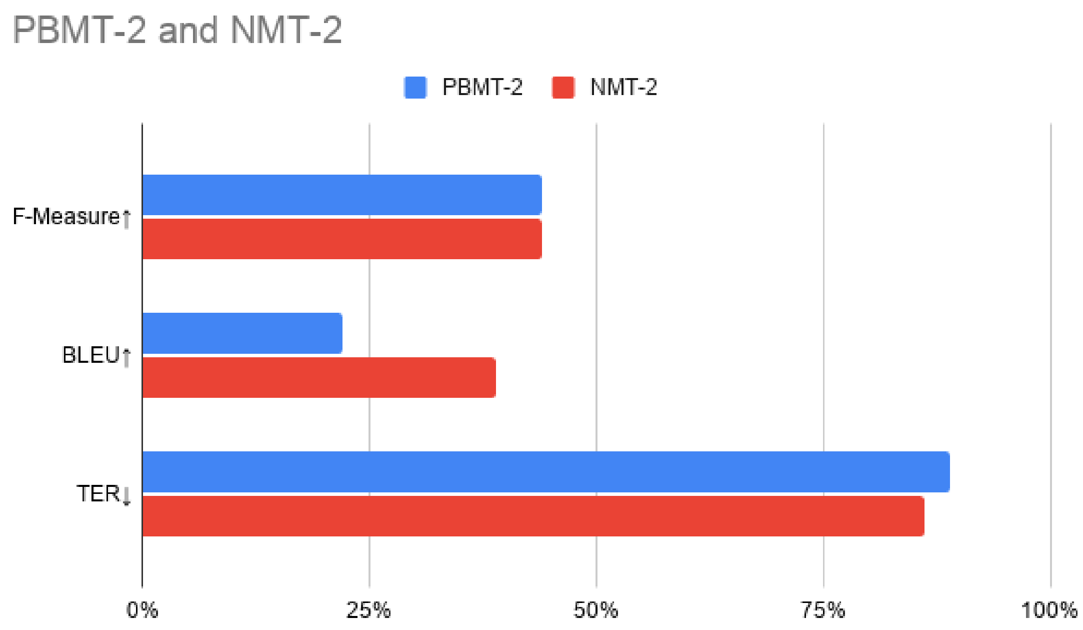


Figure 20. F-Measure, BLEU and TER scores shown in a bar chart for PBMT-2 and NMT-2.

Since F-Measure scores are below 50%, KantanMT considers that their knowledge of target domain and language are below average. Similarly, while there is a 17% difference in BLEU scores in favor of NMT-2, KantanMT points out that they will not produce highly fluent translations. There is only a 3% difference in TER scores in favor of NMT-2. Considering the three automatic metrics, NMT-2 has better overall score but all its scores are below adequate threshold in the context of KantanMT.

5.1.1. Discussion

In this section, we will compare our engines, investigate the evolution of the quality based on MT type (SMT vs NMT) and corpus size, and determine the best and worst engines.

Considering the overall scores, PBMT-1 has the highest F-Measure score with 51% and NMT-1 has the lowest F-Measure score with 32%. In terms of BLEU score, NMT-2 is the best performing engine while PBMT-2 is the worst one. Finally, again, PBMT-1 has the best TER score while NMT-1 has the worst score. While there is no one single outstanding engine, we can argue that PBMT-1 and NMT-2 are better than the other two engines according to the automatic metrics. Hence, in terms of MT system type, there does not seem to be a significant difference in the context of TR → En automatic evaluation scores. However, a fine-grained look at the change in corpus size in MT systems provides an interesting insight.

Engine Name	F-Measure↑	BLEU↑	TER↓
PBMT-1	**51%**	36%	**83%**
NMT-1	32%	25%	105%
PBMT-2	44%	22%	89%
NMT-2	44%	**39%**	86%

Table 33. Automatic evaluation scores for 4 engines. Best scores are shown in bold.

The second round of trainings with PBMT-2 and NMT-2 has included a mixed domain corpus which was 6.63 times larger than the one used in the first round with PBMT-1 and NMT-1. Furthermore, the unique word count in the second round is 7.18 times larger the first one. However, this increase in corpus size does not necessarily translate into better automatic quality scores. And there is a significant difference between NMT and PBMT engines considering the evolution in the quality. While PBMT engines are negatively affected by the corpus increase, NMT engines are positively affected.

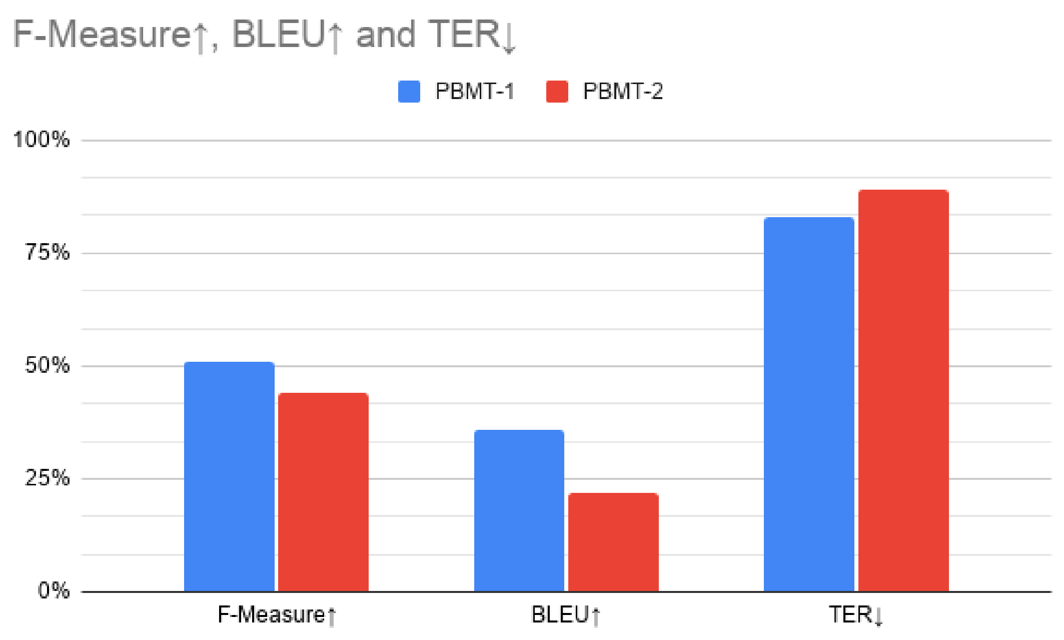


Figure 21. The change of F-Measure, BLEU and TER scores between PBMT-1 and PBMT-2.

All of the automatic evaluation scores degraded from PBMT-1 to PBMT-2. F-Measure score has decreased by %7 and BLEU score by 14% while TER score (which should decrease for better performance) increased by %6. Figure 22 illustrates these changes in a bar chart. The most striking change is in BLEU score. However, these degradations in automatic scores do not necessarily mean that the engine will perform worse in a real translation scenario. Human evaluation will be needed to make a final decision.

5. 1. Automatic Evaluation of TR → EN Specific and Mixed Domain MT Engines

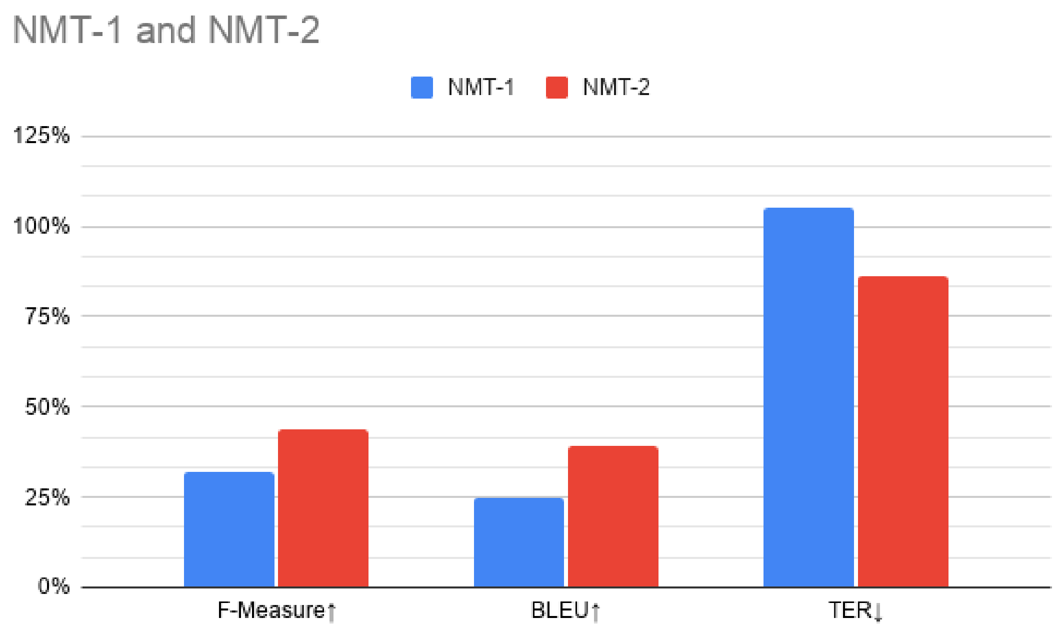


Figure 22. Change in F-Measure, BLEU and TER scores between NMT-1 and NMT-2.

When we look at NMT engines. There is a complete difference. In this case, the increase in automatic metrics positively affect each of the automatic evaluation metric. F-Measure increases by 12% and becomes 44%, BLEU increases by 14% and becomes 39%, and finally TER decreases by 19% and becomes 86%. Besides, while the amount of BLEU score change is the same (±14%) for both MT systems, the amounts of change in F-Measure and TER scores are greater in NMT engines as it can be observed from Figure 23.

5. 1. Automatic Evaluation of TR → EN Specific and Mixed Domain MT Engines

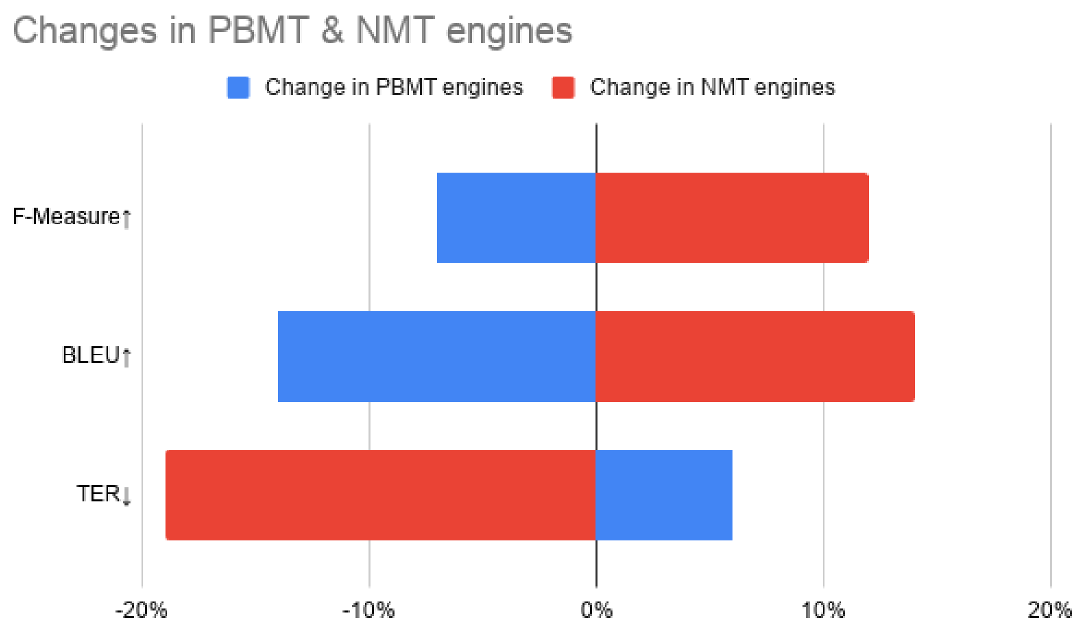


Figure 23. Amount of change in quality evaluation scores for NMT and PBMT engines.

As we have mentioned above, NMT is sensitive to the size and quality of a training corpus. In their study including incremental training data, Koehn and Knowles (2017) report that their NMT engine follows a steeper learning curve compared to their PBMT engine. We may expect a similar conduct from our engines. Besides, there also discussions of the reliability of these evaluation metrics for measuring the quality of NMT. In similar study comparing NMT and PBMT in terms of F-Measure, BLEU and TER for five language pairs, Shterinov et al. (2018, p. 8) hypothesize that "[…] F-measure, BLEU and TER underestimate the quality of NMT systems" and report that while PBMT engines yield better automatic scores, the results from human reviewers point out the contrary and give higher scores to NMT engines. This shows us that for a more complete view about the performance of an MT engine, human evaluation is necessary. In the following section, we report the human evaluation task that we performed to decide which engine is the best engine for TR → EN language pair.

5.2. Human Evaluation Results

Human evaluations of the 4 MT engines were conducted in KantanLQR platform by 5 professional translators native in Turkish. The quality evaluation is performed based on three metrics: adequacy, fluency and overall ranking.

100 Turkish segments were translated into English by our 4 different MT engines. In KantanLQR dashboard, an A/B Test project was created together with adequacy and fluency as additional key performance indicators (KPIs). Then, 100 translated sentences from each engine were imported into the project, and 5 translators were invited to connect to the dashboard for performing the evaluation. All the translators evaluated the same sentences and they did not know from which engines those translations were derived. In each window, the translators saw a source sentence, 4 different translations of this sentence (randomly ordered) and the scales for ranking, adequacy and fluency. When the quality is the same in two or more segments, they were able to assign the same score. Once they complete evaluating all the sentences, overall scores and detailed analysis of the results are shown in an analytics dashboard. Below we firstly describe the profiles of the task participants and then report the findings of this human evaluation task.

5.2.1 Profiles of the Human Evaluators

Before starting the evaluation task, the evaluators filled out a survey form related to their professional background. 5 native Turkish evaluators participated in the study. Considering their education background, %80 of them completed undergraduate studies and 20% of them also completed a master’s degree. And when it comes to their professional profile, all participants reported that they perform both translation and review services.

5.2. Human Evaluation Results

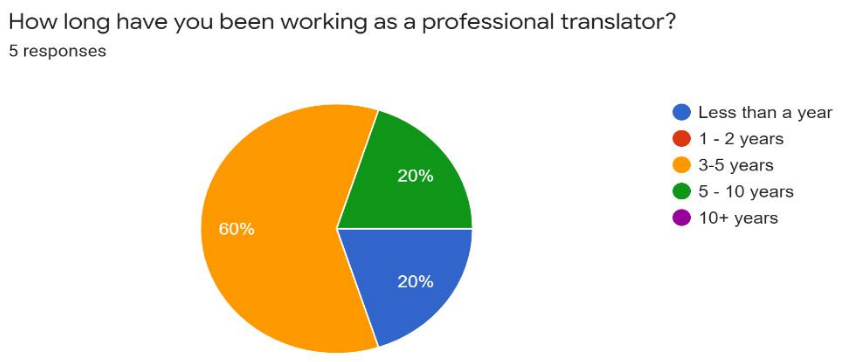


Figure 24. Experience as professional translators.

Their translation experience varies between “less than a year” to “5 – 10 years”, while 60% of them has reported that they have an experience of 3-5 years. Similarly, 60% of them reported that they use machine translation in their daily workflow.

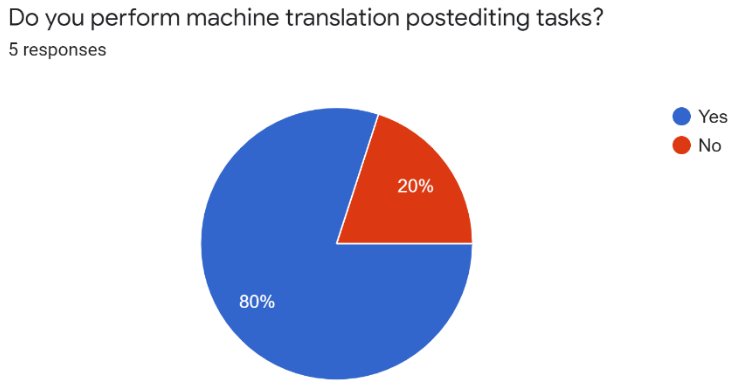


Figure 25. Answers to the question related to postediting.

Most of the evaluators (80%) also reported that they also perform postediting tasks. Lastly, since our test sample is derived from a cardiology corpus, we ask evaluators whether they have experience with medical translation. 60% of the participants said that they provide medical translation/editing/postediting services.

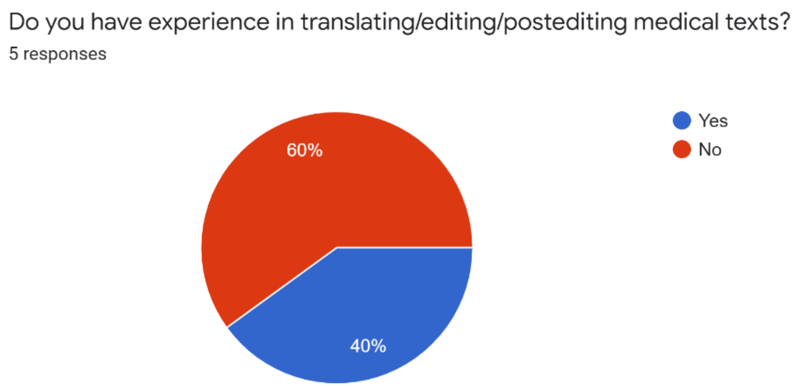


Figure 26. Medical translation/editing/postediting experience of the evaluators.

5.2.2 Evaluation Results

In this section, we report the results of the human evaluation task. 5 human reviewers completed the evaluation November 5, 2020 and November 12, 2020. Each reviewer has evaluated 100 sentences in terms of adequacy and fluency and ranked them from the best to the worst. In all the evaluation parameters, NMT-2 engine has achieved the best score. Below, we describe the results for each evaluation type.

Ranking. NMT-2 engine received the best score with a rate of 61.8% while NMT-1 is ranked as the worst engine with a rate of 32.8%. The scores for PBMT-1 and PBMT-2 in percentage form are 42.65% and 45.75%, respectively[112]. The second position is occupied by the PBMT-2 engine and the third position is occupied by PBMT-1 engine.

[112] The calculation of this percentage is provided here. There are 100 sentences and 5 reviewers; and the scoring is between 1-4. Hence, if each reviewer gives 4 points to each segment, the highest possible score for an engine is 2000. The percentages are calculated according to total score of each engine in proportion to the highest possible score.

5.2. Human Evaluation Results

Engine Name	Total Score	Percentage
PBMT-1	853/2000	42.65%
PBMT-2	915/2000	45.75%
NMT-1	656/2000	**32.8%**
NMT-2	1236/2000	**61.8%**

Table 34. Ranking scores for 4 engines.

Figure 28 shows the total ranking score each engine received from 5 reviewers. NMT-2 engine, which was trained on cardiology corpora and mixed domain corpora together, received significantly higher score compared to the other three engines. On the other hand, NMT-1 engine, which was trained on only cardiology corpora, received significantly lower score. Below we show the results of each engine with a distribution graphic starting from the worst engine to the best one.

NMT-1 engine received the lowest score 407 times (81.4%) out of 500 scoring instance[113]. It received the best score only 19 times (3.8%). Figure 27 shows the percentage distribution of the scores (1-4). This low score may be due to the low amount of training data, and having strictly specific domain data has not helped to provide high quality results.

[113] 100 sentences scored by 5 translators equal to 500 scoring instances.

5.2. Human Evaluation Results

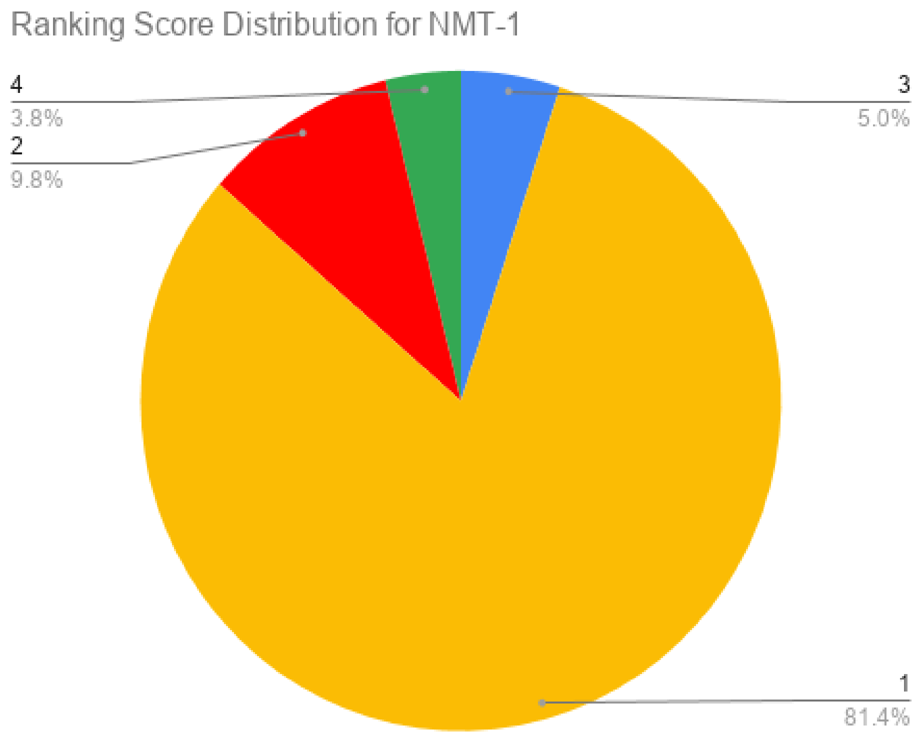


Figure 27. Ranking score distribution of NMT-1.

The engine in the third position is PBMT-1, the other engine with only cardiology corpora. This engine has also ranked very low with a ranking of the worst a total of 317

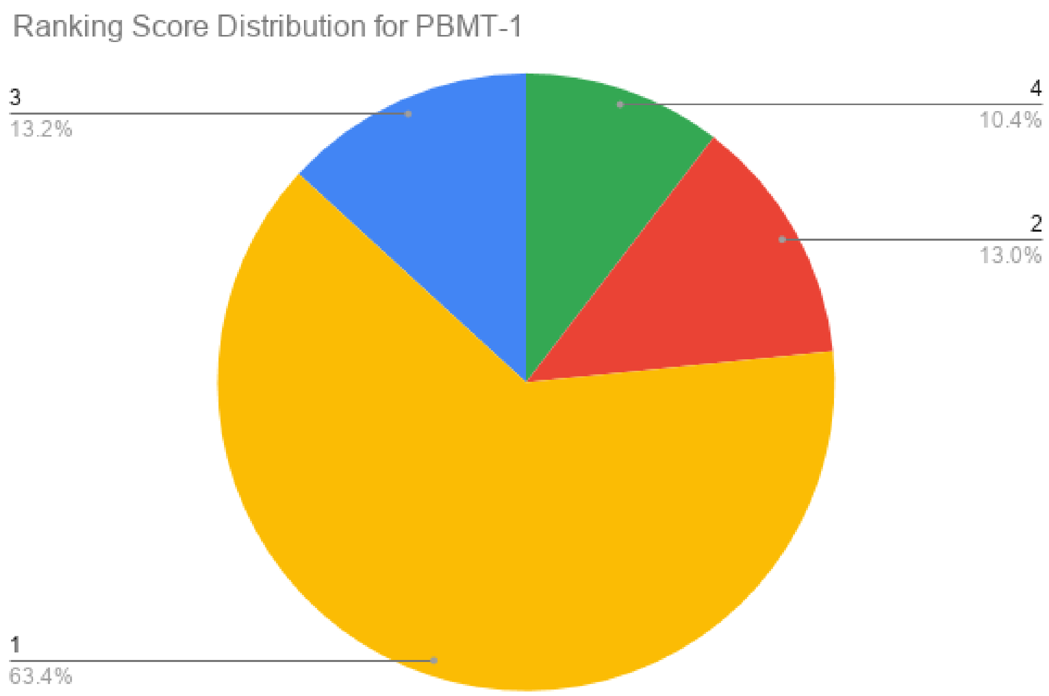


Figure 28. Ranking score distribution of PBMT-1.

times (63.4%). It only got the best ranking 54 times (10.4%). These 2 results above show that our reviewers ranked lower both of the less resourced engines.

The engine in the second position in the ranking is PBMT-2, the engine trained with cardiology corpora and mixed domain corpora on an SMT system. This engine received

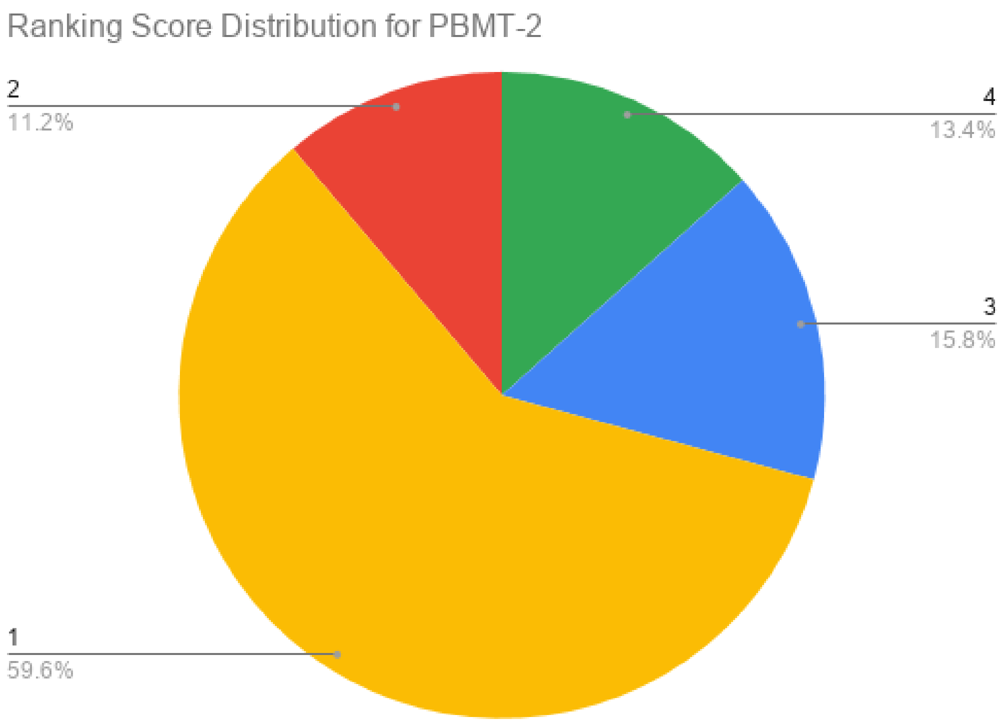


Figure 29. Ranking Score Distribution for PBMT-2.

the lowest ranking score 298 times (59.6%) and the highest ranking score 67 times (13.4%). These scores are slightly better than PBMT-1. However, they are still significantly lower than those of NMT-2.

NMT-2 engine is trained on cardiology corpora and mixed domain corpora using an NMT system. It received the best overall ranking score. Reviewers selected it as the best engine 167 times (33.4%), and as the worst engine 163 times (32.6%). Note that the ranking task was permissive in terms of assigning the same ranking score to multiple engines or not assigning the score of 4 to any engines.

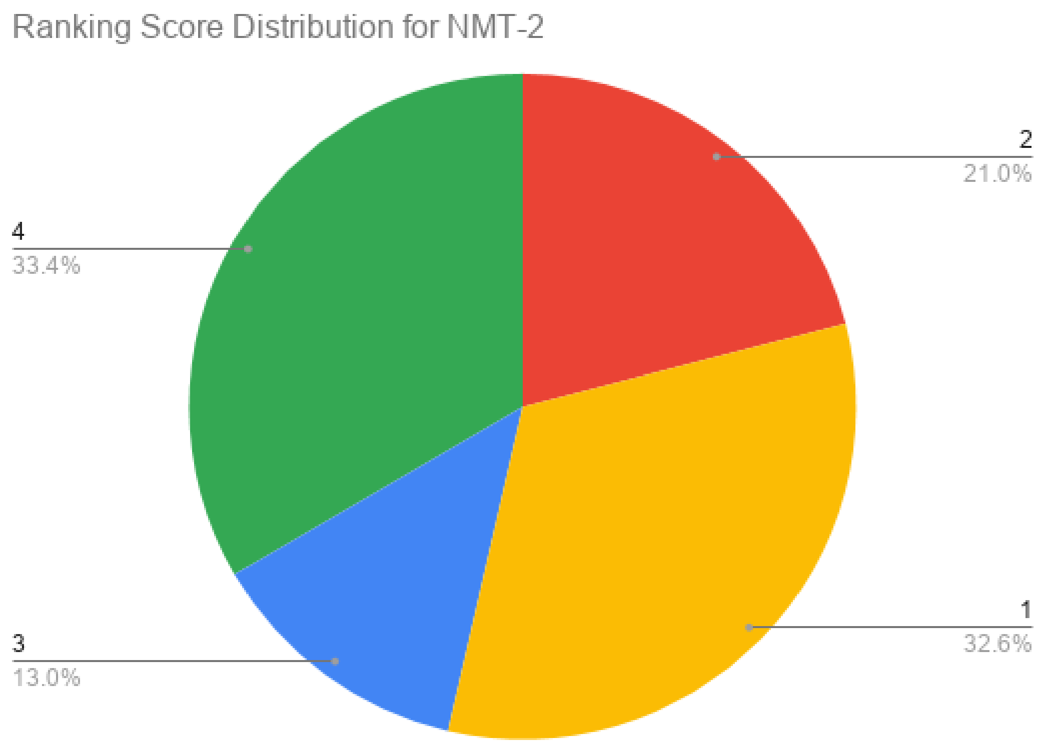


Figure 30. Ranking score distribution for NMT-2.

Finally, we look at the preferences of the reviewers based on their scores as calculated by KantanLQR. Figure 30 shows the percentage distribution of the scores given by each reviewer. Each color represents a different reviewer. It can be observed that 4 reviewers gave their highest scores to NMT-2 engine and only 1 reviewer (the reviewer in pale pink color) gave a slightly higher score to PBMT-2 compared to NMT-2 (78% to PBMT-2 and 77% to NMT-2). When it comes to the lowest ranking engine, again 4 reviewers has ranked NMT-1 as the lowest performing engine and only 1 reviewer (the reviewer in dark brown color) gave a slightly lower score to PBMT-1 and PBMT-2 (30% to NMT-1 and PBMT-2, and 29% to PBMT-1).

5.2. Human Evaluation Results

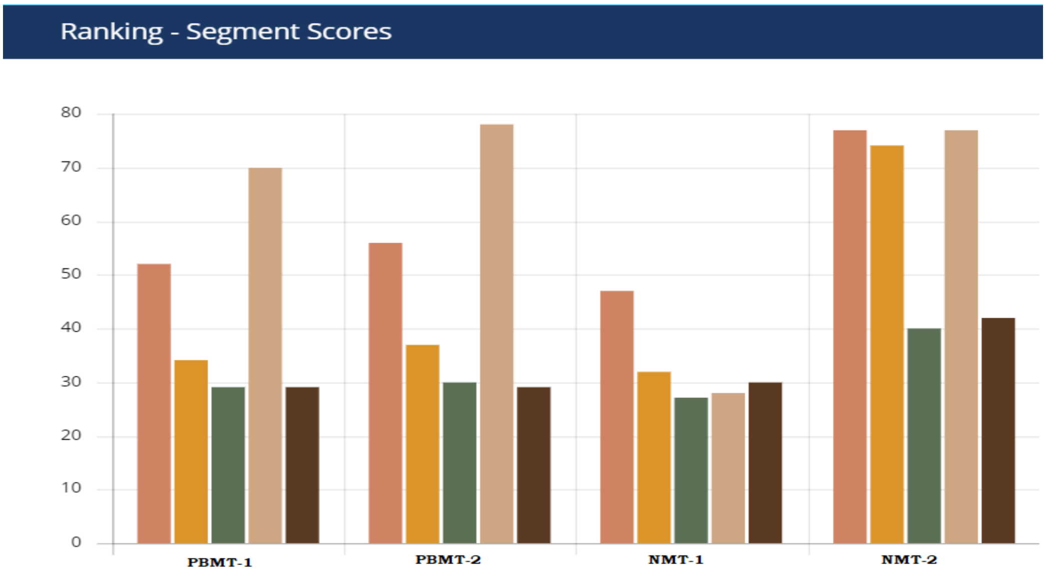


Figure 31. Ranking scores given by each reviewer, shown in percentages. Each color represents a reviewer.

The ranking task shows us that reviewers ranked higher the segments coming from the engines trained with higher amounts of data irrespective of the type of engine. It can also be inferred from this evaluation that increasing the amount of training data for NMT systems moves NMT engines from the worst to the best rank, and makes their results more preferable by reviewers.

Adequacy. The ranking task provides an overview about the performance of our MT engines from the perspectives of the human reviewers. For a fine-grained analysis, we request from the reviewers to rate the adequacy and fluency of the engines on a 1-5 star scale in the same task. Adequacy "is typically defined as the extent to which the translation transfers the meaning of the source-language unit into the target." (Castilho et al. 2018, p.18). In our operalizationalized setting, human reviwers assigned a score between 1 and 5 to the segments. A score of 1 means none of the meaning is expressed in the target sentence while a score of 5 means all the meaning is expressed in the target sentence.

5.2. Human Evaluation Results

NMT-2 received the highest score with an average adequacy score of 2.73 (54.64%)[114]. NMT-1 received the lowest score with an average adequacy score of 1.61 (32.24%). PBMT-1 and PBMT-2 received average scores of 1.90 (38.16%) and 1.99 (39.92%) respectively.

Engine Name	Total Score	Percentage
PBMT-1	954/2500	38.16%
PBMT-2	998/2500	39.92%
NMT-1	806/2500	**32.24%**
NMT-2	1366/2500	**54.64%**

Table 35. Overall adequacy scores and their percentage distribution.

It can be observed that none of the engines have a score over 4 (which means "much of the source segment meaning is expressed in the target"). Besides, as shown below, in none of the cases, the share of the score of 5 is over 10%. We show the percentage distribution of the scores (1-5) for each engine below starting with the best performing engine, NMT-2.

[114] There are 5 reviewers assigning adequacy scores of a scale of 5 for 100 sentences. Hence, the maximum numerical adequacy score that an engine can get is 2500. The percentages of our 4 engines are calculated in proportion to this maximum score. For example, NMT-1 has received a total score of 1366. The calculation of 1366 * 100 / 2500 gives us the adequacy percentage.

5.2. Human Evaluation Results

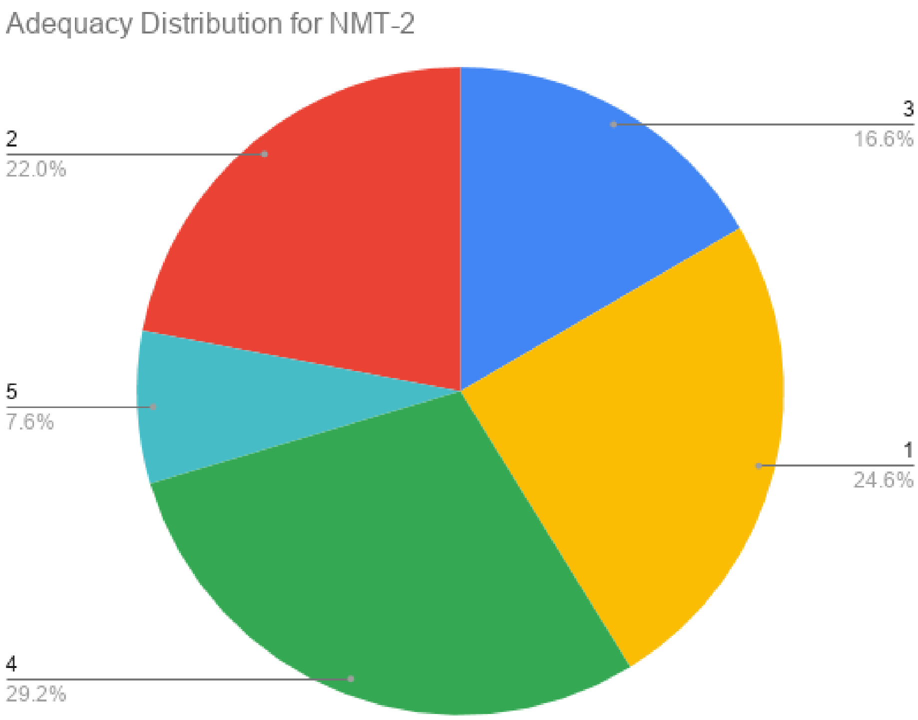


Figure 32. Percentage distributions of the adequacy scores.

Of all the total score of NMT-2, 7.6% of it is 5 stars, 29.2% of it is 4 stars, 16.6% is 3 stars, 22.0% is 2 stars and 24.6% is 1 star. It can be observed from the table that the

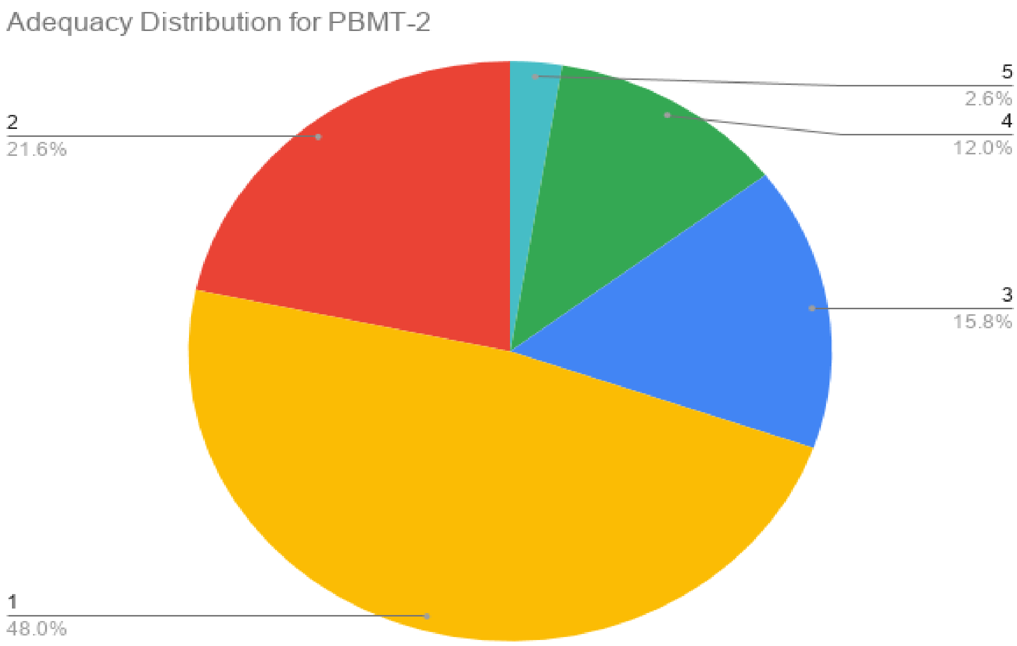


Figure 33. Percentage distribution of the adequacy scores of NMT-2 engine.

score with the highest percent is 4 stars. NMT-2 engine is followed by PBMT-2 engine. Only 2.6% of this engine received 5 stars, followed by 12.0% of 4 stars, 15.8% of 3 stars, 21.8% of 2 stars and 48.0% of 1 star. Most striking score in this engine is the score of 1 star with 48.0%, which implies that nearly half of the segments do not express the meaning of the source segments.

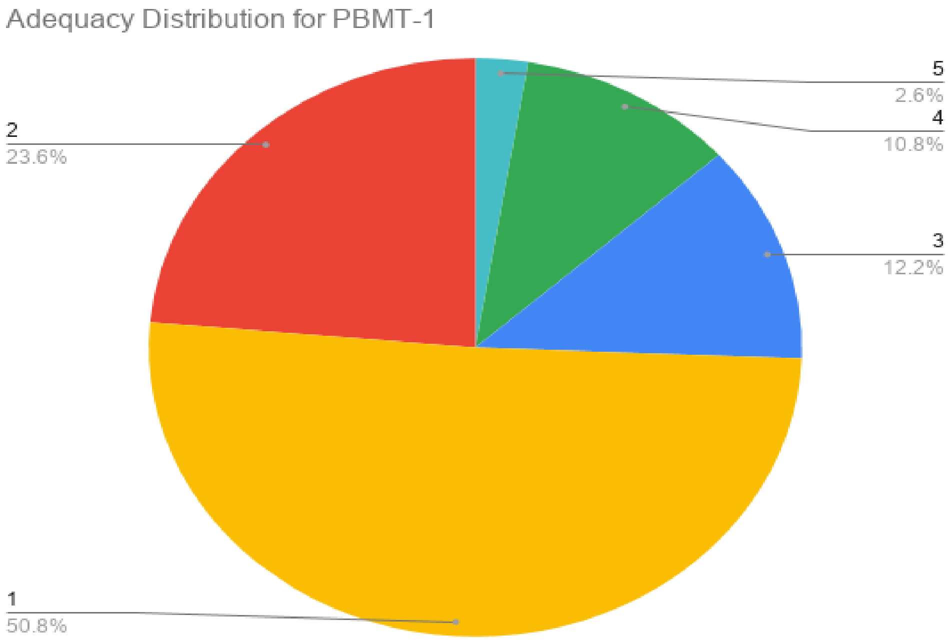


Figure 34. Percentage distribution of adequacy scores of PBMT-1 engine.

Similar to PBMT-2 engine, PBMT-1 engine also has slightly more than half of the segments with a score of 1 stars (50.8%). The percentage of 5 stars is 2.6% (same as PBMT-2), followed by 10.8% of 4 stars, 12.3% of 3 stars and 23.6% of 2 stars.

Finally, NMT-1 is in the last position with its scores. A daunting 68.4% of the scores given by human reviewers is 1 star. Only 1.0% of the score is 5 star. The 4 star has a share of 9.8% while 3 star has 7.0% and 2 star has 13.8%.

5.2. Human Evaluation Results

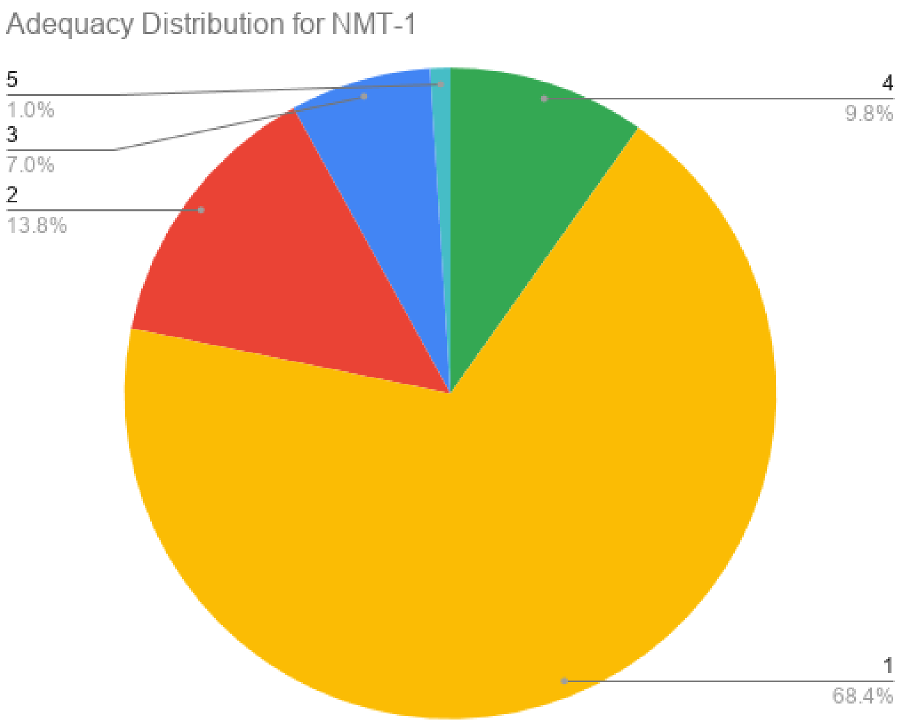


Figure 35. Percentage distribution of the adequacy scores of NMT-1.

The difference between the NMT-2's 1 star score and NMT-1's 1 star is remarkable: 24.6% and 68.4%. Considering that these two engines vary in terms of the amount of bilingual training corpora, we can conclude that increasing the amount of training data significantly improves the adequacy of NMT systems. When it comes to the difference between PBMT-2 and PBMT-1, the percentage distribution of the scores is very similar.

Fluency. Fluency tasks focus on the target sentences and consider if the target language grammar rules are followed, word choice is good and grammatical structure is appropriate. In the scale of KantanLQR, the highest score (5 stars) means "[n]ative language fluency. No grammar errors, good word choice and syntactic structure. No post-editing required." And the lowest score means "[n]o fluency. Absolutely ungrammatical and for the most part doesn't make any sense. Translation has to be re-written from scratch." (see Annex III for the scale and definitions of fluency). Table 33 shows the fluency scores of each engine.

5.2. Human Evaluation Results

Engine Name	Total Score	Percentage
PBMT-1	913/2500	**36.52%**
PBMT-2	959/2500	38.36%
NMT-1	1031/2500	41.24%
NMT-2	1428/2500	**57.12%**

Table 36. Total fluency scores of each engine. Highest and lowest percentages are shown in bold.

It can be observed that NMT engines received higher scores compared to PBMT engines. NMT-2 engine is the most fluent engine and PBMT-1 is the least fluent engine. Below we report the percentage distributions of each engine.

Although NMT-2 engine has the highest fluency score, the percentage of 5 stars is very low with a rate of 8.4%. This implies that the number of target sentences with native target language fluency is very low. However, NMT-2 engine has significantly lower number of target sentences with 1 star (20.8%) compared to other three engines (NMT-1: 46.6%, PBMT-2: 49.4%, PBMT-1: 54.0%).

5.2. Human Evaluation Results

Unlike ranking and adequacy tasks, the second position is occupied by NMT-1. NMT-

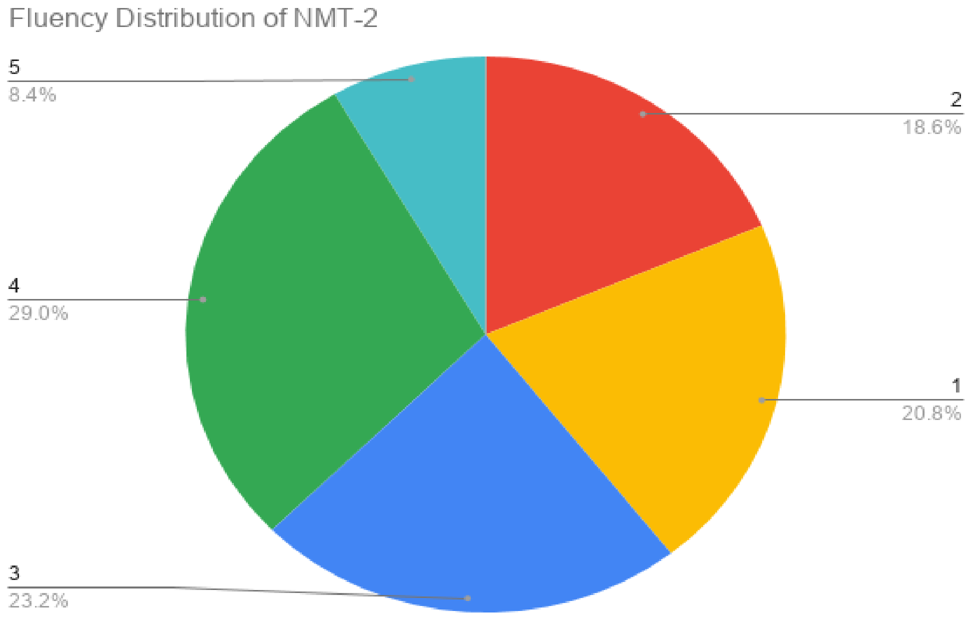


Figure 36. The percentage distribution of fluency scores for NMT-2.

1 has a 5-star rate of 1.4% but its 4-star and 3-star rates are higher than PBMT-2.

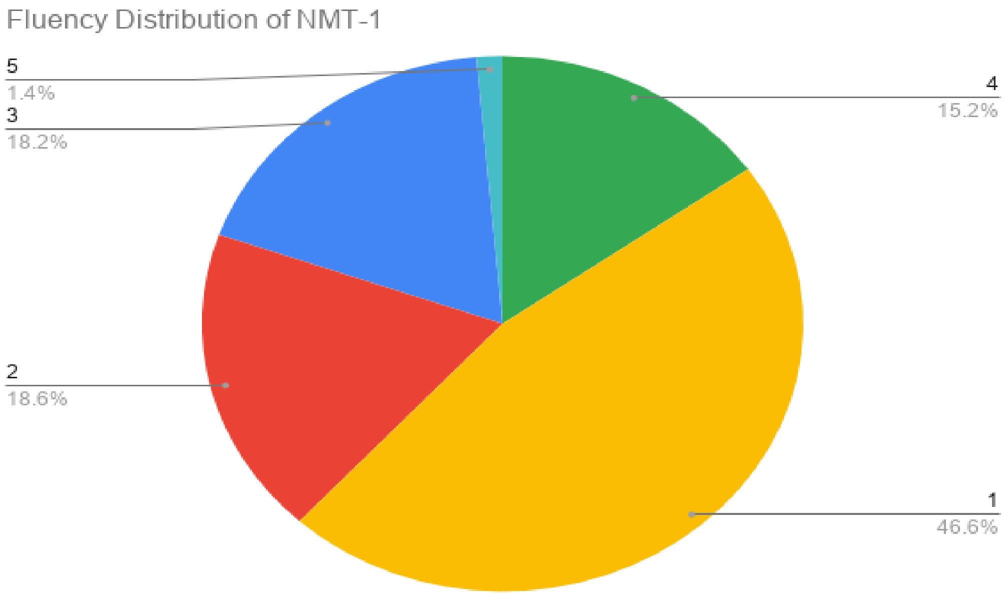


Figure 37. The percentage distribution of fluency scores for NMT-1.

5.2. Human Evaluation Results

PBMT-2 is in the third position with a 5-star rate of 2.8% and 1-star rate of 49.4%. While the percentages of this engine are close to NMT-1, PBMT-2 is slightly less fluent.

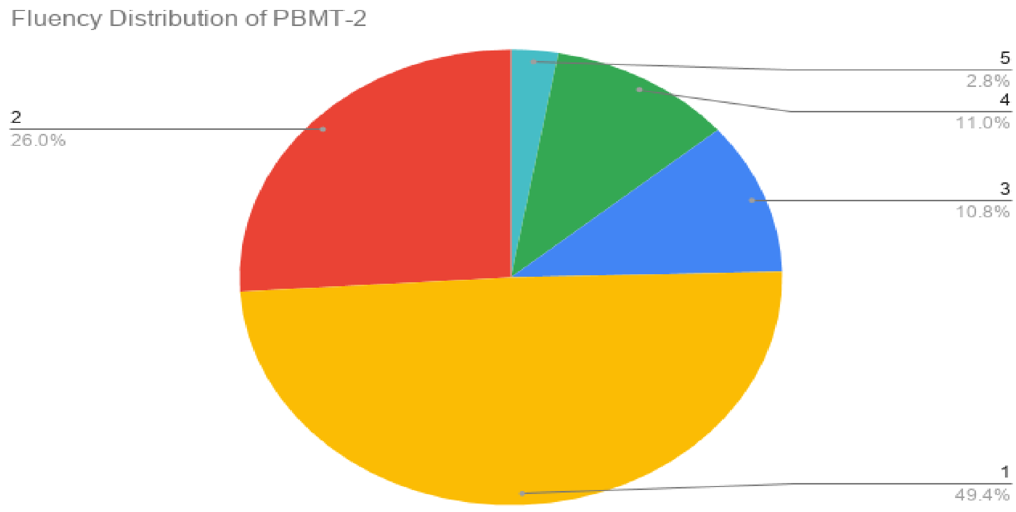


Figure 38. The percentage distribution of fluency scores for PBMT-2.

The least fluent engine is PBMT-1 engine with more than half of the target sentences (54%) receiving a score of 1-star. However, the rate of 5-star is the same as that of PBMT-2.

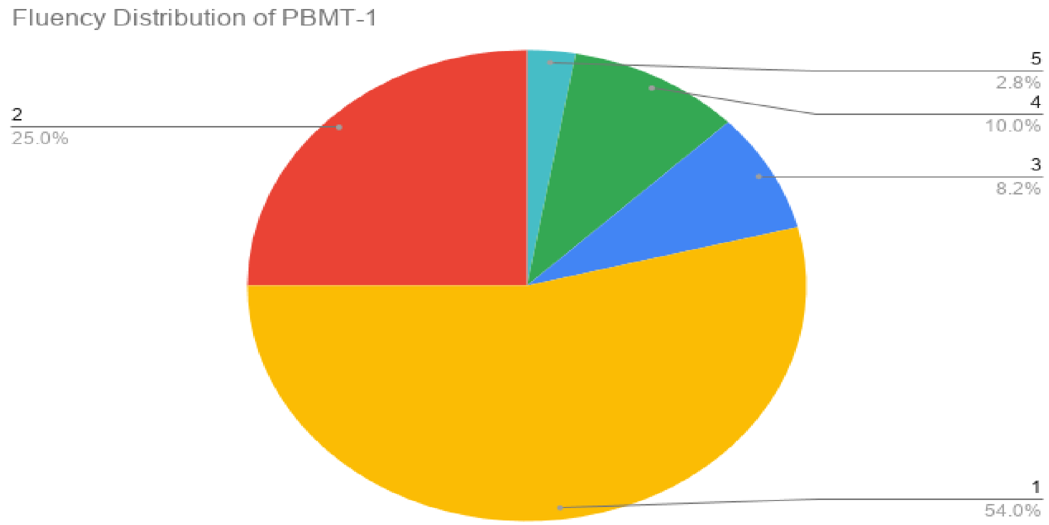


Figure 39. Percentage distribution of fluency scores for PBMT-1.

5.2.3 Final Remarks

It can be observed that in all the human evaluation tasks NMT-2 performed the best. Although this engine had mixed domain corpora, it performed better than PBMT-1 and NMT-1 engines which have only specific domain cardiology-based corpora. Besides, it performed better than the two SMT engines (PBMT-1 and NMT-1).

Engine Name	**Ranking**	**Adequacy**	**Fluency**
PBMT-1	42.65%	38.16%	**36.52%**
PBMT-2	45.75%	39.92%	38.36%
NMT-1	**32.8%**	**32.24%**	41.24%
NMT-2	**61.8%**	**54.64%**	**57.12%**

Table 37. An overview of percentages of the human evaluation scores.

There are a few observations related to corpus size, system type and corpus type that we can make based on the human evaluations. Concerning adequacy in the context of customized Turkish to English MT, it seems that corpus size is more important than system type and corpus type since NMT-2 and PBMT-2 engines perform better than the other two engines. Concerning fluency in the same context, system type is more important than the corpus size and corpus type since both NMT engines performed better than both PBMT engines. Yet, NMT-2 is better than NMT-1 and PBMT-2 is better than PBMT-1, which implies that larger corpus size leads to better fluency. And the overall ranking is compatible with adequacy results: more corpora have led to better scores, NMT-2 being the best, followed by PBMT-2.

5.3. Discussion - Final Remarks – Correlation between automatic evaluation vs human evaluation

Having conducted an automatic and human evaluation, we can now make overall observations about the evaluation results and discuss them in comparison to other studies.

The first observation is that automatic evaluation scores do not correlate with human evaluation scores in terms of best performing engine. PBMT-1 engine is the best performing engine in terms of F-Measure and TER scores and the second best performing engine in terms of BLEU score. However, PBMT-1 engine is in the third position in ranking and adequacy scores and the worst performing engine in terms of fluency score. The best performing engine in all human evaluation tasks is NMT-2 engine. This shows that our finding is compatible with the hypothesis of Shterionov, et al. (2018) claiming that automatic evaluation scores underestimate the actual quality of NMT engines. They train PBMT and NMT engines for five language pairs and evaluate the MT quality with human and automatic evaluation metrics. However, their study is conducted with engines having a corpora size of more than 35 million words at least, which is significantly larger than our corpora. Our NMT-1 engine with low amount of corpora received low scores both in human and automatic evaluation. That’s why we should add an exception to this hypothesis: automatic scores correlate with human evaluation scores when NMT engine is trained with low amount of corpora (at least in the context of Turkish → English MT).

5.3. Discussion - Final Remarks – Correlation between automatic evaluation vs human evaluation

Having the NMT system with more parallel corpora as the best and the NMT system with less parallel corpora as the worst indicates that NMT systems are very sensitive to the amount of training data. Koehn&Knowles (2017) train PBMT and NMT engines with different amounts of training data and observe that the quality of NMT follows a steeper curve compared to PBMT in terms of BLEU score. Both our human evaluation and automatic evaluation scores confirms this observation for Turkish to English MT. Considering the evolution from NMT-1 to NMT-2, the ranking, adequacy and fluency scores increased by 29%, 22% and 15.88%, respectively. There is also an improvement in PBMT when corpus size is increased; yet, the increase rate is quite low compared to that of NMT. Ranking, adequacy and fluency increased by 3.10%, 1.76% and 1.84%.

One expectation (an implicit hypothesis) of the study was that engines with strictly narrow domain corpora will perform better than the ones with mixed domain corpora. However, at least according to the human evaluation scores, this has not been the case. Besides, our specific domain engines (PBMT-1 and NMT-1) have significantly lower amount of corpora than those with our mixed domain corpora. Hence, we cannot reach a final conclusion about the effect of corpus type unless in both scenarios the engines have the same amount of parallel corpora. In other words, a future study with same size specific domain parallel corpora and mixed domain parallel corpora can be conducted in Turkish – English language pair.

Chapter 6. Terminological Errors and Their Frequencies in SMT and NMT

It can hardly be a coincidence that no language on Earth has ever produced the expression "as pretty as an airport".

-Douglas Adams

6.1. Term Annotation of the Sample Set

In this chapter, we report the results of the terminology annotation and terminology error categorization and analysis in our 4 engines. We firstly report the terminology annotation results from the Turkish → English sample set in which cardiology-related bilingual terms were extracted. In the second part, we expose the results of error categorization in two levels. The first level includes a binary classification of correct/incorrect term translations according to the methodology described in Chapter 4. In the second, fine-grained level, incorrect terms are analyzed more profoundly and are labelled with 11 error categories while correct terms are further labelled with 4 correct term levels. After completing the reporting of the results, we discuss the relative strengths and weaknesses of each engine, comment on the performance of each engine and compare our findings to other studies and report the limitations of the study methodology.

6.1. Term Annotation of the Sample Set

During the sentence-by-sentence term annotation process, 231 cardiology-related term pairs were identified. 35 source terms occurred in more than one sentence; hence, the total count of unique (considering morphological form, not the conceptual meaning) terms is 196. We separated these terms into 5 categories considering the number of n-grams that they have in order to evaluate later if n-gram size affects term translation quality in MT systems. The table below shows the n-gram distributions of source and target terms in our sample set.

	1-gram	**2-gram**	**3-gram**	**4-gram**	**>4-gram**
Turkish	43	69	68	41	10
English	46	72	67	32	14

6.1. Term Annotation of the Sample Set

Table 38. n-gram distribution of source and target terms in the sample set.

As it can be observed, majority of the terms have either 2 grams or 3 grams. Note that the term count above does not include acronyms. Aside from the 231 terms, 67 acronyms were identified. Hence, in total, 298 terminological units were studied. Full list of the terms and acronyms are given in Annex VI. In this list, each sentence had at least 1 term and the maximum number of terms in a sentence was 6.

Within this terminology set, there were some synonyms as well. We identified 8 occurrences of synonyms within the sample set. As we mentioned in Chapter 2, the use of Greco-Latin rooted terms such as the cases of “diabetes mellitus”, “inflamasyon” and “trombüs” is common in Turkish. We observed that some sentences included Turkish equivalents such as “şeker hastalığı,” “yangı” and “pıhtı” instead of these variants. The rest of the denominational variations were due to morphological variations such as “karotis arter stentleme” vs “karotis arter stentlemesi”. One case (“primer perkütan coroner girişim”) is partially abbreviated as “primary PCI” in the target translation.

1. karotis arter stentleme → carotid artery stenting karotis arter stentlemesi → carotid artery stenting
2. diabetes mellitus → diabetes mellitus şeker hastalığı → diabetes mellitus
3. kalp yetersizliği → heart failure kalp yetmezliği → heart failure
4. yangı → inflammation inflamasyon → inflammation
5. sol ventrikül fonksiyonu → left ventricular function sol ventrikül fonksiyon → left ventricular function
6. primer perkütan koroner girişim → primary PCI primer perkütan koroner girişim → primary percutaneous coronary intervention

7. ST yükselmeli miyokart enfarktüsü → ST-elevation myocardial infarction ST yükselmeli miyokart enfarktüsü → ST-segment elevation myocardial infarction ST-segment yükselmeli miyokart enfarktüsü → ST-segment elevation myocardial infarction ST-segment yükselmeli miyokart enfarktüsü → ST-segment myocardial infarction
8. pıhtı → thrombus trombüs → thrombus

Table 39. Synonyms in the sample set. Source terms and their translations are presented as they occur in the sample set.

Finally, in three cases, we had to include compound terms which consisted of more than one term as a single term unit for the purposes of term error analysis. For instance, tr. "eşzamanlı obstrüktif aortik ve mitral prostetik kapak trombozları" = en. "concurrent obstructive aortic and mitral prosthetic valve thrombosis" includes two concepts: "concurrent obstructive aortic prosthetic valve thrombosis" and "concurrent obstructive mitral prosthetic valve thrombosis". Since this phrasal unit with 2 concepts is nested in such a way that does not allow for analyzing the translation of two terms separately, we considered this phrasal unit as a single term at the error analysis level. Having only 3 cases out of 231 terms (and 67 acronyms) like this decreases the chance of affecting the overall result of the analysis. Hence, these 3 cases are included in our analysis.

We publish the annotated source and target sentences as a free and open corpus in an open repository[115] as a research material for terminology evaluation in Turkish → English MT to be used by MT researchers. The same repository also hosts the TRENCARD and GENCOR corpora.

[115] Turkish English Parallel Corpora and MT Evaluation Results. https://github.com/gokhandogru/TurkishEnglishParallelCorporaandMTEvaluation (last access: 09.02.2021)

6.2. Results for Terminology Errors in 4 MT Engines

We analyzed term errors in 4 MT engines by assigning terminology error categories to each incorrect term translations. Note that each incorrect term translation is only assigned one error category which, in some cases, let to difficulty of deciding which error category to assign when more than one error category assignation is possible. We will handle these issues in discussion part of the chapter. Below we report the results of each engine. Table 43 includes a summary of error annotation results for all the engines.

6.2.1. Term Error categorization Results for PBMT-1

Our manual evaluation firstly concentrated on a binary analysis of correct/incorrect term translation in PBMT-1, and then further analyzed the correct terms and incorrect terms to annotate each term instance with a fine-grained correct term or incorrect term class. Out of the 298 terms evaluated, 204 (68.46%) terms were translated correctly while 94 (31.54%) terms were translated incorrectly by PBMT-1.

Considering the 204 correct terms, 145 terms were equal to the terms in the reference human translation (category 1: correct term considering reference sentence). 14 terms were not equal to the reference sentence but were correct synonyms based on the reference resources we resorted to. We annotated 1 term translation as a possible equivalent and finally, 45 acronyms were correctly translated into English.

When it comes to error categories, the most common error type in the translations of PBMT-1 was the partial translation error category with 46 errors. And the second most frequent error was acronym error with 22 errors in this class. The other error types and their frequencies are as follows: Incorrect term equivalent: 6, Morphosyntactic error: 2, 6.2. Results for Terminology Errors in 4 MT Engines

Reordering error: 2, Source term insertion: 8, Term drop: 5, Term extended: 1, Term with incorrect grammatical category: 1, Literal Translation: 0, Other term errors: 0. Complete list of term error annotations are included in Annex VIII and Annex IX.

Considering the term n-gram length based translation quality of PBMT-1, correct terms versus incorrect terms are as follows: 1-gram: 30 vs. 13, 2-gram: 49 vs. 20, 3-gram: 41 vs. 27, 4-gram: 37 vs. 4 and >4-gram: 3 vs. 7.

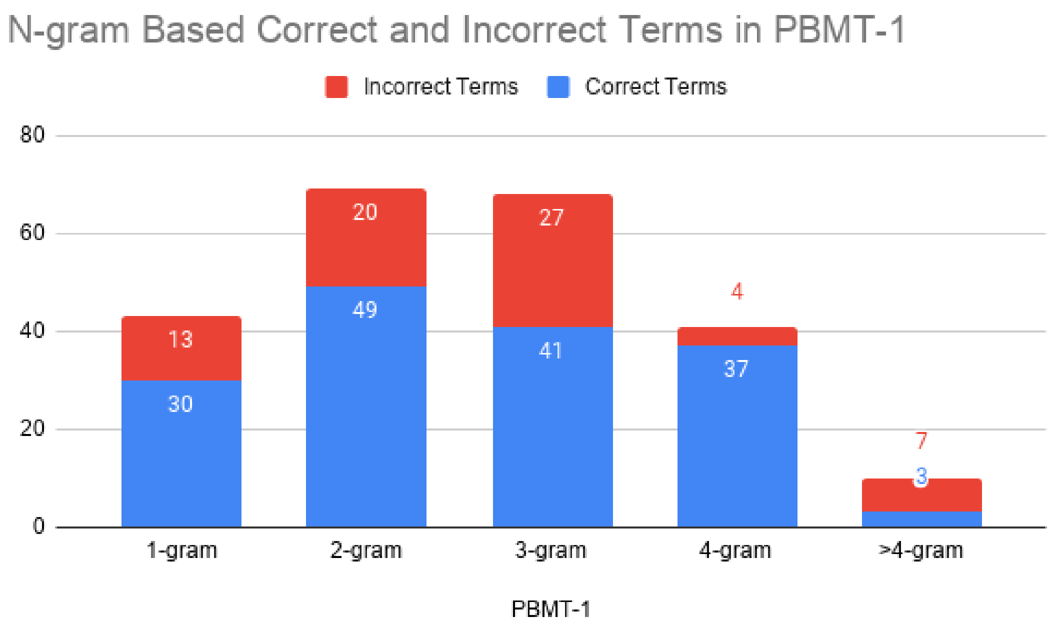


Figure 40. N-gram distribution correct and incorrect term translation in PBMT-1.

It can be observed that only in the case of >4-gram, the number of error is larger than the number of correctly translated terms. It should also be highlighted that 90.24% of the 4-gram terms are translated correctly by the engine. Percentage of correct term translation for the other n-grams are as follows: 1-gram: 30 (69.76%); 2-gram: 49 (71.01%); 3-gram: 41 (60.29%) 4-gram (90.24%), and >4-gram (30%).

Finally, we look at the correlation between sentence length and term translation. In 5-10-word sentences 26 terms were translated correctly while 13 term translations were erroneous. In 11-20-word sentences, we had 68 correct terms and 33 incorrect terms. And lastly, >20-word sentences had 66 correct terms and 25 incorrect terms.

Sen. Length	5-10 words	11-20 words	>20 words
Correct Term	26 (66.67%)	68 (67.33%)	66 (72.53%)
Incorrect Term	13 (33.33%)	33 (32.67%)	25 (27.47%)

Table 40. Term translation in PBMT-1 according to sentence length.

After completing the reporting of the results of the other engines, we will contrast the PBMT-1 results with other engines. Note that as we mentioned in Chapter 4, correct and incorrect acronym translations are not included in n-gram analyses and sentence length-based analysis in none of the engines.

6.2.2. Term Error categorization Results for PBMT-2

PBMT-2 is the second engine which we evaluated regarding the term translation quality. As described in section 3.3.1, this engine differentiates from the previous engine in view of having a mixed domain corpus together with the specific domain cardiology corpus in the training. 209 (70.13%) terms were translated correctly by this engine while 89 (29.87%) terms had errors.

Looking at the 209 correct terms, 149 terms were equal to the terms in the reference human translation. 15 terms were not equal to the reference sentence but were correct synonyms based on the reference resources we resorted to. We included 2 term translation as a possible equivalent and finally, 43 acronyms were correctly translated into English.

6.2. Results for Terminology Errors in 4 MT Engines

After correct translation categories we analyzed error category. According to this analysis, the most common error type in the translations of PBMT-2 was the partial translation error category with 42 cases. And the second most frequent error was acronym error with 24 errors in this class. The other error types and their frequencies are as follows: Incorrect term equivalent: 3, Morphosyntactic error: 2, Reordering error: 3, Source term insertion: 6, Term drop: 3, Term extended: 1, Term with incorrect grammatical category: 3, Literal Translation: 1, Other term errors: 1. Complete list of term error annotations can be found in Annex VIII and Annex IX.

The source term n-gram based analysis results for PBMT-2 comparing correct terms and incorrect terms are as follows: 1-gram: 34 vs. 9, 2-gram: 49 vs. 20, 3-gram: 44 vs. 24, 4-gram: 36 vs. 5 and >4-gram: 3 vs. 7.

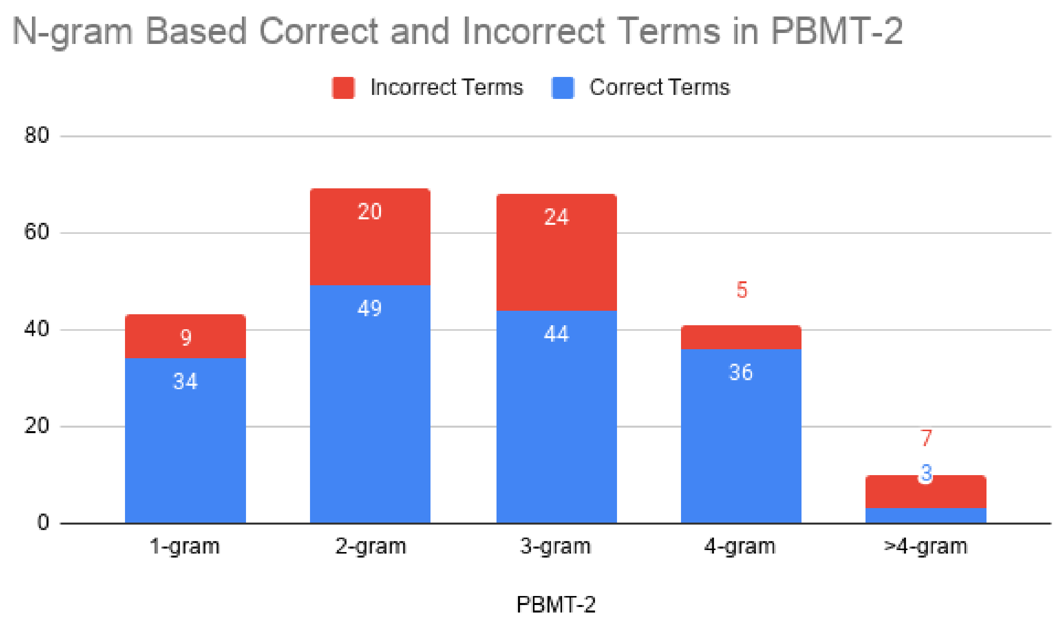


Figure 41. N-gram distribution correct and incorrect term translation in PBMT-2.

Similar to the PBMT-1, only in >4-gram terms, the incorrect term rate (70%) is greater than correct term rate (30%) in PBMT-2. Percentages of correct term translation based on n-gram are as follows: 1-gram: 34 (79.06%), 2-gram: 49 (71.01%), 3-gram: 44 (64.70%), 4-gram: 36 (87.80%), >4-gram 3 (39%).

Lastly, we report the sentence length based term translation results. We found that in 5-10-word sentences, 26 terms were translated correctly while 13 term translations were incorrect. In 11-20-word sentences, we had 70 correct terms and 31 incorrect terms. And finally, >20-word sentences had 70 correct terms and 21 incorrect terms.

Sen. Length	5-10 words	11-20 words	>20 words
Correct Term	26 (66.67%)	70 (69.31%)	70 (76.92%)
Incorrect Term	13 (33.33%)	31 (30.69%)	21 (23.08)

Table 41. Sentence length based term translation results in PBMT-2.

With these evaluation results, we completed the reporting of the term translation evaluation in SMT engines. The two sections below, we report the results for the same parameters for the 2 NMT engines.

6.2.3. Term Error categorization Results for NMT-1

NMT-1 engine was trained on solely bilingual cardiology corpus. The term translation results for this engine are significantly different. First of all, only 50 (16.78%) term translations are correct while 248 (83.22%) terms were not translated correctly.

In a more fine-grained look at correct term translations, 38 term translations were correct considering the reference terms; 2 term translations had correct synonyms, 2 were possible equivalents and 8 were correct acronyms.

There are significantly more incorrect term translations than correct ones in NMT-1. The most common error category was term drop with 125 cases. And the second most frequent error was acronym error with 59 errors in this class. The other error types and their frequencies are as follows: partial term translation: 36, Incorrect term equivalent: 11, Morphosyntactic error: 2, Reordering error: 1, Source term insertion: 0, Term extended: 3, Term with incorrect grammatical category: 0, Literal Translation: 0, Other term errors: 11.

The source term n-gram based analysis results for NMT-1 comparing correct terms and incorrect terms are as follows: 1-gram: 2 vs. 41, 2-gram: 18 vs. 51, 3-gram: 14 vs. 54, 4-gram: 33 vs. 8 and >4-gram: 0 vs. 10. Complete list of term error annotations are presented in Annex VIII and Annex IX.

In all n-gram cases, incorrect term translation outweighs correct term translations. Percentages of correct term translation based on n-gram are as follows: 1-gram: 2 (4.65%), 2-gram: 18 (26.08%), 3-gram: 14 (20.58%), 4-gram: 8 (19.51%), >4-gram: 0 (0%).

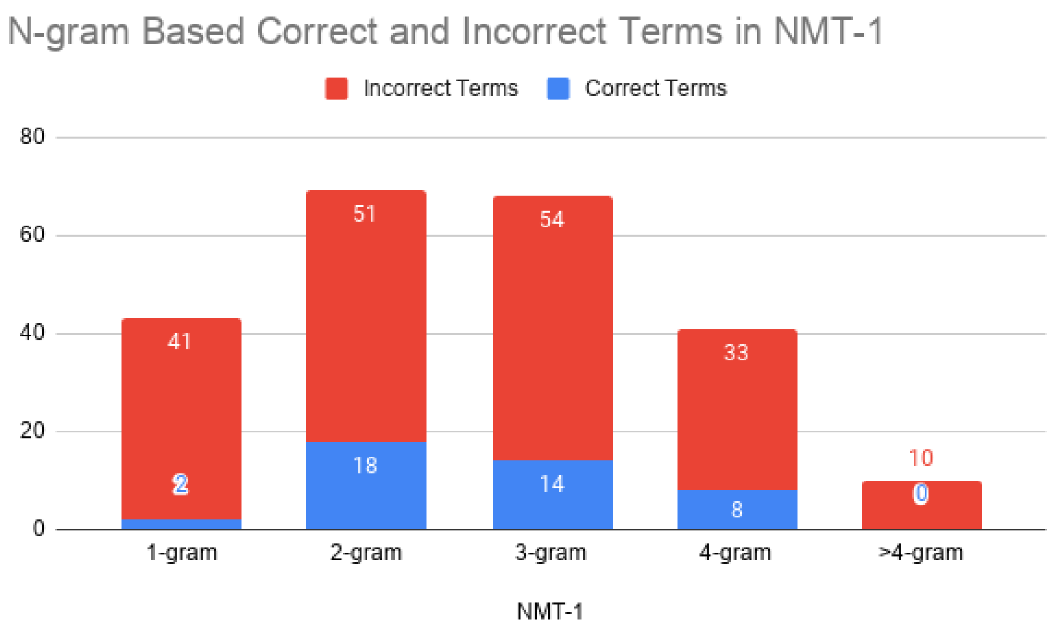


Figure 42. N-gram distribution correct and incorrect term translation in NMT-2.

Finally, we break down term translation results based on sentence length in NMT-1. We observed that in 5-10-word sentences, 5 terms were translated correctly while 34 term translations were incorrect. In 11-20-word sentences, we had 20 correct terms and 81 incorrect terms. And finally, >20-word sentences had 17 correct terms and 74 incorrect terms.

Sen. Length	5-10 words	11-20 words	>20 words
Correct Term	5 (12.82%)	20 (19.80%)	17 (18.68%)
Incorrect Term	34 (87.18%)	81 (80.20%)	74 (81.32%)

Table 42. Sentence length based term translation results in NMT-1.

As it can be observed from the table above, term translation errors are significantly higher than correct term translation in all sentence length groups.

6.2.4. Term Error categorization Results for NMT-2

We present the term translation evaluation results for NMT-2 in this section. NMT-2 engine was trained with specific domain cardiology corpora plus a mixed domain corpora. In the binary analysis, 185 (62.08%) correct term translations and 113 (37.92%) incorrect term translations were detected.

The count of correct term translations is as follows: correct term considering reference: 125, correct synonym: 22, possible equivalent: 6, correct acronym: 32. After correct translation categories we analyzed error category. Considering the term translation error categories, the most common error category in the translations of NMT-2 was term drop with 41 cases. And the second most frequent error was acronym error with 35 cases in this class. The other error types and their frequencies are as follows: partial term translation: 17, Incorrect term equivalent: 5, Morphosyntactic error: 2, Reordering error: 3, Source term insertion: 0, Term extended: 1, Term with incorrect grammatical category: 0, Literal Translation: 4, Other term errors: 5. Complete list of term error annotations are shown in Annex VIII and Annex IX.

The source term n-gram based analysis results for NMT-2 comparing correct terms and incorrect terms are as follows: 1-gram: 28 vs. 15, 2-gram: 43 vs. 26, 3-gram: 49 vs. 19, 4-gram: 28 vs. 13 and >4-gram: 5 vs. 5. Accordingly, the percentages of correct term translation based on n-gram are as follows: 1-gram: 28 (65.11%), 2-gram: 43 (62.31%), 3-gram: 49 (72.05%), 4-gram: 28 (68.29%) and >4-gram 5 (50%).

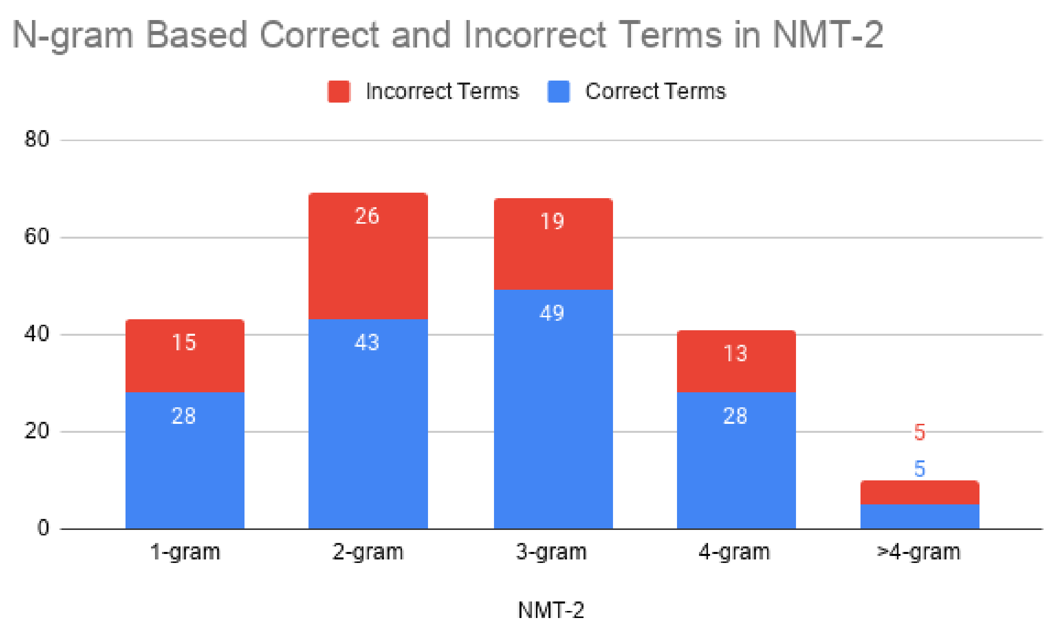


Figure 43. N-gram distribution correct and incorrect term translation in NMT-2.

Lastly, we present term translation results based on sentence length in NMT-2. We observed that in 5-10-word sentences, 28 terms were translated correctly while 11 term translations were incorrect. In 11-20-word sentences, we had 67 correct terms and 34 incorrect terms. And finally, >20-word sentences had 58 correct terms and 33 incorrect terms.

Sen. Length	5-10 words	11-20 words	>20 words
Correct Term	28 (71.79%)	67 (66.34%)	58 (63.74%)
Incorrect Term	11 (28.21%)	34 (33.66%)	33 (36.26%)

Figure 44. Sentence length based term translation results in NMT-2.

Having presented the terminology evaluation results for each engine, we compare and analyze the results from different angles including MT system type and relative corpus sizes in the next section.

6.3. Analysis of the Results and Discussion

In this section, the results from each engine are compared and interpreted. We analyze the error categories and their frequencies in the 4 MT engines based on corpus size and MT systems.

Considering the overall terminology evaluation, PBMT-2 has the highest number of correct term translations with 209 correct term instances, followed by PBMT-1 with 204 instances. The third position is occupied by NMT-2 with 185 instances and finally, NMT-1 has only 85 correct term translations. In parallel to these results, considering the term translation errors, NMT-1 has the highest number of errors with 248 term translation errors, followed by NMT-2 with 113 term translation errors, PBMT-1 with 94 errors and lastly PBMT-2 with 89 errors. In reference to these results, we can argue that in Turkish to English corpus-based MT, PBMT engines commit less term translation errors than NMT engines, PBMT-2 being the highest performing and NMT-1 being the lowest in view to term translation. In the following paragraphs, we analyze different error categories and their frequencies as committed by each engine to understand differences in translation behavior of different MT systems.

	PBMT-1	PBMT-2	NMT-1	NMT-2	Mean
a. Correct term considering reference	145	**149**	38	125	*114.25*
b. Correct synonym	14	15	2	**22**	*13.25*
c. Possible equivalent	1	2	2	**6**	*2.75*
d. Correct acronym	**45**	43	8	32	*32*
1. Partial term translation	**46**	42	36	17	*35.25*
2. Incorrect term equivalent	6	3	**11**	5	*6.25*
3. Morphosyntactic error	2	2	2	2	*2*
4. Reordering error	2	3	1	3	*2.25*
5. Source term insertion	**8**	6	0	0	*3.5*
6. Acronym mistranslation	22	24	**59**	35	*35*
7. Term drop	5	3	**125**	41	*43.5*
8. Term extended	1	1	**3**	1	*1.5*
9. Term with incorrect grammatical category	1	**3**	0	0	*1*
10. Literal Translation	0	1	0	**4**	*1.25*
11. Other term errors	0	1	**11**	5	*4.25*

Table 43. Distribution of the correct term translations and term translation errors in 4 engines and averages of each class.

The table above summarizes the fine-grained distribution of term error categories and correct term translations in 4 engines. Besides, total averages for each category is provided in the last column.

Let us begin with correct term translations. According to our results, PBMT-2 has the highest number of correct term translations (149) considering reference (a) in the table above while NMT-1 has the lowest count in this category followed by PBMT-1 with 145 correct term translations. The second correct term translation category is related to synonyms (b) and NMT-2 has the highest count while NMT-1 has the lowest one. In the possible equivalent correct term category (c), again NMT-2 has a higher count than the other engines. NMT-2's superiority with (b) and (c) may be derived from its use of sub-word units (cf. section 1.2.4 for byte pair encoding) which leads to more "creative" translations. It seems that not having the same results in NMT-1 implies that proper applicability of this feature requires sufficient training corpus. Lastly, concerning correct acronyms (d), PBMT-1 has the highest count with 44 correct acronym translations but PBMT-2 follows it with only one less correct term translation (43 correct acronym translations). NMT-1 only has 8 correct acronym translations.

Partial term error category was most common in PBMT-1. PBMT-2 occupies the second place. In both of these SMT systems, partial term error was the most widespread error type. NMT-2 had the lowest amount of partial errors. Incorrect term equivalent category was assigned to term translations that are totally incorrect, and NMT-1 was the engine with the highest count while PBMT-2 had the lowest count for this error category. Number of morphosyntactic errors were low and equal in all the engines. Reordering errors were also low (1-3) and all counts were around the average of 2.25. When it comes to source term insertion class, PBMT-1 had 8 cases and PBMT-2 had 6 cases while NMT-1 and NMT-2 did not have any cases of source term insertions. NMT-1 commits by far the highest number of acronym errors while NMT-2 is slightly below the average 35.25. PBMT-1 and PBM-2 are again very close to each other with 23 and 24 errors, respectively. Term drop error category is committed significantly more by NMT-1 with a 125 cases. The second closest count belongs to NMT-2 with 41 cases. Compared to these 2 NMT engines, PBMT-1 and PBMT-2 only have 5 and 3 term drop errors. Term extension errors in which terms had modifiers that changed their meanings radically were not very common. NMT-1 had 3 such cases and the other engines only 1 case. Term with incorrect grammatical category was not frequent either. Only PBMT engines had such errors with PBMT-1 having one such case while PBMT-2 had 3. NMT-2 had more literal translation errors (4) than the other engines. PBMT-1 and NMT-1 didn’t have any cases while PBMT-2 had only 1. Errors that did not have any of the errors above were classified under other term error. The specific domain NMT-1 engine produced the highest count of other term errors with 11 cases. NMT-2 had 5 cases while PBMT-1 had 0 and PBMT-2 had 1.

6.3. Analysis of the Results and Discussion

6.3.1. The Evolution of Term Errors Based on Corpus Size in MT Systems

In this section, we analyze how the corpus size increase has influenced term translation in each MT system type. Coupling the specific domain training corpora with mixed domain corpora had an overall positive outcome in both PBMT and NMT considering the improvement in term translation errors. The table below implies that the term translation quality has slightly improved in PBMT systems while it has significantly improved in NMT systems after the training with more corpora.

	Change from PBMT-1 to PBMT-2	Change from NMT-1 to NMT-2
a.Correct term considering reference	**4**	**87**
b.Correct synonym	**1**	**20**
c. Possible equivalent	**1**	**4**
d. Correct acronym	-2	**24**
	Error categories	
1.Partial term translation	**-4**	**-19**
2.Incorrect term equivalent	**-3**	**-6**
3.Morphosyntactic error	0	0
4.Reordering error	1	2
5.Source term insertion	**-2**	0
6.Acronym mistranslation	2	**-24**
7.Term drop	**-2**	**-84**
8.Term extended	0	**-2**
9.Term with incorrect grammatical category	2	0
10.Literal Translation	1	4
11.Other term errors	1	**-6**

Table 44. Training corpus-based change in term translation errors and correct term translations. Improvements are marked in bold.

Three of the 4 correct term categories improved at a count between 1-4 in PBMT while the count of correct acronym decrease by 1. In NMT, the improvement is very suggestive and all correct term categories are better with more corpora. Correct term considering reference improves by 87 correct terms and the other three categories are better by 4 – 24 more correct terms. While the mixed domain corpora do not necessarily include more cardiology terms, it seems that more corpora in NMT helps decode and translate better the source sentences. This implies that even including mixed domain data to an NMT engine may improve the in-domain performance of the engine.

When it comes to the error categories, the change from PBMT-1 to PBMT-2 considering term errors oscillates between -4 and 2 while the change from NMT-1 to NMT-2 oscillates between -84 and 4. Clearly, the change is more suggestive in NMT systems. Of the 11 error categories, 4 of them decrease, 2 of them remain the same and 5 of them increase slightly in PBMT. And in NMT, the number of 6 error categories decrease, 2 remain the same and 2 slightly increase. Especially the change in the number of term drops is worth highlighting. In NMT-2, there are 87 less term drop errors than NMT-1. Acronym mistranslation also drops by 24 cases and partial term translation by 19 cases.

6.3.2. Term N-gram Size Based Term Error Analysis in MT Systems

Considering the n-gram sizes of the terms, the engine with the lowest performance is NMT-1 in all 5 categories. In 1-gram, 2-gram and 4-gram categories, PBMT-1 and PBMT-2 perform very close to one another and are better than NMT-1 and NMT-2. NMT-2 has the highest correct term translations in 3-gram and <4-gram categories.

As it can be observed from the table below, increasing the corpus size significantly increases correct n-gram translations in all categories in the case of NMT engines. While there are only 2 correct term translations in 1-gram category in NMT-1, this count rises to 28 correct terms in NMT-2 after the addition of the mixed domain training corpora. The same is true for >4-gram category where 0 correct terms rises to 5 correct terms. The change from PBMT-1 to PBMT-2 is not big in comparison to that of the NMT engines. While there is a slight increase in 1-gram and 3-gram categories, the count does not change in 2-gram and >4-gram scenarios. It even falls by 1 case in 4-gram scenario. However, it should be highlighted that in 4-gram category, the performances of PBMT-1 and PBMT-2 engines are particularly high: 37/41 (90.24%)

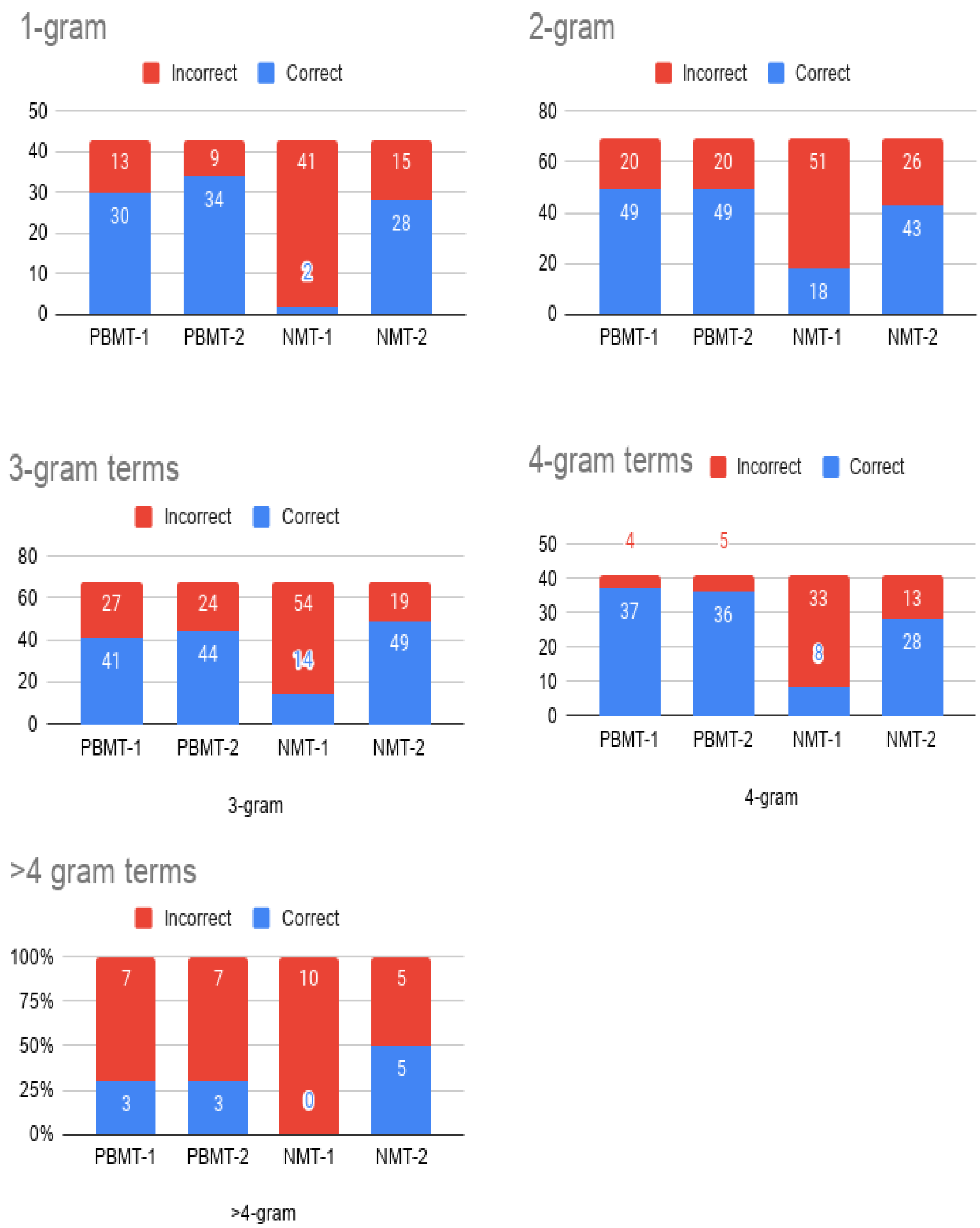


Figure 45. Term n-gram translation evaluation comparing the performance of 4 engines.

and 36/41 (87.80%) correct term translations; this is the highest frequency of correct term translations in any category.

On the other side of the scale, the lowest frequency of correct terms is yielded by NMT-1 in 1-gram terms with 2/41 (4.65%) and >4-gram terms with 0/10 (0%).

6.3.3. Sentence Length Based Term Error Analysis in MT Systems

The sentence length influences the translation of the terms by MT systems. In this section, we report our findings about this influence. NMT-1 has the highest term translation error rate in all 3 categories. In the short sentence (5-10 words) category, the best performing engine is NMT-2. However, in longer sentences (the ones with both 11-20 words and 21-29 words), PBMT-2 performs over other engines.

	5-10 words	11-20 words	21-29 words
PBMT-1	66.67%	67.33%	72.53%
PBMT-2	66.67%	69.31%	76.92%
NMT-1	12.82%	19.80%	18.68%
NMT-2	71.79%	66.34%	63.74%
	Correct Term Percentage		
	5-10 words	11-20 words	21-29 words
PBMT-1	33.33%	32.67%	27.47%
PBMT-2	33.33%	30.69%	23.08%
NMT-1	87.18%	80.20%	81.32%
NMT-2	28.21%	33.66%	36.26%
	Incorrect Term Percentage		

Table 45. Percentages of correct and incorrect term translations based on source sentence length.

PBMT-1 follows PBMT-2 with a difference of only 1.98% in mid-size sentences and a difference of 4.34% in long sentences. The difference between PBMT-2 and NMT-2 are 2.97% and 13.18% for mid-size and long sentences, respectively. These results imply that PBMT systems perform better than NMT systems when the sentence length increases. Only in short sentences of 5-10 words, NMT-2 system performs better. And as in the cases of previous parameters, the quality gap between NMT-1 and NMT-2 points to the importance of training corpus size.

6.3.5. Term Translation by PBMT and NMT Engines

In this section, we approach closely to the translation behavior of PBMT and NMT engines as observed by the term translation outputs and explain the decision-making process in assigning correct term subcategories and incorrect term subcategories.

There are 4 cases in PBMT-1 to be highlighted. The Turkish term "atriyal elektromekanik gecikme parametreleri" with reference English equivalent of "parameters of atrial electromechanical delay" was translated as "AEMD parameters" in PBMT-1 which was annotated as a correct synonym. Based on the training corpora, PBMT-1 was able to correctly infer that "AEMD" was the acronym of "atrial electromechanical delay". In another case, "sağ kalp fonksiyonu: right heart function" was translated as "right heart disfunction" which shifted the meaning completely and hence was deemed incorrect term equivalent. In the third case, "derin bradikardi: profound bradycardia" was translated as "deep bradycardia" by 3 engines except for NMT-1. This might be caused by higher frequency of "deep" compared to "profound" in the training corpus. "Deep bradycardia" was considered to be a correct synonym. The fourth example is "ibuprofen: ibuprofen". Since it is the written in the same way both in Turkish and English, it might have been inserted from the source sentence although it is not translated, but it is considered correct because it is a valid Turkish term.

There are 5 cases to be highlighted in PBMT-2. Firstly, in the case of "şeker hastalığı: diabetes mellitus", PBMT-2 translates it as "sugar disease"; which is an incorrect literal translation because the first component of the terminological unit "şeker" was literally translated into English. The same occurs with "Jude medikal mitral sığır biyoprotez kapağı: St. Jude Medical Biocor bovine bioprosthesis in mitral position" which was translated as "St. Jude Medical cattle bioprosthetic mitral valve" where "sığır" is literally translated as "cattle". And in another case, what should be translated as "left ventricular function" was translated as "left ventricular script". In these three cases, since PBMT-2 had mixed domain training corpora, more frequent out-of-domain words might have gained weight compared to specific medical terms. And lastly "high-risk transcatheter valve replacement procedures" was highlighted because all the components of the term were available in the target sentence yet they were in wrong order, which led to reordering error.

Many term translations in NMT-1 exhibited errors that were not common in other engines. Some sentences with correct terms had the same term repeated unnecessarily. For example, the target sentence including the translation of "diabetes mellitus" was like this: "However, patients with diabetes mellitus, diabetes mellitus and coronary artery disease and coronary artery disease." For more examples of repeated target terms in NMT-1, see Annex V. Another noteworthy example includes the translation of the term "major vascular complications" which was unnecessarily extended as "postoperative vascular complications" with "postoperative" modifier.

NMT-2 exhibited interesting term translation behaviors as well. Firstly, the reference term "transvenous pacemakers" ("transvenöz kalp pilleri" in Turkish) was translated as "transvenous cardiac pacemakers" with the insertion of "cardiac" which was a valid contribution that lead to the classification as correct synonym. In another example, "aortik yaprakçıklar: aortic cusps" were translated as "aortic leaflets" which is an incorrect literal translation. "Yaprakçık" literally means "leaflet" yet such a translation causes nonexistent terminological analogy in the target language. The translation of *konvülziyon* ('convulsion') as 'convoy' is an example of byte pairing encoding (cf.

section 1.2.4.) in NMT where subword units such as *kon* ('con') are used when a word that is not in the training corpus is to be translated by the engine. In this case, it does not succeed in creating a correct translation. Considering correct term translations, NMT-2 was the only engine to translate correctly "high-risk transcatheter valve replacement procedures" which was a long term. In the case of "atrial fibrillation", it correctly inserted the acronym ("AF") without the extended version. "Human immun-deficiency virus" and "prognostic nutritional index" were translated as "human immune insufficiency virus" and "prognostic dietary index", which were considered correct denominational variations, hence as correct synonyms. However, "inappropriate sinus tachycardia" ("uygunsuz sinüs taşikardisi" in Turkish) was translated as "false sinus tachycardia" and was deemed incorrect.

6.4. Conclusions and Limitations of the Terminology Evaluation Task

The terminological evaluation of Turkish → English PBMT and NMT engines has shown that PBMT and NMT systems handle terminology distinctively since the types and frequencies of terminology errors are significantly different in both systems. Besides, the 2 PBMT engines performed better than the 2 NMT engines considering overall term translations. However, the increase in training corpus has significantly increased the term translation quality of NMT-2 (though the overall corpus was a mixed domain one) and led it approach to the translation qualities of PBMT-2 and PBMT-1, which may imply that in even higher amount of training corpora, NMT-2 may surpass the quality of PBMT systems. Considering the impact of corpus increase in PBMT, the there was only a slight improvement in the overall term translation count. When it comes to term translation based on n-gram length, in general, PBMT engines were slightly better than NMT-2 engine. Finally, concerning sentence based term translation evaluation, NMT-2 performed better in short sentences with 5-10 words while PBMT engines were better in mid-size and long sentences.

Haque et al. (2019) conducted a semi-automatic term evaluation and term error categorization task on legal English-Hindi and Hindi-English NMT and PBMT engines. As explained in section 2.5, Hindi is a morphologically rich language just like Turkish, we can compare the results of Hindi-English direction with our Turkish-English MT results. Their Hindi-English PBMT engine has a term error rate of 12.9% while our PBMT-1 has a rate of 31.21% and PBMT-2 has a rate of 29.87%. On the other hand, their NMT engine has a term error rate of 11.5% while our NMT-1 has a rate of 83.22 and NMT-2 has a rate of 37.92%. The significant different between their engines and ours may be cause by two reasons: their training corpus has more than 1 million sentences and their bilingual term database is more permissive for generic law terms such as “case”, “dispute”. Considering the term error category, just like Haque et al (2019) we also observed that NMT engines commit more term drop errors; in their study, PBMT has 38 term drop errors and NMT has 86 cases of this class. Similarly, we also found that NMT engines do not commit source term insertion errors. While in their study, the most common term error type was incorrect lexical selection (roughly equal to our incorrect term equivalent error type) in both MT types, in our study it was the partial error in PBMT engines and term drop error in NMT engines. In another study, Scansani et al. (2019) make terminology evaluation study without fine-grained term error categories comparing Google Translate and their institutional academic domain Italian-English NMT system. Their overall term hit rate (correct terms) reaches 65.33% while Google Translate yields a rate of 63.72%, which are close but superior to our NMT-2 engine (62.08%). However, these rates are slightly below the results of our PBMT engines.

Our term annotation and evaluation tasks have intrinsic limitations in that both term annotation and terminology error evaluation and classification were conducted manually. However, in order to minimize the subjectivity bias, reference terminology dictionaries as well as keyword sections of the cardiology abstracts were used to identify terms in the annotation phase. Additionally, human reference translations were compared to the MT outputs. As described in section 4.3, term error categories in our study were adapted from the study of Haque et al. (2019) but they were extended to include new categories. Future lines of research could include a board of experts in medical translation, especially in the term classification task. Besides, our corpus size is still quite small compared to the sizes of current engines. In the future, instead of concentrating solely on cardiology, a medical corpus and a larger test corpora for such evaluation tasks can be created. Besides, automating the process of term annotation and term error categorization can facilitate these processes and decrease cognitive load as opposed to spreadsheets.

Conclusion and Future Work

This dissertation tracked machine translation beginning with creating Turkish to English parallel corpora from scratch to training and evaluating SMT and NMT engines. Our research aimed to answer 3 main questions. Below we reflect upon these questions.

Q1: Can large amount of high quality, specific domain, Turkish to English parallel corpora be created in semi-automatic procedures that can be used by translators?

We studied the available tools and procedures for creating parallel corpora and formulated a 8-step parallel corpora preparation strategy which requires minimal technical competence. We found that it is possible to create large amounts of domain-specific parallel corpora using tools and semi-automatic methods that may be used by translators. While projects such as Paracrawl (Esplà-Gomis et al., 2019) aim to fully automate parallel corpora preparation pipeline, our semi-automatic strategy allow for human intervention especially in alignment process for ensuring high quality sentence pairs are included. The first parallel corpus created was Turkish to English corpus from cardiology abstracts (TRENCARD) and 788,046 source words were obtained. Using the same strategy we built a domain-specific test corpus as well. This relatively small test corpus that we created for MT human evaluation and terminology evaluation had 11,015 source words. With these two corpora, we aimed to provide free and open Turkish – English specific domain parallel corpora which are scarce especially in medical domain. Besides, our strategy has the potential to help other translators and research to contribute with more parallel corpora in the future. In the last step of corpora preparation, we curated a mixed domain parallel corpora from Opus Corpus which had

5,668,129 source words. These three corpora served as the basis for training and evaluation steps of the thesis.

Q2: Are the performances of SMT and NMT engines different when they are trained with the same specific domain and mixed domain corpora?

Using the specific-domain cardiology corpora, we trained one SMT engine (PBMT-1) and one NMT engine (NMT-1). In this very-narrow-domain scenario, SMT performed significantly better than NMT engine in automatic scores. In the second scenario, when mixed domain corpora are added to the training set and one more SMT engine (PBMT-2) and one more NMT engine (NMT-2) are trained, the SMT quality decreased and NMT quality significantly improved in all automatic metrics. According to human evaluation with cardiology sample set, the change from PBMT-1 with specific domain corpora to PBMT-2 with mixed domain corpora brought a slight improvement. However, the change from NMT-1 with specific domain corpora to NMT-2 with mixed domain corpora brought a very significant improvement. In fact, NMT performed the best in ranking, adequacy and fluency metrics. Collectively, these results imply that NMT has the potential to perform better when translating specific domain content with an engine trained on mixed domain corpora. Nevertheless, when low amount of specific corpora is available, SMT may still perform better than NMT, at least in the case of Turkish to English MT.

Q3: Are there qualitative and quantitative differences between SMT and NMT engines with regard to the types and frequencies of terminology errors?

There were two main findings of the terminology evaluation task: i). SMT engines committed less terminology errors than NMT engines. While the change from NMT-1 to NMT-2 significantly decreased the amount of terminology errors, the amount of term error was still more than that of PBMT-1 and PBMT-2. ii). The types and frequencies of terminology errors in SMT and NMT engines were significantly different. To highlight some of the differences, SMT engines committed more partial term translation error while NMT engines committed more term drop errorr. When the translation of a term is not available in SMT, source term is inserted in the target. Yet, in NMT, either the term is totally dropped or its subword components are used to create possible match; hence, source term insertion was not available in NMT. In short, it can even be a matter of discussion whether term errors in NMT or the ones in SMT are more serious problems for MT quality.

The differences in frequencies and types of terminology has implications both for postediting and MT postprocessing in the runtime. Firstly, postediting guidelines may need to updated to reflect the types of terminology errors committed by NMT. Besides, posteditors may be trained to be aware of possible term errors in NMT and know what to expect in an NMT postediting task with a specific domain. New technologies such as automatic postediting may also benefit from term error patterns in MT so as to detect and modify them. MT postprocessing uses runtime glossary lists to replace term translations by the ones available in the list. But this method may still cause morphosyntactic terminology errors. Terms in the acronym form continue to be problematic for both PBMT and NMT systems. Pre-editing strategies may be applied both (i). at corpus preparation phase where acronyms are provided with their unabbreviated forms consistently throughout the corpus, and ii). before source text is processed by the MT system where acronyms are unabbreviated either automatically or manually.

Our research had many limitations because of the decisions taken throughout the process. The objective of the dissertation was to be able to control all the steps of MT training process from a translation studies perspective, which had advantages and disadvantages. While preparation of cardiology corpora from scratch instead of benefitting from readily available, open corpora was beneficial for the study and for the research community in general, it was time-consuming at first and since the domain (cardiology) was very narrow, it wasn't possible to create a parallel corpora larger than 1 millon source words. This limited the size of training corpora in the specific domain MT trainings. In the future, we would like to create a less narrow, medical parallel corpora for large scale medical MT training using the corpus preparation strategy above. A considerable limitation of the thesis is the subjective term evaluation. While certain criteria including the use of reference human translations and reference terminology resources, the term error annotation could have been performed by other human evaluators just like the general human evaluation of the MT engines. However, since there were four engines to be evaluated, and no user friendly GUI for term evaluation and annotation was available, and term error annotation task was a complex, spreadsheet based task, we decided to conduct a subjective evaluation. In the future, we aim to benefit from automatic and human term evaluation methods for quick analysis of term translation qualities of MT engines.

Bibliography

Achkasov, A. (2014). What Translators Do to Terminology: Prescriptions vs. Performance. *Journal of Siberian Federal University. Humanities & Social Sciences 2*, 210-221.

Aguilar-Amat, A. (1990). Syntactic characterization of idioms and a 'parser' proposal for a machine translation system. *Actas del Congreso de la Sociedad Española de Lingüística, XX*, 824-833.

Aguilar-Amat, A. (1993). En torno a la combinatoria del léxico los conceptos de colocación e idiomatismo. *Lenguajes naturales y lenguajes formales : actas del IX congreso de lenguajes naturales y lenguajes formales*, (pp. 267-272). Reus.

Aguilar-Amat, A. (1998). La Terminología y su documentación. *Anuari SOCADI de Documentació i d'Informació, Societat Catalana de Documentació i d'Informació*, 77-84.

Aguilar-Amat, A. (2002). Traducción, computación, utopía. *Terminologie et Traduction, Office diciembre publications oficielles diciembre Commaunatés européennes*, 15-50.

Aguilar-Amat, A. (2021). Aspectos de la traducción de neologismos. In S. T. Judit Freixa, *La neología del español: del uso al diccionario* (p. (in press)). Iberoamericana/Vervuert.

Aguilar-Amat, A. (in press). *Aspectos de la terminología: luces y sombras desde la traducción especializada*. EUMO.

Aguilar-Amat, A., & Torres-Hostench, O. (2021). Compensación humana de déficits de la traducción automática. *Revista Letras UNA*, (in press).

Aguilar-Amat, A., Mesa Lao, B., Soler, M. P., & Sánchez Gijón, P. (2009). Applying Terminology Knowledge in Translation: Problem-Based Learning in the Degree in Translation and Interpreting. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication*, 105-118.

Aguilar-Amat, A., Mesa, B., & Gijón, P. S. (2006). La formación del traductor especializado y la creación de recursos terminológicos compartidos. *Actas del I Congreso de Traducción Especializada. Colegio de Traductores Públicos de la Ciudad de Buenos Aires.*

Aguilar-Amat, A., Neunzig, W., Sánchez-Gijón, P., & Touhami, O. (2005). Contrastive Terminology: Semitic, Anglo-Germanic and Latin Concept Building through Complex Terms (Avoiding One-Way Method Procedures in Translation). *Polissema*, 237-247.

Aksan, M., & Aksan, Y. (2014). Patterns and frequency: Evidence from the Turkish National Corpus (TNC). *Proceedings of ICTL 2014.* Berlin: Harrasowitz Verlag.

Alkım, E., & Çebi, Y. (2019). Machine Translation Infrastructure for Turkic Languages (MT Turk). *The International Arab Journal of Information Technology, 16, 3.*

Anderson, P., Fernando, B., Johnson, M., & Gould, S. (2017). Guided Open Vocabulary Image Captioning with Constrained Beam Search. *Proceedings of* *the 2017 Conference on Empirical Methods in Natural Language Processing*, (pp. 936-945). Retrieved from https://www.aclweb.org/anthology/D17-1098

Arnold, D., Balkan, L., Meijer, S., Humphreys, R., & Sadler, L. (1994). *Machine Translation: an Introductory Guide.* Blackwells-NCC. Retrieved from https://www1.essex.ac.uk/linguistics/external/clmt/mtbook/

Ataman, D. (2018). Bianet: A Parallel News Corpus in Turkish, Kurdish and English. *Proceedings of the LREC 2018 Workshop MLP–Moment*, (pp. 14-17).

Ataman, D. (2018). Bianet: A Parallel News Corpus in Turkish, Kurdish and English. *Proceedings of the LREC Workshop MLP-Moment*, (pp. 14-17).

Baroni, M., Kilgarriff, A., Pomikálek, J., & Rychlý, P. (2006). WebBootCaT: a web tool for instant corpora. *Proceedings of 12th EURALEX International Congress*, (pp. 123-131).

Bektaş, E., Yılmaz, E., Mermer, C., & El-Kahlout, I. D. (2016). TÜBİTAK SMT System Submissions for WMT 2016. *Proceedings of the First Conference on Machine Translation, 2*, pp. 246–251.

Bentivogli, L., Bisazza, A., Cettolo, M., & Federico, M. (2016). Neural versus Phrase-Based Machine Translation Quality: a Case Study. *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, (pp. 257–267). doi:10.18653/v1/D16-1025

Bourigault, D., & Slodzian, M. (1998). Pour une terminologie textuelle. *Terminologies Nouvelles, 19*, 29-32.

Bowker, L. (2015). Terminology and Translation. In K. H, & S. F (Eds.), *Handbook of Terminology, Volume 1* (pp. 304-323). John Benjamins.

Bowker, L., & Pearson, J. (2002). *Working with Specialized Language: A Practical Guide to Using Corpora.* London & New York: Routledge.

Brown, P. F., Cocke, J., & Della-Pietra, S. A. (1988). A statistical approach to language translation. *Proceedings of the International Conference on Computational Linguistics (COLING).*

Burlot, F., Scherrer, Y., Ravishankar, V., Bojar, O., Gronroos, S.-A., Koponen, M., . . . Yvon, F. (2018). The WMT'18 Morpheval test suites for English-Czech, English-German, English-Finnish and Turkish-English. *Proceedings of the Third Conference on Machine Translation: Shared Task Papers,* (pp. 546–560).

Cabré, M. T. (1999). *Terminology: Theory, methods and applications.* (J. A. DeCesaris, Trans.) Amsterdam & Philadelphia: John Benjamins.

Cabré, M. T. (2003). Theories of terminology. Their description, prescription and explanation. *Terminology 9:2*, 163–199.

Cabré, M. T. (2010). Terminology and translation. In Y. Gambier, & L. (. Van Doorslaer, *Handbook of translation studies. 1* (pp. 356-365).

Campo, A. (2013). *The reception of Eugen Wsters work and the development of terminology.* Ph.D. thesis, Universite de Montreal.

Castilho, S., Doherty, S., Gaspari, F., & Moorkens, J. (2018). Approaches to human and machine translation quality assessment. In J. Moorkens, S. Castilho, F. Gaspari, & S. (. Doherty (Eds.), *Translation Quality Assessment* (pp. 9-38). Springer.

Castilho, S., Moorkens, J., & Gaspari, F. e. (2018). Evaluating MT for massive open online courses: A multifaceted comparison between PBSMT and NMT systems. *Machine Translation, 32*, 255–278. doi:https://doi.org/10.1007/s10590-018-9221-y

Castilho, S., Moorkens, J., Gaspari, F., Calixto, I., Tinsley, J., & Way, A. (2017). Is Neural Machine Translation the New State of the Art? *The Prague Bulletin of Mathematical Linguistics, 108(1)*, pp. 109-120. doi:https://doi.org/10.1515/pralin-2017-0013

Cengizhan, L., & Tanı, A. (2010). Purification of the Medical Terms in Turkish. *Procedia Social and Behavioral Sciences 2*, 5326–5330.

Ciapuscio, G. (2003). *Textos especializados y terminología.* Barcelona: IULA.

Cigan, V. (2018). TERMINOLOGICAL COLLOCATIONS AND TERM VARIATION IN MECHANICAL ENGINEERING DISCOURSE. *FLUMINENSIA : časopis za filološka istraživanja, Vol. 30 No. 2*, 91-120.

Crego, J., Kim, J., & et., a. (2016). Systran's pure neural machine translation systems. *CoRR abs/1610.05540.* Retrieved from http://arxiv.org/abs/1610.05540

Diki-Kidiri, M. (2007). Éléments de terminologie culturelle. *Cahiers du Rifal, 26*, 14-25.

Doğru, G., Martín-Mor, A., & Aguilar-Amat, A. (2018). Parallel Corpora Preparation for Machine Translation of Low-Resource Languages: Turkish to English Cardiology Corpora. *Proceedings of the LREC 2018 Workshop “MultilingualBIO: Multilingual Biomedical Text Processing”*, (pp. 12 - 15). Miyazaki.

Eisold, C. (2017). Terminology Usage in Machine Translation. Retrieved from http://kv-emptypages.blogspot.com/2017/05/terminology-usage-in-machine-translation.html

Eken, N. T. (2016). Eşdizimlerin Saptanmasına ve Betimlenmesine Yönelik Kuram ve Yaklaşımlar. *Mustafa Kemal University Journal of Graduate School of Social Sciences*, 28-47.

Eker, S. (2013). Modernleşme Sürecinde Türkiye'de Terimbilim Çalışmalarına Genel Bir Bakış. *Altaistics and Turcology, 2013/1*, 68-95.

El-Kahlout, İ. D., & Oflazer, K. (2006). Initial Explorations in English to Turkish Statistical Machine Translation. *Proceedings of the Workshop on Statistical Machine Translation*, (pp. 7-14).

El-Kahlout, İ. D., & Oflazer, K. (2010). Exploiting Morphology and Local Word Reordering in English to Turkish Phrase-based Statistical Machine Translation. *IEEE Transactions on Audio, Speech and Language Processing*, 1313-1322.

Erten, A. (2007). *Tıp Terminolojisi ve Tıp Metinleri Çevirisi.* Ankara: Seçkin Yayıncılık.

Esplà-Gomis, M., Forcada, M. L., Ramírez-Sánchez, G., & Hoang, H. (2019). ParaCrawl: Web-scale parallel corpora for the languages of the EU. *Proceedings of MT Summit XVII, volume 2*, (pp. 118 - 119).

Faber, P. (2011). The dynamics of specialized knowledge representation: Simulational reconstruction or the perception–action interface. *Terminology, 17,* 9-29.

Faber, P., Araúz, P. L., Velasco, J. A., & Reimerink, A. (2007). Linking Images and Words: the description of specialized concepts. *International Journal of Lexicography, 20, 1,* 39-65.

Faber, P., Márquez Linares, C., & Expósito, M. V. (2005). Framing Terminology: A Process-Oriented Approach. *Meta : journal des traducteurs / Meta: Translators' Journal, 50. No. 4.*

Farajian, M. A., Negri, M., Turchi, M., & Federico, M. (2018). Evaluation of terminology translation in instance-based neural MT adaptation. *Proceedings of the 21st Annual Conference of the European Association for Machine Translation.*

Forcada, M. (2017). Making sense of neural machine translation. *Translation Spaces, 6:2,* 291–309. doi:https://doi.org/10.1075/ts.6.2.06for

Forcada, M. L. (2010). Machine translation today. In Y. Gambier, & L. Doorslaer (Eds.), *Handbook of Translation Studies* (pp. 215-223). Amsterdam and Philadelphia: John Benjamins.

Freixa, J. (2006). Causes of denominative variation in terminology: A typology proposal. *Terminology 12(1)*, 51-77.

Galbrun, E. (2009). *Phrase table pruning for Statistical Machine Translationn. [Technical Report] C-2009-22.* University of Helsinki.

Gehring, J., Auli, M., Grangier, D., Yarats, D., & Dauphin, Y. (2017). Gehring, J., Auli, M., Grangier, D., Yarats, D., & Dauphin, Y. (2017). Convolutional Sequence to Sequence Learning. In *ICML.*

Guner, S. P. (2020). Translation of Medical Texts: Problems and Solutions. *The Journal of Social Sciences*, 349-351.

Gupta, R., Lambert, P., Patel, R. N., & Tinsley, J. (2019). Improving Robustness in Real-World Neural Machine Translation. *Proceedings of MT Summit XVII, volume 2*, (pp. 142-148).

Hakkani-Tür, D., Saraçlar, M., Tür, G., Oflazer, K., & Yuret, D. (2018). Morphological Disambiguation for Turkish. In O. K., & S. M. (eds), *Turkish Natural Language Processing. Theory and Applications of Natural Language Processing* (pp. 53-68). Springer.

Hamzaoglu, İ. (1993). *Machine translation from Turkish to other Turkic languages and an implementation for the Azeri language.* Bogazici University (PhD Thesis).

Haque, R., Hasanuzzaman, M., & Way, A. (2019). Investigating Terminology Translation in Statistical and Neural Machine Translation: A Case Study on English-to-Hindi and Hindi-to-English. *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019)* (pp. 437–446). INCOMA Ltd. doi:10.26615/978-954-452-056-4_052

Haque, R., Hasanuzzaman, M., & Way, A. (2019). Terminology Translation in Low-Resource Scenarios. *Information, 10*(273). doi:doi:10.3390/info10090273

Hasler, E., de Gispert, A., Iglesias, G., & Byrne, B. (2018). Neural Machine Translation Decoding with Terminology Constraints. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, (pp. 506-512). Retrieved from https://www.aclweb.org/anthology/N18-2081

Hassan, H., Aue, A., Chen, C., Chowdhary, V., Clark, J., Federmann, C., . . . Zhang, Z. &. (2018). Achieving Human Parity on Automatic Chinese to English News Translation. *Arxiv*. Retrieved from https://arxiv.org/abs/1803.05567

Hassan, H., Hearne, M., Way, A., & Sima'an, K. (2006). Syntactic phrase-based statistical machine translation. *2006 IEEE Spoken Language Technology Workshop*, (pp. 238-241). Palm Beach.

Hearne, M., & Way, A. (2011). Statistical Machine Translation: A Guide for Linguists and Translators. *Language and Linguistics Compass*, *5/5*, 205-226.

Hoffmann, L. (1979). Towards a Theory of LSP. Elements of a Methodology of LSP Analysis. *Fachsprache, 1, 1-2*, 7-12.

Hoffmann, L. (1998). *Llenguatges d'especialitat.* Barcelona: IULA.

Hokamp, C., & Liu, Q. (2017). Lexically Constrained Decoding for Sequence Generation Using Grid Beam Search. *Proceedings of the 55th Annual Meeting* *of the Association for Computational Linguistics (Volume 1: Long Papers),* (pp. 1535–1546). Retrieved from https://www.aclweb.org/anthology/P17-1141

Hutchins, W. (1986). *Machine Translation: Past, Present, Future.* Retrieved from http://www.hutchinsweb.me.uk/PPF-3.pdf

Inkpen, D., Paribakht, T., Faez, F., & Amjadian, E. (2016). Term Evaluator: A Tool for Terminology Annotation and Evaluation. *Int. J. Comput. Linguistics Appl., 7*, 145-165.

Islamaj Doğan, R., Comeau, D. C., Yeganova, L., & Wilbur, W. J. (2014). Finding abbreviations in biomedical literature: three BioC-compatible modules and four BioC-formatted corpora. *Database : the journal of biological databases and curation*. doi:https://doi.org/10.1093/database/bau044

Kageura, K. (2015). Terminology and lexicography. In *Handbook of Terminology, 1* (pp. 45-59). John Benjamins.

Karaman, B. İ. (2014). Atatürk and the Turkish Terminology Reform: The (Re-)turkification of Geography Terms. *Bilig, 69*, 89-125.

Kenny, D. (2009). Corpora. In M. B. (eds.), *Routledge encyclopedia of translation studies* (pp. 59-62). London: Routledge.

Kenny, D., & Doherty, S. (2014). Statistical machine translation in the translation curriculum: Overcoming obstacles and empowering translators. *The Interpreter and Translator Trainer*, 276-294.

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., . . . Suchomel, V. (2014). The Sketch Engine: Ten Years On. *Lexicography, 1*, 7-36.

Kocourek, R. (1992). Specialized Texts and Terminology as Objects of Linguistic and Pragmatic Analysis. *Actes du XVe Congrès international des linguistes*.

Koehn, P. (2004). Challenges in Statistical Machine Translation. Presentation at PARC. Google, ISI, MITRE, BBN, University of Montreal. Retrieved 02 18, 2020, from http://homepages.inf.ed.ac.uk/pkoehn/publications/challenges2004.pdf

Koehn, P. (2005). Europarl: A Parallel Corpus for Statistical Machine Translation. *Conference Proceedings: the tenth Machine Translation Summit*, (pp. 79-86). Phuket. Retrieved from http://homepages.inf.ed.ac.uk/pkoehn/publications/europarl-mtsummit05.pdf

Koehn, P. (2009). *Statistical Machine Translation.* Cambridge: Cambridge University Press.

Koehn, P. (2016). Koehn, Philipp. 2016. Moses User Manual and Code Guide. Retrieved from http://www.statmt.org/moses/manual/manual.pdf

Koehn, P. (2020). *Neural Machine Translation.* Cambridge University Press.

Koehn, P., & Hoang, H. (2007). Factored Translation Models. *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, (pp. 868–876). Prague.

Koehn, P., & Knowles, R. (2017). Six Challenges for Neural Machine Translation. *Proceedings of the First Workshop on Neural Machine Translation* (pp. 28–39). Association for Computational Linguistics. doi:10.18653/v1/W17-3204

Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., & B. Cowan et al., 2. (2007). Moses: Open Source Toolkit for Statistical Machine Translation. *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (ACL), Companion Volume Proceedings of the Demo and Poster Sessions* (pp. 177 - 180). Prague: ACL.

Koprulu, S. G. (2017). Tıbbi Çeviri ve Zorlukları. *Turkish Studies. International Periodical for the Languages, Literature and History of Turkish or Turkic*, 249-262.

Kraif, O. (2002). Translation Alignment and Lexical Correspondence. In B. Altenberg, & S. (. Granger, *Lexis in contrast. Corpus-based approach* (pp. 271-290). Amsterdam & Philadelphia: John Benjamins.

Lambert, P. (2018). Terminology in Neural MT. Retrieved from https://iconictranslation.com/2018/08/issue-7-terminology-in-neural-mt/

Lambert, P. (2020). Merging Terminology into Neural Machine Translation. Retrieved from https://iconictranslation.com/2020/04/issue-79-merging-terminology-into-neural-machine-translation/

Lanners, Q. (2019). Neural Machine Translation. Retrieved from https://towardsdatascience.com/neural-machine-translation-15ecf6b0b

Li, X., Tracey, J., Grimes, S., & Strassel, S. (2016). Uzbek-English and Turkish-English Morpheme Alignment Corpora. *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)* (pp. 2925–2930). European Language Resources Association (ELRA).

Loening, K., & Sonneveld, H. (1994). Introducing Terminology. *Terminology*.

Lommel, A. (2018). Metrics for Translation Quality Assessment: A Case for Standardising Error Typologies. In S. Castilho, J. Moorkens, F. Gaspari, & S. (. Doherty (Eds.), *Translation Quality Assessment: From Principles to Practice* (pp. 109-128). Springer.

Lumeras, M. A., & Way, A. (2017). On the Complementarity between Human Translators and Machine Translation. *HERMES - Journal of Language and Communication in Business*(56), 21-42. doi:https://doi.org/10.7146/hjlcb.v0i56.97200

Luong, M.-T., & Manning, C. (2015). Stanford Neural Machine Translation Systems for Spoken Language Domains. *Proceedings of the 12th International Workshop on Spoken Language Translation*, (pp. 76-79).

Martín-Mor, A. (2017). 25MTradumàtica: Statistical Machine Translation Customisation for Translators. *SKASE Journal of Translation and Interpretation*, 25-39.

Martín-Mor, A. (2017). MTradumàtica: Statistical Machine Translation Customisation for Translators. *SKASE journal of translation and* *interpretation*, 25-29. Retrieved from http://www.skase.sk/Volumes/JTI12/pdf_doc/02.pdf

Metin, S. K., & Karaoglan, B. (2010). Measuring Collocation Tendency of Words. *Journal of Quantitative Linguistics*, 175-187.

Molina, L., & Hurtado, A. (2002). Translation Techniques Revisited: A Dynamic and Functionalist Approach. *Meta 47(4)*, 498-512.

Montalt, V. (2011). Medical translation and interpreting. In Y. Gambier, & v. Doorslaer (Eds.), *Handbook of Translation Studies Volume 2* (pp. 79 - 83). John Benjamins Publishing Company. doi:https://doi.org/10.1075/hts.2.med2

Newmark, P. (1988). *A Textbook of Translation.* New York: Prentice-Hall International.

Nießen, S., Och, F. J., Leusch, G., & Ney, H. (2000). An Evaluation Tool for Machine Translation: Fast Evaluation for MT Research. *Proceedings of the Second International Conference on Language Resources and Evaluation (LREC'00).* Retrieved from http://www.lrec-conf.org/proceedings/lrec2000/pdf/278.pdf

O'Brien, S. (2012). Towards a Dynamic Quality Evaluation Model for Translation. *The Journal of Specialised Translation*, 55-77.

Oflazer, K., & El-Kahlout, I. D. (2007). Exploring different representational units in English-to-Turkish statistical machine translation. *Proceedings of the Second Workshop on Statistical Machine Translation*, (pp. 25-32). Prague.

Oflazer, K., & Saraclar, M. (2018). *Turkish Natural Language Processing.* Springer International Publishing.

Orozco-Jutorán, M. (2016). *Metodología de la traducción directa del inglés al español.* Granada: Colección Interlingua.

Ozkan, B. (2012). Türkiye Türkçesinin Eşdizim Sözlüğü. *IV. Dünya Dili Türkçe Sempozyumu*, (pp. 93-102).

Papineni, K., Roukos, S., Ward, T., & J, Z. W. (2002). BLEU: a method for automatic evaluation of machine. *Proceedings of the 40th annual meeting on association for computational linguistics,* (pp. 311–318). Philadelphia, Pennsylvania, USA.

Pena, G., & Andrade-Filho, J. (2010). Analogies in medicine: valuable for learning, reasoning, remembering and naming. *Adv in Health Sci Educ*, 609–619.

Popovic, M. (2015). ChrF: character n-gram F-score for automatic MT evaluation. *Proceedings of the 10th workshop on Statistical Machine Translation (WMT-15),* (pp. 392–395). Lisbon.

Post, M., & Vilar, D. (2018). Fast Lexically Constrained Decoding with Dynamic Beam Allocation for Neural Machine Translation. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, (pp. 1314–1324). Retrieved from https://www.aclweb.org/anthology/N18-1119

Protopopescu, D. (2013). Theories of Terminology–Past and Present. . *Studii şi cercetări de onomastică şi lexicologie SCOL 6/1-2*, 195-201.

Reynolds, P. (2015). Machine translation, translation memory and terminology management. In H. J. Kockaert, & F. Steurs (Eds.), *Handbook of Terminology, Volume 1* (pp. 276–287). John Benjamins.

Rikters, M. (2018). Impact of Corpora Quality on Neural Machine Translation. *Baltic HLT 2018*.

Sager, J. (1990). *A Practical Course in Terminology Processing.* Amsterdam/Philadelphia: Benjamins.

Sager, J. C., Dungworth, D., & McDonald, P. F. (1980). *English Special Languages. Principles and Practice in Science and Technology*. Wiesbaden: Brandstetter.

Şahin, M. (2015). Çevirmen Adaylarının Gözünden İngilizce- Türkçe Bilgisayar Çevirisi ve Bilgisayar Destekli Çeviri: Google Deneyi. *Çeviribilim ve Uygulamaları Dergisi, Journal of Translation Studies*(21), 43-60.

Saldanha, G., & O'Brien, S. (2013). *Research methodologies in translation studies.* New York: Routledge.

Santamaría Pérez, I. (2009). *La terminología y la comunicación especializada.* Retrieved from https://rua.ua.es/dspace/bitstream/10045/12766/6/TEMA_2.pdf

Sarikas, F. (2006). Problems in Translating Collocations. *Elektronik Sosyal Bilimler Dergisi*, 33-40.

Scansani, R. (2020). *Machine Translation for Institutional Academic Texts: Output Quality, Terminology Translation and Post-Editor Trust.* University of Bologna.

Scansani, R., Bentivogli, L., Bernardini, S., & Ferraresi, A. (2019). MAGMATic: A Multi-domain Academic Gold Standard with Manual Annotation of Terminology for Machine Translation Evaluation. *MTSummit.*

Sennrich, R., Haddow, B., & Birch, A. (2016). Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). *54th Annual Meeting of the Association for Computational Linguistics* (pp. 1715-1725). Association for Computational Linguistics.

Sezer, T. (2017). TS Corpus Project: An online Turkish Dictionary and TS DIY Corpus. *European Journal of Language and Literature, 9(1),* 18-24.

Shterionov, D., Superbo, R., Nagle, P., Casanellas, L., O’Dowd, T., & Way, A. (2018). Human versus automatic quality evaluation of NMT and PBSMT. *Machine Translation, 32,* 217–235. doi:https://doi.org/10.1007/s10590-018-9220-z

Tantuğ, A. C., Oflazer, K., & El-Kahlout, I. D. (2008). BLEU+: a Tool for Fine-Grained BLEU Computation. *Proceedings of the International Conference on Language Resources and Evaluation, LREC.*

Temmerman, R. (2000). *Towards new ways of terminology description. The sociocognitive.* Amsterdam/Philadelphia: John Benjamins.

Thelen, M. (2015). The Interaction between Terminology and Translation Or Where Terminology and Translation Meet. *trans-kom, 8*(2), 347-381.

Tiedemann, J. (2012). Parallel Data, Tools and Interfaces in OPUS. *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012)*, (pp. 2214 - 2218).

Tiedemann, J. (2019, December 02). *EMEA Corpus*. Retrieved from Opus Corpus: http://opus.nlpl.eu/EMEA.php

Toral, A., & Sanchez-Cartagena, V. M. (2017). A Multifaceted Evaluation Neural versus Phrase-Based Machine Translation for 9 Language Directions. *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, (pp. 1063–1073). Retrieved from https://aclanthology.info/papers/E17-1100/e17-1100

Toral, A., Castilho, S., Hu, K., & Way, A. (2018). Attaining the Unattainable? Reassessing Claims of Human Parity in Neural Machine Translation. *WMT*. Retrieved from https://www.semanticscholar.org/paper/Attaining-the-Unattainable-Reassessing-Claims-of-in-Toral-Castilho/83e08fd46208b7faae8f6b5ff96c2b5b7859c730

Türk Kardiyoloji Derneği. (2009). *Kardiyoloji (Yürek Bilimi) Türkçe Terimler Kılavuzu (Guide for Cardiology Turkish Terms).*

Türk Pediatrik Kardiyoloji ve Kalp Cerrahisi Derneği. (2017). *Pediyatrik Kardiyoloji Terimler Kılavuzu.*

Tyers, F. M., & Alperen, M. S. (2010). South-east European Times: A parallel corpus of Balkan Languages. *Proceedings of LREC 2010, Seventh International Conference on Language Resources and Evaluation.* Retrieved from http://nlp.ffzg.hr/resources/corpora/setimes/

Van Brussel, L., Tezcan, A., & Macken, L. (2018). A fine-grained error analysis of NMT, PBMT and RBMT output for English-to-Dutch. *Proceedings of the Eleventh International Conference on Language Resources and Evaluation*, (pp. 3799–3804).

Vieira, L., O'Hagan, M., & O'Sullivan, C. (2020). Understanding the societal impacts of machine translation: a critical review of the literature on medical and legal use cases. *Information, Communication & Society*, 1-18.

Vinay, J.-P., & Darbelnet, J. (1977). *Stylistique comparée du français et de l'anglais.* Georgetown University Press.

Wang, T., Kuang, S., Xiong, D., & Branco, A. (2019). Merging External Bilingual Pairs into Neural MachineTranslation. Retrieved from https://arxiv.org/pdf/1912.00567.pdf

Way, A. (2009). A Critique of Statistical Machine Translation. *Journal of translation and interpreting studies: Special Issue on Evaluation of Translation Technology*, 17-41.

Way, A. (2018). Quality Expectations of Machine Translation. In S. Castilho, J. Moorkens, & F. &. Gaspari (Eds.), *Translation Quality Assessment: From Principles to Practice* (pp. 159-178). Springer.

Way, A. (2019). Machine Translation: where are we at today? In M. E.-D. Erik Angelone (Ed.), *The Bloomsbury Companion to Language Industry Studies* (pp. 311-332). Bloomsbury Academic Publishing.

Way, A., & Hearne, M. (2011). On the Role of Translations in State-of-the-Art Statistical Machine Translation. *Language and Linguistics Compass, 5/5*, 227–248.

Wołk, K., & Marasek, K. (2014). Building Subject-aligned Comparable Corpora and Mining it for Truly Parallel Sentence Pairs. *Procedia Technology, 18*, 126-132.

Wołk, K., & Marasek, K. (2015). Polish-English Statistical Machine Translation. In Z. A, C. K, & S. A. (eds), *New Research in Multimedia and Internet Systems. Advances in Intelligent Systems and Computing* (Vol. vol 314, pp. 169 - 180). Cham: Springer.

Yazıcı, M. (2018). Akademik Tıp Makalelerinin Türkçeden İngilizceye Çevirisi. *Tıp Eğitimi Dünyası, (17) 52*, 14-26. Retrieved from https://dergipark.org.tr/tr/pub/ted/issue/38508/401082

Yonghui Wu, M. S. (2016). Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation. Retrieved from https://arxiv.org/abs/1609.08144

Zanettin, F. (2012). *Translation-Driven Corpora.* New York: Routledge.

Zanettin, F. (2013). Corpus Methods for Descriptive Translation Studies. *Procedia - Social and Behavioral Sciences, 95*, 20-32.

Annexes

Annex I. Background Trainings Steps in KantanMT

Training Steps for PBMT-1

All the background tasks for PBMT-1 are shown below. It can be observed that the overall job is completed in 16 minutes.

Task	Results	Dates
Creating instance for [157215]	Instance created [i-03870cb4783590f3b]	Jan 19th, 16:55
Dispatched Job [157215] to EC2_QUEUE	Job [157215] dispatched to EC2_QUEUE	Jan 19th, 16:55
Creating EBS Storage Devices	Created EBS [vol-04e36b6e0ca843f91} (300GB)] for [i-03870cb4783590f3b]....	Jan 19th, 16:56
Starting launch sequence	Starting launch sequence	Jan 19th, 16:56
Preparing FS	Preparing FS	Jan 19th, 16:57
Fetching KantanTools	Fetching KantanTools	Jan 19th, 16:57
Fetching KantanMT engine	Fetching KantanMT engine	Jan 19th, 16:59
Retrieving KantanISR cache	Retrieving KantanISR cache	Jan 19th, 16:59
Converting KantanISR cache	Converting KantanISR cache	Jan 19th, 16:59
Fetching training data	Fetching training data	Jan 19th, 16:59
Validating training data	Validating training data	Jan 19th, 16:59
Converting training data	Converting training data	Jan 19th, 17:00
Creating translation memory	Creating translation memory	Jan 19th, 17:00
Creating 3Ts	Creating 3Ts	Jan 19th, 17:00
Data cleansing	Data cleansing	Jan 19th, 17:00
Generating WCs	Generating WCs	Jan 19th, 17:00
Mono wordcount created	Mono wordcount created: [0]	Jan 19th, 17:00
Wordcount created	Wordcount created: [702,424]	Jan 19th, 17:00
Generating unique WC	Generating unique WC	Jan 19th, 17:00

Annex I. Background Trainings Steps in KantanMT

Unique wordcount created	Unique wordcount created: [85,877]	Jan 19th, 17:00
Generating training stats	Generating training stats	Jan 19th, 17:00
Generating language model	Generating language model	Jan 19th, 17:00
Generating analysis LM	Generating analysis LM	Jan 19th, 17:01
Building recaser	Building recaser	Jan 19th, 17:01
Building TM	Building TM	Jan 19th, 17:02
Optimising engine	Optimising engine	Jan 19th, 17:03
Backup engine	Backup engine	Jan 19th, 17:04
Storing engine	Storing engine	Jan 19th, 17:04
Calculated Bleu Score[36.27]	Calculated Bleu Score[36.27]	Jan 19th, 17:04
Calculated F-Measure Score[0.51266586248492]	Calculated F-Measure Score[0.51266586248492]	Jan 19th, 17:05
Calculated TER Score[0.83229988726043]	Calculated TER Score[0.83229988726043]	Jan 19th, 17:05
Running gap analysis	Running gap analysis	Jan 19th, 17:05
Creating analysis trace file	Creating analysis trace file	Jan 19th, 17:05
Training analysis SVM	Training analysis SVM	Jan 19th, 17:06
Packaging engine	Packaging engine	Jan 19th, 17:06
Storing engine	Storing engine	Jan 19th, 17:07
Storing engine scores	Storing engine scores	Jan 19th, 17:07
Storing engine details	Storing engine details	Jan 19th, 17:07
Storing test data	Storing test data	Jan 19th, 17:07
Storing timeline entry	Storing timeline entry	Jan 19th, 17:09
Storing training rejects	Storing training rejects	Jan 19th, 17:09
KantanMT engine built in [769s]	KantanMT engine built in [769s]	Jan 19th, 17:09
Running Cleanup	Stopped [i-03870cb4783590f3b]	Jan 19th, 17:09
Running Cleanup	Deleted [vol-04e36b6e0ca843f91]	Jan 19th, 17:09
[OK] Job completed.	[i-03870cb4783590f3b] and [vol-04e36b6e0ca843f91] removed	Jan 19th, 17:11

Training Steps for PBMT-2

All the background tasks for PBMT-2 are shown below. It can be observed that the overall job is completed in 44 minutes.

Task	Results	Dates
Dispatched Job [160726] to EC2_QUEUE	Job [160726] dispatched to EC2_QUEUE	Mar 2nd, 11:23
Creating instance for [160726]	Instance created [i-08736f163c02673b6]	Mar 2nd, 11:23
Creating EBS Storage Devices	Created EBS [vol-01eddfe05d26b3e1f} (300GB)] for [i-08736f163c02673b6]....	Mar 2nd, 11:24
Starting launch sequence	Starting launch sequence	Mar 2nd, 11:24
Preparing FS	Preparing FS	Mar 2nd, 11:24
Fetching KantanTools	Fetching KantanTools	Mar 2nd, 11:24
Fetching KantanMT engine	Fetching KantanMT engine	Mar 2nd, 11:27
Retrieving KantanISR cache	Retrieving KantanISR cache	Mar 2nd, 11:27
Converting KantanISR cache	Converting KantanISR cache	Mar 2nd, 11:27
Fetching training data	Fetching training data	Mar 2nd, 11:27
Validating training data	Validating training data	Mar 2nd, 11:27
Converting training data	Converting training data	Mar 2nd, 11:28
Creating translation memory	Creating translation memory	Mar 2nd, 11:31
Creating 3Ts	Creating 3Ts	Mar 2nd, 11:31
Data cleansing	Data cleansing	Mar 2nd, 11:31
Wordcount created	Wordcount created: [4,657,519]	Mar 2nd, 11:33
Mono wordcount created	Mono wordcount created: [0]	Mar 2nd, 11:33
Generating WCs	Generating WCs	Mar 2nd, 11:33
Unique wordcount created	Unique wordcount created: [616,231]	Mar 2nd, 11:33
Generating unique WC	Generating unique WC	Mar 2nd, 11:33

Generating training stats	Generating training stats	Mar 2nd, 11:33
Generating language model	Generating language model	Mar 2nd, 11:33
Generating analysis LM	Generating analysis LM	Mar 2nd, 11:36
Building recaser	Building recaser	Mar 2nd, 11:39
Building TM	Building TM	Mar 2nd, 11:41
Optimising engine	Optimising engine	Mar 2nd, 11:50
Backup engine	Backup engine	Mar 2nd, 11:59
Storing engine	Storing engine	Mar 2nd, 12:00
Calculated Bleu Score[21.52]	Calculated Bleu Score[21.52]	Mar 2nd, 12:00
Calculated F-Measure Score[0.4364810330912]	Calculated F-Measure Score[0.4364810330912]	Mar 2nd, 12:01
Calculated TER Score[0.88655367231638]	Calculated TER Score[0.88655367231638]	Mar 2nd, 12:01
Running gap analysis	Running gap analysis	Mar 2nd, 12:01
Creating analysis trace file	Creating analysis trace file	Mar 2nd, 12:02
Training analysis SVM	Training analysis SVM	Mar 2nd, 12:03
Packaging engine	Packaging engine	Mar 2nd, 12:04
Storing engine	Storing engine	Mar 2nd, 12:05
Storing engine scores	Storing engine scores	Mar 2nd, 12:05
Storing engine details	Storing engine details	Mar 2nd, 12:05
Storing test data	Storing test data	Mar 2nd, 12:05
Storing training rejects	Storing training rejects	Mar 2nd, 12:05
Storing timeline entry	Storing timeline entry	Mar 2nd, 12:05
KantanMT engine built in [2,484s]	KantanMT engine built in [2,484s]	Mar 2nd, 12:05
Running Cleanup	Stopped [i-08736f163c02673b6]	Mar 2nd, 12:06
Running Cleanup	Deleted [vol-01eddfe05d26b3e1f]	Mar 2nd, 12:06
[OK] Job completed.	[i-08736f163c02673b6] and [vol-01eddfe05d26b3e1f] removed	Mar 2nd, 12:07

Training Steps for NMT-1

All the background tasks for NMT-1 are shown below. It can be observed that the overall job is completed in 1 hours and 37 minutes.

Task	Results	Dates
Dispatched Job [160903] to EC2_QUEUE	Job [160903] dispatched to EC2_QUEUE	Mar 3rd, 11:52
Creating instance for [160903]	Instance created [i-0d95e1bccfff890e2]	Mar 3rd, 11:52
Creating EBS Storage Devices	Created EBS [vol-074f2a72978d59c3a} (300GB)] for [i-0d95e1bccfff890e2]....	Mar 3rd, 11:53
Starting launch sequence	Starting launch sequence	Mar 3rd, 11:54
Preparing FS	Preparing FS	Mar 3rd, 11:54
Fetching KantanTools	Fetching KantanTools	Mar 3rd, 11:54
Retrieving KantanISR cache	Retrieving KantanISR cache	Mar 3rd, 11:56
Converting KantanISR cache	Converting KantanISR cache	Mar 3rd, 11:56
Fetching training data	Fetching training data	Mar 3rd, 11:56
Validating training data	Validating training data	Mar 3rd, 11:56
Converting training data	Converting training data	Mar 3rd, 11:57
Creating translation memory	Creating translation memory	Mar 3rd, 11:57
Creating 3Ts	Creating 3Ts	Mar 3rd, 11:58
Data cleansing	Data cleansing	Mar 3rd, 11:58
Data preprocessing for NMT	Data preprocessing for NMT	Mar 3rd, 11:58
Wordcount created	Wordcount created: [702,424]	Mar 3rd, 12:00
Mono wordcount created	Mono wordcount created: [0]	Mar 3rd, 12:00
Generating WCs	Generating WCs	Mar 3rd, 12:00
Unique wordcount created	Unique wordcount created: [85,647]	Mar 3rd, 12:00
Generating unique WC	Generating unique WC	Mar 3rd, 12:00
Building TM	Building TM	Mar 3rd, 12:01
Generating training stats	Generating training stats	Mar 3rd, 12:01
NMT training status	Training cycle [2] - Perplexity: 173.58	Mar 3rd, 12:20
NMT training status	Training cycle [3] - Perplexity: 114.48	Mar 3rd, 12:29
NMT training status	Training cycle [4] - Perplexity: 87.79	Mar 3rd, 12:38
NMT training status	Training cycle [5] - Perplexity: 70.59	Mar 3rd, 12:46
NMT training status	Training cycle [6] - Perplexity: 59.18	Mar 3rd, 12:55
NMT training status	Training cycle [7] - Perplexity: 52.81	Mar 3rd, 13:04
NMT training status	Training cycle [8] - Perplexity: 47.00	Mar 3rd, 13:13

NMT training status	Training cycle [9] - Perplexity: 43.85	Mar 3rd, 13:22
Backup engine	Backup engine	Mar 3rd, 13:22
Storing engine	Storing engine	Mar 3rd, 13:23
Calculated Bleu Score[24.88]	Calculated Bleu Score[24.88]	Mar 3rd, 13:24
Calculated F-Measure Score[0.31656728026964]	Calculated F-Measure Score[0.31656728026964]	Mar 3rd, 13:25
Calculated TER Score[1.0534296028881]	Calculated TER Score[1.0534296028881]	Mar 3rd, 13:25
Calculated PSPPL Score[202.73]	Calculated PSPPL Score[202.73]	Mar 3rd, 13:25
Running gap analysis	Running gap analysis	Mar 3rd, 13:25
Packaging engine	Packaging engine	Mar 3rd, 13:25
Storing engine	Storing engine	Mar 3rd, 13:26
Storing engine scores	Storing engine scores	Mar 3rd, 13:26
Storing engine details	Storing engine details	Mar 3rd, 13:26
Storing test data	Storing test data	Mar 3rd, 13:26
Storing training rejects	Storing training rejects	Mar 3rd, 13:26
Storing timeline entry	Storing timeline entry	Mar 3rd, 13:26
KantanMT engine built in [5,567s]	KantanMT engine built in [5,567s]	Mar 3rd, 13:26
Running Cleanup	Stopped [i-0d95e1bccfff890e2]	Mar 3rd, 13:27
Running Cleanup	Deleted [vol-074f2a72978d59c3a]	Mar 3rd, 13:27
[OK] Job completed.	[i-0d95e1bccfff890e2] and [vol-074f2a72978d59c3a] removed	Mar 3rd, 13:29

Training Steps for NMT-2

All the background tasks for NMT-1 are shown below. It can be observed that the overall job is completed in 11 hours and 44 minutes.

Task	**Results**	**Dates**
Dispatched Job [160738] to EC2_QUEUE	Job [160738] dispatched to EC2_QUEUE	Mar 2nd, 12:32
Creating instance for [160738]	Instance created [i-09031c05e272dd029]	Mar 2nd, 12:32

Creating EBS Storage Devices	Created EBS [vol-0b52def2db72ec261} (300GB)] for [i-09031c05e272dd029]....	Mar 2nd, 12:33
Starting launch sequence	Starting launch sequence	Mar 2nd, 12:33
Preparing FS	Preparing FS	Mar 2nd, 12:34
Fetching KantanTools	Fetching KantanTools	Mar 2nd, 12:34
Retrieving KantanISR cache	Retrieving KantanISR cache	Mar 2nd, 12:36
Converting KantanISR cache	Converting KantanISR cache	Mar 2nd, 12:36
Fetching training data	Fetching training data	Mar 2nd, 12:36
Validating training data	Validating training data	Mar 2nd, 12:36
Converting training data	Converting training data	Mar 2nd, 12:38
Creating translation memory	Creating translation memory	Mar 2nd, 12:43
Creating 3Ts	Creating 3Ts	Mar 2nd, 12:44
Data cleansing	Data cleansing	Mar 2nd, 12:44
Data preprocessing for NMT	Data preprocessing for NMT	Mar 2nd, 12:47
Wordcount created	Wordcount created: [4,657,520]	Mar 2nd, 12:56
Generating WCs	Generating WCs	Mar 2nd, 12:56
Mono wordcount created	Mono wordcount created: [0]	Mar 2nd, 12:56
Generating unique WC	Generating unique WC	Mar 2nd, 12:56
Unique wordcount created	Unique wordcount created: [615,724]	Mar 2nd, 12:56
Generating training stats	Generating training stats	Mar 2nd, 12:57
Building TM	Building TM	Mar 2nd, 12:57
NMT training status	Training cycle [2] - Perplexity: 78.04	Mar 2nd, 15:27
NMT training status	Training cycle [3] - Perplexity: 48.34	Mar 2nd, 16:41
NMT training status	Training cycle [4] - Perplexity: 35.75	Mar 2nd, 17:55
NMT training status	Training cycle [5] - Perplexity: 29.97	Mar 2nd, 19:09
NMT training status	Training cycle [6] - Perplexity: 25.84	Mar 2nd, 20:24
NMT training status	Training cycle [7] - Perplexity: 23.78	Mar 2nd, 21:38
NMT training status	Training cycle [8] - Perplexity: 22.12	Mar 2nd, 22:52
NMT training status	Training cycle [9] - Perplexity: 21.03	Mar 3rd, 00:07
Backup engine	Backup engine	Mar 3rd, 00:07
Storing engine	Storing engine	Mar 3rd, 00:09

Calculated Bleu Score[38.53]	Calculated Bleu Score[38.53]	Mar 3rd, 00:09
Calculated F-Measure Score[0.43537047052461]	Calculated F-Measure Score[0.43537047052461]	Mar 3rd, 00:10
Calculated TER Score[0.85527328714396]	Calculated TER Score[0.85527328714396]	Mar 3rd, 00:10
Calculated PSPPL Score[5567.77]	Calculated PSPPL Score[5567.77]	Mar 3rd, 00:10
Running gap analysis	Running gap analysis	Mar 3rd, 00:10
Packaging engine	Packaging engine	Mar 3rd, 00:10
Storing engine	Storing engine	Mar 3rd, 00:12
Storing engine scores	Storing engine scores	Mar 3rd, 00:12
Storing engine details	Storing engine details	Mar 3rd, 00:12
Storing test data	Storing test data	Mar 3rd, 00:12
Storing timeline entry	Storing timeline entry	Mar 3rd, 00:13
Storing training rejects	Storing training rejects	Mar 3rd, 00:13
KantanMT engine built in [41,953s]	KantanMT engine built in [41,953s]	Mar 3rd, 00:13
Running Cleanup	Stopped [i-09031c05e272dd029]	Mar 3rd, 00:13
Running Cleanup	Deleted [vol-0b52def2db72ec261]	Mar 3rd, 00:13
[OK] Job completed.	[i-09031c05e272dd029] and [vol-0b52def2db72ec261] removed	Mar 3rd, 00:15

ANNEX II. Automatic Evaluations for MT Engines

F-Measure, BLEU and TER scores of PBMT-1

Row	Source	PBMT-1 Output	Reference/Target	F-M	BLEU	TER
1	Metabolik sendrom bulunmayan obez grupta sistolik ve diyastolik kan basınçları, açlık kan glukozu, trigliserit ve HDL kolesterol	without metabolic syndrome in obese group, systolic and diastolic blood pressures, fasting blood glucose, triglyceride, and HDL-cholesterol	Control subjects and obese subjects without MetS had similar systolic and diastolic blood pressures, fasting blood glucose, triglyceride, and HDL cholesterol levels, but all these significantly	57%	39%	78%

	düzeyleri kontrol grubuyla benzer bulunurken, bu değerler MetS'li obez grupta anlamlı farklılık gösterdi.	levels were similar with the control group, patients with MetS were obese groups was significantly different between the groups.	differed in patients with MetS.			
2	Açık kalp ameliyatı sonrası geç dönemde ortaya çıkan intraperikardiyal organize hematom ve kalp tamponadı: Manyetik rezonans görüntüleme	Open heart surgery in the late period after a of intrapericardial organized hematoma and gastric adenocarcinoma: magnetic resonance imaging	Cardiac tamponade caused by intrapericardial organized hematoma as a late complication of open heart surgery: magnetic resonance imaging	59%	7%	68%
3	56 hastadaki 57 lezyonun JO'u (% 17,5) aorto-ostiyal, 47'si (% 82,5) no11 aorto-ostiyal (dal ve yanda/) idi.	The lesion was 56 hastadaki57 jo (17,5%) aorto-ostial, 47 (82,5%) no11 aorto-ostial (branch and yanda/years).	Ten (17,5 %) Iesions were aorto-ostial and 47 (82,5 %) were non aorto-ostial (branch and side branch vessel) Jesions.	32%	18%	100%
4	Planar Egzersiz Talyum Sintigrafisinin Koroner Arter Hastalığı Tanısında Değerini Etkileyen Faktörler	outcome of planar exercise thallium scintgraphy value in the diagnosis of coronary artery disease and Affecting Factors	Factors Affecting the Value of Planar Exercise Thallium Scintigraphy in the Diagnosis of Coronary Artery Disease	85%	53%	53%
5	Anahtar Kelimeler: Koroner hastalık, mortalite/trend, Türkiye/epidemiyoloji.	Keywords: Coronary heart disease, mortality/trend, Turkey/epidemiology.	Keywords: Coronary disease, mortality/trends, Turkey/epidemiology.	73%	12%	33%
6	Homozigot AH (HoAH) bulunan olgularda çocukluk döneminden itibaren kardiovasküler olaylar gelişebilmektedir.	FH (HoFH) in patients with cardiovascular events to childhood the maneuver.	In homozygous individuals (HoFH) cardiovascular events could develop in child-hood.	38%	7%	73%
7	AMAÇ Çıkan aort anevrizması (ÇAA) mortalite ve morbiditenin önemli	Objective Ascending aortic aneurysms (surgical outcomes of) important causes of	Assessment of myocardial performance index and its association with aortic	25%	4%	100%

	nedenleri arasında yer almaktadır.	mortality and morbidity was performed.	elasticity in patients with ascending aortic aneurysm			
8	Artmış QT Disperziyonuna Yol Açan Miyokard Perfüzyon Bozukluğu ile QT Disperziyonunun Yüksek Riskli Koroner Arter Hastaları Öngörmedeki Değeri	increased QT disperziyonuna leading to myocardial perfusion abnormalities and QT disperziyonunun high predictive value for coronary artery patients	Relation of Myocardial Perfusion Abnormalities to Increased QT Dispersion and Value of QT Dispersion in Identification of High-risk Patients With Coronary Artery Disease	59%	4%	89%
9	Koroner stent (KS) lokalize darlıkların tedavisinde ve perkütan transluminal koroner anjiyoplasti (PTCA) sonrası gelişen suboptimal sonuçların giderilmesinde seçilmiş bir yöntem olup bypass cerrahisine güçlü alternatif oluşturmuştur.	Coronary stenting (CAS) for the treatment of localized darlıkların and percutaneous transluminal coronary angioplasty (PTCA) after family results giderilmesinde selected is a strong bypass method for an alternative.	Coronary artery stenting (CAS) is used in patients with localized coronary stenosis, and for suboptimal results after PTCA as an alternative treatment to bypass surgery.	49%	5%	86%
10	Ekokardiyografi ile sol atriyum ve sol ventrikülde genişleme, sol ventrikülde sistolik fonksiyon bozukluğu belirlendi.	Echocardiography was performed and left atrium and left ventricular enlargement, left ventricular systolic dysfunction.	Echocardiography revealed enlargement of the left atrium and ventricle and left ventricle systolic dysfunction.	57%	6%	57%
11	Besleyici arterlere ve dallarına önce koil sonra Onyx ile embolizasyon uyguladık.	The feeding arteries before and after implanted in the coil embolization with Onyx.	We performed embolization of the feeding arteries and their branches, with coils and Onyx.	52%	5%	92%
12	Uygulanan metoprolol tedavisi hastanın	Metoprolol therapy, the patient underwent	Control Holter recording showed very rare PVCs,	0%	7%	100%

	semptomlarında hızlı ve belirgin iyileşme sağladı.	immediate and marked improvement of symptoms.	without PR prolongation or AV block.			
13	Hastalığın seyri, kimi ailelerde ani kardiyak ölümle son bulurken; kimi ailelerde de ani kardiyak ölüme rastlanmamaktadır.	the course of the disease, overt ailelerde sudden cardiac death may complicate the last bulurken; ailelerde rastlanmamaktadır sudden cardiac death.	The natural course in certain families is ceased with sudden cardiac death, whereas in others sudden cardiac death is absent.	35%	5%	85%
14	Antibiyotik (meropenem günde 3 gram) ve analjezik (pethidin günde 50 mg) tedavisi intravenöz başlandı.	antibiotic (meropenem daily gram) and analgesic (pethidin 50 mg/day) therapy, intravenous was started.	Intravenous antibiotic therapy of 3 grams meropenem per day and analgesic of 50 mg pethidine per day were administered.	31%	5%	100%
15	Anahtar Kelimeler: Kist hidatik, İnterventriküler septum	Keywords: Hydatid cyst, interventricular septum	Keywords: Hydatid cyst, interventricular septum	100%	100%	0%
16	Hastaların 9'unda (% 45) V4R'da ST yüksekliği saptandı.	patients, 9 (%45) or in V 4R was ST elevation.	ST segment elevation was detected in V4R in 9 (45 %) of 20 patients (group A).	30%	5%	100%
17	Behçet hastalarının mitral akımında, kontrol grubuna göre VTI E/VTI total, V zirve E ve V zirve E/A değerlerinde azalma, VTI A/VTI toal ve V zirve A değerlerinde artma olmakla birlikte hiçbiri istatistiksel olarak anlamlı bulunmadı.	Patients with Behçet's disease mitral accordance, compared to the control group, total VTI e/vti, V peak E and V peak E/A, remained decreased VTI A/VTI toal and V peak a increase However none of the difference was not statistically significant.	In mitral flow, VTI E/VTI total, V peak E, V peak E/A were decreased, VTI A/VTI total, V peak A were increased in patients with Behçet's disease, but none to a statistically significant degree.	43%	27%	83%
18	Yeni P2Y12 inhibitörlerinin günlük pratiğimizde kullanımının yaygınlaşması bu	New P2Y12 inhibitors in daily practice, the use of these agents, antithrombotic the risk of bleeding, mortality	The widespread use of new P2Y12 inhibitors in daily practice will demonstrate the antithrombotic efficacy, bleeding risk, effect on	58%	28%	84%

	ilaçların antitrombotik etkinliğini, kanama riskini, mortaliteye etkisini ve hastaların uyuncunu gösterecektir.	and uyuncunu effect of patients with suspected HoFH.	mortality and patient compliance associated with these drugs.			
19	Yürütülen kontrollü çalışmalar kullanım alanları konusunda daha geniş bilgiler sağlıyacaktır.	process of controlled trials have of information about sağlıyacaktır.	It is highly effective against a wide range of arrhythmias.	11%	6%	100%
20	Miyokardiyal Bridge'lerin Değerlendirilmesi:	Myocardial bridge of group:	Assessment of Myocardial Bridges:	50%	16%	75%
21	Çalışmaya Türkiye’de 36 merkez katıldı ve 558 hasta alındı.BULGULAR Ortalama yaş 62±13 yıl olup, hastaların % 38’i kadındı.	The study included in 36 one year and 558 patients alındı.bulgular mean age 62±13 years, and 38% of the patients were female.	Total of 588 patients were enrolled from 36 participating medical centers from across the country.RESULTS Mean age was 62±13 years and 38 % of the patients were female.	48%	31%	82%
22	AMAÇ Bizim bu çalışmadaki amacımız sol ana koroner arter (LMCA) darlığı olan hastaların cerrahi revaskülarizasyon sonrası klinik seyir ve uzun dönem sağkalım sonuçlarını geriye dönük olarak incelemektir. Bu çalışmaya tek başına LMCA darlığı tanısı konulan ve cerrahi revaskülarizasyon yapılan 38 hasta (27 erkek, 11 kadın) alındı.	Objectives In this study, we aimed to left main coronary artery (LMCA) stenosis after surgical revascularization patients with the clinical course and long-term survival outcomes were retrospectively. This study, isolated LMCA stenosis, surgical revascularization was diagnosed, and the 38 patients (27 males, 11 were female) were included in the study.	OBJECTIVE The objective of this study was to retrospectively analyze the clinical course and postoperative long-term survival of patients diagnosed with isolated left main coronary artery (LMCA) stenosis after surgical revascularization.METHODS A total of 38 patients (27 males, 11 females) who were diagnosed with isolated LMCA stenosis and underwent surgical revascularization were enrolled in the study.	66%	30%	86%

23	Ameliyat sonrası değerlendirme periyodu erken ve uzun dönem takiplerinden oluşmaktaydı.	in postoperative assessment of early and long-term takiplerinden.	The postoperative assessment period included short- and long-term follow-up.	47%	7%	63%
24	Her iki hasta ameliyat sonrasmda sorunsuz olarak izlenmektedir.	Both the probabil ity of the patient was uneventfully operation.	Both patients are is well after operation.	24%	8%	80%
25	Tüm hastalarda RFKA işleminden 24 saat önce ve sonrasında ekokardiyografik inceleme yapıldı.	All patients underwent RFCA işleminden 24 h before and after the echocardiographic examination was performed.	Echocardiographic examination was performed in all the patients 24 hours before and after RFCA.	69%	29%	67%
26	Koroner arteriyovenöz fistül nadir gözlenen, koroner anjiyografilerin % 0,1- % 0,2 sinde tesadüfen rastlanan bir anomalidir.	Coronary arteriovenous fistula in a rare, coronary anjiyografilerin 0,1- %0,2% of patients is a incidentally anomaly.	Coronary arteriovenous fistula is an uncommon anomaly, occurring as an incidental finding in 0,1 % to 0,2 % of coronary angiograms.	38%	4%	100%
27	Supravalvüler aort stenozunun çoğu zaman progresif bir seyir gösterdiği, periferik pulmoner stenozun ise genellikle zaman içinde hafiflediği bildirilmektedir.	Supravalvular aortic different often a progressive course %), and peripheral pulmonary stenozun is usually hafiflediği seperate entity.	Literature knowledge points to the fact that SVAS may be progressive over the years whereas PPS generally has a better prognosis.	11%	4%	100%
28	Kardiyovasküler tutulumu olmayan Behçet hastalarında arter sertliği ve subklinik aterosklerozun değerlendirilmesi	Cardiovascular Evaluation: collaborative efforts artery stiffness and subclinical atherosclerosis.	Evaluation of arterial stiffness and subclinical atherosclerosis in patients with Behçet’s disease without cardiovascular involvement	33%	15%	100%

29	Kırmızı kan hücrelerinin dağılım genişliğinde (KHDG) artma ve yüksek ürik asit düzeyi enflamasyon için belirteç olabilir.	Red cell distribution width (RDW) and increased uric acid levels may be a marker for inflammation.	Increased red cell distribution width (RDW) and uric acid level may be indicative of an underlying inflammatory state.	65%	34%	56%
30	Arulan risk parametreleri e rkeklerde 96 crn 'lik bel geuişliğinde11 itibaren yiiksek bulrwdu (p	arulan risk parameters E rkeklerde a waist 96 crn geuişliğinde11 the high bulrwdu (p	Furtlýermore. diabetes was morefrequent by 2.2-fold in men and by 4.8-jofd in women across rlýese quintifes.	0%	7%	100%
31	Çoklu lineer regresyon analizinde, plazma BNP'nin doğal logaritması akut MY'li grupta E/Ea oranı (β=0.50, p=0.002) ve SAV (β=0.38, p=0.015) ile, kronik MY'li grupta ise sistolik pulmoner arter basıncı (β=0.60, p=0.002) ve EF (β=-0.36. p=0.039) ile anlamlı ilişki gösterdi.SONUÇ Akut MY'li hastalarda MY derecesi ekokardiyografik olarak daha belirgin olmasına rağmen, serum BNP düzeyi daha düşük bulunmuştur.	In multiple linear regression analysis showed that plasma BNP, the natural history of is of acute MR in E/Ea ratio (β=0.50, p=0.002) and LaV (β=0.38, p=0.015), chronic MR in the systolic pulmonary artery pressure (β=0.60, p=0.002) and EF (β=-0.36. P=0.039) were significantly correlated with acute MR groups.CONCLUSION in patients with MR than in the echocardiographic, despite low serum BNP levels.	In multiple linear regression analysis, the natural logarithm of BNP was significantly correlated with E/Ea ratio (β=0.50, p=0.002) and LAV (β=0.38, p=0.015) in patients with acute MR, and with systolic pulmonary artery pressure (β=0.60, p=0.002) and EF (β=-0.36, p=0.039) in patients with chronic MR.CONCLUSION Although the echocardiographic degree of MR was more pronounced in patients with acute MR, serum BNP levels tended to be lower in this group.	63%	40%	73%

32	Alfieri ve grubu tarafından, mitral yetersizlik jetinin olduğu yerde mitral kapağın iki kanadının birbirlerine bir dikiş ile birleştirilip çift orifis haline getirilmesinden uyarlanmıştır ve mitral yetersizlik derecesinde belirgin bir azalmaya yol açmaktadır.	alfieri and group by mitral regurgitation, regurgitant the yerde mitral valve two kanadının 24 suture birleştirilip with a double orifice with getirilmesinden uyarlanmıştır and mitral insufficiency regurtation a significant reduction in exposure to.	It is adapted from the surgical technique that was initially described by Dr. Alfieri and his group by placement of a suture approximating the edges of the mitral leaflets at the origin of the MR jet, leading to creation of so-called bow-tie or double orifice with significant reduction in the MR jet.	33%	2%	100%
33	Görülen aritmiler, yan etki ciddiyet sınıflandırma sistemine göre; düşük, yüksek, majör olarak sınıflandırıldı ve sıklıkları sırası ile % 2,7, % 4.3 ve % 1.2 olarak saptandı.	seen arrhythmias, side effects severity göre; classification system, high, major classified as and frequency was 2,7%, 4.3% and 1.2% .	Arrhythmia was classified as low, high, or major, according to the adverse event severity score; the rates were 2.7 % , 4.3 % , and 1.2 % , respectively.	24%	4%	100%
34	Dünyanın birçok yerindeki ve ülkemizdeki yaşlı nüfusun dramatik bir şekilde artmasıyla birlikte, sinüs düğümü işlev bozukluğu (SND) ve atriyoventriküler (AV) bloklu hastalarda eşlik eden bir artış olacaktır.	dünyanın yerindeki and many elderly population in our country to be dramatically become, sinus node dysfunction (snd) and atrioventricular (AV) block is increased in patients with comorbidities.	With the dramatic increase in the number of elderly people in most parts of the world and in our country there will be an accompanying increase in patients with sinus node dysfunction (SND) and atrioventricular (AV) block.	50%	25%	100%
35	Her iki böbreğin RFK'ları ayrı ayrı grup 2'de grup 1'e göre daha yüksekti (sırasıyla, p <0.001 ve p=0.023).	Both böbreğin rfk decreased individualized in group 2 than in group 1 had higher (p <0.001 and p=0.023).	RFC of both kidneys in Group 2 was significantly higher than results in Group 1 (p <0.001 and p=0.023, respectively).	63%	19%	72%

36	Arakidonik asit metabolizmasının lipooksijenaz yolunun ürünü olan lökotrien (LTC4), kronik kararlı angina pektoris tanımlayıp koroner arter hastalığı saptanan 26 olguda perifer arter ve ven kanında ölçüldü.	arakidonik acid metabolizmasının lipooksijenaz of submission product with lökotrien (LTC4), chronic stable angina pectoris tanımlayıp coronary artery disease was detected in 26 peripheral artery and vein kanında were measured.	The product of lipooxygenase pathway of arachidonic acid metabolism, leukotriene C4 (LTC4) levels are measured from peripheral artery and vein blood samples in 26 cases, all diagnosed as coronary artery disease and had chronic stable angina pectoris.	45%	17%	100%
37	Grup I'de SV'de trombüs saptanan 22 hasta (3 kadın, 19 erkek, yaş ortalaması 60.4 ± 12.1) yer aldı.	Group I consisted of LV thrombus was detected in 22 patients (3 females, 19 males, mean age 60.4 ±12.1 mean: 51±8).	Group I consisted to 22 patients (3 women, 19 men) with a mean age of 60.4 ± 12. 1, who had left ventricular thrombus (LVT) in TTE.	54%	4%	100%
38	AMAÇ Dilate kardiyomiyopati (DCM) sol yada her iki ventrikülün sistolik fonksiyonlarının bozulması ve genişlemesi ile karakterize bir hastalıktır.	Objective Dilated cardiomyopathy (DCM) yada both left ventricular systolic function and cheaper dilatation, epidemiyolojik çalýþmalarda ölçülmemekteydi.	Angiotensin-Converting Enzyme, Angiotensin II Receptor, Apolipoprotein E and Endothelial Constitutive Nitric Oxide Synthase Gene Polymorphisms in Dilated Cardiomyopathy	18%	4%	100%
39	Yine hasta grubuna koroner anjiografi yapılarak hastalar iskemik ve idyopatik DCM gruplarına ayrıldı.	The patient group also coronary angiography via the ischemic and idiopatic DCM groups.	Patients having normal coronary arteries were classified as 'idiopathic' and the remaining group as 'ischemic' DCM.	28%	7%	100%
40	Sonuç olarak predominan orta-ileri MD grubunda daha sık SA SEK-TR görülmesi nedeniyle daha sık SAE gelişmektedir.	In conclusion, predominant moderate to severe MS was more frequent in LA SEC and THR because of the rarity more	We conluded that the higher frequency of SAE in the moderate to severe MS group is due to the higher frequency of LA SEC and THR in this group.	44%	18%	92%

		frequent in pts with practice.				
41	Perkütan koroner girişim sonrası -mobiliteye izin veren-kontrollü baskı kemerinin kum torbasıyla karşılaştırılması: Pilot çalışma	after percutaneous coronary intervention to -mobiliteye veren-controlled compression kemerinin sand torbasıyla as A pilot study	Comparison of controlled pressure belt -allowing mobilityto sandbags after percutaneous coronary intervention: pilot study	40%	19%	69%
42	Östrojene progesteron ilavesinin östrojenik kardiyoprotektif etkilerini nasıl etkilediği de iyi bilinmemektedir.	östrojene progesterone pharmacomechanical östrojenik cardioprotective effects on the effect of is not well-known.	Also in women with a uterus, unopposed estrogen is associated with an increased risk of endometrial cancer, therefore in most women, progesterone must be added to estrogen.	15%	3%	100%
43	Etkili bir oral antikoagülan tedavi ile atriyum fibrilasyonunda inme görülme oranları azaltılabilir.	An effective oral anticoagulant therapy with atrial fibrillation the stroke rates were reduced.	Until recently warfarin, a vitamin K antagonist was the only effective OAC in this field.	14%	6%	100%
44	Depresyonun tüm nedenlere bağlı mortaliteye etkisini değerlendirmede % 95 güven aralığında risk oranlarını (HR) tahmin etme amacıyla Cox orantısal riskler modeli kullanıldı.	depression for all-cause mortality in the effect of 95% confidence 1-16 risk rates (HRs) were used to predict the Cox orantısal requires elimination model.	The Cox proportional hazards regression model was used to estimate hazard ratios (HR) with a 95 % confidence interval (CI) for the impact of depression on all-cause mortality.	38%	3%	100%
45	Depresyonu olmayan hastalar (HR=1.0) kategorik göstergelerin analizi için referans grubu olarak kabul edildi.	Depression patients without (hr=1.0) for categorical markers analysis defined as the reference method.	Patients without depression were accepted as a reference group with HR=1.0 for analysis of the categorical indicator.	60%	6%	100%

46	Grup I'deki yüksek LPa değerleri olan olgulardaki hasta damar sayısı ortalama 3.1±0.9 iken, bu diğer grupta 2.9±1.0 idi. ? % 50 darlıkların sayısı ise 1. grupta ortalama 4.2±2.4 iken, ikinci grupta 3.9±2.6 idi.	In group I with high LPa levels and the number of diseased vessels the mean 3.1±0.9, while the other group 2.9±1.0.? %50 darlıkların count in group 1, while the mean 4.2±2.4 3.9±2.6 the second group, respectively.	Mean number of diseased vessels was 3.1±0.9 in group I and 2.9±1.0 group II. The mean number of lesions with = % 50 narrowing were 4.2 ± 2.4 in gorup I and 3.9±2.6 in group II. The extent scores were 0.47±0.2 versus 0.46±0.3.	43%	12%	100%
47	(SAP grubu); diğer 15 olguda sol atriyuma yönelik herhangi bir cerrahi girişim yapılmadı (non-SAP grubu). Her iki grupta erken postoperatif dönemde izlenen hemodinamik bulgular karşılaştırıldı:	(LAP group); 15 in the left atrium during any surgical intervention application (non-LAP group) in both groups, early postoperative hemodynamic findings of karşılaştırıldı:	Two groups were compared with respect to hemodynamic windings in the early postoperative period:	27%	5%	87%
48	TEKHARF Çalışmasının temelde Marmara ve İç Anadolu bölgelerinde oturan kohortu, toplam ve koroner kalp hastalığı (KKH) mortalitesi ile yeni koroner olay prevalanslarını da değerlendirmek amacıyla, 27 ay aradan sonra 2003 yılı Ağustos ayında tarandı.	Contribution of TARF study to mainly with the Turkish, and coronary heart disease (CHD) mortality and new coronary events in order to prevalanslarını, 27 was conducted in August 2003 yýlý month.	The Risk Factor Survey of 2003 in Western Turkey Indicates Trend to Declining Coronary Mortality and Urban Overall Mortality	32%	3%	90%
49	Ülkemizde Kardiyoloji Uzmanlık Eğitimi süresi 4 yıl olarak belirlenmiştir.	in cardiology specialty training in Turkey as süresi4 years.	The duration of Cardiology Specialist Training has been defined as 4 years in our country.	33%	6%	100%

50	Kalp kapak hastalığı olan gebelerin klinik ve ekokardiyografik takibi	Heart valve disease in patients with high clinical and echocardiographic follow-up	Clinical and echocardiographic follow-up in pregnant patients with valvular heart disease	82%	40%	82%

F-Measure, BLEU and TER scores of PBMT-2

Row	Source	PBMT-2 Output	Reference/Target	F-M	BLEU	TER
1	=== IMI Micro-Uzi ===1983 yılında Uzi ailesine bir üye daha katıldı: Micro-Uzi.	= Bookmark micro-Uzi =1983 Uzi family member in a more Rus: micro-Uzi.	The Mini-Uzi is a smaller version of the regular Uzi, first introduced in 1980.	15%	7%	100%
2	Müzede, dünyanın en geniş Mondrian koleksiyonu yer alır.	It is the world's largest Mondrian collection.	It is renowned for its large Mondrian collection, the largest in the world.	50%	7%	100%
3	1998 FIFA Dünya'da Yılın Oyuncusu Ödülü 'nü Fransa'nın 1998 FIFA Dünya Kupası finalinde Brezilya'yı 3-0 yendiği karşılşamada iki kez ağları havanlandıran Zinedine Zidane kazanmıştır.	1998 FIFA World Player of the Year award was won by France for the 1998 FIFA World Cup in Brazil with a score of 3-0 in the final of the Eredivisie karşılşamada twice havanlandıran Zinedine Zidane networks.	The 1998 FIFA World Player of the Year award was won by Zinedine Zidane, after scoring twice in France's 3–0 win over Brazil in the 1998 FIFA World Cup Final.	66%	42%	65%
4	Dünya Savaşı ==1935'de Henschel Panzer I tankını üretmeye başladı.	=1935 World War II, the Henschel Panzer I was the T-44 tank.	==World War II==Early in 1935, Henschel began manufacturing Panzer I tanks.	35%	7%	83%
5	Waffen SS Dağ Kolordusu Komutanı, SS Generali Karl Pfeffer-Wildenbruch kentin savunmasından	Waffen SS mountain corps commander of the SS General Karl Pfeffer-Wildenbruch	"Waffen SS" General Karl Pfeffer-Wildenbruch, the commander of the IX Waffen SS Alpine Corps, was put in charge of the city's defences.	47%	26%	100%

	sorumlu komutan olarak atandı.	savunmasından commander of the city.				
6	Örneğin, Cesàro toplamı Grandi ıraksak serisine 1/2 değerini atamaktadır.	For example, Cesàro summation Grandi divergent series atamaktadır 1/2 Value.	For example, Cesàro summation assigns Grandi's divergent series:formula_2the value 1/2.	50%	27%	60%
7	Bu yerel izleme bilgileri sayesinde yerel televizyon istasyonları, kablo TV operatöleri ve reklamcıları, program oluşturma ve reklam verme konusunda kararlarını verirler.	Local monitoring information to the local television stations, and cable TV operatöleri reklamcıları Program for creating and advertising making decisions about immigrants.	This local viewing information provides a basis for program scheduling and advertising decisions for local television stations, cable systems, and advertisers.	56%	4%	73%
8	Lugano kentinin güneyindeki Melide (İsviçre) ile doğu kıyısındaki Bissone (İtalya) arasında gölün son derece sığ olmasından yararlanarak taştan büyük bir set yapılmış ve üzerinden Sankt Gotthard demiryolu ile bir karayolu geçirilmiştir.	Lugano, in the south the city of melide (Switzerland) and East Coast bissone (Italy) Lake be very shallow through the stone a large set of the St. Gotthard, and is a rail and highway replacement.	Places at the lake in Switzerland (CH) and in Italy (I) include (from Lugano, clockwise):==Navigation==The lake is navigable, and used by a considerable number of private vessels.	29%	6%	89%
9	Av (İngilizce: Prey), Michael Crichton'ın bir romanıdır.	AV (prey), Michael of the thriller genre of the novel.	Prey is a novel by Michael Crichton, first published in November 2002.	9%	8%	100%
10	Tainan şehri (Çince 臺南 ya da 台南; okunuşu: T'ai-nan; anlamı: Güney Tayvan) Çin	Tainan city (Chinese 臺南 or 台南; t 'ai-nan; meaning Southern Taiwan) of the	Tainan (or ; literally "Taiwan South"), officially Tainan City, is a special municipality located in southern Taiwan, facing the	23%	4%	100%

	Cumhuriyeti'nin Taipei, Kaohsiung, ve Taichung'dan sonra dördüncü büyük şehridir.	Republic of China, Lingya District, Taipei, and taichung after the fourth-largest city.	Taiwan Strait in the west and south.			
11	Angus Young albümün ismi konusunda ,okuduğu For Those About To Die, We Salute You adlı kitaptan esinlendiğini belirtmiştir.	Angus Young, he read on the album, “For Those About to die, We Salute You” Love book.	The name of the album was inspired by a book Angus Young read, entitled "For Those About to Die, We Salute You", about Roman gladiators.	38%	32%	100%
12	Top 40 Mainstream (billboard.com'da Pop Songs olarak geçer), "Billboard" dergisinin bir radyo listesidir.	the Mainstream Top 40 (billboard.com in the pop songs as “Billboard” magazine a radio list.	The Mainstream Top 40 (referred to as Pop Songs on billboard.com) is a weekly airplay chart from "Billboard" magazine.	47%	5%	80%
13	Nisan ayın ortasında günde 1.000 kadar Japon asıllı ABD'liler ulaştı ve Temmuz ayına kadar kampın nüfusu 10.000'e yaklaştı.	On the day of the month on which to 1,000 were of Japanese descent and members reached July until camps population of 10,000 approached.	By mid–April up to 1,000 Japanese Americans were arriving daily, and by July the population of the camp neared 10,000.	45%	5%	88%
14	8 Haziran 2011’de 2PM Seoul Apgujung’da yer alan Canon Plex Galeri’de gerçekleşen Coca Cola‘ nın 125.	On June 8, 2011, 2PM Seoul apgujung in the Canon plex gallery, the Coca Cola of 125.	On 8 June 2011, 2PM participated in Coca Cola‘s 125th anniversary event, held at the Canon Plex Gallery in Apgujung, Seoul.	47%	20%	100%
15	Desdemona hem kedisi hem de Otello için Tanrının merhametini diler.	Desdemona both the cat and Otello for God merhametini as blessings.	She asks God for mercy, both for her and for Otello.	36%	9%	91%

16	Takımı adına ilk golünü 22 Eylül 2007'de Fulham'a karşı attı.	team scored his first goal on 22 September 2007 against Fulham.	His first goals came on 22 September 2007 against Fulham.	76%	36%	36%
17	Wehrmacht'ın işgali altındaki bölgede Smolensk ve Roslavl civarında oldukça geniş bir kara ve demiryolu ağı vardır.	The Wehrmacht's invasion of the Smolensk and the region around Hrodna (Grodno a large black and railway network.	The "Wehrmacht" controlled a much wider network of roads and railroads, centered on Smolensk and Roslavl.	35%	6%	89%
18	İtalya'nın Ravello şehrinde yer alan Oscar Niemeyer Ravello Oditoryumu'nun inşaatı Ocak 2010'da tamamlandı.	Italy's ravello located in the Oscar Niemeyer ravello Auditorium's construction was completed in January 2010.	In January 2010, the Auditorium Oscar Niemeyer Ravello was officially opened in Ravello, Italy, on the Amalfi Coast.	48%	5%	100%
19	2008 ve 2010 yıllarında Viyana Filarmoni Orkestrası'nın geleneksel Yeni Yıl Konseri'ni yönetti.	2008 and again in 2010, the Vienna Philharmonic Orchestra in the traditional New Year's Concert in the soundtrack record.	Prêtre has conducted the Vienna New Year's Concert twice, in 2008 and in 2010.	55%	18%	79%
20	Irina, Moore ile evliliğinde güvenliği bulmuş ve kısa zamanda onun için modellik yapmaya başlamıştı.	, Moore and Irina unsatisfied with found safety and soon he started modelling for the work.	Irina found security in her marriage to Moore and was soon posing for him.	40%	5%	88%
21	Petersburg Spor ve Konser Kompleksi () St. Petersburg, Rusya'da bir arena.	Petersburg Sport Concert Complex () in St. Petersburg, Russia of the arena.	The Saint-Petersburg Sports and Concert Complex (SKK Peterburgsky) () is an arena in St. Petersburg, Russia.	50%	20%	92%
22	İşsizlik nedeniyle köyden kente göçün artması, şehir hayatının çekiciliği ve	Unemployment village a state of increased due to the city, it will appeal	Owing to unemployment, increasing rural-urban migration, the attraction of city life, and a lack of political will,	40%	3%	80%

	siyasi irade eksikliği nedeniyle Hindistan şu anda dünyada en fazla çocuk işçiye sahip ülkedir.	and political life due to a lack of India is currently the world of the child have the employee and the country.	India has developed one of the largest child labor forces in the world.			
23	Lionel Messi rekor puan toplayarak dünyada yılın en iyi futbolcusu seçildi.	Lionel Messi scored the world record the IPM the best player of the year.	Lionel Messi was announced as the World Player of the Year with a record points total.	53%	21%	71%
24	Şu anda Canonical Ltd. şirketininin sahibidir ve şirketin ücretsiz olarak çıkardığı Debian tabanlı Linux dağıtımı olan Ubuntu işletim sisteminin başındadır.	The currently Canonical Ltd. Şirketininin owner and company by Debian free Linux distribution, based on Ubuntu operating system carries today.	In 2004 he returned to the free software world by funding the development of Ubuntu, a Linux distribution based on Debian, through his company Canonical Ltd.	39%	13%	100%
25	Ülkeyi sözünü Hikmet Münir Ebcioğlu'nun yazdığı, müziğini ise Kemal Ebcioğlu'nun yaptığı, Semiha Yankı'nın seslendiği Seninle Bir Dakika temsil etmiştir.	country that Hikmet Munir ebcioğlu the music was written, Kemal ebcioğlu of his, Semiha seslendiği Seninle Bir minutes of the contest.	The country was represented by Semiha Yankı with the song “Seninle Bir Dakika” written by Hikmet Münir Ebcioğlu and composed by Kemal Ebcioğlu.	41%	5%	100%
26	Bu seriyi ve çocuk kitabının ortak yapımcılığını Noah Z. Jones yaratır ve çizer ve Alex Hirsch ve Bill Reiss tarafından televizyon için geliştirilmiştir.	This series, and co-produced by children's book Noah Z. Jones and draws creates and Alex Hirsch and Bill Reiss developed for television by.	The series is created and co-executive produced by children's book illustrator Noah Z. Jones and developed for television by Alex Hirsch and Bill Reiss.	64%	48%	70%
27	Thiès'deki Demiryolu işçileri, Senegal'de işçi	in Thiès railway workers, in Senegal	The rail workers of Thiès played a key role in the	39%	4%	87%

	haklarının önemsenmesinde önemli rol oynamışlardır.	önemsenmesinde important role in the labor rights of others.	immergence of Senegal's labor movement.			
28	* İsveç: "İsveç, diğer devletler gibi, Kudüs'ü İsrail'in başkenti olarak tanımamaktadır, bu nedenle elçilik Tel Aviv'dedir.	*Sweden: "Sweden, other states, Jerusalem, Israel"s capital of Antarctica, therefore embassy in Tel Aviv and Trygg-Hansa.	*: "Sweden, like other states, does not recognise Jerusalem as Israel's capital, which is why the embassy is in Tel Aviv.	27%	5%	100%
29	Bu yapı uzay gemilerini ve kargoları herhangi bir roket yardımı olmadan yörüngeye daha az masrafla yerleştirmek için kullanılır.	the activities of this structure space and kargoları rocket without any aid orbit than quarter to bring.	Such a structure would be used to raise payloads to orbit without having to use rockets, making it much more cost effective.	21%	5%	100%
30	== Çoklu platform desteği ==Mozilla Firefox pek çok platform ve işletim sisteminde çalışır.	#NAME?	Mozilla Firefox is also the browser of choice for a good number of smaller operating systems, such as SkyOS and ZETA.	23%	3%	100%
31	" diyerek geri çevirmiştir.Bazı kaynaklara göre Hitler Yakov'u esir düşen yeğeni Leo Raubal ile takas etmek ister fakat bu teklifi kabul edilmez.	"Back çevirmiştir.bazı sources Hitler Yakov's nephew Leo raubal captured corresponding with her exchange, but to put this proposal is unacceptable.	" According to some sources, there was another proposition as well, that Hitler wanted to exchange Yakov for Hitler's nephew Leo Raubal; this proposition was not accepted either.	21%	10%	100%
32	* Portekizce (Portekiz): Pedro Sousa ve Luís Freitas Lobo.	*Portuguese (Portugal): Pedro present and Luís Freitas Lobo.	* Pedro Sousa and Luís Freitas Lobo provide the Portuguese commentary for Portugal.	38%	24%	100%
33	ISBN 0-7837-2221-4*Drijvers, Jan Willem.	ISBN 0-7837-2221-4*drijvers, Jan Willem.	ISBN 0-7837-2221-4* Drijvers, Jan Willem.	67%	100%	50%

34	Bandar Kangan (Farsça: بندر کنگان, Bandar-e Kangān), İran'ın Buşehr Eyaleti'nde şehir.	Bandar kangan () (, also Romanized as بندر کنگان Bandar-e kangān) is Iran's Bushehr Province in the city.	Bandar Kangan (, also Romanized as Bandar-e Kangān; also known as Kangān and Kangun) is a city in and capital of Kangan County, Bushehr Province, Iran.	45%	21%	100%
35	"vijñāna", Pāli "Vijnana", Tib.	"vijñāna vijnana", Pāli ", Tib.	"vijñāna", Pāli "", Tib.	44%	17%	60%
36	Ayrıca şempanzelerin insanlarla iletişim kurmak için kasıtlı olarak el jestleri ve bunların yanı sıra ses sinyalleri ürettiği bulgusu, insan dilinin öncüllerinin hem davranışsal hem de nöroanatomikal düzeylerde var olduğu izlenimini uyandırmaktadır.	also şempanzelerin people to communicate and as El jestleri, as well as audio signals provided evidence of human language and in the absence of structural nöroanatomikal levels of existence to perform lymphatics.	Further, the data indicating that chimpanzees intentionally produce manual gestures as well as vocal signals to communicate with humans suggests that the precursors to human language are present at both the behavioral and neuronanatomical levels.	33%	2%	91%
37	Çukurbağ, Mersin ilinin Mut ilçesine bağlı bir köydür.	çukurbağ is a village in Mut district of Mersin Province, Turkey.	Çukurbağ is a village in Mut district of Mersin Province, Turkey.	100%	100%	0%
38	1907 yılında Accademia di Belle Arti'ye (Güzel Sanatlar Akademisi) girdi.	In 1907, the Accademia di Belle Arti (Fine Arts Academy).	From 1907 to 1913 he studied at the Accademia di Belle Arti di Bologna of Fine Arts of Bologna.	41%	21%	100%
39	1387 yılının baharında, Timur, Gürcistan'ı tekrar işgal etmiş fakat Gürcüleri teslim olmaya zorlamamıştır.	In the spring of 1387, Timur, Georgia, but invaded again, gürcüleri zorlamamıştır to surrender.	In the spring of 1387, Timur again invaded Georgia but could not force the Georgians to submission.	52%	29%	71%
40	Korra ve arkadaşları polis şefi Lin Beifong ve Birleşik Kuvvetler	Korra and his friends police chief Lin beifong and	Korra and her friends, aided by police chief Lin Beifong and United Forces General Iroh,	54%	33%	71%

	generali Iroh'nın da yardımıyla, Amon'u devirmeyi ve maskesini indirmeyi başarırlar.	United Forces General Iroh's help, Amon the overthrow and challenges başarırlar download.	unmask Amon as a bloodbender and Tarrlok's brother, ending the Equalists' coup.			
41	Çağdaş Torchwood sonunda "Army of Ghosts"/"Doomsday" bölümlerinde Doktor ve Rose'un ziyaretiyledir, bu noktada Londra'nın Kanarya Rıhtımı'nda olduğunu belli etmiştir ve kazara Siberadamlar'ın işgalini sağlar ve, ardından, Daleklerin.	The contemporary Torchwood “Army of Ghosts”/ “the Doomsday” sections of the Doctor and Rose's ziyaretiyledir London, at this point, the Canary harbor from the contest, and that certain accidentally siberadamlar allows the invasion and, after daleklerin.	Contemporary Torchwood is finally visited by the Doctor and Rose in "Army of Ghosts"/"Doomsday", at which point it is situated within London's Canary Wharf and accidentally allows the invasion of the Cybermen and, subsequently, the Daleks.	47%	15%	83%
42	Lancaster House, Londra'daki bir zengin dönem evi, Kral'ın konuşma yapması için yürüdüğü sahnel ile sonradan resmî fotoğraf çekimi için Buckingham Sarayı'nın içi kullanıldı; günlük kiralanması 20.000 sterlin etti.	Lancaster house in London, a term In wealthy homes, the King's speech Ahmedabad to do with the official sahnel was shot her Buckingham Palace for the day in 339 yes kiralanması £20,000.	Lancaster House, an opulent, government-owned period house in London, was used for the interiors of Buckingham Palace that the King walks through prior to making his speech and for the official photograph afterwards; it cost £20,000 a day to rent.	44%	3%	100%
43	Kitabın diğer temaları arasında ahlak, itaat, fedakarlık, kefaret, sevgi, ve hukuk bulunur.	The other themes moral, and obedience, make sacrifice, Atonement, Love, and law.	Other themes in the book include morality, obedience, sacrifice, redemption, love, and law.	64%	6%	67%

44	==== Esther ====The Book of Esther was composed in the late 4th or early 3rd century BCE among the Jews of the eastern diaspora.	= Esther =Ruth=The Book of Esther was composed in the late 4th or early 3rd century BCE among the Jews of the Eastern diaspora.	====Esther====The Book of Esther was composed in the late 4th or early 3rd century BCE among the Jews of the eastern diaspora.	91%	92%	13%
45	=== Pasifik'teki diğer bölgeler ===Depremden kısa bir süre sonra Hawaii'deki Pasifik Tsunami Uyarı Sistemi (PTWC) Pasifik Okyanusu'ndaki bazı bölgeler için tsunami uyarısı yaptı.	#NAME?	===Elsewhere across the Pacific===Shortly after the earthquake, the Pacific Tsunami Warning Center (PTWC) in Hawaii issued tsunami watches and announcements for locations in the Pacific.	42%	3%	89%
46	=== Kıtalara göre kontenjanlar ===Aşağıdaki tablo FIFA tarafından her turnuva için kıtalara ayrılan kontenjanları göstermektedir.	#NAME?	=== Qualification spots by continent ===The table below lists the numbers of spots allocated by FIFA for each continent in each tournament.	32%	5%	100%
47	Pompey MÖ 63'te Suriye'ye vardı, iki kardeş ve üçüncü bir parti diğer Haşmonayim hanedanlıklarının ortadan kaldırılma arzularını Pompey'e gönderilen temsilcilerle iletti.	Pompey 63 BC, Syria, he had two brothers and a third party to other haşmonayim hanedanlıklarının kaldırılma combining investor fervor Pompey temsilcilerle to authorities.	When Pompey arrived in Syria in 63 BC, both brothers and a third party that desired the removal of the entire dynasty, sent their delegates to Pompey, who however delayed the decision.	32%	13%	100%
48	Germanus (1267)*136.	Germanus II (1267 *136).	Germanus III (1266)*136.	29%	51%	75%

49	Akıl almaz bir intikam planını durdurmak, Carter ve Sadie'ye düşüyor.	mental no plan for a revenge stop, Carter and Sadie to fall.	Percy, Carter, Annabeth, and Sadie all appear together.	20%	6%	92%
50	==Kaynakça==Spotify 2006 yılından bu yana Stokholm, İsveç merkezli Spotify AB bünyesindeki bir ekip tarafından geliştirilmektedir.	=kaynakça=spotify since 2006 headquartered in Stockholm, Sweden, Spotify AB the progressive developed by a team.	==History==Spotify was developed in 2006 by a team at Spotify AB, in Stockholm, Sweden.	48%	7%	87%

F-Measure, BLEU and TER scores of NMT-1

Row	Source	NMT-1 Output	Reference/Target	F-M	BLEU	TER	P
1	Ayrıca, bu grupta hastanede yatış süresi boyunca (risk oranı 3.91, % 95 güven aralığı 1.12-13.6; p=0.03) ve 30 günlük takipte (risk oranı 5.08, % 95 güven aralığı 1.64-15.7; p=0.005) ölüm riskinin anlamlı derecede yüksek olduğu görüldü.	In addition, there was no significant difference in the follow-up period (95% confidence interval confidence interval, 95% CI confidence interval, 95% confidence interval [CI], p=0.03).	A significant association was found between high levels of MPO and increased mortality risk during both hospitalization (hazard ratio-HR, 3.91; 95 % CI 1.12-13.6; p=0.03) and 30-day follow-up period (HR 5.08, 95 % CI 1.64-15.7; p=0.005).	16%	2%	100 %	55%
2	Triküspid akımında ise, Behçet'lilerde VTI E/VTI A, VTI E/VTI total ve V zirve E/A değerlerinde azalma,	There was no significant difference in the presence of MR, E/A ratio, and E/A ratio, E/A ratio,	In tricuspid flow, VTI E/VTI A, VTI E/VTI total and V peak E/A decreased, VTI A,	24%	2%	100 %	79%

	VTI A, VTI A/VTI total ve V zirve A değerlerindeki artış istatistiksel olarak anlamlı bulundu (5.7±1'e karşılık 4.6±1 cm, p <0.05).	E/A ratio, and E/A ratio (p <0.05).	VTI A/VTI total and V peak A diminshed in patients with Behçet's disease, but only an increase of VTI A was significant (5.7±1 vs 4.6±1 cm, p <0.05).				
3	Yatışının yedinci gününde yapılan koroner anjiyografide, sol ana koroner arterin tam tıkalı olduğu ve sol ön inen ve sirkumfleks arterlerinin sağdan kollaterallerle TIMI III akımla dolduğu görüldü.	Coronary angiography showed the left main coronary artery in the left main coronary artery.	Coronary angiography performed on the seventh day of admission showed total occlusion of the left main coronary artery and well-developed collateral vessels extending from the right coronary artery to the left anterior descending (LAD) and circumflex arteries.	43%	7%	100 %	9%
4	Grup-C hastalarında Lp(a) seviyesi düşük olanlarda, yüksek olanlara göre patensi oranı anlamlı olarak daha yüksek bulundu (p <0.01) Grup-C hastalarının Lp(a) seviyesi, Grup-A ve B ile benzer oranlarda idi (p>0.05).	Group B was significantly higher in patients with CAD (p <0.05).	In Group C however, patency rate had been found significantly higher in cases with high Lp(a) level compared to cases with low Lp(a) (p <0,01) Lp(a) level in	32%	0%	100 %	58%

			Group C patients were similar with Group A and B (p>0.05).				
5	Angiografide SAKA kısa bir segmentten sonra tıkalı idi.	The patient was discharged from the right ventricle.	Left main coronary artery was completely occluded after a short proximal segment in one of the patients (A.Y.).	15%	2%	100 %	31%
6	AMAÇ İnternal kardiyoversiyonun (İKV), atriyal fibrilasyonlu (AF) hastalarda sinüs ritmi sağlamak için etkili bir yöntem olduğu gösterilmiştir.	Objectives We concluded that (ICD) patients with atrial fibrillation (AF) in patients with atrial fibrillation (AF).	Internal cardioversion (ICV) has been suggested as an effective method to restore sinus rhythm in patients with atrial fibrillation (AF).	33%	23%	88%	10%
7	Başvuru sırasında ölçülen ortalama ejeksiyon fraksiyonu (EF) % 33±13 idi.	The mean ejection fraction (EF) was 3% .	Mean ejection fraction (EF) measured at admission was 33±13 % .	63%	33%	75%	9%
8	Postoperatif erken dönemde 3 hasta (% 8,1), geç dönemde ise 2 hasta (% 5,9) kaybedildi.	Two patients (1%) died in the early postoperative period.	Three patients (8.1 %) were lost in the postoperative early period and 2 patients in the late follow up.	40%	3%	100 %	12%
9	Anahtar Kelimeler: Anjiyoplasti, translüminal, perkütan koroner, kalp kateterizasyonu/enstrüm antasyon/yöntem; koroner anjiyografi;	Keywords: Angioplasty, balloon, percutaneous coronary intervention, coronary	Keywords: Angioplasty, transluminal, percutaneous coronary, heart catheterization/ins trumentation/meth	46%	4%	60%	11%

	koroner darlık/tedaviBüyük Arter Transpozisyonunda Arteryel Switch Operasyonu (Jatene Prosedürü)	angiography; coronary angiography; coronary artery anomalies/complications; coronary artery/stents/stents	ods; coronary angiography; coronary stenosis/therapy				
10	"Arteryel switch" ameliyalı anatomik düzelmıe sağlaması ve uzun dönem sonuçlarımn atriyal düzeyde düzeltme sağlayan ameliymlara oranla daha iyi olması nedeniyle, günümüzde birçok merkezde TGA 'n.m cerrahi fedavisinde ilk seçenek o/muşfllr.	It is concluded that there is a significant difference between the treatment and long-term follow-up period.	Currently "arterial switch" procedure is the treatment of choice for the transposition of great arteries in most of the major surgical centers dealing with congenital heart disease.	14%	2%	100 %	100%
11	Kompleks transpozisyon grubundaki hasta/ann ortalama yaşı 9 aydı.	The mean age of the patients was 9 months.	The mean age in complex TGA group was 9 months (18 days to 2 years).	42%	6%	100 %	26%
12	Ancak 15 günden biiyiik basit TGA 'lı hastalarda ekokardiyografik olarak uygun sol ventrikül m01jolojisi tespit edilmeyen olgularda jatene morta/itesi önemli oranda artmaktadır. Bu hastalarda morta/itesi diişük bulduğumuz iki aşamalı tamiri önernıekleyiz.	However, there was no significant difference between the left ventricular outflow tract and a decrease in the left ventricular outflow tract.	We believe the morta lity significantly increases in patients operated later than 15 days of age especially if the echocardiographic evaluation reveals unfavorable left ventricle morphology to recommend the "two-stage approach" in this	14%	3%	100 %	100%

			patient population.				
13	Anahtar Kelimeler: Ekokardiyografi, left ventrikül, mitral-aortik intervalvüler; psödoanevrizma.	Keywords: Echocardiography, ventricular dysfunction, mitral valve implantation.	Keywords: Echocardiography , left ventricle, mitral-aortic intervalvular; pseudoaneurysm.	29%	8%	71%	29%
14	İnfarktüs alanında canlı miyokard ile kollateral akım ilişkisini değerlendirmek ve kollateral dolaşımın sol ventrikül fonksiyonu üzerine etkisini irdelemek amacı ile ilk kez miyokard infarktüsü (Mİ) geçiren ve sol ön inen koroner arter (LAD) proksimali tam tıkalı 20 anteriyor Mİ olgusu incelendi.	The aim of this study was to investigate the effect of myocardial ischemia in the left anterior descending coronary artery (LAD) and left anterior descending coronary artery (LAD) and left anterior descending coronary artery (LAD).	This study attempted to determine the assodation between viable myocardium in the infarct area and collateral flow and also to assess the role of collateral flow on preservation of left ventricular function. We studied 20 patients with a first recent anterior myocardial infaretion who had total occlusion of the proximal part of the left anterior descending coronary artery.	39%	10%	100 %	29%
15	Arulan risk parametreleri e rkeklerde 96 crn 'lik bel geuişliğinde11 itibaren yiiksek bulrwdu (p	It was concluded that there was a significant correlation between the risk factors and	Furtlýermore. diabetes was morefrequent by 2.2-fold in men and by 4.8-jofd in	12%	6%	100 %	100%

		the presence of hypertension.	women across rlýese quintifes.				
16	Serbest radikallere bağlı lipid peroksidasyonu, kalp operasyonları sonrasında görülen metabolik ve ventriküler fonksiyon bozukluklarının sebeplerindendir.	It is the most common cause of myocardial ischemia and ventricular dysfunction in patients with acute coronary syndromes.	Free radical lipid peroxidation contributes to the abnormal metabolism and ventricular function frequently seen after cardiac operations.	17%	4%	89%	100%
17	Hastaların 11'inde kardiyopulmoner canlandırma yapılması gerekti.	None of the patients had cardiac arrest.	Eleven patients required cardiopulmonary resuscitation.	17%	8%	86%	19%
18	HoAH için lomitapidin ve mipomersenin ruhsatlandırılmasından sonra tedavi yaklaşımlarının sayısı artmıştır.	It is concluded that the success rate and the use of this intervention is important to reduce the success and treatment.	The number of therapeutic approaches has increased following approval of lomitapide and mipomersen for HoFH.	17%	6%	95%	100%
19	Koroner arter hastalığı olan grupta normal koroner akım vardı. Tüm hastalarda koroner akım düzeltilmiş TIMI kare sayısı ile değerlendirildi ve serum GGT düzeyleri ölçüldü.	Coronary flow was performed in all patients with coronary artery disease.	Coronary flow was quantified using the corrected TIMI frame count (TFC) method and serum levels of gamma-glutamyltransferase were measured.	20%	2%	100 %	17%
20	Bu tedaviyle ateş iki hastada gerileme gösterdi; üç hastaya ise yüksek doz	Cardiac transplantation was administered in all patients in two patients.	The remaining three patients received high-dose	12%	7%	90%	43%

	metilprednizolon verildi.		methylprednisolone.				
21	Kronik (Kr.)AF'Iu romatizmal kalp hastalığı olanlarda kardiyologların % 38.1'i warfarini kullanıyor, % 22.1 'i kullanmıyordu. Bu oranlar iç hastalıkları uzmanlarında sırasıyla % 27.3 ve % 52.2 bulundu.	Atrial fibrillation in patients with rheumatic heart disease (AF) is more important in patients with congestive heart disease.	In patients with rheumatic heart disease and chronic atrial fibrillation, the rate of regular and conditional warfarin prescription was reported to be 38.1 % and 39.8 % , respectively for cardiologists, whereas these rates were 27.3 % and 52.2 % for internists, respectively.	23%	12%	100 %	35%
22	Ancak ana ventrikül morfolojisi ile kardiyak patolojinin ventrikül şeklinin değerlendirilmesinde kullanılan çap/uzunluk oranında istatistiki anlamlılık taşıyan fark yaratmadığı gözlendi.	However, there was no significant difference between the two groups between the patients with left ventricular outflow tract and cardiac function.	Also the ventricular shape in patients with Fontan operation was more like an ellipsoid than other patient groups. The main ventricular morphology and cardiac pathology had no influence on this finding.	35%	3%	100 %	100%
23	Dobutamin Stres Ekokardiyografisinin İskemik Kalp Hastalıklarında Tanı Değeri	Effect of Exercise in Myocardial Infarction	Usefulness of Dobutamine Stress Echocardiography for Detecting Ischemic Heart Disease	13%	5%	100 %	43%

24	Konu ve Yazar Dizini	Analysis of Experience	Subject and Author Index	0%	34%	100 %	100%
25	On bir yaşındaki kız çocuğu halsizlik, efor dispnesi ve göğüs ağrısı şikayetleri ile hastanemize başvurdu.	A 53-year-old female patient was admitted to our hospital with dyspnea and dyspnea.	An 11-year-old girl was admitted to our hospital with a history of fatigue, effort dyspnea, and chest pain.	52%	31%	92%	7%
26	Tek değişkenli analizde MS’li hastaların aort halkası ve STB’si daha ufaktı (sırasıyla, p=0.003 ve p=0.043).	Logistic regression analysis showed that the presence of aortic valve and type B levels were found to be associated with the presence of aortic regurgitation.	In univariate analysis, patients with MS had a smaller annulus and STJ (p=0.003 and p=0.043, respectively).	10%	5%	96%	16%
27	Kor triatriatum sinister: Olgu serisi	Cor triatriatum sinister is a case of pericardial effusion	Cor triatriatum sinister: a case series	53%	8%	56%	72%
28	Hastalar ortalama 4 yıl boyunca semptomsuz olarak takip edildi.	The patients were followed up for 1 years.	Patients were followed up for a median of 4 years and were symptom free.	45%	27%	100 %	22%
29	Renkli akım Doppler incelemesinde trabeküller içinde kan akımı görüldü.	Color Doppler echocardiography showed normal blood flow.	Color flow Doppler examination confirmed the presence of blood flow within the trabeculae.	30%	4%	100 %	49%
30	Kırgız toplumunda esansiyel hipertansiyonlu hastalarda iskemik inmenin risk faktörü olarak anjiyotensin dönüştürücü enzim	The risk of risk factors in patients with essential hypertension in patients with essential hypertension	The I/D polymorphism of the angiotensin converting enzyme gene as a risk factor for ischemic stroke in patients with	46%	18%	100 %	25%

	geninde I/D polimorfizminin rolü		essential hypertension in Kyrgyz population				
31	1-3 (1.4±0.6) Kompresyon zamanı :	Average median sternotomy (range:	1.4 ±0.6); mean compression time:	0%	15%	100 %	100%
32	Aile desteği, akut koroner sendrom sonrası morbidite ve mortaliteyi azaltmada önemli bir role sahiptir.	It is an important role in morbidity and mortality in patients with acute coronary syndrome.	Family support has a critical role in decreasing morbidity and mortality after acute coronary syndrome.	53%	7%	60%	13%
33	Anahtar Kelimeler: Non Hodgkin Lenfoma, Miyokardiyal İnfarktüs, Metastaz	Keywords: Myocardial infarction, Myocardial bridge, hepatitis B	Cardiac metastasis of non-Hodgkin's lymphoma presenting with acute antero-lateral myocardial infarction with ST-segment elevation	10%	2%	100 %	100%
34	Sonuç olarak çalışmamız, KAH olan hastalarda, kaptopril'in Lv diyastolik fonksiyonunda belirgin iyileşme sağladığını ve bu iyileşmenin miyokardın oksijen gereksiniminin azalmasına ve muhtemelen miyokardın kan akımındaki artışa bağlı olduğunu düşündürmektedir.	In conclusion, there was no significant difference in the diastolic function in patients with CAD.	This study suggested that captopril significantly improved LV diastolic function in patients with CAD and that this improvement was due to a decrease in myocardial oxygen requirement and presumably to an increase in myocardial blod flow.	31%	9%	100 %	52%

35	İnferior AMİ grubunda 21 hasta (1 kadın, 20 erkek; ort. yaş 56), anterior AMİ grubunda 18 hasta (2 kadın, 16 erkek; ort. yaş 56) vardı.	The patients were divided into two groups according to the ST segment elevation in patients with acute myocardial infarction.	Inferior and anterior AMI groups included 21 patients (1 female, 20 males; mean age 56 years) and 18 patients (2 females, 16 males; mean age 56 years), respectively.	13%	6%	100 %	15%
36	Çokdeğişkenli lineer regresyon analizinde, sol ventrikül MPİ (b=0.450, p=0.001) ve sağ ventrikül izovolümetrik gevşeme zamanı (b=0.507, p <0.001) sağ ventrikül MPİ ile bağımsız ilişki gösteren değişkenlerdi.	In multivariate logistic regression analysis, left ventricular ejection fraction (EF), and right ventricular outflow tract (LVMI) were found to be independent predictors of left ventricular hypertrophy.	In multivariate regression analysis, the left ventricle MPI (b=0.450, p=0.001) and right ventricular isovolumetric relaxation time (b=0.507, p <0.001) showed independent relationship with the right ventricle MPI.	34%	3%	73%	13%
37	Anahtar Kelimeler: Avrupa Kardiyoloji Derneği, eğitim, kardiyoloji uzmanlık eğitimi, Türk Kardiyoloji Derneği	Keywords: European Society of Cardiology, Turkey, Turkey, Turkey	Keywords: European Society of Cardiology, training, cardiology specialty training, Turkish Society of Cardiology	48%	28%	100 %	9%
38	Demografik özellikler, gebelik sırasında geçirilen kardiyak girişim hikâyesi, akciğer ödemi, anne	Demographic characteristics of the patients were evaluated with respect to clinical	Demographic characteristics including history of cardiac intervention	32%	3%	100 %	100%

	veya çocuğa ait ölüm bilgileri, hastane bilgi işlem, kardiyoloji ve doğum kliniklerinin kayıtlarından toplandı. Ekokardiyografik inceleme başvuru anında, üçüncü trimesterde ve doğumdan bir ay sonra yapıldı.	characteristics, cardiac catheterization, and cardiac arrest, and the patient were followed up after surgery.	performed during pregnancy, pulmonary edema, and maternal and fetal mortality, and cesarean section (C/S) history were collected from the hospital database and clinical records of the cardiology and obstetrics departments.				
39	Genetik testlerde hastanın heterozigot metilentetrahidrofolat redüktaz (MTHFR) taşıyıcısı olduğu saptandı.	Keywords:Türk Kardiyol Dern Arþ 2004; 32: Türk Arþ 2004; 32:89)	Genetic tests revealed that the patient was a heterozygous carrier of the methylenetetrahy drofolate reductase (MTHFR) gene mutation.	0%	2%	100 %	100%
40	Anahtar Kelimeler: pnömoperikardiyum, pnömotoraks, dispne	Keywords: pneumothorax, pneumothorax, pneumothorax	Keywords: pneumopericardiu m, pneumothorax, dyspnea	50%	16%	50%	7%
41	Uzun QT sendromu (UQTS) EKG'de QT intervalinde uzama ve "torsade de pointes" (TdP) tipi ventriküler aritmi ve tekrarlayan senkop veya ani ölümle karakterize kardiyak iyon kanalı bozukluğudur.	Magnetic resonance syndrome (MRI) is an important role in the diagnosis of metabolic syndrome (MRI) in the diagnosis of ventricular arrhythmias and sudden death.	Long QT syndrome (LQTS) is a disorder of cardiac ion channels that affect repolarization and is characterized by QT interval prolongation on the ECG and	27%	2%	100 %	80%

			torsade de pointes arrhythmias leading to recurrent syncope or sudden death.				
42	Yavaş koroller aktnıı olan hastalardaki göğüs ağnstllt. n oluşumunda endotelin-i artışı sorumlu olabilir.	It is concluded that the frequency of the myocardium is the most important role in the treatment of the myocardium.	The increase in ET-1 levels may be responsible for the development of the chest pain in patients with slow coronary flow. Additionaly much more elinical studies are necessary to eletermine the exact role of ET - 1 in the development of microvascular clysfunction.	29%	2%	100 %	100%
43	Ambulatuvar 24 saatlik Holter takibinde, ortalama 74 atım/dk kalp hızı ile esas olarak sinüs ritminde olan hastada geçici ikinci derece AV blok ve 10.6 saniyeye kadar varabilen ileri derecede AV blok atakları izlendi.	After 24-hour Holter monitoring, the mean age of the patients were measured by 24-hour Holter monitoring, the mean age of the patients with sinus rhythm was found to be normal.	Ambulatory 24-hour Holter monitoring showed predominant sinus rhythm with a mean heart rate of 74 bpm, intermittent second-degree AV block, and high-degree AV block with pauses of up to 10.6 seconds.	30%	3%	90%	84%
44	Perkütan transvenöz endomiyokardiyal biyopsi materyalinin	In order to evaluate the purpose of the study, the European	Percutaneous transvenous endomyocardial	21%	3%	100 %	100%

	histolojik incelemesinde, ISHLT (International Society of Heart and Lung Transplantation) sınıflamasına göre derece IA doku reddi saptandı.	Society of Cardiology was used to evaluate the European Society of Cardiology.	biopsy yielded a histologic diagnosis of grade IA rejection according to the ISHLT (International Society of Heart and Lung Transplantation) scoring system.				
45	Sonuç olarak, STİ'nin standart egzersiz parametrelerinin yetersiz olduğu hastalarda ÇDH tanısında kullanılabilecek yeni bir kriter olduğu kanısına varıldı.	In conclusion, it was concluded that the use of the use of exercise test in patients with exercise levels were found to be used in patients with exercise.	In conclusion, SRI may be used as a new index for the identification of MVD in patients, in whom the accuracy of standard electrocardiographic criteria is unsatisfactory.	36%	5%	89%	29%
46	Ancak bu durumun klinik önemi ile ilgili net veri bulunmamaktadır.	However, there are no clinical findings.	However, there are scarce data regarding the clinical significance.	53%	7%	83%	18%
47	Tirofiban tedavisine başlanan hastada göğüs ağrısı tekrarlamadı ve 72. saatte yapılan koroner anjiyografide septal arter içindeki trombüsün erimiş olduğu görüldü; ancak sol ön inen arterde yavaş akım devam ediyordu.	Coronary angiography was performed in a patient with a history of chest pain and the left anterior descending artery in the left anterior descending artery.	Tirofiban administration resulted in relief of chest pain. Control coronary angiography performed at 72 hours showed complete disappearance of the intracoronary thrombus, but slow flow in the	45%	17%	100 %	50%

			left anterior descending coronary artery persisted.				
48	Anahtar Kelimeler: Anemi/komplikasyon/tedavi, anemi, demir yetmezlkli/ilaç tedavisi; eritropoietin/terapötik kullanım; kalp yetersizliği konjestif; hematokrit; hemoglobin; demir; böbrek yetersizliği, kronik; risk faktörü.	Keywords: Adrenergic beta-antagonists/therapeutic, heart failure, iron deficiency/drug therapy; drug therapy; heart failure.	Keywords: Anemia/complications/therapy, anemia, iron-deficiency/drug therapy; erythropoietin/therapeutic use; heart failure, congestive; hematocrit; hemoglobins; iron; kidney failure, chronic; risk factors.	27%	13%	100 %	100%
49	Bu hastaların üçü (3/10, % 30) yan dal olmayan, üçü de yan dalı açık olan (3/4, % 75) gruptandı.SONUÇ Sol internal mamaryan arter yan dal varlığında miyokard iskemisini değerlendirmede LIMA akım hızı yol gösterici olabilir ve bunun objektif değerlendirilmesi için TIMI kare sayısı kullanılabilir.	In this study, we aimed to investigate the relationship between the left main coronary artery and the presence of coronary artery disease (CHD), and the TIMI frame count was used to determine the TIMI frame count.	The incidence of anterior ischemia was 30 % (3/10) for those without a side branch, and 75 % (3/4) for those with a side branch.CONCLUSION We suggest that, in the presence of a LIMA side branch, LIMA flow may be used in assessing myocardial ischemia, and TIMI frame count is an objective means of	28%	2%	100 %	74%

			measuring LIMA flow.				
50	Çalışmamızın amacı, "Kısmi önemlilik endeksi (KÖE)" adını verdiğimiz bu endeksin klinik sonlanım ve sol ventrikül (SV) sistolik fonksiyonlarındaki düşme ile ilişkisini araştırmaktır.ÇALIŞMA PLANI Daha önce akut koroner sendrom hikayesi olmayan ve ön duvar ME'si geçiren ardışık 123 hasta ileriye dönük bir şekilde çalışmaya alındı.	The aim of this study was to investigate the relationship between the presence of coronary syndrome and left ventricular (LV) systolic function in patients with acute coronary syndrome.	The aim of the study was to assess the predictive value of RII in left ventricle (LV) systolic function reduction and its relation to adverse clinical outcome.STUDY DESIGN One hundred twenty-three acute anterior MI patients with their first acute coronary syndrome incident were consecutively and prospectively enrolled in to the study.	51%	3%	100 %	41%

F-Measure, BLEU and TER scores of NMT-2

Row	Source	NMT-2 Output	Reference/Target	F-M	BLEU	TER	P
1	Çalışma Dizisi Yükle	Load Worksheet	Working Set	0%	47%	100%	100%
2	Tamamlanmamış ya da Geçersiz Yapılandırma	Incomplete or Invalid Configuration	Incomplete or Invalid Configuration	100%	100%	0%	2%
3	Biçim alanı	Format space	Style area	0%	47%	100%	5%
4	Öznitelik durumu	Attribute state	Attribute case	50%	51%	50%	5%

5	. Ağaç Dosyalarında üst seviye klasörleri kaldırabilir ve yeni yerel klasörler ekleyebildiğiniz. RMB Yeni Üst Klasörleri. Kullanın ve bir yerel klasör seçin ya da uzak klasör seçin be ftp: // kullanıcı@ sunucu nun formu içinde uzak sunucu adı girin.	. that you can remove top-level folders and add new local folders. You can use RMB New Top Folder. Use a local folder and select a local folder. Select a local folder and choose the remote server name inside ftp:/user@server.	. that you can add new local and remote top level folders to the Files Tree? Use RMB New Top Folder. and select a local folder or enter the remote server name in form of ftp: / /user@server and select the remote folder.	63%	34%	63%	38%
6	dakika	minutes	min	0%	47%	100%	21%
7	Renklendirilmiş nesne, değişik dizinlerde değişik dosya türüne sahip. Ne yapılacağını seçiniz.	The Colored object contains different file types in different folders. Select this.	The highlighted item has a different type in the different directories. Select what to do.	37%	7%	92%	13%
8	Bellekten çağır	Memory from Memory	Recall from memory	67%	59%	33%	100%
9	damla	drops	drop	0%	47%	100%	100%
10	Herhangi birini kullanır:	In any of:	By any of:	67%	59%	33%	4%
11	Üst dizin	Parent folder	Parent directory	50%	51%	50%	2%
12	Hızlı- ara	Match-search	Quicksearch	0%	47%	100%	100%
13	Tanımlanmış Bağlantılar	Defined Links	Defined Links	100%	100%	0%	3%
14	Dizin yolu adresi bozulmuş!	Directory path address is corrupt!	The directory path URL is malformed!	55%	11%	60%	15%
15	Tema Yüklenemiyor@ title: window	Unable to Load Theme	Cannot Load Skin	29%	14%	75%	7%
16	Fare Aracı Hakkında	About Mouse Tool	About KMouseTool	40%	46%	67%	6%
17	Metin Okuma	Text-to-Speech	Text-to-Speech	100%	100%	0%	1%
18	Bu düğme ile seçili sözlük bir dosyaya aktarılır.	With this button, the selected dictionary will be exported to a file.	With this button you export the selected dictionary to a file.	70%	11%	50%	3%
19	ZıplayanİnekName	Bouncing Indoor	BouncingCow	0%	47%	100%	60%

20	KumppaName	Kumppa	Kumppa	100%	100%	0%	3%
21	XMMS Oynatıcı penceresiComment	XMMS Player window	XMMS Player window	100%	100%	0%	2%
22	Arama seçeneklerini kapat@ label	Close search options	Close search options	100%	100%	0%	1%
23	Önizleme: @ info: tooltip	Preview	Preview:	0%	100%	100%	8%
24	Eksik hata ayıklama sembolleri paketleri yüklenmek isteniyor. @ info	Missing debugging symbols are requested to install packages.	Requesting installation of missing debug symbols packages.	40%	10%	88%	96%
25	Shift tuşu şimdi etkin değil.	The Shift key is now inactive.	The Shift key is now inactive.	100%	100%	0%	1%
26	Zayıf	Weak	Slight	0%	47%	100%	100%
27	Resim Bayt Sırası	Image Byte Order	Location	0%	47%	100%	100%
28	Pencereyi 14. Masaüstüne Gönder	Window to Desktop 14	Window to Desktop 17	75%	100%	25%	3%
29	Bu seçenek, bir dosyayı "silmek" istediğinizde Konqueror' un size sorup sormayacağını belirler. Çöpe Kutusuna Taşı: dosyayı, kolayca geri alabileceğiniz çöp dizinine taşır. Sil: dosyayı tamamen siler. @ option: check Ask for confirmation when moving to trash	This option determines you to ask if you want to remove a file. If you want to delete the file, you will move the file to the trash directory. Sil: The file is completely deleted.	This option tells Konqueror whether to ask for a confirmation when you "delete" a file. Move To Trash: moves the file to your trash folder, from where it can be recovered very easily. Delete: simply deletes the file.	38%	3%	86%	50%
30	Tarih değerlerinin nasıl görüntüleneceğini gösterir.	Shows how the date values will be displayed.	This is how date values will be displayed.	75%	54%	38%	4%
31	Pozitif işareti:	Positive sign:	Positive sign:	100%	100%	0%	2%
32	Efekti panellere uygula	Apply Effect	Apply effect to panels	67%	29%	100%	4%
33	hafif sağanakweather forecast	light showers	light shower	50%	51%	50%	3%

34	Çiselemeli Buz Parçacıklarıweather condition	Drizzle Ice Pellets	Drizzle Ice Pellets	100%	100%	0%	1%
35	Şiddetli Sağanak Yağmurlu Sis/ Pusweather condition	Heavy Showers Rain Fog/Mist	Heavy Showers Rain Fog/ Mist	67%	100%	50%	1%
36	Veri kaybını ve diğer zararları önlemek için sisteminizin askıya alma ya da uyku kipini desteklemesi gerekir ki kazara sisteminizi boş pil ile çalıştıramayasınız. Makinenizi yapılandırılmış eylemlerle çalıştırmak için aşağıdaki bölümden birkaç dakikada yapılandırın.	Configure the data loss to the system to prevent the systems and other losses to prevent the systems, if you need to suspend the system, and you cannot run your system. Configure the following part of a few minutes.	To prevent data loss or other damage, you can have the system suspend or hibernate, so you do not run accidentally out of battery power. Configure the number of minutes below which the machine will run the configured action.	41%	5%	90%	22%
37	Sesi AzaltA button on a Remote Control	Volume Reduction	Volume Down	50%	51%	50%	100%
38	2 B ÇizimComment	2 D Drawing	Plot2d	0%	51%	100%	100%
39	Çözücü miktarı daima hacim cinsinden belirtilir.	The amount of solvent is always specified in terms of volume.	Amount of solvent is always specified in terms of volume.	95%	89%	9%	2%
40	Alaşımlar	Subscription	Alloys	0%	47%	100%	100%
41	Hal	Hal	Non-Metals	0%	17%	100%	100%
42	Aracajubrazil. kgm	Taraju	Aracaju	0%	47%	100%	32%
43	Marsan Dağıfrance. kgm	Marsan Dam	Mont de Marsan	40%	31%	100%	100%
44	Langjökull Glaciericeland_ regions. kgm	Langjökull Glacier	Langjökull Glacier	100%	100%	0%	3%
45	Bresciaitalyprovince. kgm	Breslau	Brescia	0%	47%	100%	12%

46	Güney Hamgyŏngnorth_ korea. kgm	South Hamgyŏ n	South Hamgyŏ ng	67%	59%	33%	2%
47	Serravallesan- marino. kgm	Serravalles	Serravalle	0%	47%	100%	36%
48	Blekinge İlisweden. kgm	Blkingge County	Blekinge County	50%	51%	50%	3%
49	Lufwanyamazambia_ districts. kgm	Luwanyz	Lufwanyama	0%	47%	100%	100%
50	Nakondezambia_ districts. kgm	Nakondi	Nakonde	0%	47%	100%	25%

ANNEX III. Survey Answers for Professional Reviewers/Translators

	Timestamp	**Which languages do you speak? (Divide languages by comma)**	**What is your native language?**
Reviewer 1	29/10/2020 19:02:59	Turkish, English, German	Turkish
Reviewer 2	30/10/2020 00:55:29	Turkish (Native) , English (C1) , Spanish (A2)	Turkish
Reviewer 3	30/10/2020 10:43:20	Turkish, English, Spanish, Catalan	Turkish
Reviewer 4	30/10/2020 10:45:37	English, Turkish	Turkish
Reviewer 5	30/10/2020 11:23:30	Turkish, English, Spanish	Turkish
	What is the latest higher education level that you have completed?	**How long have you been working as a**	**Do you use machine translation in your daily workflow?**

	(Undergraduate, master, PhD)	**professional translator?**	
Reviewer 1	Undergraduate	3-5 years	No
Reviewer 2	Undergraduate	3-5 years	Yes
Reviewer 3	Master's Degree	5 - 10 years	No
Reviewer 4	Undergraduate	3-5 years	Yes
Reviewer 5	Undergraduate	Less than a year	Yes
	Do you perform human translation review tasks?	**Do you perform machine translation postediting tasks?**	**Do you have experience in translating/editing/postediting medical texts?**
Reviewer 1	Yes	Yes	No
Reviewer 2	Yes	Yes	Yes
Reviewer 3	Yes	No	No
Reviewer 4	Yes	Yes	Yes
Reviewer 5	Yes	Yes	No

ANNEX IV. Task Instructions for Translators

Guidelines for performing the TR - EN Evaluation Task

Title:	Machine Translation Quality Evaluation for Turkish - English Statistical MT and Neural MT engines
Objective:	The task consists of evaluating the translation quality of 4 Machine Translation Engines by human reviewers
Task facilitator:	Gokhan Dogru (Predoctoral Researcher, UAB)
Organization:	Facultat de Traducció i d'Interpretació, Universitat Autònoma de Barcelona
Supervisors:	Anna d'Aguilar-Amat Castillo & Adrià Martín
Approval	This experiment is approved by the Ethics Commission on Animal and Human Experimentation with the number of 20190927CEEEAH.

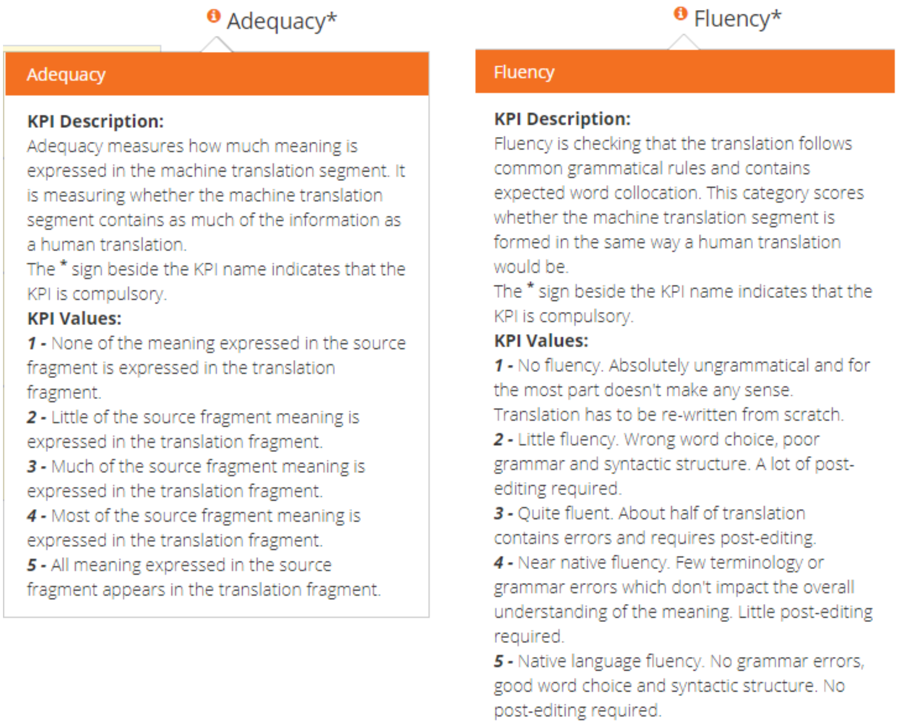
ANNEX IV. Task Instructions for Translators

Task Guidelines

1. Please fill out the anonymous, short survey aiming at collecting professional details of the participants. It should take less than 3 minutes to complete. Form link is here: https://forms.gle/Bnyvg1RmP8ZNfDa29
2. Once you complete the survey, we will send you a link to your email address to connect to KantanLQR platform. You will need to enter with your email and create a password.
3. Once you login to the platform, you will see the dashboard with the task. You should first click on the "?" to accept the task. Once you accept the task, you can click on the pen symbol and begin the evaluation task.
4. In the upper left corner, you will see the source sentence, and below it there will be its 4 machine translations by different MT engines. The order of these translations is randomized in each step to avoid bias.
5. There are 2 evaluation criteria: adequacy and fluency. In a nutshell, adequacy measures the accuracy of the translation compared to the source sentence while fluency measures how grammatically correct the translation. You will have a scale of 5 stars. More stars mean better adequacy or fluency. The "i" symbol near each title gives hints about the meaning of each star and the respective definitions. See the image below.
6. Finally, you are expected to rank each engine from the best to the worst. Again, more stars mean better quality. Hence, the best translation result should get 4 stars while the worst one should get 1 star. Note that if you think two engines are equal, you can assign the same number of stars to them.
7. You can press "Finish" to pause and leave the task before finishing and return later to complete it. There are 100 source sentences to be evaluated.
8. If you need more help about the task, please send an email to gokhan.dogru@uab.cat.

9. A simulation of the steps from the reviewer's perspective are also available in a video: https://drive.google.com/file/d/1bNbTVUhdvJDvenVryHuHlyFjoUFkgBh6/view?usp=sharing

Image: Tips for understanding Adequacy and Fluency available in KantanLQR



ANNEX V. MT Evaluation Resource Documents

Turkish source segments, their human reference translations, and machine translation of the segments by 4 TR → EN engines (PBMT-1, PBMT-2, NMT-1, NMT-2) are provided in this annex.

Turkish Source Segments

Turkish source segments are provided below. These segments are extracted from the source corpus. And their length is between 5 – 29 words. There are 4 segments from each length.

ID	Turkish Source	WC
1	Hiçbir rabdomiyoliz veya miyopati gözlenmedi.	5
2	Karotis stentleme sonrası gelişen psödoanevrizma:	5
3	Periferik embolinin az rastlanan nedeni:	5
4	Çoklu Safen Ven Greft Anevrizması	5
5	Fizik muayenede hipotansiyon ve taşikardi saptandı.	6
6	Hastaların klinik ve anjiyografik özellikleri değerlendirildi.	6
7	İnvazif nöroradyoloji konsültasyonu medikal takip önerdi.	6
8	Karotis arter stentlemesinin etkinliği ve güvenliği:	6
9	Genç kadın hastada arteryel torasik çıkış sendromu	7
10	İmplantasyondan hemen önce asistoli ve konvülziyon gelişti.	7
11	Koroner arter ektazisinde trombofilik gen mutasyon analizi	7
12	Sağ ventriküle açılan post-infarkt sol ventrikül anevrizması	7
13	Düşük kardiyak debi gelişimini değerlendirmek için DKDSs kullanıldı.	8
14	Gebelik sırasında bir hastada spontan aort kökü rüptürü	8
15	Kalp yetersizliği tedavisinde sodyum glikoz ko-transporter 2 inhibitörleri	8
16	Prognostik nutrisyonel indeks enfektif endokardit hastalarında mortaliteyi öngördürür	8
17	Gerçek yaşamda uzun süreli fibrat kullanımı etkili ve güvenlidir.	9
18	Jude medikal mitral sığır biyoprotez kapağın 27 yıllık dayanıklılığı	9
19	Koroner yavaş akım fenomeninde hücresel iskemi için yeni gösterge:	9
20	Sağ posteriyor serebral artere embolize olan koroner stent olgusu	9
21	Akut kalp yetersizliği en sık görülen kardiyak acil durumlardan biridir.	10
22	Bu alanda radyofrekans ablasyon nadir ancak ciddi komplikasyonlara sebep olabilir.	10
23	Koroner arter hastalarında CETP geni rs289714 varyasyonunun metabolik etkilerinin araştırılması:	10
24	Ventriküler erken vurular (VEV), klinikte en sık karşılaşılan aritmilerden birisidir.	10

25	Çocuklarda transkateter atrial septal defekt kapatılmasının sol kalp fonksiyonları üzerine etkisi	11
26	Karotis arter PA’sı, travma, spontan enfeksiyon, vaskülit veya iyatrojenik nedenlerden kaynaklanabilir.	11
27	Vasküler sistemin floroskopik taramasında stentin sağ posteriyor serebral arterde olduğu saptandı.	11
28	Ventriküler ekstrasistol ilişkili kardiyomiyopatili bir adölesanda başarılı sol koroner kusp kriyoablasyonu	11
29	İdiyopatik pulmoner fibrozun akut alevlenmesine bağlı olarak hipoksinin seyrindeki çeşitli ekokardiyografik değişiklikler	12
30	Kronik obstrüktif akciğer hastalarında atriyal elektromekanik gecikmenin ve P dalga dispersiyonun değerlendirilmesi	12
31	Protez kapak trombozu (PKT) gelişimi protez kalp kapaklı hastalarda ciddi bir komplikasyondur.	12
32	Serum laktat seviyesi, akut dekompanse kalp yetersizliğinde akut böbrek hasarının gelişmesini öngördürebilir	12
33	Aralık 2018-Nisan 2019 tarihleri arasında koroner anjiyografi uygulanan kararlı anjina pektorisli hastalar değerlendirildi.	13
34	Global kronik böbrek hastalığı yükü ve Türk kalp yetersizliği hastalarında böbrek fonksiyonlarında azalma	13
35	Majör vasküler komplikasyonlar ve kardiyak tamponadın düşük PNİ grubunda daha yüksek olduğu bulundu.	13
36	Primer perkütan koroner girişimden 2 ay sonra hastaya başarılı koroner baypas cerrahisi yapıldı.	13
37	COVID-19 pandemisinin ST-segment yükselmeli miyokart enfarktüsü nedeniyle yapılan primer perkütan koroner girişim zamanlamasına etkisi	14
38	Eşzamanlı obstrüktif aortik ve mitral prostetik kapak trombozları tanısında farklı görüntüleme yöntemlerinin tamamlayıcı rolü	14
39	Patent duktus arteriyozuslu hastaların tedavisinde eş zamanlı asetaminofen ve ibuprofen kullanımına yönelik artmış eğilim	14

40	Protez kalp kapak trombozu olan hastalarda trombolitik tedavinin hemen sonrasında endotel fonksiyonlarında gözlenen iyileşme	14
41	Atriyal fibrilasyon ve sinüs ritminde olan korunmuş ejeksiyon fraksiyonlu kalp yetersizliği hastalarının klinik özelliklerinin karşılaştırılması:	15
42	Aynı kardiyovasküler sonlanıma sahip ailevi hiperkolesterolemi ve yüksek lipoprotein (a) düzeyi olan tek yumurta ikizleri:	15
43	Çocuk yoğun bakım ünitesinde kardiyak cerrahi sonrası hastaların değerlendirilmesinde kullanılan düşük kardiyak debi sendromu skorlaması	15
44	ST yükselmeli miyokart enfarktüsü hastalarında çekilen pıhtının histopatolojik özelliklerinin uzun dönem sol ventrikül fonksiyonuyla ilişkisi	15
45	Bu olgu raporu, 73 yaşında bir erkekte 2 hipoksik atak epizodunda gözlenen farklı ekokardiyografik değişiklikleri göstermektedir.	16
46	Elli sekiz yaşında kadın hasta iki hafta önce başka bir hastanede atipik pnömoni nedeniyle hastaneye yatırılmış.	16
47	Kemik ajanları ile yapılan sintigrafi, TTR-KA tanısında oldukça güvenilir ve uygulaması kolay bir yöntem olarak gösterilmektedir.	16
48	Transkateter ASD kapatılması ile sol atriyum sistolik ve sol ventrikül diyastolik fonksiyonlarında olumlu yönde etkilenme izlenmiştir.	16
49	Brugada sendromu (BrS) konjenital bir kanalopati olup malign ventriküler aritmi gelişimi ve ani kardiyak ölüm ile ilişkilidir.	17
50	Elektrokardiyografide hızlı idiyoventriküler ritim, yakalama ve füzyon vurularıyla birlikte DII, DIII, aVF derivasyonlarında belirgin J dalgası izlendi.	17
51	Karotis arter stentlemesi, karotis arter stenozunun tedavisinde son 30 yıldır yaygın olarak kullanılan girişimsel bir tedavi yöntemidir.	17
52	Obez hastalarda sağ ventriküler disfonksiyonunun saptanmasında miyokardiyal performans indeksi ile isovolumetrik kasılma fazındaki sağ ventriküler akselerasyonun karşılaştırılması	17
53	Bu yazıda, akut miyokart enfarktüsü geçirmesinden bir hafta sonra dispne ile acil servise başvuran diyabetik bir hastayı sunuyoruz.	18

54	Bu, kardiyovasküler olay gelişimi eşiğini aşan kümülatif LDL maruziyetinin benzer zamanlaması nedeniyle AH ailelerinin önemli bir özelliği olabilir.	18
55	Düşük kardiyak debi sendromu (DKDS) tanımlanması ve bu hasta grubunun takibi ile ilgili net bir fikir birliği bulunmamaktadır.	18
56	Kardiyovasküler hastalıklar (KVH) tüm dünya genelinde morbidite ve mortalitenin başlıca sebeplerinde biridir ve sağlık sistemine yüksek maliyet getirmektedir.	18
57	Bir yıl önce, sol ventrikül destek cihazı (SVDC) implantasyon öyküsü olan 58 yaşında erkek hasta istirahat dispnesi ile başvurdu.	19
58	Bununla birlikte, AF mevcut olmayan hastalar daha yüksek diabetes mellitus, obstrüktif uyku apnesi ve koroner arter hastalığı yüküne sahipti.	19
59	Sıklıkla hipermobil eklemler veya kronik yorgunluk sendromu, kronik karın ağrısı, migren, baş ağrısı ve şeker hastalığı gibi durumlarla ilişkilidir.	19
60	Son derece aktif antiretroviral tedavi, insan immün yetmezliği virüsü (HIV) pozitif hastalar için daha uzun yaşam beklentisini mümkün kılmıştır.	19
61	Bu yazıda, çok sık nonsustained ventriküler taşikardi atakları ve prematür ventriküler kontraksiyon ilişkili kardiyomiyopatisi olan 17 yaşındaki erkek hasta sunuldu.	20
62	Genellikle iyi huylu olarak kabul edilmesine rağmen, kardiyomiyopati gelişimine sebep olabilir ve nadiren de ani kardiyak ölüm ile ilişkili olabilir.	20
63	Kablosuz kalp pilleri, yüksek riskli transkateter kapak değiştirme prosedürleri geçiren hastalar için geleneksel transvenöz kalp pillerine potansiyel bir alternatif sunar.	20
64	Uygunsuz sinüs taşikardisi (UST) herhangi bir ikincil neden olmadan artmış istirahat kalp hızı ve rahatsız edici semptomlarla karakterize bir sendromdur.	20
65	Bizim vakamızda, 41 yaşında erkek hastada evde çarpıntı, bilinç bulanıklığı ve ardından bayılma gelişti. Bilinç kaybının ardından kalp durması meydana geldi.	21

66	Çalışmada değerlendirilen hemostatik değişkenlerin serum seviyeleri ve demografik ve klinik parametreler, ikili lojistik regresyon analizinde SAA trombüsünü belirleme konusunda anlamlı bulunmadı.	21
67	Koroner yavaş akım fenomeni (KYAF) en az bir majör epikardiyal koroner arterde kan akımının distal damar yatağına geç ulaşması olarak tanımlanır.	21
68	Örneğin L/O yapısı, taze pıhtı varlığına kıyasla daha iyi taburculuk öncesi ve uzun dönem SV hacimleri ve ejeksiyon fraksiyonu değerleriyle ilişkiliydi.	21
69	Akut miyokardit tanısında GKT ve/veya ödem varlığı tanı kriteri olarak kabul edildiğinde Lake Louise kriterlerine göre daha yüksek tanısal doğruluk elde edildi.	22
70	Koroner yaprakçıklardan kaynaklanan ventriküler aritmi nadir olmayan bir durumdur. Ancak aortik yaprakçıklardan kaynaklanan çıkım yolu ventriküler aritmilerinin haritalaması ve ablasyonu uğraştırıcı olabilir.	22
71	Obezitenin kalp yetersizliği için bir risk faktörü olduğu bilinmesine rağmen, bu konuda yapılmış çalışmaların çoğunda obezitenin sol kalp fonksiyonları üzerine etkisi incelenmiştir.	22
72	Serum albümin ve lenfosit konsantrasyonlarını baz alan prognostik nutrisyonel indeks (PNİ) çeşitli hasta popülasyonlarında prognozu gösteren yeni bir enflamasyon temelli risk skorudur.	22
73	Biz bu çalışmada, COVID-19 döneminde ST-segment yükselmeli miyokart enfarktüsü (STYME) tanısı alan hastaların primer perkütan koroner girişim (PPKG) yapılana kadar geçen sürelerini değerlendirdik.	23
74	Bu çalışmanın amacı, koroner arter hastalığı (KAH) olan hastalarda CETP rs289714 gen polimorfizminin, serum lipit profili ve diğer metabolik parametreler üzerindeki etkilerinin araştırılmasıdır.	23
75	Monozigotik ikizler intrauterin çevreyi, yaşı ve tüm genlerini ortak paylaştıklarından, AH’nin nedenlerini ve doğal seyrini araştırmak için çok özel bir kaynağı temsil edebilirler.	23

76	Örnekte artmış lökosit sayısı ve taze pıhtı varlığı, klinik ve işlemle ilişkili özelliklerden bağımsız olarak SV fonksiyonlarında uzun dönemde bozulma ile ilişkili olabilir.	23
77	Ancak yeni veriler korunmuş ejeksiyon fraksiyonlu kalp yetersizliği olgularının %13 ve yüksek riske sahip ciddi aort darlığı bulunan yaşlı olguların %16'sında TTR-KA bulunduğunu göstermektedir.	24
78	Bu çalışma, bir lipit kliniğinin hipertrigliseridemi hastalarında fibrat tedavisinin etkinliği, güvenirliği ile kardiyovasküler, diyabetik sonlanım noktalarına etkisini gerçek yaşam verilerine dayanarak ortaya koymayı amaçlamıştır.	24
79	KEF-KY hastaları; AF ritminde olan KEF-KY hastaları ve sinüs ritminde (SR) olan KEF-KY hastaları olarak iki gruba ayrılarak, bu hastaların klinik karakteristik özellikleri karşılaştırıldı.	24
80	Zaman içinde kompleks ilaç etkileşimlerinin yönetim zorluklarına vurgu yapan bu rapor 2006 yılında kalp nakli yapılan HIV seropozitif bir hastanın 11 yıllık takip raporudur.	24
81	Atriyal fibrilasyonu (AF) bulunan korunmuş ejeksiyon fraksiyonlu kalp yetersizliği (KEF-KY) hastalarının klinik özelliklerini değerlendirmeyi ve bu hastaların klinik özelliklerini AF'si bulunmayan KEF-KY hastalarıyla karşılaştırmayı amaçladık.	25
82	Hastada dakikalar içersinde hipotansiyon, sağ ventrikül ve inferiyor vena kavada dilatasyon, sol atriyum ve sol ventrikülde kollaps gelişti ve SCDC hızı azaltılmasına rağmen devam etti.	25
83	Kronik obstrüktif akciğer hastalığı (KOAH) ile atriyal fibrilasyonu (AF) ilişkidir ve azalmış birinci saniye zorlu ekspiratuvar volümü (ZEV1) yeni başlangıçlı AF için bağımsız bir etkendir.	25
84	SAA trombüs (+) grubu, kalp yetmezliği, periferik arter hastalığı, koroner arter hastalığı ve kronik obstrüktif akciğer hastalığı oranlarında anlamlı olarak daha yüksek oranlara sahipti ($p<0.05$).	25

85	Alınan örneğin ortalama lökosit sayımı (127.5±86.0 ve 227.2±120.7; p=0.026) ve belirgin yangısı olan hastaların sıklığı (%35 ve %75; p=0.046) SVYB gözlenen grupta anlamlı olarak daha yüksekti.	26
86	Postural ortostatik taşikardi sendromu (POTS) baş dönmesi, çarpıntı, presenkop, senkop ve özellikle bacaklarda güçsüzlük veya ağırlık gibi heterojen semptomlarla karakterize edilen, kronik ve güçsüzleştirici bir durumdur.	26
87	Semptomsuz olgularda ve PKT'li bazı semptomlu hastalarda, birinci basamak görüntüleme aracı olan transtorasik ekokardiyografi, uygun tedavi seçeneğini seçmek için zorunlu olan trombüs gösterimi açısından yetersiz kalabilir.	26
88	Transkateter aort kapak değişimi ve mitral kapak kapakçığı prosedürünü takiben derin bradikardi gelişen 51 yaşında bir kadın hastada, başarılı bir kablosuz kalp pili uygulaması olgusu sunuldu.	26
89	Bu çalışmada, bilinen hastalığı olmayan sağlıklı bireylerde obezitenin sağ kalp fonksiyonlarını anlamlı derecede etkilediği ve obezite derecesi ile sağ kalp fonksiyonlarındaki bozulma arasında anlamlı korelasyon olduğu gösterilmiştir.	27
90	Koroner arter ektazisi (KAE), koroner arterin lümeninin lokalize veya yaygın olarak dilatasyonu olup, sağlıklı komşu damar segment çapınına göre en az 1.5 kat dilate olması olarak tanımlanmıştır.	27
91	Sıvı ve elektrolit replasman tedavisinden sonra elektrokardiyografide sırasıyla atriyal fibrilasyon ile birlikte dar QRS ve sonra düzeltilmiş QT değeri 490 msn ile birlikte normal sinüs ritmi görüldü.	27
92	Toplam 1865 ardışık hasta değerlendirildi ve hastaların %32'si yetersiz kayıt süresi veya kalitesi, atriyal fibrilasyon atakları, atriyoventriküler blok ve >%1 atriyal veya ventriküler ekstra sistoller nedeniyle dışlandı.	27
93	Amacımız hastada kötü perfüzyonu gösteren parametreleri içeren ve daha önce tanımlanmış olan bir skorlama ile kardiyak cerrahi geçiren kritik çocuk hasta bakımında ortak bir dil kullanımına dikkat çekmekti.	28

94	Bu olgu serisinde EKG'de normal QT ve düzeltilmiş QT aralıkları olan ve ailesinde ani kalp ölümü öyküsü olan üç bireyde gizli UQTS tanısı için epinefrin provokasyon testi kullanıldı.	28
95	Koroner yavaş akım (KYA) patofizyolojisi tam manasıyla açıklığa kavuşturulamamıştır, dahası artmış fruktoz tüketiminin, KYA patofizyolojisinde rol oynayabilecek oksidatif stres ve inflamasyon ile ilişkili olduğunu gösteren birçok çalışma vardır.	28
96	Sonuç olarak, yüksek Lpa seviyeleri ve hemen hemen aynı erken koroner arter tutulumu olan monozigotik HoAH ikizler, ömür boyu kolesterol yükü / maruziyeti hipotezine doğrulayacı bir kanıt sayılabilir.	28
97	Bu çalışmada, ST yükselmeli miyokart enfarktüsü (STYME) sonrası emme yöntemiyle alınan pıhtının yaşı ve yangı düzeyinin (YD), taburculuk öncesi ve uzun dönem sol ventrikül (SV) fonksiyonlarıyla olası ilişkisi araştırıldı.	29
98	Bu yazıda AH ve yüksek lipoprotein a (Lpa) düzeylerine bağlı olarak hemen hemen aynı erken koroner arter tutulumu olan 36 yaşında monozigotik ikiz kardeşleri literatür derlemesi ile birlikte sunuyoruz.	29
99	Hem P dalga dispersiyonu hem atriyal elektromekanik gecikme parametreleri hastalığın erken dönemlerinden başlamak üzere yapısal ve fonksiyonel kardiyak hastalığı olmayan KOAH hastalarında AF'ye artmış eğilimin göstergesi olarak anlamlı uzundu.	29
100	Semptomatik hastalarda koroner girişimler daha çok KAS sonrası (%38.9) yapılırken, asemptomatik grupta ise KAS öncesi (%25.9) ve sonrasında (%25.9) benzer oranlarda yapıldığı görüldü ama gruplar arasında istatistiksel fark yoktu.	29

Human Reference Translation

Human translated reference translations are provided below. They are extracted from the test corpus.

ID	English Reference Translation	WC
1	No rhabdomyolysis or myopathy was seen.	6
2	Pseudoaneurysm after carotid stenting:	4
3	Infrequent origin of a peripheral embolism:	6
4	Multiple Saphenous Vein Graft Aneurysm	5
5	A physical evaluation revealed hypotension and tachycardia.	7
6	Clinical and angiographic characteristics of the patients were assessed.	9
7	Consultant invasive neuroradiologist recommended medical follow-up.	6
8	Efficacy and safety of carotid artery stenting:	7
9	Arterial thoracic outlet syndrome in a young woman	8
10	Just before the stent implantation, asystole developed, followed by convulsions.	10
11	Analysis of thrombophilic gene mutations in coronary artery ectasia	9
12	Post-infarction aneurysm of left ventricle perforating the right ventricle	9
13	The LCOSs was used to evaluate the deve-lopment of low cardiac output.	12
14	Spontaneous aortic root rupture during pregnancy	6
15	Sodium glucose co-transporter 2 inhibitors in heart failure therapy	9
16	Prognostic nutritional index predicts mortality in infective endocarditis	8
17	In real life, long-term fibrate use is effective and safe.	10
18	Twenty-seven-year durability of St. Jude Medical Biocor bovine bioprosthesis in mitral position	12
19	New indicator of cellular ischemia in coronary slow-flow phenomenon:	9
20	Coronary stent embolism to the right posterior cerebral artery	9
21	Acute heart failure (HF) is one of the most common cardiac emergencies.	12
22	Radiofrequency ablation of this area can cause rare but serious complications.	11
23	Investigation of metabolic effects of CETP gene rs289714 variation in coronary artery patients:	13

24	Premature ventricular complexes (PVCs) are one of the most common arrhythmias seen in daily practice.	15
25	The effect of transcatheter atrial septal defect closure on left heart function in pediatric patients	15
26	A carotid artery PA can be caused by trauma, spontaneous infection, vasculitis, or it may be iatrogenic.	17
27	Fluoroscopic scanning of the vascular system showed that the coronary stent was in the right posterior cerebral artery.	18
28	Ventricular arrhythmias arising from coronary cusps are not uncommon.	9
29	Diverse echocardiographic changes in the course of hypoxia due to acute exacerbation of idiopathic pulmonary fibrosis	16
30	Assessment of atrial electromechanical delay and P wave dispersion in patients with chronic obstructive pulmonary disease	16
31	Prosthetic valve thrombosis (PVT) is a serious complication among patients with prosthetic heart valves.	14
32	Serum lactate level may predict the development of acute kidney injury in acute decompensated heart failure	16
33	Patients with stable angina pectoris who underwent coronary angiography between December 2018 and April 2019 were evaluated for this study.	20
34	Global burden of chronic kidney disease and decreased kidney function in Turkish heart failure patients	15
35	Major vascular complications and cardiac tamponade were significantly more frequent in the lower PNI group.	15
36	Two months after the primary PCI, a successful coronary artery bypass graft operation was performed.	15
37	Effect of the COVID-19 pandemic on access to primary percutaneous coronary intervention for ST-segment elevation myocardial infarction	17
38	Complementary role of different imaging modalities in the diagnosis of concurrent obstructive aortic and mitral prosthetic valve thrombosis	18

39	Positive tendency toward synchronous use of acetaminophen and ibuprofen in treating patients with patent ductus arteriosus	16
40	Improvement of endothelial function early after thrombolytic therapy in patients with prosthetic heart valve thrombosis	15
41	Comparison of clinical characteristics of patients with heart failure and preserved ejection fraction with atrial fibrillation versus sinus rhythm:	19
42	Monozygotic twins with familial hypercholesterolemia and high lipoprotein(a) levels leading to identical cardiovascular outcomes:	14
43	Low cardiac output syndrome score to evaluate postoperative cardiac surgery patients in a pediatric intensive care unit	17
44	Relationship between histopathological features of aspirated thrombi and long-term left ventricular function in patients with ST-segment elevation myocardial infarction	19
45	This case report describes the different echocardiographic changes observed in 2 episodes of hypoxic attack in a 73-year-old man.	19
46	A 58-year-old female had been hospitalized in another hospital 2 weeks prior to the currently described presentation due to atypical pneumonia.	21
47	Scintigraphy using bone isotopes is considered a highly reliable and easy-to-use method in the diagnosis of TTR-CA.	17
48	Improvement in the LV diastolic and left atrial systolic functions was observed in children who underwent transcatheter closure of an ASD.	21
49	Brugada syndrome (BrS) is a congenital channelopathy associated with the development of malignant ventricular arrhythmias and sudden cardiac death.	19
50	Electrocardiography (ECG) showed a fast idioventricular rhythm with capture and fusion beats and evident J waves in leads DII, DIII, and aVF.	22
51	Carotid artery stenting has been a widely used interventional treatment method for the last 3 decades in the treatment of carotid artery stenosis.	23
52	Comparison of myocardial performance index and right ventricular myocardial acceleration during isovolumic contraction in detection of right ventricular dysfunction in obese patients	22

53	This case report describes a diabetic patient admitted with dyspnea one week after she suffered an acute myocardial infarction.	19
54	This could be an important aspect of FH families as a result of the similar timing of cumulative LDL exposure exceeding the threshold of CVD events.	26
55	There is no clear consensus regarding the definition of low cardiac output syndrome (LCOS) or the follow-up of this patient group.	21
56	Cardiovascular disease (CVD) is the leading cause of morbidity and mortality throughout the world and contributes a considerable burden to healthcare costs.	22
57	A 58-year-old man with a left ventricular assist device (LVAD), which had been implanted 1 year earlier, presented with rest dyspnea.	21
58	Patients without AF had a higher burden of diabetes mellitus, obstructive sleep apnea, and coronary artery disease.	17
59	It is frequently associated with hypermobile joints or conditions such as chronic fatigue syndrome, chronic abdominal pain, migraine headache, and diabetes mellitus.	22
60	Highly active antiretroviral therapy has led to greater life expectancy for human immun-deficiency virus (HIV)-positive patients.	16
61	This was a report of a 17-year-old male patient with very frequent, nonsustained ventricular tachycardia attacks and premature ventricular contraction-induced cardiomyopathy.	21
62	Although PVCs are generally considered benign, they can lead to the development of cardiomyopathy and, rarely, can be associated with sudden cardiac death.	23
63	Leadless pacemakers provide a potential alternative to conventional transvenous pacemakers for patients undergoing high-risk transcatheter valve replacement procedures.	18
64	Inappropriate sinus tachycardia (IST) is a syndrome characterized by an elevated resting heart rate with distressing symptoms and no secondary cause of sinus tachycardia.	24

65	In our case, a 41-year-old male patient developed palpitation, confusion, loss of consciousness, and cardiac arrest while at home.	19
66	Neither the serum levels of the study markers nor demographic and clinical parameters were predictive of an LAA thrombus in binary logistic regression analysis.	24
67	Coronary slow-flow phenomenon (CSFP) is defined as the delayed arrival of coronary blood flow to the distal vascular bed in at least 1 major epicardial coronary artery.	27
68	An L/O thrombus was related to better pre-discharge and long-term LV volumes and ejection fraction values compared with a fresh thrombus.	21
69	LGE and/or edema as a sole criterion for the diagnosis of acute myocarditis demonstrated better diagnostic accuracy than the Lake Louise criteria.	22
70	However, the mapping and ablation of outflow tract ventricular arrhythmias originating from aortic cusps can be challenging.	17
71	Although obesity is a risk factor for heart failure, studies analyzing the effect of obesity on heart functions have primarily examined the left side of the heart.	27
72	The prognostic nutritional index (PNI), based on serum albumin and lymphocyte concentration, is an inflammation-based nutritional score that has been shown to be a prognostic determinant in several populations.	29
73	This is an investigation of access to primary percutaneous coronary intervention (PPCI) for patients diagnosed with ST-segment myocardial infarction (STEMI) during the pandemic.	23
74	The aim of this study was to investigate the effects of the CETP gene rs289714 polymorphism on the serum lipid profile and other metabolic parameters in Turkish patients with coronary artery disease (CAD).	33
75	Since monozygotic twins share the intrauterine environment and have the same age and gene profile, they could represent a very special resource for the investigation of the causes and the natural course of FH.	34

76	A significant increase in the leukocyte count in the aspirate and a fresh thrombus might predict long-term LV functional deterioration irrespective of the clinical and procedure-related characteristics.	27
77	However, recent reports have suggested that 13% of heart failure patients with a preserved ejection fraction and 16% of advanced-age patients with severe aortic stenosis have TTR-CA.	27
78	This study was designed to evaluate the real-life efficacy and side effects of fibrate treatment for hypertriglyceridemia seen in a lipid clinic, as well as cardiovascular and diabetic outcomes.	29
79	HFpEF with AF and HFpEF with sinus rhythm (SR), and the clinical characteristics of the groups were compared.	18
80	This was a report of 11 years of follow-up of an HIV-seropositive patient who underwent heart transplantation in 2006, with emphasis on the management challenges of complex drug interactions over time.	31
81	The aim of this study was to assess the clinical characteristics of patients with heart failure and preserved ejection fraction (HFpEF) and atrial fibrillation (AF) and compare them with those of HFpEF patients without AF.	35
82	Within minutes, hypotension, RV and inferior vena cava dilatation, and left atrial (LA) and left ventricular (LV) collapse occurred and persisted despite LVAD speed reduction.	25
83	Chronic obstructive pulmonary disease (COPD) is associated with atrial fibrillation (AF) and reduced forced expiratory volume (FEV1) is an independent predictor for new onset AF.	25
84	The LAA thrombus (+) group had significantly higher rates of heart failure, peripheral artery disease, coronary artery disease, and chronic obstructive pulmonary disease ($p<0.05$).	24
85	The mean leukocyte count of the aspirate (127.5±86.0 vs 227.2±120.7; $p=0.026$) and frequency of significant inflammation (35% vs 75%; $p=0.046$) were significantly higher in the group with LVR.	28

86	Postural orthostatic tachycardia syndrome (POTS) is a chronic, debilitating condition characterized by heterogeneous symptoms, such as lightheadedness, palpitations, pre-syncope, syncope, and weakness or heaviness, especially of the legs.	28
87	In asymptomatic cases, as well as certain symptomatic patients with PVT, the results of the first-line imaging tool, transthoracic echocardiography, may be inconclusive in terms of illustrating the thrombus, which is necessary in order to select the proper treatment option.	40
88	This is a description of a successful leadless pacemaker implantation in a 51-year-old woman who developed profound bradycardia following a transcatheter aortic valve replacement and mitral valve-in-valve procedure.	28
89	Obesity significantly affected right heart function and there was a significant correlation between the degree of obesity and right heart functional deterioration.	22
90	Coronary artery ectasia (CAE) is defined as localized or diffuse dilatation in the coronary artery lumen of at least 1.5 times the diameter of adjacent healthy reference segments.	28
91	After fluids and electrolyte replacement therapy were administered, the ECG results revealed narrow QRS complex atrial fibrillation followed by a normal sinus rhythm with a 490 ms corrected QT interval.	30
92	A total of 1865 consecutive patients were evaluated and 32% were excluded due to an inadequate Holter recording period or insufficient quality, atrial fibrillation episodes, atrioventricular block, or >1% atrial or ventricular extrasystoles.	33
93	The objective of this study was to draw attention to the potential use of a common language in the care of critical pediatric patients undergoing cardiac surgery with a previously defined scoring method that includes parameters indicating poor perfusion in the patient.	42
94	This case report describes the use of an epinephrine provocation test to diagnose hidden LQTS in 3 patients who had normal QT interval and	37

	corrected QT interval on ECG and a family history of sudden cardiac death.	
95	The precise pathophysiology of slow coronary flow (SCF) has not yet been clarified; however, many studies have shown that significant fructose consumption is associated with oxidative stress and inflammation, which may play a role in the pathophysiology of SCF.	39
96	In conclusion, this first report of monozygotic HoFH twins with elevated Lpa levels and almost identical early coronary artery involvement at the same age provides evidence to substantiate the hypothesis of lifetime cholesterol burden/exposure.	34
97	This study was an investigation of the severity of inflammation (SOI) in aspirated material and thrombus age to examine any association with pre-discharge and long-term left ventricular (LV) function after ST-elevation myocardial infarction (STEMI).	34
98	This report is a description of 36-year-old monozygotic twin brothers with almost identical early coronary artery involvement due to FH concomitant with high lipoprotein(a) (Lpa) levels and a review of the literature.	32
99	Both P wave dispersion and AEMD parameters were significantly longer in COPD patients without any established structural or functional cardiac abnormalities, indicating an increased tendency for AF development, beginning from the initial stages of the disease.	36
100	A percutaneous coronary intervention was performed after CAS more often in symptomatic patients (38.9%), while it was observed at the same rate both before (25.9%) and after (25.9%) CAS in the asymptomatic group, but the difference between the groups was not statistically significant.	43

Translation by the PBMT-1 Engine

100 segments translated by PBMT-1 engine which is an SMT engine trained on only TRENCARD cardiology corpus.

ID	PBMT-1
1	No rhabdomyolysis or myopathy was observed.
2	Carotid artery stenting after pseudoaneurysm:
3	Peripheral source is less failure:
4	Multiple Saphenous Vein Graft Aneurysm
5	On physical examination, hypotension and tachycardia.
6	The clinical and angiographic characteristics of the patients were evaluated.
7	If the invasive nöroradyoloji önerdi medical follow-up.
8	Carotid artery stentlemesinin efficacy and device:
9	Young female patient Arterial thoracic outlet syndrome
10	parameters before and asystole konvülziyon.
11	Coronary artery ektazisinde trombofilik gene mutation analysis
12	The right ventricle post-infarct left ventricular aneurysm
13	Low cardiac output in order to detect DKDSs were used.
14	During pregnancy in a patient with spontaneous aortic root rupture
15	The treatment of heart failure, Sodium glucose co-transporter 2 inhibitors
16	Prognostic nutrisyonel index öngördürür mortality in patients with infective endocarditis
17	True embrional remnant long-term use of fibrates be effective and safe.
18	Transesophageal echocardiography; St. Jude Medical mitral bovine bioprosthetic valve 27-year dayanıklılığı
19	Coronary slow flow fenomeninde for the cellular ischemia new gösterge:
20	The right posterior cerebral artery embolized coronary stent.
21	Acute heart failure is the most common cardiac emergency conditions.
22	In this field radiofrequency ablation is a rare but serious complications orifice.
23	Coronary artery disease CETP gene rs289714 varyasyonunun metabolic heart:
24	Ventricular premature (VPB), is the most common the treatment.
25	In children transcatheter atrial septal defect closure on left ventricular function

26	Carotid artery PA had, trauma, spontaneous infection, vasculitis or iatrogenic young parenchyma.
27	Fluoroscopic stent vascular system in the right posterior cerebral artery.
28	Premature ventricular contractions in a successful associated with adölesanda left coronary kusp ablation
29	Idiopathic pulmonary fibrozun acute alevlenmesine as a result of prilocaine-associated role of various echocardiographic changes
30	Chronic obstructive pulmonary disease delays of atrial electromechanical coupling and P wave dispersiyonun.
31	Prosthetic valve thrombosis (n=13) development in patients with prosthetic heart valves is a serious complication.
32	Serum lactate level, acute decompensated heart failure prevent acute kidney injury öngördürebilir
33	December ≥65 years) between April 2019 who underwent coronary angiography with stable angina pectoris of the patients were evaluated.
34	Chronic renal disease and Global burden and the Turkish heart failure patients with reduced renal functions
35	The major vascular complications and cardiac tamponadın low PNİ higher in, respectively.
36	Primary percutaneous coronary intervention (PCI) with 2 months after the patient underwent successful coronary artery bypass surgery.
37	COVID-19 pandemisinin ST-segment elevation myocardial infarction for primary percutaneous coronary intervention was performed on the timing of
38	Concomitant obstructive aortic and mitral prosthetic valve interactions diagnosis of Complementary in different imaging techniques
39	Patent ductus arteriyozuslu patients simultaneous asetaminofen and ibuprofen for the use of increased trend
40	Prosthetic heart valve thrombosis, thrombolytic therapy in patients with just after the endothelial function improved

41	Sinus rhythm and atrial fibrillation with heart failure with preserved ejection fraction in patients with the clinical rabbits:
42	The same cardiovascular sonlanıma with familial hypercholesterolemia and higher lipoprotein (a) levels with single Egg ikizleri:
43	in pediatric cardiac intensive care unit was used for the evaluation of the low cardiac output syndrome score
44	ST-elevation myocardial infarction revealed pıhtının Histopathological assessment of long-term Left Ventricular fonksiyonuyla relationship
45	In this case report, a 73-year-old men and 2 gastric attack epizodunda in different Echocardiographic changes.
46	A 58-year-old female patient with a two-week ago in another hospital because of atypical pneumonia hospital presentation.
47	Bone agents with sintigrafi, TTR-diagnosis of CA is a safe and easy method is a gösterilmektedir.
48	Transcatheter ASD closure with left atrial systolic and diastolic functions of left ventricle were positive the years.
49	Brugada syndrome (BrS) is a congenital kanalopati development of malignant ventricular arrhythmias and sudden cardiac death, associated with.
50	Electrocardiography showed rapid thalassemia intermedia, yakalama and However, the first vurularıyla elevations, DIII, aVF leads improved J wave) was noted.
51	Carotid artery stenting, carotid artery different for the past 30 years was used as an interventional treatment modality.
52	In obese patients right ventricular dysfunction and for myocardial performance index isovolumetrik contraction cyclus Comparison of right ventricular acceleration
53	In this article, we report a geçirmesinden acute myocardial infarction after one week dyspnea was admitted to the emergency service with diabetic one patient.
54	This, cardiovascular events eşiğini exceding cumulative LDL maruziyetinin similar to ailelerinin timing of FH is an important.

55	The low cardiac output syndrome (DKDS) inheritance and this patient group with a follow-up by no consensus.
56	Cardiovascular diseases (CVD) is the morbidity and mortality is one of the main sebeplerinde and high cost on the health care system.
57	A year ago, left ventricular assist device (SVDC) implantation with a history of a 58-year-old male patient was admitted with resting dyspnea.
58	However, the patients were higher than without diabetes mellitus, obstructive sleep apnea and the use of coronary artery disease.
59	Frequently hipermobil eklemler or chronic fatigue syndrome, chronic abdominal pain, connections, headache, and şeker disease such as the upper extremity vessels.
60	Extremely active antiretroviral therapy (Haart), immune failure and human immunodeficiency virus (HIV) positive for longer life beklentisini kılmıştır possible.
61	In this article, we present a very common in nonsustained ventricular tachycardia attacks and ventricular premature contraction cardiomyopathy associated with 17-year-old male patient.
62	As usually benign, but may be due to the development of cardiomyopathy and it is associated with sudden cardiac death.
63	Leadless arteriosis, high-risk patients who received transcatheter valve procedures for conventional transvenous cardiac pillerine a potential alternative sunar.
64	Inappropriate sinus tachycardia (IST) without any secondary increased Resting heart rate and theraphy:a syndrome characterized by symptoms.
65	Our catheter, 41-year-old male patient presented with complaints of palpitations at home, instability and consciousness, syncope. Loss of consciousness, cardiac arrest occurred.
66	study aims to evaluate the hemostatic variables with serum levels and demographic and clinical parameters, binary logistic regression analysis showed that LAA trombüsünü for the prediction of was not significant.

67	Coronary slow flow phenomenon (CSFP) is at least one major epicardial coronary artery blood flow distal vessel yatağına defined as late reach older.
68	e. G/L and structure, fresh than the presence of pre-discharge and long-term LV volumes and ejection fraction values.
69	The diagnosis of acute myocarditis GKT and/or edema diagnostic criterion for the presence of to be Lake Louise criteria of higher diagnostic accuracy.
70	Coronary yaprakçıklardan of ventricular arrhythmias is a rare condition. However, aortic yaprakçıklardan originating from the outflow tract ventricular arrhythmias uğraştırıcı mapping and ablation.
71	Obesity heart failure is a risk identify artýþ, but only a handful of most of obesity on left ventricular function üzerine.
72	Serum albumin and lymphocyte konsantrasyonlarını base nutrisyonel the prognostic index (PNİ) in various patient with popülasyonlarında is a new Inflammation-based risk skorudur.
73	This study aims to investigate the COVID-19ST -segment elevation myocardial infarction (Stemi) patients who underwent primary percutaneous coronary intervention (PPKG) yapılana to the exact pulse numbers.
74	The aim of this study is to coronary artery disease (CAD) in patients with CETP rs289714 gene polymorphism, serum lipid profile and other metabolic effects on using.
75	Monozigotik defects delivery çevreyi, age and all genlerini common paylaştıklarından, FH is causes, and to determine the natural course of specific in a source edebilirler.
76	Samples of Increased leucocyte count and fresh the presence of thrombus, clinical and procedure-related özelliklerden LV function independently associated with deterioration in the long-term.
77	However, the new data heart failure with preserved ejection fraction of %13 and high risk patients with severe aortic stenosis in elderly patients with %16 patients (TTR-CA or vice versa.

78	In this study, a lipid concluded that fibrates are underused in the treatment of hypertriglyceridemia the effectiveness, reliability and cardiovascular, diabetic outcome points effect of real-life data of koymayı clinic.
79	Not Rule Out the hastaları; With AF rhythm not rule out the patients, and sinus rhythm (SR) (not rule out the patients were divided into two groups, clinical characteristics of the patients were compared.
80	Time in complex drug interactions management zorluklarına systemic this report the heart transplantation in 2006 in the HIV seropozitif the raporudur 11-year follow-up.
81	Atrial fibrillation (AF) in patients with preserved ejection fraction is not rule out heart failure (HF) patients by clinical characteristics and clinical characteristics of the patients without AF were not rule out the hastalarıyla the procedure.
82	100 patients dakikalar hypotension, right ventricular dilatation, and inferior vena cava, left atrial diameter and left ventricular collapse developed and SCDC rate unity, continued.
83	Chronic obstructive pulmonary disease (COPD) patients with atrial fibrillation (AF) and decreased ilişkidir forced expiratory volume (ekspiratuvar ZEV1) The new-onset AF an independent risk factor for disease.
84	LAA thrombus (+) group, heart failure, peripheral artery disease, coronary artery disease and chronic obstructive pulmonary disease rates were significantly higher in had oranlara ($p < 0.05$).
85	The mean leukocyte count (e.g. 127.5±86.0-and-227.2±120.7; p=0.026) and the prevalence of yangısı patients (%35 and %75; p=0.046). SVYB in were significantly higher.
86	Postural orthostatic tachycardia syndrome (POTS), dizziness, inciurling palpitations, syncope, especially in legs and weakness or weight characterized by heterogeneous symptoms such as, chronic and güçsüzleştirici entity.
87	Asymptomatic and subsequent in some symptomatic patients, the first-line tool in the imaging by transthoracic echocardiography, appropriate treatment for offering seçmek compulsorily thrombus gösterimi with respect to insufficient.

88	Transcatheter aortic valve replacement and mitral valve, followed by other cases have undergone prosedürünü deep bradycardia of a 51-year-old female patient, a successful Leadless pacemaker implantation.
89	In this study, we aimed to investigate the disease-known that obesity healthy individuals right heart function significantly affected and the degree of obesity right heart dysfunction significant correlation between has been shown.
90	Coronary artery ectasia (CAE), coronary artery arising as a localized or diffuse lümeninin dilatation, normal vessel segment adjacent to çapınına of at least 1.5 times as in dilated cardiomyopathy.
91	Fluid and electrolyte replacement after were electrocardiogram showed atrial fibrillation with a narrow QRS complex and after the corrected QT değeri490 msec with normal sinus rhythm.
92	A total of 1865 consecutive patients were evaluated and %32 of registry or inadequate duration quality, atrial fibrillation, atrioventricular block and > %1 atrial or ventricular interruption sistoller to stable.
93	Aim of patients with poor perfusion parameters, and previous defined a scoring with critical children undergoing cardiac surgery, patient care is a common language use of çekmekti.
94	This case series described ECG normal QT and corrected QT intervals, and family history of sudden cardiac death in patients with three k were diagnosed as LQTS and concealed for epinefrin provokasyon test was used.
95	Slow coronary flow (SCF) pathophysiology of recurrent trans manasıyla kavuşturulamamıştır, Morover, increased fruktoz consumption, SCF pathophysiology cardioprotective role of oxidative stress and inflammation, is associated with several studies.
96	In conclusion, high Lpa levels and nearly the same early coronary artery involvement monozigotik twins with HoFH life-cholesterol burden/relevant hipotezine doğrulayacı a evidence.
97	In this study, we aimed to investigate the ST-segment elevation myocardial infarction (Stemi) is not suction pıhtının by after the age and inflammation level

	(PBV), before discharge and long-term Left ventricular (LV) fonksiyonlarıyla possible relationship between.
98	In this article, a FH and higher lipoprotein (Lpa) levels in nearly the same early coronary artery involvement 36-year-old monozigotik twin kardeşleri with the review of the literature.
99	And P wave dispersion and parameters of atrial electromechanical delay and early dönemlerinden başlamak structural and functional cardiac disease without COPD patients with AF as a marker of increased eğilimin significantly longer.
100	In patients with symptomatic coronary interventions are more muscle (%38.9) afterDuring, asymptomatic group (%25.9) before muscleand after (%25.9) An but similar cannulation was no statistical difference between the groups.

Translation by the PBMT-2 Engine

100 segments translated by PBMT-2 engine which is an SMT engine trained on TRENCARD + GENCOR corpora.

ID	PBMT-2
1	No rhabdomyolysis or myopathy was observed.
2	Carotid artery stenting after pseudoaneurysm:
3	Peripheral is treated less visible
4	Multiple Saphenous Vein Graft Aneurysm
5	On physical examination, hypotension and tachycardia.
6	The clinical and angiographic characteristics of the patients were evaluated.
7	invasive nöroradyoloji konsültasyonu proposed medical follow-up.
8	The effectiveness of carotid artery stentlemesinin and device:
9	The young woman Arterial thoracic outlet syndrome
10	parameters before and asystole konvülziyon.
11	Coronary artery ektazisinde trombofilik gene mutation analysis

12	The right ventricle post-infarct left ventricular aneurysm
13	Low cardiac output in order to detect DKDSs.
14	During pregnancy in a patient with spontaneous aortic root rupture
15	The treatment of heart failure, Sodium glucose co-transporter 2 inhibitors
16	Prognostic nutritional index öngördürür mortality in patients with infective endocarditis
17	In real life in the long-term use of fibrates effective and safe.
18	St. Jude Medical cattle bioprosthetic mitral valve 27 years of tolerance
19	In slow coronary flow cellular ischemia new indicators:
20	The right posterior cerebral artery case of embolized coronary stent
21	Acute heart failure is the most common cardiac emergency conditions.
22	In this area radiofrequency ablation rare but can cause severe complications.
23	In patients with coronary artery disease which lowers CETP gene rs289714 varyasyonunun metabolic heart:
24	Ventricular premature (VPB) is the most common clinical aritmilerden of.
25	children transcatheter atrial septal defect closure in effect on left ventricular function
26	Carotid artery PA of, trauma, spontaneous infection, vasculitis or iatrogenic causes kaynaklanabilir.
27	Fluoroscopic vascular system showed stent in the right posterior cerebral arteries, respectively.
28	Associated with premature ventricular contractions in a adölesanda successful ablation left coronary kusp
29	Idiopathic pulmonary fibrozun Acute exacerbation impairs a role of various echocardiographic changes in
30	Chronic obstructive pulmonary disease atrial electromechanical delay and P-wave dispersiyonun evaluation
31	Prosthetic valve thrombosis (n=13) in patients with prosthetic heart of a serious complication.

32	Serum lactate levels in acute decompensated heart failure öngördürebilir development of acute renal injury
33	December 2018 between April 2019 who underwent coronary angiography in patients with stable angina pectoris.
34	Global chronic kidney disease burden and the Turkish heart failure patients with reduced renal function
35	The major vascular complications and cardiac tamponadın low PNİ significantly higher in, respectively.
36	Primary percutaneous coronary intervention (PCI) 2 months later, the patient underwent successful coronary artery bypass surgery.
37	COVID-19 pandemisinin ST-segment elevation myocardial infarction for primary percutaneous coronary intervention in or zoom effect
38	Concomitant obstructive aortic and mitral prosthetic valve trombozları diagnosis of complementary role in different imaging techniques
39	Patent ductus arteriyozuslu the treatment of patients with simultaneous acetaminophen and ibuprofen use increased trend
40	Prosthetic heart valve thrombosis, thrombolytic therapy in patients with immediately after recovery of endothelial function
41	In sinus rhythm and atrial fibrillation, heart failure with preserved ejection fraction clinical characteristics of patients surgery:
42	The same cardiovascular sonlanıma with familial hypercholesterolemia and higher lipoprotein (a) levels with one egg ikizleri:
43	pediatric cardiac intensive care unit after surgery patients were used for the evaluation of the low cardiac output syndrome score
44	In patients with ST-elevation myocardial infarction revealed pıhtının histopathological of the long-term relationship with the left ventricular Script
45	In this case report, a 73-year-old male 2 hypoxic epizodunda attack of various echocardiographic changes.
46	A 58-year-old female patient with two weeks ago in another hospital because of atypical pneumonia, CIMT the hospital.

47	Bone agents with sintigrafi, TTR-CA diagnosis is a safe and easy method for the application of Vogue.
48	Transcatheter ASD closure with left atrial systolic and diastolic functions of left ventricle were positively being viewed.
49	Brugada syndrome (BrS) is a congenital kanalopati of malignant ventricular arrhythmias and sudden cardiac death, associated with.
50	The electrocardiogram showed rapid thalassemia intermedia, capture and fusion with vurularıyla elevations, DIII, aVF leads wave significant J.
51	Carotid artery stenting, carotid artery stenozunun in the last 30 years commonly used as a noninvasive method of treatment.
52	In obese patients was right ventricular dysfunction of myocardial performance index and isovolumetrik contraction cyclus comparison of right ventricular acceleration
53	In this article, we report a geçirmesinden acute myocardial infarction after a few weeks of dyspnea was admitted to the emergency department with a diabetic patients in the literature.
54	This, cardiovascular events over threshold cumulative LDL maruziyetinin similar timing of FH is an important feature of their families.
55	Low cardiac output syndrome (DKDS) defining the aims and tracking in this patient group with a clear consensus is lacking.
56	Cardiovascular diseases (CVD) are all morbidity and mortality worldwide sebeplerinde of the health system and high cost and dispersal.
57	The year before, left ventricular assist device (SVDC) implantation with a history of a 58-year-old male patient was admitted with resting dyspnea.
58	However, AF current higher than patients without diabetes mellitus, obstructive sleep apnea and coronary artery disease in charge.
59	Often hipermobil joints or chronic fatigue syndrome, chronic abdominal pain, migraine, headache, and sugar disease, Upper extremity vessels.
60	Highly active antiretroviral therapy, human immune failure virus (HIV), the patients with positive for longer possible life as possible as well.

61	In this article, we present a very common and nonsustained ventricular tachycardia attacks of ventricular premature contraction cardiomyopathy associated with a 17-year-old male patient.
62	Often considered benign, although it may cause the development of cardiomyopathy and rarely associated with sudden cardiac death.
63	Leadless pacemaker, transcatheter aortic valve replacement in high-risk patients with procedures for conventional transvenous cardiac pillerine a potential alternative.
64	Inappropriate sinus tachycardia (IST) without any secondary cause of increased resting heart rate and disturbing a syndrome characterized by symptoms.
65	Our catheter, 41-year-old male patient with palpitations at home, consciousness Blur and syncope. The loss of consciousness after cardiac arrest occurred.
66	study of the hemostatic variables with serum levels and demographic and clinical parameters, binary logistic regression analysis showed that trombüsünü to determine the LAA was not significant.
67	Koroner yavaþ akým fenomeni (KYAF), at least one major epicardial coronary artery blood flow distal vessel to bed defined as late reach.
68	For example, the structure of L/O, fresh thrombus more frequently compared to the pre-discharge and long-term Left ventricular volumes and ejection fraction values associated.
69	The diagnosis of acute myocarditis GKT and/or edema of noninvazive diagnostic criteria of Lake Louise when According to a higher diagnostic accuracy were obtained.
70	Coronary yaprakçıklardan of ventricular arrhythmias is not a rare condition. However, aortic yaprakçıklardan originating from the outflow tract ventricular arrhythmias uğraştırıcı mapping and ablation.
71	Obesity heart failure to identify a risk that despite, a handful of studies of obesity on left ventricular function üzerine studied.
72	Serum albumin and lymphocyte konsantrasyonlarını base field prognostic nutritional index (PNİ) in various patient populations prognosis, a new Inflammation based risk dunks.

73	This study aims to investigate the COVID-19ST -segment elevation myocardial infarction (Stemi) patients with the diagnosis of primary percutaneous coronary intervention (PPKG), until the time of pulse numbers.
74	The aim of this study was to coronary artery disease (CAD) in patients with which lowers CETP rs289714 gene polymorphism, serum lipid profile and metabolic parameters impacts araştırılmasıdır.
75	Monozigotik twins intrauterine environment, age and all genes common paylaştıklarından, FH's and natural course to investigate the reasons for a special source can represent.
76	Example Increased leucocyte count and fresh the presence of thrombus, clinical, and procedure-related, independent of left ventricular function in the long-term associated with impairment.
77	However, recent data heart failure with preserved ejection fraction %13 and high-risk patients with severe aortic stenosis in elderly patients in %16 that TTR-CA.
78	In this study, a lipid varying in patients with hypertriglyceridemia, percutaneous treatment effectiveness, reliability and cardiovascular, diabetic outcome its effect on the basis of data of real life in order to Google.
79	KEF-HF hastaları; AF rhythm KEF-HF) patients, and sinus rhythm (SR) (KEF-HF patients are divided into two groups as the clinical characteristics of the patients were compared.
80	Over time the complex drug interactions zorluklarına this report "which has laid emphasis in 2006, heart transplantation HIV seropozitif an 11-year follow-up of the patient in the series.
81	Atrial fibrillation (AF) in heart failure with preserved ejection fraction (KEF-HF) patients to evaluate the clinical characteristics and clinical characteristics of the patients were in AF without KEF-HF clients the procedure.
82	Patients minutes içersinde hypotension, right ventricular dilatation, and inferior vena cava, left atrial diameter and left ventricular collapse developed and despite SCDC rate reduction.

83	Chronic obstructive pulmonary disease (COPD) patients with atrial fibrillation (AF) each other and decreased ekspiratuvar forced expiratory volume in the first second (ZEV1) The new-onset AF an independent predictor for the disease.
84	LAA thrombus (+) group, heart failure, peripheral artery disease, coronary artery disease and chronic obstructive pulmonary disease was found to be significantly higher had Ağırdır ($p < 0.05$).
85	For example, mean that the leukocyte count (127.5±86.0-and-227.2±120.7; p=0.026) and the prevalence of significant yangısı patients (%35 and %75; p=0.046) SVYB was observed to be significantly higher.
86	Postural orthostatic tachycardia syndrome (POTS), dizziness, palpitations, presyncope and syncope, and especially the upper extremities and weakness or weight, characterized by heterogeneous symptoms, which is a chronic and güçsüzleştirici.
87	Asymptomatic patients with (n=13) and some of the symptomatic patients, the first-line tool in the imaging by transthoracic echocardiography, appropriate treatment option to compulsory for thrombus may remain inadequate representation.
88	Transcatheter aortic valve replacement and mitral valve following kapakçığı prosedürünü deep bradycardia of a 51-year-old female patient with a successful application of leadless pacing case.
89	In this study, we aimed to investigate the disease is known that obesity in Healthy Individuals and right heart function significantly affected and obesity with deterioration of right heart dysfunction is a significant correlation between.
90	Coronary artery ectasia (CAE), coronary artery arising as a localized or diffuse lümeninin dilatation, normal vessel segment adjacent to çapınına of at least 1.5 times be defined as a dilated cardiomyopathy.
91	Fluid and electrolyte replacement after treatment, respectively, electrocardiogram showed atrial fibrillation with a narrow QRS complex and later corrected QT değeri490 msec with normal sinus rhythm.

92	Total 1865 consecutive patients were evaluated and %32 of record time or inadequate quality, atrial fibrillation, atrioventricular block, and > and in atrial or ventricular extra sistoller was excluded due to.
93	The aim of the patients with poor perfusion parameters and previously defined a scoring with critical children undergoing cardiac surgery, patient care to use a common language çekmekti.
94	This case series on ECG normal QT and corrected QT intervals and family history of sudden cardiac death of three K secret were diagnosed as LQTS and epinephrine for provocation test was used.
95	Slow coronary flow (SCF) pathophysiology full refer to recurrent kavuşturulamamıştır, further increased fructose consumption, play a role in the pathophysiology of SCF and inflammation, or oxidative stress associated with many of the study.
96	In conclusion, high Lpa levels and almost the same early coronary artery involvement with HoFH monozigotik twins, lifelong cholesterol load/plaques hipotezine doğrulayacı evidence, and others.
97	This study aims to evaluate patients with ST-elevation myocardial infarction (Stemi) after taken by sucking pıhtının age and inflammation levels (110–120 YD) and prior to discharge and long-term Left ventricular (LV) fonksiyonlarıyla investigated the possible relationship.
98	In this article, FH and higher lipoprotein (Lpa levels of almost the same early coronary artery involvement of 36-year-old monozigotik twin brothers who presented with the review of the literature.
99	And P wave dispersion and atrial electromechanical delay parameters obscure early stages of the disease, starting structural and functional cardiac disease without COPD patients with AF increased trend to be significantly longer.
100	Symptomatic coronary interventions in patients with more muscle (%38.9) afterwith asymptomatic group (muscle-%25.9) and (%25.9) of the similar, but was no statistical difference between the groups.

Translation by the NMT-1 Engine

100 segments translated by NMT-1 engine which is an NMT engine trained on only TRENCARD cardiology corpus

ID	NMT-1
1	No complications were observed.
2	A case of chest pain
3	Chronic vascular thrombosis
4	Saphenous Vein Graft; Saphenous Vein
5	Physical examination showed cardiomegaly and tachycardia.
6	Clinical and angiographic characteristics of the patients were evaluated.
7	Each of the articles were followed up.
8	Clinical and safety of carotid artery bypass grafting
9	A high-year-old female patient with a high coronary vein
10	One of the patients with complaints of symptoms and edema.
11	Coronary artery bypass graft; oxidative stress
12	Anomalous left ventricular aneurysm in the right atrium
13	Cardiac output was used for the patients with cardiac arrest.
14	Transcatheter aortic dissection in one patient
15	The effect of angiotensin-blocker therapy in patients with heart failure
16	Use of autonomic nervous system in patients with prosthetic valve endocarditis
17	The long-term treatment of long-term survival is discussed.
18	Three-year follow-up patient with mitral valve replacement
19	A new marker of myocardial ischemia in the slow flow
20	A single-year-old male patient with a right bundle branch block
21	Acute heart failure is one of the most common cardiac events.
22	In this case, ventricular septal defect is rare.
23	The aim of this study was to investigate the role of the frequency of serum Vit acid in patients with coronary artery disease

24	The most common pacemaker was seen in the most common pacemaker.
25	The effect of left atrial septal defect in atrial septal defect with atrial septal defect
26	Pulmonary artery stenosis in the carotid artery is a rare disease.
27	One of the femoral artery was detected in the right coronary artery.
28	Giant left coronary artery in a patient with left ventricular outflow tract
29	Pulmonary embolism in the diagnosis of pulmonary atresia with pulmonary embolism
30	Evaluation of atrial fibrillation in patients with chronic pulmonary hypertension with chronic pulmonary hypertension
31	Prosthetic valve thrombosis (AF) is a serious complication in patients with mechanical valve thrombosis.
32	Serum gamma-glutamyltransferase level, acute heart failure and acute kidney failure
33	Study Design The study included 30 patients with stable angina pectoris who underwent coronary angiography.
34	Chronic renal failure and mortality in patients with chronic renal disease
35	Postoperative vascular complications were significantly higher in group 1 in group I.
36	One patient underwent coronary bypass surgery after percutaneous coronary intervention.
37	The role of a new role in the primary percutaneous coronary intervention in Turkish Adults:
38	Evaluation of the evaluation of the use of aortic and mitral valve replacement in patients with aortic stenosis
39	The most commonly used in the treatment of the patients with sinus rhythm, and the most important role in the treatment of AF.
40	The effect of thrombolytic therapy in patients with heart failure in patients with heart failure.

41	Clinical characteristics of patients with congestive heart failure in patients with atrial fibrillation
42	A comparison of high-density lipoprotein (a) in patients with cardiovascular disease and high-density lipoprotein (a)
43	Evaluation of cardiac resynchronization therapy in patients with cardiac surgery after cardiac surgery
44	Evaluation of left ventricular function in patients with ST elevation myocardial infarction in patients with ST-segment elevation myocardial infarction
45	In this study, we present two patients with acute myocardial infarction due to acute myocardial infarction.
46	A 39-year-old female patient was admitted to the hospital who was admitted to our hospital with a hospital .
47	Magnetic resonance imaging is a safe and safe method to be a safe and safe method.
48	Left atrial systolic and left ventricular systolic functions and left ventricular diastolic functions were recorded.
49	Brugada syndrome (ICD) is associated with a congenital heart failure and sudden cardiac death.
50	The electrocardiogram showed been found to be normal in the presence of hypertension.
51	Arterial arterial flow is an alternative treatment for the treatment of carotid artery disease.
52	The aim of this study was to investigate the effect of myocardial infarction in patients with left ventricular myocardial infarction (AMI).
53	In this article, we present a patient with acute myocardial infarction who had acute myocardial infarction who had acute myocardial infarction.
54	It is concluded that the prevalence of coronary artery disease may be more important in the pathogenesis of cardiovascular disease.
55	Cardiac resynchronization syndrome (ICD) is the most common condition in the patient with a history of symptoms.

56	Cardiovascular disease (CVD) is the most important cause of morbidity and mortality.
57	A 53-year-old male patient was admitted with a complaint of dyspnea due to left ventricular assist device (AV) block.
58	However, patients with diabetes mellitus, diabetes mellitus and coronary artery disease and coronary artery disease.
59	It is often associated with the presence of acute heart failure and chronic heart failure, which is associated with the presence of acute kidney failure and diabetes mellitus.
60	It has been shown that the use of angiotensin converting enzyme (angiotensin) gene, and the quality of life.
61	In this article, we present a 53-year-old male patient with ventricular tachycardia and ventricular tachycardia associated with ventricular tachycardia.
62	However, it is an effective and effective treatment of sudden cardiac death.
63	In addition, it is an alternative to an alternative therapy for the treatment of prosthetic mitral valve replacement.
64	Atrioventricular nodal reentrant tachycardia (MI) is associated with an increase in heart rate and heart failure.
65	In this study, we present a 12-year-old male patient who was admitted to our hospital with a history of dyspnea and dyspnea. Cardiac arrest was observed in all patients.
66	In this study, there was no significant difference between serum uric acid and clinical characteristics, and clinical characteristics of the patients.
67	Coronary slow flow (FFR) is an important role in the coronary artery in the coronary artery.
68	In addition, there was no significant differences in the presence of MetS and left ventricular ejection fraction.
69	It was concluded that and/or/or/or/or/or/or/or/or/or/or/or/or/or/or the diagnosis.
70	Myocardial infarction is a rare condition. However, there is a result of the right ventricular outflow tract and ventricular tachycardia.

71	The aim of this study was to investigate the effect of cardiovascular risk factors in patients with heart failure.
72	Serum uric acid levels have been shown to be an important role in the neutrophil-to-lymphocyte level in patients with acute coronary syndrome.
73	In this study, we aimed to evaluate the results of the primary percutaneous coronary intervention (PCI) in patients with ST-segment elevation myocardial infarction (Stemi).
74	The aim of this study was to investigate the effects of the patients with coronary artery disease (CAD) in patients with coronary artery disease (CAD).
75	It is concluded that the prevalence of cardiovascular disease, and the prevalence of cardiovascular disease, the prevalence of cardiovascular disease, and the prevalence of cardiovascular disease is well as well.
76	It is concluded that there is no significant differences in the level of bleeding and morbidity and mortality.
77	In this study, the prevalence of ejection fraction in patients with severe heart failure was detected in 16 patients.
78	In this study, we aimed to investigate the efficacy of the use of the use of coronary artery disease in patients with essential risk factors in the treatment of cardiovascular disease.
79	Essential HypertensionThe clinical characteristics of patients with atrial fibrillation were evaluated in two groups.
80	Patients who had undergone coronary artery disease who had undergone coronary intervention were evaluated in our clinic.
81	The aim of this study was to investigate the efficacy of heart failure in patients with atrial fibrillation (AF) in patients with atrial fibrillation (AF), and in patients with atrial fibrillation.
82	It was concluded that left ventricular outflow tract and inferior vena cava, left atrial appendage, left atrial appendage and right ventricle.
83	Congenital atrial fibrillation (AF) is associated with atrial fibrillation. Atrial fibrillation is an important predictor of AF.

84	There was no significant differences in heart failure, heart failure, pulmonary artery disease, coronary artery disease, coronary artery disease, coronary artery disease, coronary artery disease
85	The mean blood pressure level was 78±4.10 mg/dl, respectively. 6565The incidence of type (%58-and-95% CI: 6165There was no significant difference in Group B.
86	Cardiac tachycardia (ICD) is one of the patient with a history of tachycardia, dizziness, dizziness, and one of them, and the patient who had been admitted with the presence of dyspnea, dizziness.
87	In conclusion, in patients with AF is associated with a high follow-up of the patients with AF, which is an effective treatment for the treatment of the treatment of the treatment of the treatment of the treatment.
88	A 54-year-old female patient with aortic valve replacement with aortic valve replacement with mitral valve replacement with mitral valve replacement
89	In this study, there was a significant correlation between the presence of metabolic syndrome and the presence of heart failure in patients with metabolic syndrome.
90	Coronary artery ectasia (CAD) was defined as the proximal portion of the left coronary artery (CAD).
91	In addition, the QRS value of the QRS and treatment of atrial fibrillation were measured by a mean QRS value of 50 msec.
92	Twenty-eight consecutive patients who were admitted to our clinic with the quality of life and quality of life, and the quality of life of atrial fibrillation, atrial fibrillation, atrioventricular block, atrioventricular block, atrioventricular block, atrioventricular block, atrioventricular block, atrioventricular block, atrioventricular block, atrioventricular block, atrioventricular block, atrioventricular block, atrioventricular block, atrioventricular block, AV block
93	In this study, we aimed to evaluate the use of the use of the use of the patients with cardiac arrest and a new risk factor for the patient.
94	In this study, the ECG was used for the diagnosis of electrocardiography and QT dispersion.

95	Coronary flow reserve (SCF) is an important role in the presence of coronary arteries.
96	It is concluded that the levels of low-density lipoprotein levels in patients with coronary artery disease is very important.
97	In this study, we aimed to investigate the relationship between the age and relationship between the end-term and long-term follow-up of the patients with ST-elevation myocardial infarction (Stemi).
98	In this article, we present the clinical value of a 50-year-old male patient with a high-density lipoprotein (C) .STUDY Design The study included a 50-year-old male patient who underwent coronary artery disease.
99	The incidence of atrial fibrillation and atrial fibrillation were found to be significantly different in patients with chronic heart disease.
100	All patients with coronary artery disease (Group II). All patients were divided into two groups. (n=11) There was no difference between the two groups in the two groups.

Translation by the NMT-2 Engine

100 segments translated by NMT-2 engine which is an NMT engine trained on TRENCARD + GENCOR corpora.

ID	**NMT-2**
1	No rhabdomyolysis or myopathy were not observed.
2	Pseudoaneurysm of carotid stent after carotid stent
3	Chronic cause of peripheral embolism
4	Multiple Safen Vein Graft
5	Physical examination revealed hypotension and tachycardia.
6	Clinical and angiographic features were evaluated.
7	He proposed medical follow-up to the medical follow-up.

8	Efficacy and safety of carotid artery stenting
9	Arterial thoracic outlet syndrome in young women
10	Immediately after implantation, he developed a postmates and convoy.
11	Thrombophilic gene mutation in coronary artery ectasia
12	Postinfarction-in-myocardial infarction aneurysm in the right ventricle
13	DCOS was used to evaluate low cardiac output development.
14	Spontaneous aortic root rupture in a patient with pregnancy
15	Sodium glucose ko-transports 2 inhibitors in the treatment of heart failure
16	Prognostic dietary index of infective endocarditis in patients with infective endocarditis
17	The use of long-term fibrate in real life is effective and safe.
18	Twenty-year of the 27-year durability of St. Jude Medical
19	New markers for cellular ischemia in coronary slow flow phenomenon
20	A case of coronary stent originating from the right posterior cerebral artery
21	Acute heart failure is one of the most common cardiac emergency cases.
22	Radiofrequency ablation of radiofrequency ablation may cause serious complications.
23	Evaluation of metabolic effects in the CETP gene in patients with coronary artery disease
24	Ventricular premature beats (VEV) is one of the most common arrhythmias.
25	The effect of transcatheter atrial septal defect closure on left heart functions
26	Carotid artery PA is related to trauma, trauma, spontaneous infection, vasculitis or iatrogenic causes.
27	The vascular system was found to be the right posterior lesion of the stent.
28	Successful left coronary defects in a patient with dilated cardiomyopathy
29	Various echocardiographic changes in hypoxemia due to idiopathic pulmonary fibrosis
30	Assessment of atrial electromechanical delay and P wave dispersion in chronic obstructive pulmonary disease

31	Prosthetic valve thrombosis (PVT) is a serious complication in patients with prosthetic heart valve.
32	Serum lactate level may predict acute kidney injury in acute decompensated heart failure.
33	Patients with stable angina pectoris who underwent coronary angiography between December 2018 and April 2019 were evaluated.
34	Global chronic renal disease burden and renal functions in Turkish heart failure patients
35	Major vascular complications and cardiac tamponade were significantly higher in the low PCC group.
36	Successful coronary bypass surgery was performed in 2 months after primary percutaneous coronary intervention.
37	The effect of primary percutaneous coronary intervention due to ST-segment elevation myocardial infarction due to ST-segment elevation myocardial infarction
38	Evaluation of complementary imaging modalities in the diagnosis of obstructive aortic and mitral prosthetic valve thrombosis
39	A major trend in the treatment of patients with Patent ductus arteriosus and increased trend in the treatment of patients with Patent ductus arteriosus
40	The improvement of thrombolytic therapy after thrombolytic therapy in patients with prosthetic heart valve thrombosis
41	Clinical characteristics of preserved heart failure patients with preserved ejection fraction in sinus rhythm
42	A single egg with familial hypercholesterolemia and high lipoprotein (a) levels in the same cardiovascular end
43	Low cardiac output syndrome score used to evaluate cardiac surgery after cardiac surgery after cardiac surgery
44	A long-term relationship with a long-term left ventricular function in patients with ST-elevation myocardial infarction

45	This case report shows different echocardiographic changes in a 73-year-old man with 2 hypoxic attack.
46	A 52-year-old female patient was hospitalized for atypical pneumonia due to atypical pneumonia.
47	The scintigraphy of the TTHR and the diagnosis of TR-A is an easy and alternative method in the diagnosis of TR-CA.
48	Transcatheter ASD closure of left atrial systolic and left ventricular diastolic functions were affected by positive response.
49	Brugada syndrome (BrS) is a congenital distribution of congenital ventricular arrhythmias and is associated with sudden cardiac death.
50	The electrocardiogram showed rapid idioventricular rhythm, capture and fusion of the DII, DIII, and aVF in the precordial leads.
51	The carotid artery stenting is an interventional treatment modality in the treatment of carotid artery stenosis.
52	A comparison of myocardial performance index with myocardial performance index in patients with obstructive ventricular dysfunction
53	In this article, we report a diabetic patient who was admitted to emergency department with dyspnea in the emergency department.
54	This may be an important feature of FH families due to the prediction of the risk of cardiovascular events.
55	There is no clear consensus about the follow-up of the patients with low cardiac output syndrome (DKS).
56	Cardiovascular diseases (CVD) is one of the major causes of morbidity and mortality in all the world and high cost of cardiovascular disease.
57	A 58-year-old male patient with left ventricular assist device (SVDC) was admitted with rest dyspnea.
58	However, patients without AF had higher diabetes mellitus, obstructive sleep apnea and coronary artery disease.
59	It is often associated with chronic fatigue syndrome or chronic fatigue syndrome, chronic abdominal pain, migraine, and sugar disease.

60	Last degree has been used to have long life expectancy for antiimmune therapy, human immune insufficiency virus (HIV).
61	In this article, we report a 17-year-old male patient who was admitted with non-obese ventricular contraction associated with premature ventricular contraction and premature ventricular contraction.
62	Despite a good role, cardiomyopathy may cause cardiomyopathy, and rarely associated with sudden cardiac death.
63	It provides a potential alternative to conventional transvenous cardiac pacemakers for patients with high-risk transcatheter valve replacement procedures.
64	A false sinus tachycardia (UST) is a syndrome characterized by increased resting heart rate and uncomfortable symptoms.
65	A 41-year-old male patient developed complaints of palpitation, consciousness, and loss of consciousness. Cardiac arrest occurred after the death.
66	In this study, serum levels and demographic parameters of the hemostatic variables were found to be significantly different in the LAA thrombus.
67	Coronary slow flow phenomenon (CAAF) is defined as the distal vessel in the distal vessel coronary artery.
68	For example, L/O structure was associated with better discharge and long-term LV volumes and ejection fraction.
69	In the diagnosis of acute myocarditis, the presence of GKT and/or edema were obtained in the diagnosis of Lake Louise retrospectively than Lake Louise criteria.
70	Ventricular arrhythmias arising from coronary leaflets are a rare condition. However, the outflow tract arising from the aortic leaflets may be the map and ablation of ablation.
71	Although there is a risk factor for heart failure, there is a risk factor for the left heart functions in most of these subjects.
72	Prognostic levels of Serum albumin and lymphocyte concentrations are high risk score (PBI) in various patients in various patient populations.

73	This study aims to evaluate the results of primary percutaneous coronary intervention (PPCI) in patients with ST-segment elevation myocardial infarction (Stemi).
74	The aim of this study is to investigate the effects of CETP rs289714 gene polymorphism in patients with coronary artery disease (CAD).
75	Monosigotic splines can represent a special source of a particular source of FH, the age of the age and all genes, and to investigate the natural course of FH.
76	The presence of increased leukocyte count and the presence of fresh thrombus, clinical and procedural characteristics may be associated with the long-term deterioration in LV function.
77	However, the new data suggest that patients with preserved ejection fraction in patients with preserved ejection fraction (%13 and high risk) were found in %16 patients with severe aortic stenosis.
78	This study aims to determine the efficacy of fibrate therapy, reliability of fibrate therapy, the reliability of fibrate therapy and the effect of diabetic endpoints.
79	KEF-HF PatientsIn patients with AF, RV-HF patients were divided into two groups, and the clinical characteristics of the patients were compared with the clinical characteristics of these patients.
80	This report discusses the management challenges of HIV seroposis in 2006, a 11-year follow-up report of HIV-transplant.
81	The aim of this study was to evaluate the clinical characteristics of the patients with preserved ejection fraction (AF) and the clinical characteristics of the patients who were without AF.
82	During the follow-up of minutes, hypotension, right ventricular and inferior vena cava were developed, left atrial and left ventricle, and the SCDC rate continued despite reduction .
83	Chronic obstructive pulmonary disease (AF) with atrial fibrillation (AF) and decreased first second-second challenging expiratory volume (ZEVAR) It is an independent predictor of AF.

84	LAA thrombus (+) group, heart failure, peripheral arterial disease, coronary artery disease and chronic obstructive pulmonary disease were significantly higher (p. 0.05).
85	The mean leukocyte count (127.5±86.0-and-227.2±120.7) p=0.026) The frequency (%35 and %75, respectively) and the frequency of the patientsp=0.046) The patients were significantly higher in the group.
86	Postural orthostatic tachycardia syndrome (POTS) is a chronic and strong condition characterized by heterogeneous symptoms such as dizziness, palpitation, prescope, syncope, syncope, and especially the legs of the legs.
87	Transthoracic echocardiography, a first-step imaging tool for symptomatic patients with PKT, may be inadequate to choose the appropriate treatment option to select appropriate treatment option.
88	A 51-year-old woman with deep bradycardia and mitral valve replacement procedure was present in a 51-year-old female patient with deep bradycardia following the procedure.
89	In this study, there was a significant correlation between the right heart function and the severity of obesity in healthy subjects and the severity of obesity was significantly correlated with the severity of obesity.
90	Coronary artery ectasia (CAE) was defined as localized or common dilatation of the lesion of the coronary artery, with less than 1.5 times.
91	A normal sinus rhythm was observed in patients with atrial fibrillation with a liquid and electrolyte replacement therapy, and normal sinus rhythm was detected with 490 msec.
92	A total of 1865 consecutive patients were evaluated with insufficient registries or quality, atrial fibrillation attacks, atrioventricular block, and > %1 atrial or ventricular extra systole.
93	The aim of our study was to highlight the use of a common language in the care of patients with a single scoring of patients who underwent cardiac surgery with the parameters of poor perfusion.
94	This case was used to determine the normal QT and corrected QT interval in three individuals with a history of sudden and sudden cardiac death.

95	The pathophysiology of coronary slow flow (CFR) has been reported to determine the pathophysiology of the pathophysiology, which may be associated with oxidative stress and inflammation.
96	As a result, the high Lappa levels and the presence of monosigotic HoFH are an important evidence for life-density cholesterol burden/marine hypothesis.
97	This study aims to investigate the possible relationship between the age and right ventricular (LV) function in patients with ST-elevation myocardial infarction (Stemi), postoperative and long-term left ventricular (LV) functions.
98	In this article, we present a 36-year-old monosigic twin-brother of a 36-year-old monosigic twin brothers who had the same early coronary artery involvement.
99	Both P wave dispersion and atrial electromechanical delay parameters were significantly longer in COPD patients with structural and functional cardiac disease.
100	Coronary interventions in symptomatic patients (%39) In the asymptomatic group, the patient was in asymptomatic group (%29). (%29) There was no statistically significant difference between the groups.

ANNEX VI. Bilingual Terms Extracted from the Sample Set

231 bilingual terms extracted from the sample set sorted by source language n-gram counts.

ID	Turkish	English	Source N-grams	Target N-grams
1	asetaminofen	acetaminophen	1	1
2	aritmi	arrhythmia	1	1
3	asistoli	asystole	1	1
4	kardiyomiyopati	cardiomyopathy	1	1
5	kardiyomiyopati	cardiomyopathy	1	1
6	kardiyovasküler	cardiovascular	1	1
7	konvülziyon	convulsion	1	1
8	COVID-19	COVID-19	1	1

9	diyabetik	diabetic	1	1
10	disapne	dyspnea	1	1
11	elektrokardiyografi	Electrocardiography	1	1
12	fibrat	fibrate	1	1
13	hiperkolesterolemi	hypercholesterolemia	1	1
14	hipertrigliseridemi	hypertriglyceridemia	1	1
15	hipotansiyon	hypotension	1	1
16	hipotansiyon	hypotension	1	1
17	hipoksi	hypoxia	1	1
18	iyatrojenik	iatrogenic	1	1
19	ibuprofen	ibuprofen	1	1
20	implantasyon	implantation	1	1
21	yangı	inflammation	1	1
22	inflamasyon	inflammation	1	1
23	lökosit	leukocyte	1	1
24	lökosit	leukocyte	1	1
25	migren	migraine	1	1
26	miyopati	myopathy	1	1
27	çarpıntı	palpitation	1	1
28	çarpıntı	palpitation	1	1
29	patofizyoloji	pathophysiology	1	1
30	perfüzyon	perfusion	1	1
31	presenkop	pre-syncope	1	1
32	psödoanevrizma	pseudoaneurysm	1	1
33	rabdomiyoliz	rhabdomyolysis	1	1
34	sintigrafi	scintigraphy	1	1
35	senkop	syncope	1	1
36	taşikardi	tachycardia	1	1
37	pıhtı	thrombus	1	1
38	pıhtı	thrombus	1	1
39	pıhtı	thrombus	1	1
40	pıhtı	thrombus	1	1
41	vaskülit	vasculitis	1	1
42	trombüs	thrombus	1	1
43	yangı	inflammation	1	1

44	akut miyokardit	acute myocarditis	2	2
45	antiretroviral tedavi	antiretroviral therapy	2	2
46	aortik yaprakçıklar	aortic cusps	2	2
47	aort darlığı	aortic stenosis	2	2
48	atriyal fibrilasyon	atrial fibrillation	2	2
49	Atriyal fibrilasyon	atrial fibrillation	2	2
50	atriyal fibrilasyon	atrial fibrillation	2	2
51	atriyal fibrilasyon	atrial fibrillation	2	2
52	atriyoventriküler blok	atrioventricular block	2	2
53	atipik pnömoni	atypical pneumonia	2	2
54	kalp durması	cardiac arrest	2	2
55	kardiyak cerrahi	cardiac surgery	2	2
56	kardiyak cerrahi	cardiac surgery	2	2
57	kardiyak tamponad	cardiac tamponade	2	2
58	kardiyovasküler hastalıklar	cardiovascular disease	2	2
59	kardiyovasküler sonlanım	cardiovascular outcome	2	2
60	karotis stentleme	carotid stenting	2	2
61	hücresel iskemi	cellular ischemia	2	2
62	CETP geni	CETP gene	2	2
63	konjenital kanalopati	congenital channelopathy	2	2
64	koroner anjiyografi	coronary angiography	2	2
65	koroner arter	coronary artery	2	2
66	koroner kusp	coronary cusp	2	2
67	koroner girişim	coronary intervention	2	2
68	COVID-19 pandemisi	COVID-19 pandemic	2	2
69	kardiyovasküler olay	CVD event	2	2
70	diabetes mellitus	diabetes mellitus	2	2
71	şeker hastalığı	diabetes mellitus	2	2
72	diyabetik sonlanım	diabetic outcome	2	2
73	TTR-KA tanısı	diagnosis of TTR-CA	2	3
74	ekokardiyografik değişiklikler	echocardiographic changes	2	2
75	ekokardiyografik değişiklikler	echocardiographic changes	2	2
76	ejeksiyon fraksiyonu	ejection fraction	2	2
77	endotel fonksiyon	endothelial function	2	2
78	fibrat tedavisi	fibrate treatment	2	2

79	floroskopik tarama	fluoroscopic scanning	2	2
80	kalp yetersizliği	heart failure	2	2
81	kalp yetersizliği	heart failure	2	2
82	kalp yetersizliği	heart failure	2	2
83	kalp yetersizliği	heart failure	2	2
84	kalp yetersizliği	heart failure	2	2
85	kalp yetmezliği	heart failure	2	2
86	kalp yetersizliği	heart failure	2	2
87	kalp nakli	heart transplantation	2	2
88	yüksek lipoprotein	high lipoprotein	2	2
89	histopatolojik özellik	histopathological feature	2	2
90	HIV seropozitif	HIV-seropositive	2	1
91	hipermobil eklem	hypermobile joint	2	2
92	enfektif endokardit	infective endocarditis	2	2
93	intrauterin çevre	intrauterine environment	2	2
94	isovolumetrik kasılma	isovolumic contraction	2	2
95	SAA trombüsü	LAA thrombus	2	2
96	baş dönmesi	lightheadedness	2	1
97	lenfosit konsantrasyon	lymphocyte concentration	2	2
98	baş ağrısı	headache	2	1
99	oksidatif stres	oxidative stress	2	2
100	periferik emboli	peripheral embolism	2	2
101	derin bradikardi	profound bradycardia	2	2
102	radyofrekans ablasyon	radiofrequency ablation	2	2
103	istirahat dispnesi	rest dyspnea	2	2
104	sağ ventrikül	right ventricle	2	2
105	rs289714 varyasyonu	rs289714 variation	2	2
106	serum albümin	serum albumin	2	2
107	sinüs ritmi	sinus rhythm	2	2
108	sinüs ritmi	sinus rhythm	2	2
109	spontan enfeksiyon	spontaneous infection	2	2
110	trombolitik tedavi	thrombolytic therapy	2	2
111	transtorasik ekokardiyografi	transthoracic echocardiography	2	2
112	vasküler sistem	vascular system	2	2
113	akut kalp yetersizliği	acute heart failure	3	3

114	akut böbrek hasarı	acute kidney injury	3	3
115	akut miyokart enfarktüsü	acute myocardial infarction	3	3
116	atriyal elektromekanik gecikme	atrial electromechanical delay	3	3
117	atriyal fibrilasyon atakları	atrial fibrillation episodes	3	3
118	Brugada sendromu (BrS)	Brugada syndrome	3	2
119	kardiyak acil durum	cardiac emergencies	3	2
120	Karotis arter PA’sı	carotid artery PA	3	3
121	karotis arter stenozu	carotid artery stenosis	3	3
122	karotis arter stentleme	carotid artery stenting	3	3
123	Karotis arter stentlemesi	Carotid artery stenting	3	3
124	kronik karın ağrısı	chronic abdominal pain	3	3
125	kronik yorgunluk sendromu	chronic fatigue syndrome	3	3
126	kronik böbrek hastalığı	chronic kidney disease	3	3
127	invazif nöroradyoloji konsültasyonu	consultant invasive neuroradiologist	3	3
128	koroner baypas cerrahisi	coronary artery bypass graft operation	3	5
129	koroner arter hastalığı	coronary artery disease	3	3
130	koroner arter hastalığı	coronary artery disease	3	3
131	koroner arter hastalığı	coronary artery disease	3	3
132	koroner arter ektazisi	coronary artery ectasia	3	3
133	koroner arter ektazisi	coronary artery ectasia	3	3
134	koroner arterin lümeni	coronary artery lumen	3	3
135	kümülatif LDL maruziyeti	cumulative LDL exposure	3	3
136	distal damar yatağı	distal vascular bed	3	3
137	epinefrin provokasyon testi	epinephrine provocation test	3	3
138	hipoksik atak epizodu	episode of hypoxic attack	3	4
139	hızlı idiyoventriküler ritim	fast idioventricular rhythm	3	3
140	zorlu ekspiratuvar volümü	forced expiratory volume	3	3
141	yüksek lipoprotein a	high lipoprotein (a)	3	3
142	uygunsuz sinüs taşikardisi	inappropriate sinus tachycardia	3	3
143	kablosuz kalp pili	leadless pacemaker	3	2
144	Kablosuz kalp pilleri	leadless pacemakers	3	2
145	sol kalp fonksiyonu	left heart function	3	3
146	sol ventrikülde kollaps	left ventricular collapse	3	3
147	sol ventrikül fonksiyonu	left ventricular function	3	3

148	sol ventrikül fonksiyon	left ventricular function	3	3
149	düşük kardiyak debi	low cardiac output	3	3
150	majör vasküler komplikasyonlar	major vascular complications	3	3
151	malign ventriküler aritmi	malignant ventricular arrhythmia	3	3
152	miyokardiyal performans indeksi	myocardial performance index	3	3
153	normal sinüs ritmi	normal sinus rhythm	3	3
154	obstrüktif uyku apnesi	obstructive sleep apnea	3	3
155	P dalga dispersiyonu	P wave dispersion	3	3
156	P dalga dispersiyonu	P wave dispersion	3	3
157	patent duktus arteriyozus	patent ductus arteriosus	3	3
158	periferik arter hastalığı	peripheral artery disease	3	3
159	Ventriküler erken vurular	premature ventricular complexes	3	3
160	prematür ventriküler kontraksiyon	premature ventricular contraction	3	3
161	korunmuş ejeksiyon fraksiyonu	preserved ejection fraction	3	3
162	korunmuş ejeksiyon fraksiyonu	preserved ejection fraction	3	3
163	korunmuş ejeksiyon fraksiyonu	preserved ejection fraction	3	3
164	prognostik nutrisyonel indeks	prognostic nutritional index	3	3
165	prognostik nutrisyonel indeks	prognostic nutritional index	3	3
166	protez kalp kapağı	prosthetic heart valve	3	3
167	Protez kapak trombozu	Prosthetic valve thrombosis	3	3
168	istirahat kalp hızı	resting heart rate	3	3
169	sağ kalp fonksiyonu	right heart function	3	3
170	sağ ventriküler disfonksiyonu	right ventricular dysfunction	3	3
171	sağ ventriküler akselerasyon	right ventricular myocardial acceleration	3	4
172	Serum laktat seviyesi	serum lactate level	3	3
173	serum lipit profili	serum lipid profile	3	3
174	artmış fruktoz tüketimi	significant fructose consumption	3	3
175	koroner yavaş akım	slow coronary flow	3	3
176	kararlı anjina pektoris	stable angina pectoris	3	3
177	ani kardiyak ölüm	sudden cardiac death	3	3
178	ani kardiyak ölüm	sudden cardiac death	3	3
179	Transkateter ASD kapatılması	transcatheter closure of an ASD	3	5
180	transvenöz kalp pilleri	transvenous pacemakers	3	2

181	akut dekompanse kalp yetersizliği	acute decompensated heart failure	4	4
182	atriyal elektromekanik gecikme parametreleri	AEMD parameters	4	2
183	trombofilik gen mutasyon analizi	analysis of thrombophilic gene mutations	4	5
184	arteryel torasik çıkış sendromu	arterial thoracic outlet syndrome	4	4
185	CETP rs289714 gen polimorfizmi	CETP gene rs289714 polymorphism	4	4
186	kronik obstrüktif akciğer hastalığı	chronic obstructive pulmonary disease	4	4
187	kronik obstrüktif akciğer hastalığı	chronic obstructive pulmonary disease	4	4
188	kronik obstrüktif akciğer hastalığı	chronic obstructive pulmonary disease	4	4
189	klinik ve anjiyografik özellikler	clinical and angiographic characteristics	4	4
190	koroner yavaş akım fenomeni	coronary slow-flow phenomenon	4	3
191	koroner yavaş akım fenomeni	coronary slow-flow phenomenon	4	3
192	embolize olan koroner stent	Coronary stent embolism	4	3
193	erken koroner arter tutulumu	early coronary artery involvement	4	4
194	erken koroner arter tutulumu	early coronary artery involvement	4	4
195	ani kalp ölümü öyküsü	history of sudden cardiac death	4	5
196	insan immün yetmezliği virüsü	human immun-deficiency virus	4	3
197	inferiyor vena kavada dilatasyon	inferior vena cava dilatation	4	4
198	SAA trombüs (+) grubu	LAA thrombus (+) group	4	4
199	sol atriyum sistolik fonksiyonları	left atrial systolic functions	4	4
200	sol ventrikül destek cihazı	left ventricular assist device	4	4
201	Düşük kardiyak debi sendromu	low cardiac output syndrome	4	4
202	düşük kardiyak debi sendromu	low cardiac output syndrome	4	4
203	majör epikardiyal koroner arter	major epicardial coronary artery	4	4
204	mitral kapak kapakçığı prosedürü	mitral valve-in-valve procedure	4	3
205	nonsustained ventriküler taşikardi atakları	nonsustained ventricular tachycardia attacks	4	4
206	çıkım yolu ventriküler aritmileri	outflow tract ventricular arrhythmias	4	4
207	post-infarkt sol ventrikül anevrizması	post-infarction aneurysm of left ventricle	4	5
208	postural ortostatik taşikardi sendromu	postural orthostatic tachycardia syndrome	4	4
209	primer perkütan koroner girişim	primary PCI	4	2

210	primer perkütan koroner girişim	primary percutaneous coronary intervention	4	4
211	primer perkütan koroner girişim	primary percutaneous coronary intervention	4	4
212	protez kalp kapak trombozu	prosthetic heart valve thrombosis	4	4
213	sağ posteriyor serebral arter	right posterior cerebral artery	4	4
214	sağ posteriyor serebral arter	right posterior cerebral artery	4	4
215	spontan aort kökü rüptürü	spontaneous aortic root rupture	4	4
216	ST yükselmeli miyokart enfarktüsü	ST-elevation myocardial infarction	4	3
217	ST yükselmeli miyokart enfarktüsü	ST-segment elevation myocardial infarction	4	4
218	ST-segment yükselmeli miyokart enfarktüsü	ST-segment elevation myocardial infarction	4	4
219	ST-segment yükselmeli miyokart enfarktüsü	ST-segment myocardial infarction	4	3
220	transkateter atrial septal defekt	transcatheter atrial septal defect	4	4
221	transkateter aort kapak değişimi	transcatheter aortic valve replacement	4	4
222	İdiyopatik pulmoner fibrozun akut alevlenmesi	acute exacerbation of idiopathic pulmonary fibrosis	5	6
223	atriyal veya ventriküler ekstra sistoller	atrial or ventricular extrasystoles	5	4
224	sağlıklı komşu damar segment çapı	diameter of adjacent healthy reference segments	5	6
225	Sıvı ve elektrolit replasman tedavisi	fluids and electrolyte replacement therapy	5	5
226	çoklu safen ven greft anevrizması	multiple saphenous vein graft aneurysm	5	5
227	sodyum glikoz ko-transporter 2 inhibitörleri	Sodium glucose co-transporter 2 inhibitors	5	5
228	yapısal ve fonksiyonel kardiyak hastalık	structural or functional cardiac abnormalities	5	5
229	yüksek riskli transkateter kapak değiştirme prosedürleri	high-risk transcatheter valve replacement procedures	6	5
230	Jude medikal mitral sığır biyoprotez kapağı	St. Jude Medical Biocor bovine bioprosthesis in mitral position	6	9
231	eşzamanlı obstrüktif aortik ve mitral prostetik kapak trombozları	concurrent obstructive aortic and mitral prosthetic valve thrombosis	8	8

ANNEX VII. Acronyms in the Sample Set

List of the acronyms in our sample set is given below.

ANNEX VII. Acronyms in the Sample Set

ID	Turkish	English
1	AF	AF
2	AF	AF
3	AF	AF
4	AF	AF
5	AF	AF
6	AH	FH
7	AH	FH
8	AH	FH
9	ASD	ASD
10	aVF	aVF
11	BrS	BrS
12	CETP	CETP
13	CETP	CETP
14	COVID-19	COVID-19
15	DII	DII
16	DIII	DIII
17	DKDS	LCOS
18	DKDSs	LCOSs
19	EKG	ECG
20	GKT	LGE
21	HIV	HIV
22	HIV	HIV
23	HoAH	HoFH
24	KAE	CAE
25	KAH	CAD
26	KAS	CAS
27	KEF-KY	HFpEF
28	KEF-KY	HFpEF
29	KOAH	COPD
30	KOAH	COPD
31	KVH	CVD
32	KYA	SCF
33	KYAF	CSFP
34	L/O	L/O

35	LDL	LDL
36	Lpa	Lpa
37	Lpa	Lpa
38	PA	PA
39	PKT	PVT
40	PKT	PVT
41	PNİ	PNI
42	PNİ	PNI
43	POTS	POTS
44	PPKG	PPCI
45	QRS	QRS
46	QT	QT
47	QT	QT
48	SAA	LAA
49	SAA	LAA
50	SCDC	LVAD
51	SR	SR
52	ST	ST
53	ST	ST
54	ST	ST
55	STYME	STYMI
56	SV	LV
57	SV	LV
58	SV	LV
59	SVDC	LVAD
60	SVYB	LVR
61	TTR-KA	TTR-CA
62	TTR-KA	TTR-CA
63	UQTS	LQTS
64	UST	IST
65	VEV	PVC
66	YD	SOI
67	ZEV1	FEV1

Annex VIII. Term Error Annotation in 4 Engines

The binary (correct/incorrect) term annotation as well as term error category assignations for each target term are provided below for reference. Annex V includes Turkish source sentences and target reference translations as well as the translations by PBMT-1, PBMT-2, NMT-1 and NMT-2 engines. Comparisons can be made directly by consulting the term translation in relevant sentences. For a more convenient view, term error annotation is also presented in an open multi-sheet spreadsheet in GitHub repository[116]. An e-mail can be sent to gokhan.dogru@uab.cat for requesting the spreadsheets used in the term error annotation process as well.

i. Term error annotation in PBMT-1

Term	Binary Analysis	Term Error category
rhabdomyolysis	Correct Term	a.CT-Correct term considering reference
carotid stenting	Incorrect	8.Term extended
peripheral embolism	Incorrect	1.Partial term translation
multiple saphenous vein graft aneurysm	Correct Term	a.CT-Correct term considering reference
hypotension	Correct Term	a.CT-Correct term considering reference
clinical and angiographic characteristics	Correct Term	a.CT-Correct term considering reference
consultant invasive neuroradiologist	Incorrect	1.Partial term translation
carotid artery stenting	Incorrect	1.Partial term translation
arterial thoracic outlet syndrome	Correct Term	a.CT-Correct term considering reference
implantation	Incorrect	7.Term drop
coronary artery ectasia	Incorrect	1.Partial term translation
right ventricle	Correct Term	a.CT-Correct term considering reference
low cardiac output	Correct Term	a.CT-Correct term considering reference

[116] https://github.com/gokhandogru/TurkishEnglishParallelCorporaandMTEvaluation

spontaneous aortic root rupture	Correct Term	a.CT-Correct term considering reference
heart failure	Correct Term	a.CT-Correct term considering reference
prognostic nutritional index	Incorrect	1.Partial term translation
fibrate	Correct Term	a.CT-Correct term considering reference
St. Jude Medical Biocor bovine bioprosthesis in mitral position	Incorrect	1.Partial term translation
coronary slow-flow phenomenon	Incorrect	1.Partial term translation
right posterior cerebral artery	Correct Term	a.CT-Correct term considering reference
acute heart failure	Correct Term	a.CT-Correct term considering reference
radiofrequency ablation	Correct Term	a.CT-Correct term considering reference
coronary artery	Correct Term	a.CT-Correct term considering reference
premature ventricular complexes	Incorrect	1.Partial term translation
transcatheter atrial septal defect	Correct Term	a.CT-Correct term considering reference
carotid artery PA	Correct Term	a.CT-Correct term considering reference
vascular system	Correct Term	a.CT-Correct term considering reference
coronary cusp	Incorrect	1.Partial term translation
acute exacerbation of idiopathic pulmonary fibrosis	Incorrect	1.Partial term translation
chronic obstructive pulmonary disease	Correct Term	a.CT-Correct term considering reference
Prosthetic valve thrombosis	Correct Term	a.CT-Correct term considering reference
serum lactate level	Correct Term	a.CT-Correct term considering reference
coronary angiography	Correct Term	a.CT-Correct term considering reference
chronic kidney disease	Correct Term	b.CT-Correct synonym
major vascular complications	Correct Term	a.CT-Correct term considering reference
primary PCI	Correct Term	b.CT-Correct synonym
COVID-19 pandemic	Incorrect	1.Partial term translation
concurrent obstructive aortic and mitral prosthetic valve thrombosis	Incorrect	1.Partial term translation
patent ductus arteriosus	Incorrect	1.Partial term translation
prosthetic heart valve thrombosis	Correct Term	a.CT-Correct term considering reference
atrial fibrillation	Correct Term	a.CT-Correct term considering reference

Annex VIII. Term Error Annotation in 4 Engines

cardiovascular outcome	Incorrect	1.Partial term translation
cardiac surgery	Incorrect	1.Partial term translation
ST-segment elevation myocardial infarction	Correct Term	b.CT-Correct synonym
episode of hypoxic attack	Incorrect	2.Incorrect term equivalent
atypical pneumonia	Correct Term	a.CT-Correct term considering reference
scintigraphy	Incorrect	5.Source term insertion
transcatheter closure of an ASD	Correct Term	b.CT-Correct synonym
Brugada syndrome	Correct Term	a.CT-Correct term considering reference
Electrocardiography	Correct Term	a.CT-Correct term considering reference
Carotid artery stenting	Correct Term	a.CT-Correct term considering reference
right ventricular dysfunction	Correct Term	a.CT-Correct term considering reference
acute myocardial infarction	Correct Term	a.CT-Correct term considering reference
CVD event	Correct Term	b.CT-Correct synonym
low cardiac output syndrome	Correct Term	a.CT-Correct term considering reference
cardiovascular disease	Correct Term	a.CT-Correct term considering reference
left ventricular assist device	Correct Term	a.CT-Correct term considering reference
diabetes mellitus	Correct Term	a.CT-Correct term considering reference
hypermobile joint	Incorrect	5.Source term insertion
antiretroviral therapy	Correct Term	a.CT-Correct term considering reference
nonsustained ventricular tachycardia attacks	Correct Term	a.CT-Correct term considering reference
cardiomyopathy	Correct Term	a.CT-Correct term considering reference
leadless pacemakers	Incorrect	1.Partial term translation
inappropriate sinus tachycardia	Correct Term	a.CT-Correct term considering reference
palpitation	Correct Term	a.CT-Correct term considering reference
LAA thrombus	Incorrect	1.Partial term translation
coronary slow-flow phenomenon	Correct Term	a.CT-Correct term considering reference
ejection fraction	Correct Term	a.CT-Correct term considering reference
acute myocarditis	Correct Term	a.CT-Correct term considering reference

outflow tract ventricular arrhythmias	Correct Term	a.CT-Correct term considering reference
heart failure	Correct Term	a.CT-Correct term considering reference
serum albumin	Correct Term	a.CT-Correct term considering reference
COVID-19	Correct Term	a.CT-Correct term considering reference
coronary artery disease	Correct Term	a.CT-Correct term considering reference
intrauterine environment	Incorrect	2.Incorrect term equivalent
leukocyte	Correct Term	a.CT-Correct term considering reference
heart failure	Correct Term	a.CT-Correct term considering reference
hypertriglyceridemia	Correct Term	a.CT-Correct term considering reference
sinus rhythm	Correct Term	a.CT-Correct term considering reference
heart transplantation	Correct Term	a.CT-Correct term considering reference
atrial fibrillation	Correct Term	a.CT-Correct term considering reference
hypotension	Correct Term	a.CT-Correct term considering reference
chronic obstructive pulmonary disease	Correct Term	a.CT-Correct term considering reference
LAA thrombus (+) group	Correct Term	a.CT-Correct term considering reference
leukocyte	Correct Term	a.CT-Correct term considering reference
postural orthostatic tachycardia syndrome	Correct Term	a.CT-Correct term considering reference
transthoracic echocardiography	Correct Term	a.CT-Correct term considering reference
transcatheter aortic valve replacement	Correct Term	a.CT-Correct term considering reference
right heart function	Incorrect	2.Incorrect term equivalent
coronary artery ectasia	Correct Term	a.CT-Correct term considering reference
fluids and electrolyte replacement therapy	Incorrect	1.Partial term translation
atrial fibrillation episodes	Incorrect	1.Partial term translation
perfusion	Correct Term	a.CT-Correct term considering reference
history of sudden cardiac death	Correct Term	a.CT-Correct term considering reference
slow coronary flow	Correct Term	a.CT-Correct term considering reference
early coronary artery involvement	Correct Term	a.CT-Correct term considering reference

ST-elevation myocardial infarction	Correct Term	b.CT-Correct synonym
high lipoprotein (a)	Incorrect	3.Morphosyntactic error
P wave dispersion	Correct Term	a.CT-Correct term considering reference
coronary intervention	Correct Term	a.CT-Correct term considering reference
myopathy	Correct Term	a.CT-Correct term considering reference
pseudoaneurysm	Correct Term	a.CT-Correct term considering reference
tachycardia	Correct Term	a.CT-Correct term considering reference
asystole	Correct Term	a.CT-Correct term considering reference
analysis of thrombophilic gene mutations	Incorrect	1.Partial term translation
post-infarction aneurysm of left ventricle	Correct Term	a.CT-Correct term considering reference
Sodium glucose co-transporter 2 inhibitors	Correct Term	a.CT-Correct term considering reference
infective endocarditis	Correct Term	a.CT-Correct term considering reference
cellular ischemia	Correct Term	a.CT-Correct term considering reference
Coronary stent embolism	Correct Term	b.CT-Correct synonym
cardiac emergencies	Correct Term	a.CT-Correct term considering reference
CETP gene	Correct Term	a.CT-Correct term considering reference
arrhythmia	Incorrect	7.Term drop
left heart function	Incorrect	1.Partial term translation
spontaneous infection	Correct Term	a.CT-Correct term considering reference
fluoroscopic scanning	Incorrect	1.Partial term translation
hypoxia	Incorrect	7.Term drop
atrial electromechanical delay	Incorrect	1.Partial term translation
prosthetic heart valve	Correct Term	a.CT-Correct term considering reference
acute decompensated heart failure	Correct Term	a.CT-Correct term considering reference
stable angina pectoris	Correct Term	a.CT-Correct term considering reference
heart failure	Correct Term	a.CT-Correct term considering reference
cardiac tamponade	Incorrect	1.Partial term translation
coronary artery bypass graft operation	Incorrect	1.Partial term translation

ST-segment elevation myocardial infarction	Correct Term	a.CT-Correct term considering reference
acetaminophen	Incorrect	5.Source term insertion
thrombolytic therapy	Correct Term	a.CT-Correct term considering reference
sinus rhythm	Correct Term	a.CT-Correct term considering reference
hypercholesterolemia	Correct Term	a.CT-Correct term considering reference
low cardiac output syndrome	Correct Term	a.CT-Correct term considering reference
histopathological feature	Incorrect	1.Partial term translation
echocardiographic changes	Correct Term	a.CT-Correct term considering reference
diagnosis of TTR-CA	Correct Term	b.CT-Correct synonym
left atrial systolic functions	Correct Term	a.CT-Correct term considering reference
congenital channelopathy	Incorrect	1.Partial term translation
fast idioventricular rhythm	Incorrect	7.Term drop
carotid artery stenosis	Incorrect	1.Partial term translation
myocardial performance index	Correct Term	a.CT-Correct term considering reference
dyspnea	Correct Term	a.CT-Correct term considering reference
cumulative LDL exposure	Incorrect	1.Partial term translation
rest dyspnea	Correct Term	b.CT-Correct synonym
obstructive sleep apnea	Correct Term	a.CT-Correct term considering reference
chronic fatigue syndrome	Correct Term	a.CT-Correct term considering reference
human immun-deficiency virus	Correct Term	a.CT-Correct term considering reference
premature ventricular contraction	Correct Term	b.CT-Correct synonym
sudden cardiac death	Correct Term	a.CT-Correct term considering reference
high-risk transcatheter valve replacement procedures	Incorrect	1.Partial term translation
resting heart rate	Correct Term	a.CT-Correct term considering reference
cardiac arrest	Correct Term	a.CT-Correct term considering reference
major epicardial coronary artery	Correct Term	a.CT-Correct term considering reference
thrombus	Incorrect	7.Term drop
aortic cusps	Incorrect	1.Partial term translation
lymphocyte concentration	Incorrect	1.Partial term translation

ST-segment myocardial infarction	Correct Term	a.CT-Correct term considering reference
CETP gene rs289714 polymorphism	Correct Term	b.CT-Correct synonym
thrombus	Correct Term	a.CT-Correct term considering reference
aortic stenosis	Correct Term	a.CT-Correct term considering reference
fibrate treatment	Incorrect	4.Reordering error
HIV-seropositive	Incorrect	1.Partial term translation
heart failure	Correct Term	a.CT-Correct term considering reference
inferior vena cava dilatation	Incorrect	4.Reordering error
atrial fibrillation	Correct Term	a.CT-Correct term considering reference
heart failure	Correct Term	a.CT-Correct term considering reference
inflammation	Incorrect	5.Source term insertion
lightheadedness	Correct Term	b.CT-Correct synonym
thrombus	Correct Term	a.CT-Correct term considering reference
mitral valve-in-valve procedure	Incorrect	1.Partial term translation
coronary artery lumen	Incorrect	1.Partial term translation
atrial fibrillation	Correct Term	a.CT-Correct term considering reference
atrioventricular block	Correct Term	a.CT-Correct term considering reference
cardiac surgery	Correct Term	a.CT-Correct term considering reference
epinephrine provocation test	Incorrect	1.Partial term translation
pathophysiology	Correct Term	a.CT-Correct term considering reference
thrombus	Incorrect	5.Source term insertion
early coronary artery involvement	Correct Term	a.CT-Correct term considering reference
AEMD parameters	Correct Term	b.CT-Correct synonym
convulsion	Incorrect	5.Source term insertion
rs289714 variation	Correct Term	a.CT-Correct term considering reference
vasculitis	Correct Term	a.CT-Correct term considering reference
right posterior cerebral artery	Correct Term	a.CT-Correct term considering reference
echocardiographic changes	Correct Term	a.CT-Correct term considering reference
P wave dispersion	Incorrect	1.Partial term translation

acute kidney injury	Correct Term	a.CT-Correct term considering reference
primary percutaneous coronary intervention	Correct Term	a.CT-Correct term considering reference
ibuprofen	Correct Term	a.CT-Correct term considering reference
endothelial function	Correct Term	a.CT-Correct term considering reference
preserved ejection fraction	Correct Term	a.CT-Correct term considering reference
high lipoprotein	Incorrect	3.Morphosyntactic error
left ventricular function	Incorrect	1.Partial term translation
malignant ventricular arrhythmia	Correct Term	a.CT-Correct term considering reference
isovolumic contraction	Incorrect	1.Partial term translation
diabetic	Correct Term	a.CT-Correct term considering reference
coronary artery disease	Correct Term	a.CT-Correct term considering reference
chronic abdominal pain	Correct Term	a.CT-Correct term considering reference
cardiomyopathy	Correct Term	a.CT-Correct term considering reference
transvenous pacemakers	Incorrect	1.Partial term translation
distal vascular bed	Incorrect	1.Partial term translation
prognostic nutritional index	Incorrect	1.Partial term translation
primary percutaneous coronary intervention	Correct Term	a.CT-Correct term considering reference
serum lipid profile	Correct Term	a.CT-Correct term considering reference
preserved ejection fraction	Correct Term	a.CT-Correct term considering reference
cardiovascular	Correct Term	a.CT-Correct term considering reference
preserved ejection fraction	Correct Term	a.CT-Correct term considering reference
left ventricular collapse	Correct Term	a.CT-Correct term considering reference
forced expiratory volume	Correct Term	a.CT-Correct term considering reference
peripheral artery disease	Correct Term	a.CT-Correct term considering reference
palpitation	Correct Term	a.CT-Correct term considering reference
profound bradycardia	Correct Term	b.CT-Correct synonym
diameter of adjacent healthy reference segments	Incorrect	2.Incorrect term equivalent

normal sinus rhythm	Correct Term	a.CT-Correct term considering reference
atrial or ventricular extrasystoles	Incorrect	1.Partial term translation
significant fructose consumption	Incorrect	1.Partial term translation
inflammation	Incorrect	9.Term with incorrect grammatical category
structural or functional cardiac abnormalities	Correct Term	c.CT-Possible equivalent
iatrogenic	Correct Term	a.CT-Correct term considering reference
heart failure	Correct Term	a.CT-Correct term considering reference
thrombus	Incorrect	5.Source term insertion
sudden cardiac death	Correct Term	a.CT-Correct term considering reference
right ventricular myocardial acceleration	Incorrect	1.Partial term translation
migraine	Incorrect	2.Incorrect term equivalent
diabetic outcome	Correct Term	a.CT-Correct term considering reference
coronary artery disease	Correct Term	a.CT-Correct term considering reference
pre-syncope	Incorrect	2.Incorrect term equivalent
leadless pacemaker	Correct Term	a.CT-Correct term considering reference
oxidative stress	Correct Term	a.CT-Correct term considering reference
left ventricular function	Incorrect	1.Partial term translation
headache	Correct Term	a.CT-Correct term considering reference
chronic obstructive pulmonary disease	Correct Term	a.CT-Correct term considering reference
syncope	Correct Term	a.CT-Correct term considering reference
inflammation	Correct Term	a.CT-Correct term considering reference
diabetes mellitus	Incorrect	5.Source term insertion

ii. Term error annotation in PBMT-2

Term	**Binary Analysis**	**Term Error category**
acetaminophen	Correct Term	a.CT-Correct term considering reference
acute decompensated heart failure	Correct Term	a.CT-Correct term considering reference

acute exacerbation of idiopathic pulmonary fibrosis	Incorrect	1.Partial term translation
acute heart failure	Correct Term	a.CT-Correct term considering reference
acute kidney injury	Correct Term	b.CT-Correct synonym
acute myocardial infarction	Correct Term	a.CT-Correct term considering reference
acute myocarditis	Correct Term	a.CT-Correct term considering reference
AEMD parameters	Correct Term	b.CT-Correct synonym
analysis of thrombophilic gene mutations	Incorrect	1.Partial term translation
antiretroviral therapy	Correct Term	a.CT-Correct term considering reference
aortic cusps	Incorrect	1.Partial term translation
aortic stenosis	Correct Term	a.CT-Correct term considering reference
arrhythmia	Incorrect	5.Source term insertion
arterial thoracic outlet syndrome	Correct Term	a.CT-Correct term considering reference
asystole	Correct Term	a.CT-Correct term considering reference
atrial electromechanical delay	Correct Term	a.CT-Correct term considering reference
atrial fibrillation	Correct Term	a.CT-Correct term considering reference
atrial fibrillation	Correct Term	a.CT-Correct term considering reference
atrial fibrillation	Correct Term	a.CT-Correct term considering reference
atrial fibrillation	Correct Term	a.CT-Correct term considering reference
atrial fibrillation episodes	Incorrect	1.Partial term translation
atrial or ventricular extrasystoles	Incorrect	1.Partial term translation
atrioventricular block	Correct Term	a.CT-Correct term considering reference
atypical pneumonia	Correct Term	a.CT-Correct term considering reference
Brugada syndrome	Correct Term	a.CT-Correct term considering reference
cardiac arrest	Correct Term	a.CT-Correct term considering reference
cardiac emergencies	Correct Term	a.CT-Correct term considering reference
cardiac surgery	Incorrect	1.Partial term translation
cardiac surgery	Correct Term	a.CT-Correct term considering reference
cardiac tamponade	Incorrect	1.Partial term translation

cardiomyopathy	Correct Term	a.CT-Correct term considering reference
cardiomyopathy	Correct Term	a.CT-Correct term considering reference
cardiovascular	Correct Term	a.CT-Correct term considering reference
cardiovascular disease	Correct Term	a.CT-Correct term considering reference
cardiovascular outcome	Incorrect	1.Partial term translation
carotid artery PA	Correct Term	a.CT-Correct term considering reference
carotid artery stenosis	Incorrect	1.Partial term translation
carotid artery stenting	Incorrect	1.Partial term translation
Carotid artery stenting	Correct Term	a.CT-Correct term considering reference
carotid stenting	Incorrect	8.Term extended
cellular ischemia	Correct Term	a.CT-Correct term considering reference
CETP gene	Correct Term	a.CT-Correct term considering reference
CETP gene rs289714 polymorphism	Correct Term	b.CT-Correct synonym
chronic abdominal pain	Correct Term	a.CT-Correct term considering reference
chronic fatigue syndrome	Correct Term	a.CT-Correct term considering reference
chronic kidney disease	Correct Term	a.CT-Correct term considering reference
chronic obstructive pulmonary disease	Correct Term	a.CT-Correct term considering reference
chronic obstructive pulmonary disease	Correct Term	a.CT-Correct term considering reference
chronic obstructive pulmonary disease	Correct Term	a.CT-Correct term considering reference
clinical and angiographic characteristics	Correct Term	a.CT-Correct term considering reference
concurrent obstructive aortic and mitral prosthetic valve thrombosis	Incorrect	1.Partial term translation
congenital channelopathy	Incorrect	1.Partial term translation
consultant invasive neuroradiologist	Incorrect	1.Partial term translation
convulsion	Incorrect	5.Source term insertion
coronary angiography	Correct Term	a.CT-Correct term considering reference
coronary artery	Correct Term	a.CT-Correct term considering reference
coronary artery bypass graft operation	Incorrect	1.Partial term translation
coronary artery disease	Correct Term	a.CT-Correct term considering reference

coronary artery disease	Correct Term	a.CT-Correct term considering reference
coronary artery disease	Correct Term	a.CT-Correct term considering reference
coronary artery ectasia	Incorrect	1.Partial term translation
coronary artery ectasia	Correct Term	a.CT-Correct term considering reference
coronary artery lumen	Incorrect	1.Partial term translation
coronary cusp	Incorrect	1.Partial term translation
coronary intervention	Correct Term	a.CT-Correct term considering reference
coronary slow-flow phenomenon	Incorrect	1.Partial term translation
coronary slow-flow phenomenon	Incorrect	3.Morphosyntactic error
Coronary stent embolism	Correct Term	b.CT-Correct synonym
COVID-19	Correct Term	a.CT-Correct term considering reference
COVID-19 pandemic	Incorrect	1.Partial term translation
cumulative LDL exposure	Incorrect	1.Partial term translation
CVD event	Correct Term	b.CT-Correct synonym
diabetes mellitus	Correct Term	a.CT-Correct term considering reference
diabetes mellitus	Incorrect	10.Literal Translation
diabetic	Correct Term	a.CT-Correct term considering reference
diabetic outcome	Correct Term	a.CT-Correct term considering reference
diagnosis of TTR-CA	Correct Term	b.CT-Correct synonym
diameter of adjacent healthy reference segments	Incorrect	2.Incorrect term equivalent
distal vascular bed	Incorrect	9.Term with incorrect grammatical category
dyspnea	Correct Term	a.CT-Correct term considering reference
early coronary artery involvement	Correct Term	a.CT-Correct term considering reference
early coronary artery involvement	Correct Term	a.CT-Correct term considering reference
echocardiographic changes	Correct Term	a.CT-Correct term considering reference
echocardiographic changes	Correct Term	a.CT-Correct term considering reference
ejection fraction	Correct Term	a.CT-Correct term considering reference
Electrocardiography	Incorrect	2.Incorrect term equivalent
endothelial function	Correct Term	a.CT-Correct term considering reference

epinephrine provocation test	Incorrect	9.Term with incorrect grammatical category
episode of hypoxic attack	Incorrect	1.Partial term translation
fast idioventricular rhythm	Incorrect	7.Term drop
fibrate	Correct Term	a.CT-Correct term considering reference
fibrate treatment	Incorrect	1.Partial term translation
fluids and electrolyte replacement therapy	Incorrect	4.Reordering error
fluoroscopic scanning	Incorrect	1.Partial term translation
forced expiratory volume	Correct Term	a.CT-Correct term considering reference
headache	Correct Term	a.CT-Correct term considering reference
heart failure	Correct Term	a.CT-Correct term considering reference
heart failure	Correct Term	a.CT-Correct term considering reference
heart failure	Correct Term	a.CT-Correct term considering reference
heart failure	Correct Term	a.CT-Correct term considering reference
heart failure	Correct Term	a.CT-Correct term considering reference
heart failure	Correct Term	a.CT-Correct term considering reference
heart failure	Correct Term	a.CT-Correct term considering reference
heart transplantation	Correct Term	a.CT-Correct term considering reference
high lipoprotein	Incorrect	3.Morphosyntactic error
high lipoprotein (a)	Incorrect	1.Partial term translation
high-risk transcatheter valve replacement procedures	Incorrect	4.Reordering error
histopathological feature	Incorrect	1.Partial term translation
history of sudden cardiac death	Correct Term	a.CT-Correct term considering reference
HIV-seropositive	Incorrect	1.Partial term translation
human immun-deficiency virus	Correct Term	c.CT-Possible equivalent
hypercholesterolemia	Correct Term	a.CT-Correct term considering reference
hypermobile joint	Incorrect	1.Partial term translation
hypertriglyceridemia	Correct Term	a.CT-Correct term considering reference
hypotension	Correct Term	a.CT-Correct term considering reference

hypotension	Correct Term	a.CT-Correct term considering reference
hypoxia	Incorrect	7.Term drop
iatrogenic	Correct Term	a.CT-Correct term considering reference
ibuprofen	Correct Term	a.CT-Correct term considering reference
implantation	Incorrect	7.Term drop
inappropriate sinus tachycardia	Correct Term	a.CT-Correct term considering reference
infective endocarditis	Correct Term	a.CT-Correct term considering reference
inferior vena cava dilatation	Incorrect	4.Reordering error
inflammation	Incorrect	5.Source term insertion
inflammation	Correct Term	a.CT-Correct term considering reference
inflammation	Correct Term	a.CT-Correct term considering reference
intrauterine environment	Correct Term	a.CT-Correct term considering reference
isovolumic contraction	Incorrect	1.Partial term translation
LAA thrombus	Incorrect	1.Partial term translation
LAA thrombus (+) group	Correct Term	a.CT-Correct term considering reference
leadless pacemaker	Incorrect	9.Term with incorrect grammatical category
leadless pacemakers	Correct Term	a.CT-Correct term considering reference
left atrial systolic functions	Correct Term	a.CT-Correct term considering reference
left heart function	Incorrect	1.Partial term translation
left ventricular assist device	Correct Term	a.CT-Correct term considering reference
left ventricular collapse	Correct Term	a.CT-Correct term considering reference
left ventricular function	Incorrect	1.Partial term translation
left ventricular function	Incorrect	1.Partial term translation
leukocyte	Correct Term	a.CT-Correct term considering reference
leukocyte	Correct Term	a.CT-Correct term considering reference
lightheadedness	Correct Term	b.CT-Correct synonym
low cardiac output	Correct Term	a.CT-Correct term considering reference
low cardiac output syndrome	Correct Term	a.CT-Correct term considering reference

low cardiac output syndrome	Correct Term	a.CT-Correct term considering reference
lymphocyte concentration	Incorrect	1.Partial term translation
major epicardial coronary artery	Correct Term	a.CT-Correct term considering reference
major vascular complications	Correct Term	a.CT-Correct term considering reference
malignant ventricular arrhythmia	Correct Term	a.CT-Correct term considering reference
migraine	Correct Term	a.CT-Correct term considering reference
mitral valve-in-valve procedure	Incorrect	11.Other term errors
multiple saphenous vein graft aneurysm	Correct Term	a.CT-Correct term considering reference
myocardial performance index	Correct Term	a.CT-Correct term considering reference
myopathy	Correct Term	a.CT-Correct term considering reference
nonsustained ventricular tachycardia attacks	Correct Term	a.CT-Correct term considering reference
normal sinus rhythm	Correct Term	a.CT-Correct term considering reference
obstructive sleep apnea	Correct Term	a.CT-Correct term considering reference
outflow tract ventricular arrhythmias	Correct Term	a.CT-Correct term considering reference
oxidative stress	Correct Term	a.CT-Correct term considering reference
P wave dispersion	Correct Term	a.CT-Correct term considering reference
P wave dispersion	Incorrect	1.Partial term translation
palpitation	Correct Term	a.CT-Correct term considering reference
palpitation	Correct Term	a.CT-Correct term considering reference
patent ductus arteriosus	Incorrect	1.Partial term translation
pathophysiology	Correct Term	a.CT-Correct term considering reference
perfusion	Correct Term	a.CT-Correct term considering reference
peripheral artery disease	Correct Term	a.CT-Correct term considering reference
peripheral embolism	Incorrect	1.Partial term translation
post-infarction aneurysm of left ventricle	Correct Term	b.CT-Correct synonym
postural orthostatic tachycardia syndrome	Correct Term	a.CT-Correct term considering reference
pre-syncope	Correct Term	a.CT-Correct term considering reference

premature ventricular complexes	Incorrect	1.Partial term translation
premature ventricular contraction	Correct Term	b.CT-Correct synonym
preserved ejection fraction	Correct Term	a.CT-Correct term considering reference
preserved ejection fraction	Correct Term	a.CT-Correct term considering reference
preserved ejection fraction	Correct Term	a.CT-Correct term considering reference
primary PCI	Correct Term	b.CT-Correct synonym
primary percutaneous coronary intervention	Correct Term	a.CT-Correct term considering reference
primary percutaneous coronary intervention	Correct Term	a.CT-Correct term considering reference
profound bradycardia	Correct Term	b.CT-Correct synonym
prognostic nutritional index	Correct Term	a.CT-Correct term considering reference
prognostic nutritional index	Correct Term	a.CT-Correct term considering reference
prosthetic heart valve	Incorrect	1.Partial term translation
prosthetic heart valve thrombosis	Correct Term	a.CT-Correct term considering reference
Prosthetic valve thrombosis	Correct Term	a.CT-Correct term considering reference
pseudoaneurysm	Correct Term	a.CT-Correct term considering reference
radiofrequency ablation	Correct Term	a.CT-Correct term considering reference
rest dyspnea	Correct Term	b.CT-Correct synonym
resting heart rate	Correct Term	a.CT-Correct term considering reference
rhabdomyolysis	Correct Term	a.CT-Correct term considering reference
right heart function	Incorrect	2.Incorrect term equivalent
right posterior cerebral artery	Correct Term	a.CT-Correct term considering reference
right posterior cerebral artery	Correct Term	a.CT-Correct term considering reference
right ventricle	Correct Term	a.CT-Correct term considering reference
right ventricular dysfunction	Correct Term	a.CT-Correct term considering reference
right ventricular myocardial acceleration	Incorrect	1.Partial term translation
rs289714 variation	Incorrect	1.Partial term translation
scintigraphy	Incorrect	5.Source term insertion
serum albumin	Correct Term	a.CT-Correct term considering reference

serum lactate level	Correct Term	a.CT-Correct term considering reference
serum lipid profile	Correct Term	a.CT-Correct term considering reference
significant fructose consumption	Correct Term	b.CT-Correct synonym
sinus rhythm	Correct Term	a.CT-Correct term considering reference
sinus rhythm	Correct Term	a.CT-Correct term considering reference
slow coronary flow	Correct Term	a.CT-Correct term considering reference
Sodium glucose co-transporter 2 inhibitors	Correct Term	a.CT-Correct term considering reference
spontaneous aortic root rupture	Correct Term	a.CT-Correct term considering reference
spontaneous infection	Correct Term	a.CT-Correct term considering reference
ST-elevation myocardial infarction	Correct Term	a.CT-Correct term considering reference
ST-segment elevation myocardial infarction	Correct Term	b.CT-Correct synonym
ST-segment elevation myocardial infarction	Correct Term	a.CT-Correct term considering reference
ST-segment myocardial infarction	Correct Term	a.CT-Correct term considering reference
St. Jude Medical Biocor bovine bioprosthesis in mitral position	Incorrect	1.Partial term translation
stable angina pectoris	Correct Term	a.CT-Correct term considering reference
structural or functional cardiac abnormalities	Correct Term	c.CT-Possible equivalent
sudden cardiac death	Correct Term	a.CT-Correct term considering reference
sudden cardiac death	Correct Term	a.CT-Correct term considering reference
syncope	Correct Term	a.CT-Correct term considering reference
tachycardia	Correct Term	a.CT-Correct term considering reference
thrombolytic therapy	Correct Term	a.CT-Correct term considering reference
thrombus	Correct Term	a.CT-Correct term considering reference
thrombus	Correct Term	a.CT-Correct term considering reference
thrombus	Correct Term	a.CT-Correct term considering reference
thrombus	Incorrect	5.Source term insertion
thrombus	Incorrect	5.Source term insertion
transcatheter aortic valve replacement	Correct Term	a.CT-Correct term considering reference

transcatheter atrial septal defect	Correct Term	a.CT-Correct term considering reference
transcatheter closure of an ASD	Correct Term	b.CT-Correct synonym
transthoracic echocardiography	Correct Term	a.CT-Correct term considering reference
transvenous pacemakers	Incorrect	1.Partial term translation
vascular system	Correct Term	a.CT-Correct term considering reference
vasculitis	Correct Term	a.CT-Correct term considering reference

iii. Term error annotation in NMT-1

Term	Binary Analysis	Term Error category
acetaminophen	Incorrect	7.Term drop
acute decompensated heart failure	Incorrect	1.Partial term translation
acute exacerbation of idiopathic pulmonary fibrosis	Incorrect	2.Incorrect term equivalent
acute heart failure	Correct Term	a.CT-Correct term considering reference
acute kidney injury	Incorrect	1.Partial term translation
acute myocardial infarction	Correct Term	a.CT-Correct term considering reference
acute myocarditis	Incorrect	7.Term drop
AEMD parameters	Incorrect	7.Term drop
analysis of thrombophilic gene mutations	Incorrect	7.Term drop
antiretroviral therapy	Incorrect	7.Term drop
aortic cusps	Incorrect	7.Term drop
aortic stenosis	Incorrect	7.Term drop
arrhythmia	Incorrect	7.Term drop
arterial thoracic outlet syndrome	Incorrect	7.Term drop
asystole	Incorrect	7.Term drop
atrial electromechanical delay	Incorrect	7.Term drop
atrial fibrillation	Correct Term	a.CT-Correct term considering reference
atrial fibrillation	Correct Term	a.CT-Correct term considering reference
atrial fibrillation	Incorrect	8.Term extended
atrial fibrillation	Correct Term	a.CT-Correct term considering reference

atrial fibrillation episodes	Incorrect	1.Partial term translation
atrial or ventricular extrasystoles	Incorrect	11.Other term errors
atrioventricular block	Correct Term	a.CT-Correct term considering reference
atypical pneumonia	Incorrect	7.Term drop
Brugada syndrome	Correct Term	a.CT-Correct term considering reference
cardiac arrest	Correct Term	a.CT-Correct term considering reference
cardiac emergencies	Correct Term	c.CT-Possible equivalent
cardiac surgery	Correct Term	a.CT-Correct term considering reference
cardiac surgery	Incorrect	7.Term drop
cardiac tamponade	Incorrect	7.Term drop
cardiomyopathy	Incorrect	7.Term drop
cardiomyopathy	Incorrect	7.Term drop
cardiovascular	Correct Term	a.CT-Correct term considering reference
cardiovascular disease	Correct Term	a.CT-Correct term considering reference
cardiovascular outcome	Incorrect	1.Partial term translation
carotid artery PA	Incorrect	1.Partial term translation
carotid artery stenosis	Incorrect	1.Partial term translation
carotid artery stenting	Incorrect	1.Partial term translation
Carotid artery stenting	Incorrect	1.Partial term translation
carotid stenting	Incorrect	7.Term drop
cellular ischemia	Incorrect	1.Partial term translation
CETP gene	Incorrect	7.Term drop
CETP gene rs289714 polymorphism	Incorrect	7.Term drop
chronic abdominal pain	Incorrect	7.Term drop
chronic fatigue syndrome	Incorrect	7.Term drop
chronic kidney disease	Correct Term	b.CT-Correct synonym
chronic obstructive pulmonary disease	Incorrect	11.Other term errors
chronic obstructive pulmonary disease	Incorrect	7.Term drop
chronic obstructive pulmonary disease	Incorrect	11.Other term errors
clinical and angiographic characteristics	Correct Term	a.CT-Correct term considering reference
concurrent obstructive aortic and mitral prosthetic valve thrombosis	Incorrect	11.Other term errors
congenital channelopathy	Incorrect	1.Partial term translation

consultant invasive neuroradiologist	Incorrect	7.Term drop
convulsion	Incorrect	7.Term drop
coronary angiography	Correct Term	a.CT-Correct term considering reference
coronary artery	Incorrect	7.Term drop
coronary artery bypass graft operation	Incorrect	1.Partial term translation
coronary artery disease	Correct Term	a.CT-Correct term considering reference
coronary artery disease	Correct Term	a.CT-Correct term considering reference
coronary artery disease	Correct Term	a.CT-Correct term considering reference
coronary artery ectasia	Incorrect	1.Partial term translation
coronary artery ectasia	Correct Term	a.CT-Correct term considering reference
coronary artery lumen	Incorrect	1.Partial term translation
coronary cusp	Incorrect	1.Partial term translation
coronary intervention	Incorrect	2.Incorrect term equivalent
coronary slow-flow phenomenon	Incorrect	1.Partial term translation
coronary slow-flow phenomenon	Correct Term	c.CT-Possible equivalent
Coronary stent embolism	Incorrect	7.Term drop
COVID-19	Incorrect	7.Term drop
COVID-19 pandemic	Incorrect	7.Term drop
cumulative LDL exposure	Incorrect	7.Term drop
CVD event	Incorrect	7.Term drop
diabetes mellitus	Correct Term	a.CT-Correct term considering reference
diabetes mellitus	Correct Term	a.CT-Correct term considering reference
diabetic	Incorrect	7.Term drop
diabetic outcome	Incorrect	7.Term drop
diagnosis of TTR-CA	Incorrect	7.Term drop
diameter of adjacent healthy reference segments	Incorrect	7.Term drop
distal vascular bed	Incorrect	7.Term drop
dyspnea	Incorrect	7.Term drop
early coronary artery involvement	Incorrect	1.Partial term translation
early coronary artery involvement	Incorrect	1.Partial term translation
echocardiographic changes	Incorrect	7.Term drop
echocardiographic changes	Incorrect	7.Term drop

ejection fraction	Correct Term	a.CT-Correct term considering reference
Electrocardiography	Incorrect	2.Incorrect term equivalent
endothelial function	Incorrect	7.Term drop
epinephrine provocation test	Incorrect	7.Term drop
episode of hypoxic attack	Incorrect	7.Term drop
fast idioventricular rhythm	Incorrect	7.Term drop
fibrate	Incorrect	7.Term drop
fibrate treatment	Incorrect	1.Partial term translation
fluids and electrolyte replacement therapy	Incorrect	7.Term drop
fluoroscopic scanning	Incorrect	7.Term drop
forced expiratory volume	Incorrect	7.Term drop
headache	Incorrect	7.Term drop
heart failure	Correct Term	a.CT-Correct term considering reference
heart failure	Correct Term	a.CT-Correct term considering reference
heart failure	Incorrect	8.Term extended
heart failure	Incorrect	2.Incorrect term equivalent
heart failure	Correct Term	a.CT-Correct term considering reference
heart failure	Correct Term	a.CT-Correct term considering reference
heart failure	Correct Term	a.CT-Correct term considering reference
heart transplantation	Incorrect	7.Term drop
high lipoprotein	Incorrect	3.Morphosyntactic error
high lipoprotein (a)	Incorrect	3.Morphosyntactic error
high-risk transcatheter valve replacement procedures	Incorrect	11.Other term errors
histopathological feature	Incorrect	7.Term drop
history of sudden cardiac death	Incorrect	7.Term drop
HIV-seropositive	Incorrect	7.Term drop
human immun-deficiency virus	Incorrect	7.Term drop
hypercholesterolemia	Incorrect	7.Term drop
hypermobile joint	Incorrect	7.Term drop
hypertriglyceridemia	Incorrect	7.Term drop
hypotension	Incorrect	7.Term drop
hypotension	Incorrect	7.Term drop
hypoxia	Incorrect	7.Term drop

iatrogenic	Incorrect	7.Term drop
ibuprofen	Incorrect	7.Term drop
implantation	Incorrect	7.Term drop
inappropriate sinus tachycardia	Incorrect	7.Term drop
infective endocarditis	Incorrect	1.Partial term translation
inferior vena cava dilatation	Incorrect	1.Partial term translation
inflammation	Incorrect	7.Term drop
inflammation	Incorrect	7.Term drop
inflammation	Incorrect	7.Term drop
intrauterine environment	Incorrect	7.Term drop
isovolumic contraction	Incorrect	7.Term drop
LAA thrombus	Incorrect	7.Term drop
LAA thrombus (+) group	Incorrect	7.Term drop
leadless pacemaker	Incorrect	7.Term drop
leadless pacemakers	Incorrect	7.Term drop
left atrial systolic functions	Incorrect	4.Reordering error
left heart function	Incorrect	7.Term drop
left ventricular assist device	Correct Term	a.CT-Correct term considering reference
left ventricular collapse	Incorrect	1.Partial term translation
left ventricular function	Correct Term	a.CT-Correct term considering reference
left ventricular function	Incorrect	7.Term drop
leukocyte	Incorrect	7.Term drop
leukocyte	Incorrect	7.Term drop
lightheadedness	Correct Term	b.CT-Correct synonym
low cardiac output	Incorrect	1.Partial term translation
low cardiac output syndrome	Incorrect	11.Other term errors
low cardiac output syndrome	Incorrect	7.Term drop
lymphocyte concentration	Incorrect	1.Partial term translation
major epicardial coronary artery	Incorrect	1.Partial term translation
major vascular complications	Incorrect	8.Term extended
malignant ventricular arrhythmia	Incorrect	7.Term drop
migraine	Incorrect	7.Term drop
mitral valve-in-valve procedure	Incorrect	7.Term drop
multiple saphenous vein graft aneurysm	Incorrect	1.Partial term translation

myocardial performance index	Incorrect	11.Other term errors
myopathy	Incorrect	7.Term drop
nonsustained ventricular tachycardia attacks	Incorrect	1.Partial term translation
normal sinus rhythm	Incorrect	7.Term drop
obstructive sleep apnea	Incorrect	7.Term drop
outflow tract ventricular arrhythmias	Incorrect	1.Partial term translation
oxidative stress	Incorrect	7.Term drop
P wave dispersion	Incorrect	7.Term drop
P wave dispersion	Incorrect	7.Term drop
palpitation	Incorrect	7.Term drop
palpitation	Incorrect	7.Term drop
patent ductus arteriosus	Incorrect	7.Term drop
pathophysiology	Incorrect	7.Term drop
perfusion	Incorrect	7.Term drop
peripheral artery disease	Incorrect	1.Partial term translation
peripheral embolism	Incorrect	7.Term drop
post-infarction aneurysm of left ventricle	Incorrect	1.Partial term translation
postural orthostatic tachycardia syndrome	Incorrect	2.Incorrect term equivalent
pre-syncope	Incorrect	7.Term drop
premature ventricular complexes	Incorrect	7.Term drop
premature ventricular contraction	Incorrect	11.Other term errors
preserved ejection fraction	Incorrect	7.Term drop
preserved ejection fraction	Incorrect	1.Partial term translation
preserved ejection fraction	Incorrect	7.Term drop
primary PCI	Incorrect	1.Partial term translation
primary percutaneous coronary intervention	Correct Term	a.CT-Correct term considering reference
primary percutaneous coronary intervention	Correct Term	a.CT-Correct term considering reference
profound bradycardia	Incorrect	7.Term drop
prognostic nutritional index	Incorrect	7.Term drop
prognostic nutritional index	Incorrect	7.Term drop
prosthetic heart valve	Incorrect	2.Incorrect term equivalent
prosthetic heart valve thrombosis	Incorrect	7.Term drop
Prosthetic valve thrombosis	Correct Term	a.CT-Correct term considering reference
pseudoaneurysm	Incorrect	7.Term drop

radiofrequency ablation	Incorrect	7.Term drop
rest dyspnea	Incorrect	1.Partial term translation
resting heart rate	Incorrect	1.Partial term translation
rhabdomyolysis	Incorrect	7.Term drop
right heart function	Incorrect	2.Incorrect term equivalent
right posterior cerebral artery	Incorrect	7.Term drop
right posterior cerebral artery	Incorrect	2.Incorrect term equivalent
right ventricle	Incorrect	11.Other term errors
right ventricular dysfunction	Incorrect	11.Other term errors
right ventricular myocardial acceleration	Incorrect	11.Other term errors
rs289714 variation	Incorrect	7.Term drop
scintigraphy	Incorrect	7.Term drop
serum albumin	Incorrect	2.Incorrect term equivalent
serum lactate level	Incorrect	1.Partial term translation
serum lipid profile	Incorrect	7.Term drop
significant fructose consumption	Incorrect	7.Term drop
sinus rhythm	Incorrect	7.Term drop
sinus rhythm	Incorrect	7.Term drop
slow coronary flow	Incorrect	2.Incorrect term equivalent
Sodium glucose co-transporter 2 inhibitors	Incorrect	7.Term drop
spontaneous aortic root rupture	Incorrect	2.Incorrect term equivalent
spontaneous infection	Incorrect	7.Term drop
ST-elevation myocardial infarction	Correct Term	a.CT-Correct term considering reference
ST-segment elevation myocardial infarction	Correct Term	a.CT-Correct term considering reference
ST-segment elevation myocardial infarction	Incorrect	7.Term drop
ST-segment myocardial infarction	Correct Term	a.CT-Correct term considering reference
St. Jude Medical Biocor bovine bioprosthesis in mitral position	Incorrect	7.Term drop
stable angina pectoris	Correct Term	a.CT-Correct term considering reference
structural or functional cardiac abnormalities	Incorrect	7.Term drop
sudden cardiac death	Correct Term	a.CT-Correct term considering reference
sudden cardiac death	Correct Term	a.CT-Correct term considering reference
syncope	Incorrect	7.Term drop

tachycardia	Correct Term	a.CT-Correct term considering reference
thrombolytic therapy	Correct Term	a.CT-Correct term considering reference
thrombus	Incorrect	7.Term drop
thrombus	Incorrect	7.Term drop
thrombus	Incorrect	7.Term drop
thrombus	Incorrect	7.Term drop
thrombus	Incorrect	7.Term drop
transcatheter aortic valve replacement	Incorrect	1.Partial term translation
transcatheter atrial septal defect	Incorrect	1.Partial term translation
transcatheter closure of an ASD	Incorrect	7.Term drop
transthoracic echocardiography	Incorrect	7.Term drop
transvenous pacemakers	Incorrect	7.Term drop
vascular system	Incorrect	7.Term drop
vasculitis	Incorrect	7.Term drop

iv. Term error annotation in NMT-2

Term	**Binary Analysis**	**Term Error category**
acetaminophen	Incorrect	7.Term drop
acute decompensated heart failure	Correct Term	a.CT-Correct term considering reference
acute exacerbation of idiopathic pulmonary fibrosis	Incorrect	1.Partial term translation
acute heart failure	Correct Term	a.CT-Correct term considering reference
acute kidney injury	Correct Term	a.CT-Correct term considering reference
acute myocardial infarction	Incorrect	7.Term drop
acute myocarditis	Correct Term	a.CT-Correct term considering reference
AEMD parameters	Correct Term	b.CT-Correct synonym
analysis of thrombophilic gene mutations	Incorrect	1.Partial term translation
antiretroviral therapy	Incorrect	2.Incorrect term equivalent
aortic cusps	Incorrect	10.Literal Translation
aortic stenosis	Correct Term	a.CT-Correct term considering reference

arrhythmia	Correct Term	a.CT-Correct term considering reference
arterial thoracic outlet syndrome	Correct Term	a.CT-Correct term considering reference
asystole	Incorrect	7.Term drop
atrial electromechanical delay	Correct Term	a.CT-Correct term considering reference
atrial fibrillation	Incorrect	7.Term drop
atrial fibrillation	Correct Term	b.CT-Correct synonym
atrial fibrillation	Correct Term	a.CT-Correct term considering reference
atrial fibrillation	Correct Term	a.CT-Correct term considering reference
atrial fibrillation episodes	Correct Term	b.CT-Correct synonym
atrial or ventricular extrasystoles	Correct Term	b.CT-Correct synonym
atrioventricular block	Correct Term	a.CT-Correct term considering reference
atypical pneumonia	Correct Term	a.CT-Correct term considering reference
Brugada syndrome	Correct Term	a.CT-Correct term considering reference
cardiac arrest	Correct Term	a.CT-Correct term considering reference
cardiac emergencies	Correct Term	b.CT-Correct synonym
cardiac surgery	Correct Term	a.CT-Correct term considering reference
cardiac surgery	Correct Term	a.CT-Correct term considering reference
cardiac tamponade	Correct Term	a.CT-Correct term considering reference
cardiomyopathy	Correct Term	a.CT-Correct term considering reference
cardiomyopathy	Incorrect	7.Term drop
cardiovascular	Incorrect	7.Term drop
cardiovascular disease	Correct Term	a.CT-Correct term considering reference
cardiovascular outcome	Incorrect	10.Literal Translation
carotid artery PA	Correct Term	a.CT-Correct term considering reference
carotid artery stenosis	Correct Term	a.CT-Correct term considering reference
carotid artery stenting	Correct Term	a.CT-Correct term considering reference
Carotid artery stenting	Correct Term	a.CT-Correct term considering reference
carotid stenting	Incorrect	3.Morphosyntactic error

cellular ischemia	Correct Term	a.CT-Correct term considering reference
CETP gene	Correct Term	a.CT-Correct term considering reference
CETP gene rs289714 polymorphism	Correct Term	b.CT-Correct synonym
chronic abdominal pain	Correct Term	a.CT-Correct term considering reference
chronic fatigue syndrome	Correct Term	a.CT-Correct term considering reference
chronic kidney disease	Correct Term	b.CT-Correct synonym
chronic obstructive pulmonary disease	Correct Term	a.CT-Correct term considering reference
chronic obstructive pulmonary disease	Correct Term	a.CT-Correct term considering reference
chronic obstructive pulmonary disease	Correct Term	a.CT-Correct term considering reference
clinical and angiographic characteristics	Correct Term	b.CT-Correct synonym
concurrent obstructive aortic and mitral prosthetic valve thrombosis	Incorrect	1.Partial term translation
congenital channelopathy	Incorrect	1.Partial term translation
consultant invasive neuroradiologist	Incorrect	7.Term drop
convulsion	Incorrect	7.Term drop
coronary angiography	Correct Term	a.CT-Correct term considering reference
coronary artery	Correct Term	a.CT-Correct term considering reference
coronary artery bypass graft operation	Correct Term	c.CT-Possible equivalent
coronary artery disease	Correct Term	a.CT-Correct term considering reference
coronary artery disease	Correct Term	a.CT-Correct term considering reference
coronary artery disease	Correct Term	a.CT-Correct term considering reference
coronary artery ectasia	Correct Term	a.CT-Correct term considering reference
coronary artery ectasia	Correct Term	a.CT-Correct term considering reference
coronary artery lumen	Incorrect	1.Partial term translation
coronary cusp	Incorrect	2.Incorrect term equivalent
coronary intervention	Correct Term	a.CT-Correct term considering reference
coronary slow-flow phenomenon	Correct Term	b.CT-Correct synonym
coronary slow-flow phenomenon	Correct Term	a.CT-Correct term considering reference
Coronary stent embolism	Incorrect	1.Partial term translation
COVID-19	Incorrect	7.Term drop

COVID-19 pandemic	Incorrect	7.Term drop
cumulative LDL exposure	Incorrect	7.Term drop
CVD event	Correct Term	b.CT-Correct synonym
diabetes mellitus	Correct Term	a.CT-Correct term considering reference
diabetes mellitus	Incorrect	10.Literal Translation
diabetic	Correct Term	a.CT-Correct term considering reference
diabetic outcome	Incorrect	10.Literal Translation
diagnosis of TTR-CA	Incorrect	4.Reordering error
diameter of adjacent healthy reference segments	Incorrect	7.Term drop
distal vascular bed	Incorrect	1.Partial term translation
dyspnea	Correct Term	a.CT-Correct term considering reference
early coronary artery involvement	Incorrect	7.Term drop
early coronary artery involvement	Correct Term	a.CT-Correct term considering reference
echocardiographic changes	Correct Term	a.CT-Correct term considering reference
echocardiographic changes	Correct Term	a.CT-Correct term considering reference
ejection fraction	Correct Term	a.CT-Correct term considering reference
Electrocardiography	Incorrect	2.Incorrect term equivalent
endothelial function	Incorrect	7.Term drop
epinephrine provocation test	Incorrect	7.Term drop
episode of hypoxic attack	Incorrect	1.Partial term translation
fast idioventricular rhythm	Correct Term	b.CT-Correct synonym
fibrate	Correct Term	a.CT-Correct term considering reference
fibrate treatment	Correct Term	b.CT-Correct synonym
fluids and electrolyte replacement therapy	Correct Term	b.CT-Correct synonym
fluoroscopic scanning	Incorrect	7.Term drop
forced expiratory volume	Correct Term	c.CT-Possible equivalent
headache	Incorrect	7.Term drop
heart failure	Correct Term	a.CT-Correct term considering reference
heart failure	Correct Term	a.CT-Correct term considering reference
heart failure	Incorrect	7.Term drop
heart failure	Correct Term	a.CT-Correct term considering reference

heart failure	Incorrect	7.Term drop
heart failure	Correct Term	a.CT-Correct term considering reference
heart failure	Incorrect	8.Term extended
heart transplantation	Incorrect	2.Incorrect term equivalent
high lipoprotein	Correct Term	a.CT-Correct term considering reference
high lipoprotein (a)	Incorrect	7.Term drop
high-risk transcatheter valve replacement procedures	Correct Term	a.CT-Correct term considering reference
histopathological feature	Incorrect	7.Term drop
history of sudden cardiac death	Incorrect	3.Morphosyntactic error
HIV-seropositive	Incorrect	1.Partial term translation
human immun-deficiency virus	Correct Term	c.CT-Possible equivalent
hypercholesterolemia	Correct Term	a.CT-Correct term considering reference
hypermobile joint	Incorrect	7.Term drop
hypertriglyceridemia	Incorrect	7.Term drop
hypotension	Correct Term	a.CT-Correct term considering reference
hypotension	Correct Term	a.CT-Correct term considering reference
hypoxia	Correct Term	a.CT-Correct term considering reference
iatrogenic	Correct Term	a.CT-Correct term considering reference
ibuprofen	Incorrect	7.Term drop
implantation	Correct Term	a.CT-Correct term considering reference
inappropriate sinus tachycardia	Incorrect	2.Incorrect term equivalent
infective endocarditis	Correct Term	a.CT-Correct term considering reference
inferior vena cava dilatation	Incorrect	1.Partial term translation
inflammation	Incorrect	7.Term drop
inflammation	Incorrect	7.Term drop
inflammation	Correct Term	a.CT-Correct term considering reference
intrauterine environment	Incorrect	7.Term drop
isovolumic contraction	Incorrect	7.Term drop
LAA thrombus	Correct Term	a.CT-Correct term considering reference
LAA thrombus (+) group	Correct Term	a.CT-Correct term considering reference

leadless pacemaker	Incorrect	7.Term drop
leadless pacemakers	Incorrect	1.Partial term translation
left atrial systolic functions	Incorrect	4.Reordering error
left heart function	Correct Term	a.CT-Correct term considering reference
left ventricular assist device	Correct Term	a.CT-Correct term considering reference
left ventricular collapse	Incorrect	1.Partial term translation
left ventricular function	Correct Term	a.CT-Correct term considering reference
left ventricular function	Correct Term	a.CT-Correct term considering reference
leukocyte	Correct Term	a.CT-Correct term considering reference
leukocyte	Correct Term	a.CT-Correct term considering reference
lightheadedness	Correct Term	b.CT-Correct synonym
low cardiac output	Correct Term	a.CT-Correct term considering reference
low cardiac output syndrome	Correct Term	a.CT-Correct term considering reference
low cardiac output syndrome	Correct Term	a.CT-Correct term considering reference
lymphocyte concentration	Correct Term	a.CT-Correct term considering reference
major epicardial coronary artery	Incorrect	1.Partial term translation
major vascular complications	Correct Term	a.CT-Correct term considering reference
malignant ventricular arrhythmia	Incorrect	1.Partial term translation
migraine	Correct Term	a.CT-Correct term considering reference
mitral valve-in-valve procedure	Incorrect	11.Other term errors
multiple saphenous vein graft aneurysm	Incorrect	1.Partial term translation
myocardial performance index	Correct Term	a.CT-Correct term considering reference
myopathy	Correct Term	a.CT-Correct term considering reference
nonsustained ventricular tachycardia attacks	Incorrect	7.Term drop
normal sinus rhythm	Correct Term	a.CT-Correct term considering reference
obstructive sleep apnea	Correct Term	a.CT-Correct term considering reference
outflow tract ventricular arrhythmias	Incorrect	4.Reordering error
oxidative stress	Correct Term	a.CT-Correct term considering reference

P wave dispersion	Correct Term	a.CT-Correct term considering reference
P wave dispersion	Correct Term	a.CT-Correct term considering reference
palpitation	Correct Term	a.CT-Correct term considering reference
palpitation	Correct Term	a.CT-Correct term considering reference
patent ductus arteriosus	Correct Term	a.CT-Correct term considering reference
pathophysiology	Correct Term	a.CT-Correct term considering reference
perfusion	Correct Term	a.CT-Correct term considering reference
peripheral artery disease	Correct Term	b.CT-Correct synonym
peripheral embolism	Correct Term	a.CT-Correct term considering reference
post-infarction aneurysm of left ventricle	Incorrect	11.Other term errors
postural orthostatic tachycardia syndrome	Correct Term	a.CT-Correct term considering reference
pre-syncope	Correct Term	a.CT-Correct term considering reference
premature ventricular complexes	Incorrect	1.Partial term translation
premature ventricular contraction	Correct Term	a.CT-Correct term considering reference
preserved ejection fraction	Correct Term	a.CT-Correct term considering reference
preserved ejection fraction	Correct Term	a.CT-Correct term considering reference
preserved ejection fraction	Correct Term	a.CT-Correct term considering reference
primary PCI	Correct Term	b.CT-Correct synonym
primary percutaneous coronary intervention	Correct Term	a.CT-Correct term considering reference
primary percutaneous coronary intervention	Correct Term	a.CT-Correct term considering reference
profound bradycardia	Correct Term	b.CT-Correct synonym
prognostic nutritional index	Correct Term	c.CT-Possible equivalent
prognostic nutritional index	Incorrect	7.Term drop
prosthetic heart valve	Correct Term	a.CT-Correct term considering reference
prosthetic heart valve thrombosis	Correct Term	a.CT-Correct term considering reference
Prosthetic valve thrombosis	Correct Term	a.CT-Correct term considering reference
pseudoaneurysm	Correct Term	a.CT-Correct term considering reference

radiofrequency ablation	Correct Term	a.CT-Correct term considering reference
rest dyspnea	Correct Term	a.CT-Correct term considering reference
resting heart rate	Correct Term	a.CT-Correct term considering reference
rhabdomyolysis	Correct Term	a.CT-Correct term considering reference
right heart function	Correct Term	a.CT-Correct term considering reference
right posterior cerebral artery	Correct Term	a.CT-Correct term considering reference
right posterior cerebral artery	Incorrect	11.Other term errors
right ventricle	Correct Term	a.CT-Correct term considering reference
right ventricular dysfunction	Incorrect	1.Partial term translation
right ventricular myocardial acceleration	Incorrect	7.Term drop
rs289714 variation	Incorrect	7.Term drop
scintigraphy	Correct Term	a.CT-Correct term considering reference
serum albumin	Correct Term	a.CT-Correct term considering reference
serum lactate level	Correct Term	a.CT-Correct term considering reference
serum lipid profile	Incorrect	7.Term drop
significant fructose consumption	Incorrect	7.Term drop
sinus rhythm	Incorrect	7.Term drop
sinus rhythm	Incorrect	7.Term drop
slow coronary flow	Correct Term	b.CT-Correct synonym
Sodium glucose co-transporter 2 inhibitors	Correct Term	c.CT-Possible equivalent
spontaneous aortic root rupture	Correct Term	a.CT-Correct term considering reference
spontaneous infection	Correct Term	a.CT-Correct term considering reference
ST-elevation myocardial infarction	Correct Term	a.CT-Correct term considering reference
ST-segment elevation myocardial infarction	Correct Term	b.CT-Correct synonym
ST-segment elevation myocardial infarction	Correct Term	a.CT-Correct term considering reference
ST-segment myocardial infarction	Correct Term	b.CT-Correct synonym
St. Jude Medical Biocor bovine bioprosthesis in mitral position	Incorrect	11.Other term errors
stable angina pectoris	Correct Term	a.CT-Correct term considering reference
structural or functional cardiac abnormalities	Correct Term	c.CT-Possible equivalent

sudden cardiac death	Correct Term	a.CT-Correct term considering reference
sudden cardiac death	Correct Term	a.CT-Correct term considering reference
syncope	Correct Term	a.CT-Correct term considering reference
tachycardia	Correct Term	a.CT-Correct term considering reference
thrombolytic therapy	Correct Term	a.CT-Correct term considering reference
thrombus	Incorrect	7.Term drop
thrombus	Correct Term	a.CT-Correct term considering reference
thrombus	Incorrect	7.Term drop
thrombus	Incorrect	7.Term drop
thrombus	Incorrect	7.Term drop
transcatheter aortic valve replacement	Incorrect	11.Other term errors
transcatheter atrial septal defect	Correct Term	a.CT-Correct term considering reference
transcatheter closure of an ASD	Correct Term	b.CT-Correct synonym
transthoracic echocardiography	Correct Term	a.CT-Correct term considering reference
transvenous pacemakers	Correct Term	b.CT-Correct synonym
vascular system	Correct Term	a.CT-Correct term considering reference
vasculitis	Correct Term	a.CT-Correct term considering reference

Annex IX. Binary Acronym Error Annotations

The binary (correct/incorrect) acronym annotation is provided below for reference. Annex V includes Turkish source sentences and target reference translations as well as the translations by PBMT-1, PBMT-2, NMT-1 and NMT-2 engines. Comparisons can be made directly by consulting the acronym translation in relevant sentences. For a more convenient view, acronym annotations are also presented in an open multi-sheet spreadsheet in GitHub repository[117]. An e-mail can be sent to gokhan.dogru@uab.cat for requesting the spreadsheets used in the acronym annotation process as well.

PBMT-1		PBMT-2		NMT-1		NMT-2	
Acronym	**Binary Analysis**	**Acronym**	**Binary Analysis**	**Acronym**	**Binary Analysis**	**Acronym**	**Binary Analysis**
AF	Incorrect Acr	AF	Correct Acr	AF	Incorrect Acr	AF	Correct Acr
AF	Correct Acr	AF	Correct Acr	AF	Correct Acr	AF	Correct Acr
AF	Correct Acr	AF	Correct Acr	AF	Incorrect Acr	AF	Correct Acr
AF	Correct Acr	AF	Correct Acr	AF	Correct Acr	AF	Correct Acr
AF	Correct Acr	AF	Correct Acr	AF	Incorrect Acr	AF	Incorrect Acr
ASD	Correct Acr	ASD	Correct Acr	ASD	Incorrect Acr	ASD	Correct Acr
aVF	Correct Acr	aVF	Correct Acr	aVF	Incorrect Acr	aVF	Correct Acr
BrS	Correct Acr	BrS	Correct Acr	BrS	Incorrect Acr	BrS	Correct Acr
CAD	Correct Acr	CAD	Correct Acr	CAD	Correct Acr	CAD	Correct Acr
CAE	Correct Acr	CAE	Correct Acr	CAE	Incorrect Acr	CAE	Correct Acr
CAS	Incorrect Acr	CAS	Incorrect Acr	CAS	Incorrect Acr	CAS	Incorrect Acr
CETP	Correct Acr	CETP	Correct Acr	CETP	Incorrect Acr	CETP	Correct Acr
CETP	Correct Acr	CETP	Correct Acr	CETP	Incorrect Acr	CETP	Correct Acr

[117] https://github.com/gokhandogru/TurkishEnglishParallelCorporaandMTEvaluation

COPD	Correct Acr	COPD	Correct Acr	COPD	Incorrect Acr	COPD	Incorrect Acr
COPD	Correct Acr	COPD	Correct Acr	COPD	Incorrect Acr	COPD	Correct Acr
COVID-19	Correct Acr	COVID-19	Correct Acr	COVID-19	Incorrect Acr	COVID-19	Incorrect Acr
CSFP	Correct Acr	CSFP	Incorrect Acr	CSFP	Incorrect Acr	CSFP	Incorrect Acr
CVD	Correct Acr	CVD	Correct Acr	CVD	Incorrect Acr	CVD	Correct Acr
DII	Incorrect Acr	DII	Incorrect Acr	DII	Incorrect Acr	DII	Correct Acr
DIII	Correct Acr	DIII	Correct Acr	DIII	Incorrect Acr	DIII	Correct Acr
ECG	Correct Acr	ECG	Correct Acr	ECG	Incorrect Acr	ECG	Incorrect Acr
FEV1	Incorrect Acr	FEV1	Incorrect Acr	FEV1	Incorrect Acr	FEV1	Incorrect Acr
FH	Correct Acr	FH	Correct Acr	FH	Incorrect Acr	FH	Correct Acr
FH	Correct Acr	FH	Correct Acr	FH	Incorrect Acr	FH	Incorrect Acr
FH	Correct Acr	FH	Correct Acr	FH	Incorrect Acr	FH	Correct Acr
HFpEF	Incorrect Acr	HFpEF	Incorrect Acr	HFpEF	Incorrect Acr	HFpEF	Incorrect Acr
HFpEF	Incorrect Acr	HFpEF	Incorrect Acr	HFpEF	Incorrect Acr	HFpEF	Incorrect Acr
HIV	Correct Acr	HIV	Correct Acr	HIV	Incorrect Acr	HIV	Correct Acr
HIV	Correct Acr	HIV	Correct Acr	HIV	Incorrect Acr	HIV	Correct Acr
HoFH	Correct Acr	HoFH	Correct Acr	HoFH	Incorrect Acr	HoFH	Correct Acr
IST	Correct Acr	IST	Correct Acr	IST	Incorrect Acr	IST	Incorrect Acr
L/O	Incorrect Acr	L/O	Incorrect Acr	L/O	Incorrect Acr	L/O	Correct Acr
LAA	Correct Acr	LAA	Correct Acr	LAA	Incorrect Acr	LAA	Correct Acr
LAA	Correct Acr	LAA	Correct Acr	LAA	Incorrect Acr	LAA	Correct Acr
LCOS	Incorrect Acr	LCOS	Incorrect Acr	LCOS	Incorrect Acr	LCOS	Incorrect Acr
LCOSs	Incorrect Acr	LCOSs	Incorrect Acr	LCOSs	Incorrect Acr	LCOSs	Incorrect Acr
LDL	Correct Acr	LDL	Correct Acr	LDL	Incorrect Acr	LDL	Incorrect Acr
LGE	Incorrect Acr	LGE	Incorrect Acr	LGE	Incorrect Acr	LGE	Incorrect Acr

Lpa	Correct Acr	Lpa	Correct Acr	Lpa	Incorrect Acr	Lpa	Incorrect Acr
Lpa	Correct Acr	Lpa	Correct Acr	Lpa	Incorrect Acr	Lpa	Incorrect Acr
LQTS	Correct Acr	LQTS	Correct Acr	LQTS	Incorrect Acr	LQTS	Incorrect Acr
LV	Correct Acr	LV	Incorrect Acr	LV	Incorrect Acr	LV	Correct Acr
LV	Correct Acr	LV	Incorrect Acr	LV	Incorrect Acr	LV	Correct Acr
LV	Correct Acr	LV	Incorrect Acr	LV	Incorrect Acr	LV	Incorrect Acr
LVAD	Incorrect Acr	LVAD	Incorrect Acr	LVAD	Incorrect Acr	LVAD	Incorrect Acr
LVAD	Incorrect Acr	LVAD	Incorrect Acr	LVAD	Incorrect Acr	LVAD	Incorrect Acr
LVR	Incorrect Acr	LVR	Incorrect Acr	LVR	Incorrect Acr	LVR	Incorrect Acr
PA	Correct Acr	PA	Correct Acr	PA	Incorrect Acr	PA	Correct Acr
PNI	Incorrect Acr	PNI	Incorrect Acr	PNI	Incorrect Acr	PNI	Incorrect Acr
PNI	Incorrect Acr	PNI	Incorrect Acr	PNI	Incorrect Acr	PNI	Incorrect Acr
POTS	Correct Acr	POTS	Correct Acr	POTS	Incorrect Acr	POTS	Correct Acr
PPCI	Incorrect Acr	PPCI	Incorrect Acr	PPCI	Incorrect Acr	PPCI	Correct Acr
PVC	Incorrect Acr	PVC	Incorrect Acr	PVC	Incorrect Acr	PVC	Incorrect Acr
PVT	Incorrect Acr	PVT	Incorrect Acr	PVT	Incorrect Acr	PVT	Correct Acr
PVT	Incorrect Acr	PVT	Incorrect Acr	PVT	Incorrect Acr	PVT	Incorrect Acr
QRS	Correct Acr	QRS	Correct Acr	QRS	Correct Acr	QRS	Incorrect Acr
QT	Correct Acr	QT	Correct Acr	QT	Incorrect Acr	QT	Incorrect Acr
QT	Correct Acr	QT	Correct Acr	QT	Correct Acr	QT	Correct Acr
SCF	Correct Acr	SCF	Correct Acr	SCF	Correct Acr	SCF	Incorrect Acr
SOI	Incorrect Acr	SOI	Incorrect Acr	SOI	Incorrect Acr	SOI	Incorrect Acr
SR	Correct Acr	SR	Correct Acr	SR	Incorrect Acr	SR	Incorrect Acr
ST	Correct Acr	ST	Correct Acr	ST	Correct Acr	ST	Correct Acr
ST	Correct Acr	ST	Correct Acr	ST	Correct Acr	ST	Incorrect Acr

ST	Correct Acr	ST	Correct Acr	ST	Incorrect Acr	ST	Correct Acr
STYMI	Incorrect Acr	STYMI	Incorrect Acr	STYMI	Incorrect Acr	STYMI	Incorrect Acr
TTR-CA	Incorrect Acr	TTR-CA	Correct Acr	TTR-CA	Incorrect Acr	TTR-CA	Incorrect Acr
TTR-CA	Correct Acr	TTR-CA	Correct Acr	TTR-CA	Incorrect Acr	TTR-CA	Incorrect Acr