THE RELATIONSHIP BETWEEN LINGUISTIC DISTANCE AND NEURAL MACHINE TRANSLATION QUALITY

ATA LEBLEBİCİ

BOĞAZİÇİ UNIVERSITY

2023

THE RELATIONSHIP BETWEEN LINGUISTIC DISTANCE AND NEURAL MACHINE TRANSLATION QUALITY

Thesis submitted to the

Institute for Graduate Studies in Social Sciences

in partial fulfillment of the requirement for the degree of

Master of Arts

in

Translation

by

Ata Leblebici

Boğaziçi University

2023

The Relationship Between Linguistic Distance

and Neural Machine Translation Quality

The thesis of Ata Leblebici

has been approved by:

Prof. Mehmet Şahin ____________________________
(Thesis Advisor)

Assist. Prof. Ena Hodzik ____________________________

Assist. Prof. Ümit Atlamaz ____________________________

Assoc. Prof. Müge Işıklar-Koçak ____________________________
(External Member)

Assoc. Prof. Caner Çetiner ____________________________
(External Member)

December 2022

DECLARATION OF ORIGINALITY

I, Ata Leblebici, certify that

- I am the sole author of this thesis and that I have fully acknowledged and documented in my thesis all sources of ideas and words, including digital resources, which have been produced or published by another person or institution;
- this thesis contains no material that has been submitted or accepted for a degree or diploma in any other educational institution;
- this is a true copy of the thesis approved by my advisor and thesis committee at Boğaziçi University, including final revisions required by them.

Signature…………………………………………….......

Date ……………………………………………...............

ABSTRACT

The Relationship between Linguistic Distance and
Neural Machine Translation Quality

Among all the factors that may contribute to the output quality of a translation, linguistic distance between the source language and target language had been largely cast aside. Relatively recent developments in linguistic distance research, away from lexical approaches and toward syntactic approaches, have made it possible to apply linguistic distance more methodically. This thesis aims to answer the question whether the neural machine translation quality drops as translated languages get more linguistically distant. To reach this answer in relation to machine translation, a survey was conducted in which participants were asked to evaluate machine translation outputs from different software and on different texts based on questions relating to different error types. Different participants who spoke both the source language Turkish and also increasingly more distant languages to Turkish at an advanced level were found, in order to capture the effect of a wide spectrum of language distance. The results from a relationship between linguistic distance and machine translation quality provide an experimental background for future research regarding this relatively unexplored relationship by raising specific questions about sensitivity towards linguistic distance in building machine translation tools.

ÖZET

Dil Mesafesi ile Nöral Makine Çevirisi Kalitesi Arasındaki İlişki

Bir çeviri çıktısının kalitesini etkileyebilecek türlü etkenler arasında kaynak dil ile erek dil arasındaki dil mesafesinin etkisi genellikle görmezden gelinmiştir. Dil mesafesi üzerine yapılan araştırmalarda, sözcüksel yaklaşımlardan sözdizimsel yaklaşımlara doğru ilerleyen yenilikler, dil mesafesi kavramının daha bilimsel bir şekilde uygulanabilmesine olanak sağlamıştır. Bu tez, bir çeviriye dâhil olan diller arasındaki dil mesafesi arttıkça, nöral makine çevirisinin kalitesinin ne derecede değiştiği sorusunu cevaplamayı hedeflemektedir. Bu sorunun cevabına makine çevirisi özelinde erişebilmek için, Türkçe ve Türkçeden giderek uzaklaşan dilleri ileri seviyede bilen çeşitli katılımcılara, dört metnin dörder makine çevirisinin sunulduğu bir anket hazırlanmış, katılımcılardan bu çevirileri hata türlerine denk gelen sorular doğrultusunda değerlendirmeleri istenmiştir. Dil mesafesi ile çeviri kalitesi arasındaki olası ilişki, makine çevirisi yazılımlarının kurulumlarının özellikle dil mesafesine hassasiyet gösterebilmelerine dair sorular doğurarak önceden derinlemesine işlenmemiş bu ilişki üzerine daha fazla araştırmalar yapabilmek adına deneysel bir altyapı hazırlamaktadır.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

One of the most significant turning points in the field of translation, coming in with the internet age, is the mass adoption of machine translation. While machine translation in partial capacity has been around since the 1930s (Somers, 2005, p. 140), the broad availability of machine translation to anyone in the public only came along with the 21st century. The process of translation, pertaining to the very human arena of language, became democratized and mainstream. It “has been ahead of most others in terms of casualisation, globalisation, and digitalisation” (Moorkens & Rocchi, 2021, p. 5). Democratized means availability, where every individual with internet access has machine translation tools available to them; mainstream means common use, since interlingual translating acts have expanded from areas that traditionally required them – law, medicine, literature – to the daily lives of individuals, especially in a globalized world. Most sophisticated machine-learning translation engines, those developed or used by multinational companies, are now at the stage of producing translation that meets some baseline level of adequacy (Specia, Hajlaoui, Hallett, & Aziz, 2011). Despite these, the increased prevalence of machine translation tools has also spelled out some negatives for translators, who now face reduction in pay and potential replacement (Moorkens & Rocchi, 2021). The question of “to what extent can translators be sufficiently replaced by machine translation tools” has real impact on employment and compensation of people working in the field. Understanding under which conditions machine translation can be as effective and have as high quality as human translation, if at all, aids in answering this question. Therefore, as machine translation becomes more pervasive, its remaining insufficiencies increase in importance as well.

In Translation Studies, both in human and machine translation, principles of translation can frequently be assumed to have a universal quality despite being derived from the study of relatively few and related languages. Historically, Western Translation Studies theories tended to put significant focus on Western European language pairs and extended their conclusions to apply to all languages around the globe. Examples from the history of the field of Translation Studies can be found starting even with the naming of the field by James Stratton Holmes, who was proficient in Dutch and English (Holmes, 1988). Another example includes the tradition of formal equivalence (Nida, 1964). Eugene Nida makes a distinction between formal equivalence, where during translation of words it is sought to find "similar orthographic or phonological features" (Baker, 2005, p. 77) and dynamic equivalence, where "complete naturalness of expression" (Nida, 1964, p. 159) is aimed for. Thinkers that had steeped themselves into the goal of formal equivalence had strived to find direct equivalents in target languages for words in their own language, no matter how distant. Especially during the early stages of the Translation Studies field, research was coming from, and focusing on languages of the Netherlands (Dutch-English-German) (Holmes, 1988) and Canada (French-English), and although some amount of focus was given to Hebrew and Yiddish as well (Even-Zohar, 1990), the distance between these languages had rarely been regarded as a potential setback for the universal applicability of translation principles. European scholars had instead focused on cultural differences (Nida, 1964). When studying languages from the Indo-European (IE) language family that make up the majority of geographically-European languages (Figure 1), careful attention is required to check the universalizability of principles, especially when applying these principles to languages from other primary families. For European scholars, the commonality of their linguistic roots was taken for granted, while their cultural differences were recognized.

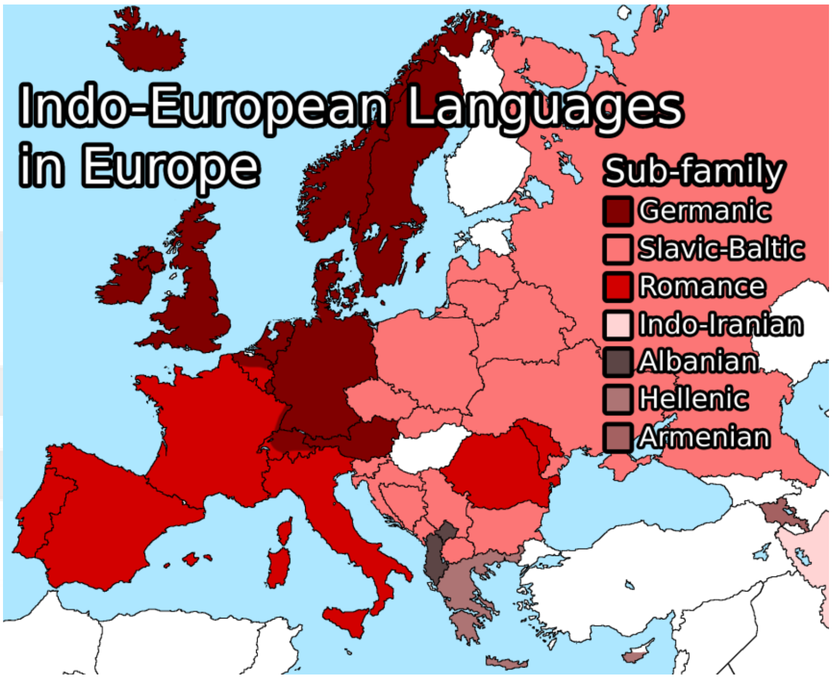


Figure 1. Indo-European languages in Europe and their subordinate language families

Stolze underlines that the heightened focus on achieving some standard of “equivalence” is directly related to developments in mathematics and logic (Stolze 2020, p. 146). This argument is well supported by logicians and language philosophers of 20th-century Europe, concerning themselves with the “mental image” that words represent in thought (Nida, 1964, p. 33), or supposing definitions of words being definitive rules of translation of word units (Wittgenstein, 2005, p. 59). It makes sense that the Western philosophical sphere after World War I, enveloped in logical positivism, with frequent scientific and mathematical breakthroughs, directing their gaze to the area of language would be keen on discovering or “inventing” clearly defined rules, methods, or principles by which translation works.

Despite these, contrary evidence is found in the different philosophies of translation around the globe. For example, older traditions of translation elsewhere had freer standards of translation when the traditions of Western Europe concerned themselves with a standard of formal equivalence. In medieval Bulgaria, translations had more freedom in the syntax and word order, even though semantic equivalence was still preferred:

> The idea that what matters is the translation of the meaning rather than mere sounds lay at the heart of the first Bulgarian and Slavonic theory of translation expounded by John Exarch. He rejected word-for-word translation and verbose explanations as deviations from the original and urged translators to aim for equivalence of meaning. (Baker, 2005, p. 349)

In the Chinese tradition of translating Buddhist religious texts, mass interpreting was crucial in understanding the original text, undertaken by “ . . . scores, sometimes hundreds, of Chinese monks and lay scholars who recorded in note form the foreign monk's explication” (Baker, 2005, p. 367). In the Japanese tradition, translation could be quite liberated in even describing the objects and concepts from the source text, where they were free to choose the closest Japanese equivalent of a particular object or concept originally foreign to them (Baker, 2005, p. 467). In the tradition of translation of the Ottoman Empire, many different forms of translation practices existed, some putting more importance on the meaning of the source text, while other practices focusing on style or poetics (Aksoy, 2005). The various global traditions of thousands of years, unfamiliar to the scholarly eyes of Western Europeans of the early 20th century, serve as an indication that translation norms have not been universal.

Despite global scholarly circles moving away from standards of equivalence and respecting agency in modern translations more, in public and professional spheres this change has not seeped through. The assumption of universality in equivalence leads to confusion, and issues in translation. Individuals might feel pressured to look for single-word phrases in the target language to match to single-word phrases in the source language. The absence of a one-to-one equivalence, which is more expected to be the case when languages belong to different primary families, would lead to an increase in loanwords in order to compensate. Arguably this can lead to changes in vocabulary, where local phrases are replaced by loanwords in the long run.

The same principle can apply to machine translation as well. Machine translation bore its roots in early 20th century attempts at code-breaking and mathematical computing (Baker, 2005), much like the mathematical approach to language philosophy at the time. It gave fruit to tools that carried marks of a standard of direct equivalence in its binary veins, using direct structures such as “dictionary-based direct replacement” (Baker, 2005, p. 140). If machine translation algorithms are trained to search and prioritize one-to-one relationships as they find in corpora like dictionaries, then they could be more likely to suggest loanwords as translational solutions. This could even occur in more modern, neural networks if the source material that software is trained upon includes loanwords. On a single term search basis, this might not create an issue for an individual who is not a translator or to someone predisposed to preferring loanwords; however, over the course of an entire text, it can cause problems in quality.

When attempting to compare how the act of translation might behave differently based on different language pairs, there needs to be a measure by which to evaluate language pairs in relation to one another. For the purposes of this research, an attribute of linguistic distance was considered. Here, linguistic distance will be defined as the overall difference between two languages or dialects, in terms of each language's lexical, syntactic, semantic, phonological, morphological, and etymological qualities.

However, a number of studies attempting to quantify this linguistic distance, based on these complex and dynamic qualities, have shown that it is a highly difficult task (Chiswick & Miller, 2004, p. 3). "Although the concept is well known among linguists, the prevailing view is that it cannot be measured. That is, no scalar measure can be developed for linguistic distance" (Chiswick & Miller, 2004, p. 10). Therefore, approximations need to be employed.

The question posed by this thesis is: "How does neural machine translation quality change based on the linguistic distance between the source and target languages?" This question arises from the aforementioned centrality of IE languages in the field of translation studies, and relates to two main areas of inquiry: whether linguistic distance between translated languages requires accommodation, and whether or not machine translation can provide for this accommodation, if it exists. These inquiries have implications for both human translation and machine translation. For machine translation, it might be that translation tools are professionally less reliable for translating languages of greater distance. Additional work might be required to improve the quality of translation tools when required to translate between two distant languages. It could be that machine translation software might behave well on a single search term basis regardless of linguistic distance, but decline in quality as the text length increases. Alternatively, linguistic distance might not be related to the machine translation quality, specifically due to the sophistication of neural network structures present in modern tools.

The hypothesis of this thesis is that machine translation quality does vary with linguistic distance with an inverse relationship. It is expected that as linguistic distance increases, the machine translation quality drops. The reasoning of this hypothesis is explained through the sub-questions of the thesis. Translation between distant languages might make it more challenging for human and machine translation to provide the same quality, given the same resources. It could be that modern machine translation tools, with complex structures like neural networks are well equipped to achieve equivalence in quality, regardless of linguistic distance. Even if a challenge exists when translating between distant languages, perhaps it reflects not in translation quality, but in translation speed or computer resources used.

If it is found to be the case that machine translation quality varies, it should signal to workers and scholars who professionally use machine translation to reconsider using software when translating distant languages, or be more selective in the type of software used. In addition, creators of such software would need to improve learning and training algorithms to adjust accordingly. The implications relating to human translation and required work may also have further implications about translator compensation, workload, and expected quality.

CHAPTER 2

LITERATURE REVIEW

2.1 Previous research into linguistic distance

When discussing literature in regard to the present thesis's discussion, it is important to distinguish between research into linguistic distance and research into machine translation. The application of linguistic distance largely focused on either ethnographical analysis, such as genetic diversity (Sokal, 1988), or socio-economic status of immigrants (Isphording & Otten, 2011; Piazzalunga, Strøm, Venturini, & Villosio, 2018).

While previous considerations were undertaken in relation to linguistic relativism in the first half of the 20th century, significant attempts at determining linguistic distance only began with the second half, with the American linguist Morris Swadesh. Swadesh attempted to map language divergence by analyzing lexical differences in word inventories of multiple languages (Swadesh, 1950, p. 161), by subjectively assembling a list of "principle" or "basic" words (Swadesh, 1971, p. 283). The words were chosen with attention paid to their fundamentality for any language, as well as theoretical resistance to change and borrowing. Swadesh's initial 100-word list (Swadesh, 1971) has been expanded upon by later research to a 207-word list from different languages across the world (Pool, 2022). While Swadesh's purpose was to chart a timeline for language divergence – also known as glottochronology – the method of comparing word lists is nevertheless useful for assessing linguistic distance. Swadesh's method and similar methods using word inventories can be classified as *lexical approaches* to linguistic distance.

Lexical approaches to linguistic distance classification have been undertaken for Indo-European languages extensively, mainly by building upon Swadesh's work. With the content of the word lists being of utmost importance for any Swadesh-like analysis, a significant portion of improvement, as well as controversy, around Swadesh-list analysis has been focused on the chosen words. Sergej Yakhontov reduced the number of words on the original Swadesh list from 100 down to 35 (Yakhontov, 1991). Citing various insufficiencies with the Yakhontov-35 list, Cecil Brown and Søren Wichmann began to develop a 40-word list, published under the research project "Automated Similarity Judgment Program" (Brown, Holman, Wichmann, & Velupillai, 2008). While offering the possibility of doing a lexical comparison using fewer words, the ASJP-40 database unfortunately suffers from a lack of quality when it comes to certain languages. In particular the Turkish list – the language central to this thesis – exhibits large inaccuracies in its words. Numerous examples can be given, such as the English negation word "not" being given in Turkish as "deyil" instead of "değil" (outdated and incorrect spelling), incorrect attributions of words like Turkish "akrep" (scorpion) for English "hand" (possibly in connection to one of the hands of an analog clock), and inconsistent conjugations for words (taking the roots "ye-" (to eat) and "bil-" (to know) correctly but not taking the root of the verb "to drink" as "iç-", instead using the noun "içki" (drink)) (Wichmann, 2020). Overall, the absence of Turkish characters "ç, ş, ı, ö, ü, ğ" also contribute to words being spelled either incorrectly or as if using a foreign keyboard: "biyik" instead of "bıyık", "kicik" instead of "küçük", "ay3z" instead of "ağız" and others (Wichmann, 2020). In Turkish specifically, the ASJP database falls highly short of the quality standard that a researcher ought to look for in their word lists and thus should not be used for lexical comparison in its current state.

Presently, lexical approaches were applied to the Turkic language family in a limited capacity. One example of these applications is the one conducted by Gerard Clauson, comparing various Turkic languages to Tungusic languages of Mongolian and Manchu (Clauson, 2005). Oktay Selim Karaca focused within the Turkic language family itself, comparing Swadesh lists of Turkish, Azerbaijani, Turkmen, Uzbek, Kazakh, Kyrgyz, and Tatar, creating a similarity matrix between them (Karaca, 2011).

Most of the research using lexical approaches to linguistic distance relies on the researcher’s own knowledge and diligence regarding the etymologies of the words. Unfortunately then, the research is prone to false attributions of word origins, since the relationship between items on a wordlist are dependent on the researcher’s own knowledge of that language. In some scenarios, in desiring to avoid committing an error, the researcher dismisses a certain item in the word lists altogether due to an etymology unknown to them. Examples can be seen in a study by Ceolin (2019), where within the notes of the appendix for the wordlists, there are misconceptions of Turkish etymologies. Ceolin ignores certain items in the wordlists based on them being loanwords from another language, or in an attempt to avoid covariance due to a shared etymology with another item. While the notes are generally accurate, on occasion there are errors such as assuming the homophonic roots of the Turkish word “düşünmek” (to think) and the Turkish verb “düş-” (to fall) signals a shared etymology between these words, or omissions such as the sole preference of the Turkish word for “fire”, “od” instead of the more commonly used “ateş” (Ceolin, 2019, p. 336). These errors and omissions at the very least cause some otherwise valuable information to be lost or remain unaccounted for, if the researcher is prudent enough to avoid false positives. Thus, clearer methodologies for systematic classification of linguistic distance are preferred:

> The premise is that one should not make any prior assumptions about whether the languages compared are related to each other. In fact, a major motivation for automated language classification is precisely that no such assumptions need to be made, such that the enterprise is independent of other methods. (Wichmann, 2010, p. 3633)

Taken as a whole, lexical approaches tend to offer similar advantages and disadvantages regardless of the word list used. For languages within the same language family, lexical comparisons could be used to shine some light onto linguistic relatedness. However, the precise distance between these languages is heavily dependent on the methodology that measures similarity of component words in the word lists. Etymological methods are restricted by the researcher's own knowledge of the languages in question and extent of their research into each word's etymology. Meanwhile, other letter-based or morpheme-based similarity metrics are susceptible to misrepresenting linguistic distance due to chance resemblances or differences between words. For languages that belong to different language families, lexical approaches are largely inappropriate. An approach using a Swadesh-100 or a Swadesh-207 list is almost entirely futile at creating any relation between languages of different families, due to the fact that the fundamental words are almost entirely unique between languages of different families. Swadesh lists of Turkish and any Indo-European language show almost no commonality etymologically. An example can be seen in Table 1 below, showing a selection of Swadesh list words from Indo-European languages and Turkish.

Table 1. Six Sample Words in Swadesh Lists of Turkish and Select Indo-European Languages

Turkish	Italian	Spanish	French	English	German
büyük	grande	grande	grand	big	groß
uzun	lungo	largo	long	long	lang
geniş	largo	ancho	large	wide	breit, weit
kalın	spesso	grueso	épais	thick	dick
ağır	pesante	pesado	lourd	heavy	schwer
küçük	piccolo	pequeño	petit	small	klein

There are some exceptions to this in alternate words that have become common usage, such as Persian loanwords in Turkish for "father" and "fire" (Pool, 2022; Ceolin, 2019). Most other possible resemblances between Turkish and an Indo-European language are then entirely coincidental, thereby clouding the integrity of such approaches. One could imagine linguistic relatedness as a relational mapping in three-dimensional space and a lexical approach only being able to determine linguistic distance two dimensionally (in a planar manner). Languages of the same family are located on the same two-dimensional plane and are thus fit for a lexical (planar) approach; but languages of different families are located on different, non-intersecting two-dimensional planes and therefore lexical approaches are unable to establish a connection.

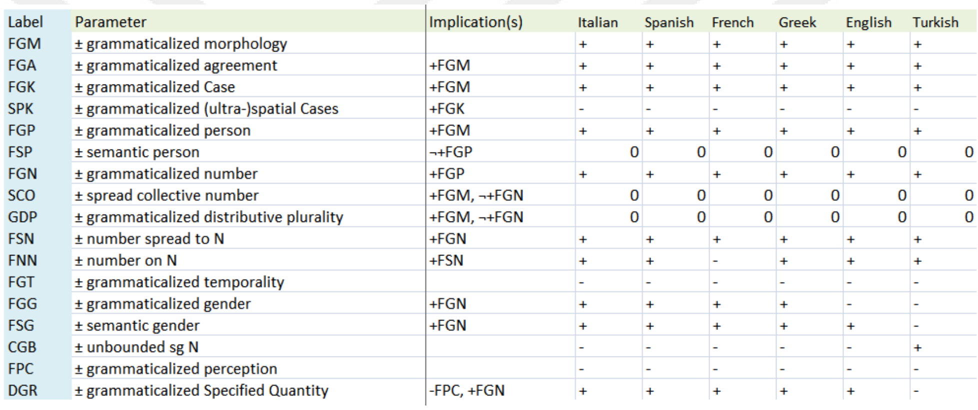
A researcher can also try to quantify linguistic distance using a *morphological approach*. The largest source of morphological information is the World Atlas of Language Structures (WALS) database (Dryer, 2013). WALS database offers a collection of language features, drawn from various other published sources. It is then theoretically possible for a researcher to use WALS and its language feature categories as a way to compare languages by. Problems arise however, when one considers exactly how these features are quantified. For a particular language in WALS, each feature is assigned a certain number, on a scale which is inconsistent across languages. For example, phonological feature 1A “Consonant Inventories” is a 5-point scale measuring the number of consonant sounds in a language’s alphabet, while morphological feature 22A “Inflectional Synthesis of the Verb” is a 7-point scale measuring the number of affixes that can be attached at the end of a verb (Dryer, 2013). Thus, comparisons become difficult when attempted between features with different number scales. The difficulty is only magnified when, such as in the case of feature 30A “Number of Genders”, one category within the feature is the null category (e.g. “no genders”) or a category does not match in number to the other ones (e.g. “five or more” when the other categories are “two”, “three”, “four”) (Dryer, 2013). Extensively retrofitting each language feature to the same point scale is an unavailable solution, due to inherent differences in the ways these features can, and often should, be quantified. Therefore, at the present state, a morphological approach based on the WALS database appears unfeasible.

Recent research provides another opportunity to measure linguistic distance. Longobardi and Guardiano attempted to determine linguistic distance using a *syntactical approach* instead (Longobardi & Guardiano, 2009). In this method, dubbed as the Parametric Comparison Method (PCM), they draw inspiration from previous Universal Grammar (UG) theories and set out to outline some syntactic parameters; using which, they can assess the features and qualities of a language (Longobardi & Guardiano, 2009). In the initial paper, they set out to prove that a syntactical approach can provide as useful of a comparison as a lexical approach. Their findings indicated that syntactical approaches can also be used to measure linguistic distance, as an alternative to lexical approaches, while also avoiding the dimensionality problem of the lexical approaches: “Finally, PCM promises to make a new tool for the investigation of our linguistic past, hopefully able to overcome the limits of the classical comparative method . . . ” (Longobardi & Guardiano, 2009, p. 1696). The difference between PCM and a WALS-based morphological approach is in how language features are quantified. Longobardi and others use a binary evaluation, marking a “+” value whenever a parameter is necessarily existent in the grammatical features of a language and with a “-” value otherwise. As delineated in a later study, a “-” value denotes the syntactic linguistic feature not being present for the mind of that language’s speaker: “cognitively, just “+” is viewed as an addition to the initial state of the mind. The “-” state of a parameter is not an entity attributed to the speaker’s mind, though it is used by the PCM as a symbol to code a difference with “+” at that parameter in another language” (Ceolin, Longobardi, Guardiano, & Irimia, 2020). For example, for a language without gendered nouns like Turkish, a “-” value for the parameter that tests the necessity of noun genders would indicate that a speaker of Turkish does not necessarily have a conception of noun genders when speaking their language, as opposed to a speaker of a language that does, such as French. Stemming from the theoretical basis of UG, they suppose that certain parameters exist deterministically alongside others. In other words, existence of certain parameters in a language presupposes the existence of other parameters (Longobardi & Guardiano, 2009). In order to account for this, PCM opts to use a null value “0” for these “implications” wherever they exist (Longobardi & Guardiano, 2009). This way, PCM aims to avoid exacerbating or diminishing the suggested distance between two languages, arising from too many instances where syntactically related parameters are marked with a “-” value. In statistical terms, it aims to avoid covariance between related parameters.

While parameter setting for all languages is an area susceptible to researcher bias (or optimistically, provides opportunity for further research), the results of the syntactical approach are convincing. Marcolli focuses on the robustness of the data, especially the parameters, but concludes that it is “preferable to exclude from the PCM all those parameters that are entailed and made irrelevant by other parameters” (Marcolli, 2016, p. 15). Crisma, Guardiano, and Longobardi detail how parameters are determined in a PCM approach, and how positive values are considered (Crisma, Guardiano, & Longobardi, 2020). Particularly, certain phrases are taken as examples from native speakers – or are presented to them – with each presented phrase exhibiting a certain parameter; these are dubbed as the *p-expressions* (Crisma et al., 2020). The p-expressions that are deemed to exemplify a particular parameter to the point that it signals a grammatical necessity are then included as part of the *Restricted List*. Phrases and p-expressions in the Restricted List are used in the paper to bring forth 94 parameters to use for language comparison (Crisma et al., 2020).

The largest piece of research using a syntactical approach is the ongoing study by Ceolin et al. (2020). This study employs the 94 parameters previously set by Crisma and others, applying them to 69 languages in Europe and Asia (Ceolin et al., 2020). Once the positive, negative, and null values are identified in each parameter for each language, languages are compared based on their values for each parameter. In order to do this, all the values of one language are concatenated into one string, and compared against the concatenated string of another language using a metric string distance measure. Special attention must be paid to the fact that the string distance measures here are different than a lexical approach. In a lexical approach, character-based measures were used to compare individual, corresponding words in word lists between two languages, whereas in a syntactical approach string distance measures are used to compare parameter values, arranged in a string form (a string consisting of “+”s, “0”s, and “-”s such as “+--+++-00++”). A section of the table of values from this paper can be found in the Table 2 below, in order to serve as a visual example to see how these values are mapped out against parameters on the left side (Ceolin, 2021).

Table 2. Sample Parameters and Respective Values in PCM



Label	Parameter	Implication(s)	Italian	Spanish	French	Greek	English	Turkish
FGM	± grammaticalized morphology		+	+	+	+	+	+
FGA	± grammaticalized agreement	+FGM	+	+	+	+	+	+
FGK	± grammaticalized Case	+FGM	+	+	+	+	+	+
SPK	± grammaticalized (ultra-)spatial Cases	+FGK	-	-	-	-	-	-
FGP	± grammaticalized person	+FGM	+	+	+	+	+	+
FSP	± semantic person	¬+FGP	0	0	0	0	0	0
FGN	± grammaticalized number	+FGP	+	+	+	+	+	+
SCO	± spread collective number	+FGM, ¬+FGN	0	0	0	0	0	0
GDP	± grammaticalized distributive plurality	+FGM, ¬+FGN	0	0	0	0	0	0
FSN	± number spread to N	+FGN	+	+	+	+	+	+
FNN	± number on N	+FSN	+	+	-	+	+	+
FGT	± grammaticalized temporality		-	-	-	-	-	-
FGG	± grammaticalized gender	+FGN	+	+	+	+	-	-
FSG	± semantic gender	+FGN	+	+	+	+	+	-
CGB	± unbounded sg N		-	-	-	-	-	+
FPC	± grammaticalized perception		-	-	-	-	-	-
DGR	± grammaticalized Specified Quantity	-FPC, +FGN	+	+	+	+	+	-

The particular string distance metric employed in the study by Ceolin et al. (2020) is called a Jaccard distance metric. Jaccard distance metric is one which counts the number of positive identities in relation to the total number of corresponding value pairs between the two strings. For example, between two strings “+ + - - + +” and “+ - + + - +” the Jaccard distance would be 2/6, since the first and last characters match. In pairs where at least one language has a null value for a parameter, that null value and the corresponding value for the other language are ignored for that parameter only.

Challenges for syntactical approaches, beyond parameter setting, reside in weighting. Initially, the researcher concedes an equal weighting between every parameter; in the case of PCM for example, the parameter “Null Possessive with Kinship Nouns” is deemed as equally important as a parameter that evaluates a language having gender cases (Ceolin et al., 2020). Another question of weighting is voiced in the paper: whether identities of parameter values should be equally weighted as the differences (Ceolin et al., 2020). Put another way, the researchers themselves wonder whether for a particular parameter a matching value pair between language A and language B should be equal in weight to language C and language D having different values in the same parameter. It could reasonably be posited that identities imply a definite link between the syntactic structure of two languages, while differences do not necessarily imply a relation between languages to the same degree, and therefore should be valued less than the occurrence of an identity. For some parameters that denote more idiosyncratic features, identities might be rarer to come by, such as feature like vowel harmony. Therefore the existence of a rare identity might be weighted more heavily. In addition, one could even consider giving some weighing to null and non-null value pairs, as that could imply a certain kind of difference as well.

One other previous approximation of linguistic distance has been the ease of mutual intelligibility. In other words, studies aim to make conclusions about linguistic distance between two languages L_1 and L_2, based on how easy it is for individual speakers of L_1 to learn and speak L_2, or vice-versa. Chiswick and Miller combine a previous report by Hart-Gonzalez and Lindemann on language learning and the Ethnologue Language Family Index published by Grimes and Grimes, to bring together a scale of linguistic distance of languages from English (Chiswick &

Miller, 2004). Their approach, in line with the nomenclature of the previous approaches, could be called an *educational* or *acquisitional approach*. These previous studies specifically follow the "ability of Americans to learn a variety of languages in fixed periods of time" (Chiswick & Miller, 2004, p. 10). They apply these observations to English proficiency levels of immigrants moving to United States and Canada, and argue, based on empirical evidence, " . . . that the greater the distance between an immigrant's origin language and English, the lower is the level of the immigrant's English language proficiency, when other relevant variables are the same" (Chiswick & Miller, 2004, p. 10). While this paper is not comprehensive enough to include other English-speaking countries such as United Kingdom, New Zealand, and Australia, one could argue that even though the geographical distribution of English speakers or the languages these speakers are most exposed to outside of English vary, their importance in determining language acquisition is auxiliary to the linguistic distance itself. Furthermore, the resultant linguistic distance table is generally found wanting compared to those from other approaches. Chiswick and Miller use a point scale, with the lowest point value at 1.00 for closest languages and the highest value at 3.00 for the most distant. With the scale incrementing only with discrete 0.25 value steps, the resultant table is highly limited in the way it can map linguistic relations. For example, according to their table, English sits at an equal distance from Turkish, Thai, Polish, Mongolian, Amharic, and Indonesian at 2.00 distance value (Chiswick & Miller, 2004). It is understandable that the discrete point system is a consequence of the educational approach, as opposed to the other methods which yield continuous scales. It would seem more reasonable that a realistic linguistic distance scale would exhibit a continuous scale, since it is unlikely that languages would all be positioned at discrete distances away from each other, especially as languages evolve over time.

2.2 Previous research into machine translation and quality

The other area of related research outside of linguistic distance is focused on assessing machine translation quality. Since directly assessing machine translation quality is a relative, qualitative task, various different approaches and proxies have been used to determine it. Poibeau summarizes the issue as such: “It is clearly difficult to evaluate the quality of a translation, since any evaluation involves some degree of subjectivity and strongly depends on the needs and point of view of the user.” (Poibeau, 2017, p. 130)

Before delving into assessing machine translation quality, it is important to understand the principles by which machine translation works. Initially, machine translation can be broken down to two broad structures, as *rule-based machine translation* and *data-driven machine translation.* Rule-based approaches, as their name suggests, generally employ a set of external rules, or an outside framework that the machine translation tool adheres to when translating between two languages, where software is provided with “ . . . a list of all the words in each of the source and the target languages, along with rules on how they can combine to create well-formed structures” (Kenny, 2022, p. 35). Perhaps the simplest structure under a rule-based approach for machine translation is direct translation. In direct translation, each word in the source language is translated to its mapped equivalent in the target language, and therefore this approach resembles what is colloquially known as a “word-for-word translation” (Hutchins, 2007, p. 4). A more complex approach than direct translation is the transfer-based approach, where the source language is abstracted into some sort of intermediate representation which then gets applied to the target language (Hutchins, 2007). A standard example of the transfer-based approach is using parse trees that are common in linguistics and grammar. Transfer-based approaches can produce better translations than direct approaches in general, due to more sensitivity towards the grammar structure of languages, even though it still largely depends on the intermediary representation used. Rule-based translation approaches in general, and transfer-based approaches in specific require " . . . highly skilled linguists to write the rules for each language pair . . . " (Kenny, 2022, p. 35) and suffers from the drawback that it is " . . . simply impossible in many cases to anticipate all the knowledge necessary to make RBMT systems work as desired" (Kenny, 2022, p. 35).

Data-driven machine translation is widely used instead of rule-based approaches in the modern machine translation landscape, due to the cognitive issue noted above and due to data-driven approaches generally offering lower costs and greater flexibility. As opposed to rule-based approaches, data-driven approaches feed on previously translated material from the source and target language to build their own rules or method of translation. A commonly used data-driven structure is called statistical machine translation (Kenny, 2022, p. 36). In statistical machine translation, the tool makes an index of all the words and phrase structures it observes in its training material, and calculates how often certain words and phrases seem to be paired up together. Among several shortcomings of statistical machine translation, its particularly poor performance in translating agglutinative languages is important to note (Kenny, 2022, p. 37), considering that Turkish also is an agglutinative language.

In the past decade, more sophisticated tools have instead moved into *neural machine translation* or a "deep learning" structure. These are node-based, hierarchical structures that allow software to “learn” from a presented set of data (Poibeau, 2017, p. 122). Deep learning in machine translation context means creating the hierarchical node structures, introducing to this structure some source material and their translations in another language, and letting it develop a system of translation based on this bilingual corpus. The larger the corpus represented, the more accurate one can expect the deep learning structure to be when suggesting a new translation. The advantage presented by neural networks is not having to devise and code-up a system of translation manually, and instead letting the software produce the weighing of operations performs on each node. Poibeau explains, “In the case of machine translation, deep learning makes it possible to envision systems where very few elements are specified manually, the idea being to let the system infer by itself the best representation from the data” (Poibeau, 2017, p. 123). While creating machine translation software is one challenging task, devising evaluation systems of machine translation software is an entirely different, yet also encumbering one. For lengthier studies with large swathes of data, researchers might prefer to employ automatic evaluation models, as opposed to human evaluation. Papineni and others present the BLEU method for cases where a quicker method of evaluation is sought (Papineni, Roukos, Ward & Zhu, 2002). BLEU uses an approach where candidate words for the translation of a particular word in the source text are selected based on common occurrence of each candidate word in the context of the source text. Another automatic evaluation framework is FEMTI, a method that uses context-based evaluation, where researchers need to map out complex structures and varieties of potential contexts to evaluate machine translation outputs by (Hovy, King, & Popescu-Belis, 2002). Automatic evaluation methods are best employed alongside human evaluation, as the former offers convenience of use, the latter allows for more sophisticated evaluations. The most glaring drawback to a data-driven machine translation approach is related to the existence, quality, and context of the data presented to the software. Poor quality of presented corpora could result in poorer performance in machine translation quality.

Assessing machine translation quality is notably difficult, requiring quantifying a seemingly subjective, and by definition, a qualitative feature. Toral and Guerberof-Arenas conducted a study on machine translation and creativity (Guerberof-Arenas & Toral, 2020). Their research is relevant for the methodology used in assessing the quality of translations through surveys. The two aims of their study were: to explore how creativity in translation differs based on the output of different translation modes (namely human translation, machine translation, and post-edited machine translation), and how differences in creativity affect reader experience. Despite desiring to assess creativity rather than quality, Guerberof-Arenas and Toral's methodology nevertheless provides one solution on how a study might attempt to quantify an abstract concept. Guerberof-Arenas and Toral, with the help of two professional reviewers, quantify acceptability of translation via the number and type of errors present (Guerberof-Arenas & Toral, 2020). This error-focused method of quality assessment is among the advised methods from Blatz et al. as well (Blatz et al., 2004). However, the drawbacks of using few reviewers for assessing acceptability are also recognized: "The analysis should be done by more than one expert reviewer and it could be more exhaustive . . . " (Guerberof-Arenas & Toral, 2020, p. 23). Guerberof-Arenas and Toral ask their reviewers a series of questions relating to translation quality, accuracy, and speed (Guerberof-Arenas & Toral, 2020, p. 8). These questions are evaluated on a 7-point Likert scale by the reviewers, thus quantifying the qualitative attributes of each translation. They assess the quality of the translation by referring to the amount and types of errors present in the translation, as they relate to the "Acceptability" criterion they set under creativity. For them, a creative translation must meet a standard of adequacy – an implied level of quality (Guerberof-Arenas & Toral, 2020). The error types they refer to stem from a standardized error typology framework named the "Dynamic Quality Framework-Multidimensional Quality Metrics (DQF-MQM) Error Typology" (Lommel et al., 2015). The specifics of this framework are explained in further detail under the Methodology section.

Guerberof-Arenas and Toral's approach in evaluating adequacy as part of creativity, and doing so by asking translators to rate creativity in translation by using a Likert scale, provides for a way for this thesis to assess the outcome of machine translation quality as well. A questionnaire-based approach, broadly resembling that of Guerberof-Arenas and Toral, with more focused questions relating directly to error types and sub-types of the DQF-MQM Error Typology framework are employed in this thesis.

A similar study employing the use of qualitative questions to assess translation accuracy was done by Şahin and Duman (Şahin & Duman, 2013). In this study, English and Russian chat logs' machine translations are assessed on intelligibility and accuracy, by using qualitative questions measured on Likert scales to evaluate each quality. In another study evaluating machine translation quality directly, one can turn to Şahin and Gürses, and their paper that analyzes machine translation quality of passages from Charles Dickens by professional and amateur translators. Similarly in this thesis, translators and other language professionals are consulted as the evaluators/reviewers for analyzing the output of a variety of literary texts (Şahin & Gürses, 2021). Besacier and Schwartz have also used a reader-survey method to evaluate quality on a machine translation output (Besacier & Schwartz, 2015). While their aim was to assess a post-edited machine translation output of a particular literary text, the types of questions used as evaluation criteria provide helpful a guide for qualitative questions in this thesis.

CHAPTER 3

DETERMINING LINGUISTIC DISTANCE

3.1 Syntactic approach to linguistic similarity

The most recent and extensive work undertaken in regard to syntactic linguistic distance is the development of the PCM framework (Ceolin et al., 2020). Despite being a syntactic approach, the PCM paper acknowledges the value in lexical approaches to linguistic distance. “Character-based algorithms, on the contrary, are the closest automatic analog to the linguists’ consolidated procedure of reconstructing all ancestral states (e.g., sounds and etymologies) and changes, and of postulating taxa on this basis” (Ceolin et al., 2020, p. 5). One particular challenge of using Swadesh lists for lexical relatedness arises from the fact that the words considered are chosen particularly for their commonality and resistance to change. A brief inspection of the words on the list would yield equivalents of pronouns “I”, “you”, “we”, indicators “this”, “that”, question words “who”, “where”, “what”, adjectives like “heavy”, “thick”, verbs like “to think”, “to cut”, “to fight”, body parts, etc. Therefore, it follows that languages which show any similarity in these “basic” words are the only ones belonging in the same language family. This is outlined earlier as the dimensionality problem of lexical approaches. That is to say, Swadesh list comparisons are able to draw relations within language families, while implying that there can be no language relation between languages from different families. Furthermore, as outlined earlier, lexical approaches rely heavily on the researcher’s own knowledge and intuition regarding the etymologies of words in Swadesh lists, and are thus prone to making errors. Other, more standardized approaches to lexical comparison, such as letter-based comparison of words require extensive amount of text manipulation in order to compensate for the lack of alphabetical unity, especially for languages in which Swadesh lists have not been transcribed into the International Phonetic Alphabet (IPA) yet (Kessler, 2007). Further questions relating to word selection, and whether selected words can be considered fundamental across all languages render lexical approaches generally undesirable. Therefore, a syntactic approach is preferred in the present thesis due to the problems of lexical approaches and in order to keep the discussion more focused on the relationship of linguistic distance and translation quality, as opposed to shifting the focus onto linguistic distance.

Despite the breadth of the original PCM paper, more languages could be introduced that would not only increase the scope, but also provide valuable connections in relation to Turkish, such as Semitic languages. This saw the addition of Standard Arabic and Hebrew in order to have a pair of Semitic languages represented. The most recent syntactic data for these languages are taken from the data repository attached to a subsequent study by the same team (Ceolin, 2021). The same repository also was the source for the latest 94 parameters which are originally drawn from Crisma to replicate the most recent version of the PCM study.

The distances based on the PCM values are calculated similar to the original study, by applying a Jaccard string distance metric on the concatenated value strings of languages. As explained in an earlier section, the Jaccard distance notes the number of positive identities (in PCM, “+/+” value matches) on a string of characters, in proportion to the possible pair matches of the string. Therefore, for the purposes of syntactic comparison, Jaccard distance can be represented as:

$$Jaccarddistance(d_J) = 1 - \frac{\#\, ofpositiveidentities\ (+/+)}{\#\, ofdifferences + \#\, ofpositiveidentities}$$

One possible alternative to Jaccard distance metric would be to consider negative identities, a “-/-” value match between two languages in PCM. This approach could also be considered a version of Hamming distance, and can be represented for the present purposes as:

$$Hammingdistance(d_H) = 1 - \frac{\#\ ofidentities\ (+/+\ or\ -/-)}{\#\ ofdifferences + \#\ ofidentities\ (+/+\ or\ -/-)}$$

It seems that with Hamming distance, since a “-” value represents the absence of a particular parameter in the speaker’s mind, counting a “-/-” value match within language pair as an identity would be unnatural; as it would be equivalent in weight to a “+/+” identity signaling the presence of said parameter in the speaker’s mind. It could also skew potential similarities in an undesired way. It is desirable to have languages be considered similar when their speakers positively exhibit certain parameters, and not to be considered similar based on cognitive absences. The Hamming distance is tested against Jaccard distance to confirm or deny these intuitions.

Once a matrix is created based on the distance derived from each metric across all parameters and languages, it is fed into a clustering algorithm in order to generate language trees using dendrograms. The particular clustering method is known as complete-linkage clustering (Sørensen, 1948). Complete-linkage is an agglomerative method for hierarchical clustering. It works by pairing two closest clusters together to form a new cluster, and it differs from other forms of hierarchical clustering methods by choosing the farthest distances to consider when establishing distances for the newly formed cluster. Table 3 below illustrates this with an example from a hypothetical linguistic distance matrix alongside Figure 2, which shows what the resultant tree dendrogram would be like.

Table 3. Complete-linkage Clustering Example

Base Distances

Language	Italian	Spanish	Japanese	Korean
Italian	0.00	0.17	0.55	0.50
Spanish	0.17	0.00	0.45	0.40
Japanese	0.55	0.45	0.00	0.18
Korean	0.50	0.40	0.18	0.00

Step 1

Language	It-Sp	Jp	Ko
It-Sp	0.00	0.55	0.50
Japanese	0.55	0.00	0.18
Korean	0.50	0.18	0.00

Step 2

Language	It-Sp	Jp-Ko
It-Sp	0.00	0.55
Jp-Ko	0.55	0.00

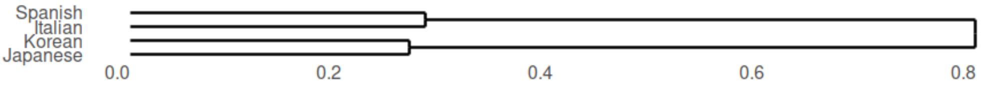


Figure 2. Complete-linkage clustering example: Sample tree

In the table above, the two closest languages – Italian and Spanish – are connected to form a new Italian-Spanish cluster. The unique aspect of complete linkage that sets it apart from other clustering algorithms is at determining the new distance of the Italian-Spanish combined cluster in “Step 1”. Italian has a farther distance to Japanese and Korean than Spanish; therefore the combined cluster takes the distances from Italian. Once again in “Step 2” it can be seen that when Japanese and Korean merge to create the Japanese-Korean cluster, the distance between two combined clusters at 0.55 is taken from Japanese, since it was farther away than Korean from the Italian-Spanish combined cluster. Complete-linkage produces different trees than other clustering algorithms, because each combined cluster sits farther apart from each other, due to the “farthest distance” prioritization. This suits well to analysis in linguistic distance since it is desirable to have clusters of independent language families as far away from each other as possible when producing a dendrogram. After all pairings are done, the tree-like structures that are created are used to assess the quality of the linguistic distance method employed. Tree structures that manage to exhibit language connections consistent with previous knowledge of language families indicate a more preferable method of string distance metric.

3.2 Linguistic distance results: Syntactic approach

Whilst reconstructing PCM for linguistic distance purposes, first it was tried to get rid of the null values “0” in the PCM approach, following the recommendation made by Marcolli (2016). By combing over, and reverse engineering the implications of the original PCM value table, it was found that the implications of the original paper are used in order to avoid excessive negative values when certain parameters are related. A direct, and relatively simple example of this can be seen in Mandarin and Cantonese, which lack (have negative values for) the parameter “FGM – Grammaticalized Morphology”, meaning that these languages do not conjugate semantic units (words/characters). Since these languages do not conjugate they, by definition, do not conjugate semantic units based on number agreement, case, or person. By extension this would mean Mandarin and Cantonese would have been given negative “-” values for the parameters “FGA – Grammaticalized Agreement”, “FGK – Grammaticalized Case” and “FGP – Grammaticalized Person”, due to having a negative “-” value for “FGM – Grammaticalized Morphology”. While the original paper recognizes this correlation, or “implication”, and replaces the implied negative values with null “0” values, the Marcolli recommendation would revert these null “0” values into the negative “-” values indicating the parametric feature not being cognitively present for the speaker (which is also identified by the Ceolin paper as the “default value”) (Ceolin et al., 2020). In the following paragraphs and figures, the original PCM approach, where the null “0” values are used, will be referred to as PCM-0 and the “non-implied” trial with “-” values replacing null “0” values will be referred to as PCM-1.

Separate trees can be obtained from evaluating PCM-0 and PCM-1 matrices with a Jaccard distance (only positive identities are counted) and a Hamming distance (both negative and positive identities are counted) metric. Four matrices are obtained overall by these methods, named PCM-0J, PCM-0H, PCM-1J, and PCM-1H, with the last letter corresponding to either the Jaccard or Hamming distance metric.

Overall, matrices obtained from Hamming distance metrics immediately draw attention, as it can be seen that the resultant heatmaps position languages much more closely to each other. This is in line with the expectation that similarities would be exaggerated in a Hamming distance method, as two languages would have a greater number of identities when negative identities are also accounted for. This effect is exaggerated in PCM-1H matrix where null identities, “0/0” value pairs, which were previously unaccounted for, now become“-/-” negative identities. On the other hand, as previously ignored “+/0” value pairs in PCM-1J become counted “+/-” differences, the pairings that were previously “0/-” that become “-/-” identities do not count as identities to offset this effect, since PCM-1J uses a Jaccard distance metric. Figure 3 below contrasts the four heatmaps (A larger, more readable version of these heatmaps can be found in Appendix A). In the figure, blue hues represent languages that are closer to one another, while orange hues represent languages that are comparatively more distant. Language families can somewhat be discerned, especially in the Jaccard heatmaps. Blue conglomerations at the top-left represent the IE language family, while blue groupings near the bottom-right represent the Turkic language family. The extent of the Turkic grouping is a representation of the extent the model incorporates Uralic languages into this group, visible in PCM-0J but less so in PCM-1J. It can be easily seen that the Hamming distance heatmaps significantly exacerbate the similarities between languages, producing heatmaps that are blue throughout. Hamming distance heatmaps also reduce the total range of distance values, making the differences between distances harder to discern. Due to the exaggerated similarities arising from negative identities, the Hamming method did not produce trees accurate enough to compete with those obtained from Jaccard trees.

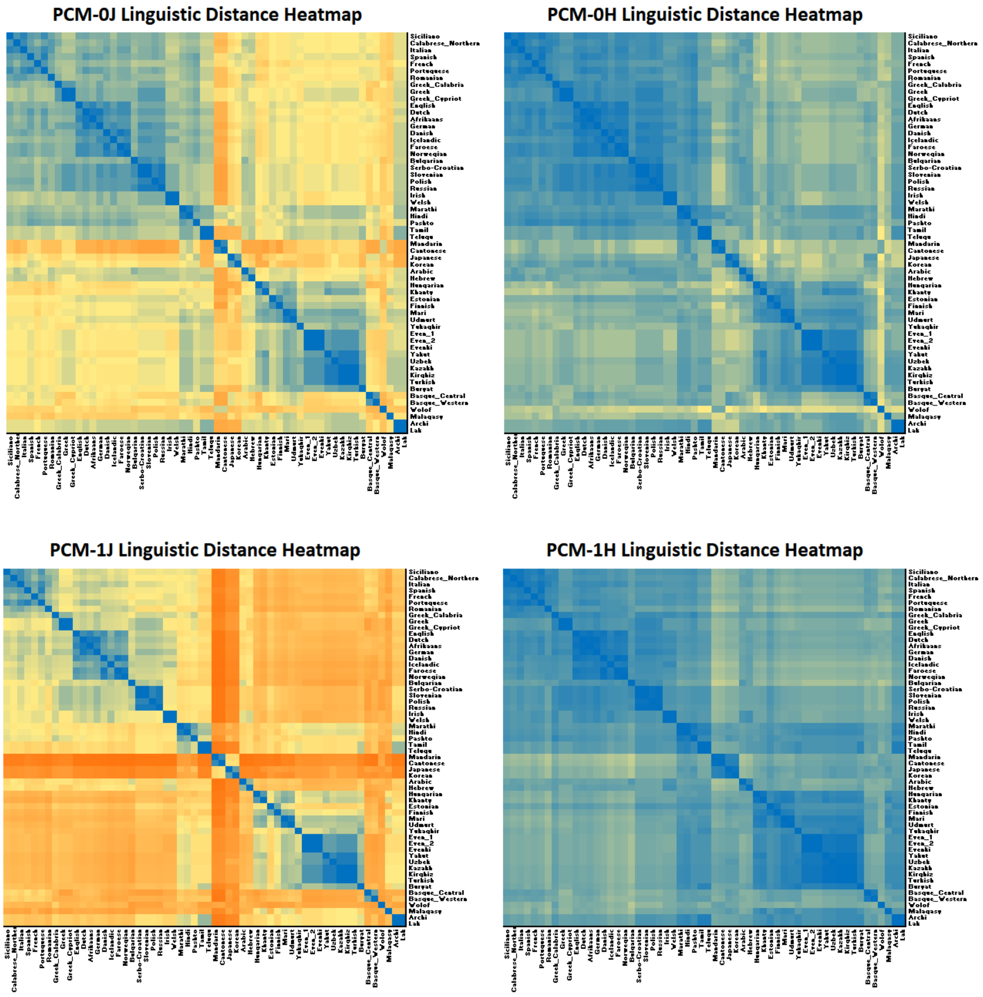


Figure 3. Heatmap comparison of PCM-0J, PCM-0H, PCM-1J, and PCM-1H

Between the two trees obtained from Jaccard distance matrices, the PCM-0J and PCM-1J, the PCM-1J trial showed slightly worse results on the resulting clustered tree. The main difference was the PCM-0J tree's ability to connect Hindi-Marathi-Pashto grouping to the rest of the IE language family, representing the necessary Indo-Aryan branch. The "non-implied" PCM-1J failed to achieve this connection. It can be seen from the heatmaps that replacing the ambiguous value "0" with the negative / default value "-" enhances the differences when using the Jaccard distance metric for value string comparison. This is because while in PCM-0J a "0/+" value pair between two languages would be ignored due to the presence of the null value, in PCM-1J the same value pair becomes “-/+”, indicating a difference. At the end, due to the failure of the PCM-1J tree to connect the Indo-Aryan branch into the IE family in the clustered tree, and exaggerated differences due to newly arising “+/-” differences from earlier “+/0” pairs, PCM-0J approach represented the selected matrix for this thesis. The final PCM-0J tree is given in Appendix B.

The Ceolin paper talks about the limitations of their clustered language tree by identifying the ambiguous placement of Malagasy near Uralic languages and Basque dialects (Ceolin et al., 2020, p. 9). In PCM-0J, while Basque dialects are properly at the outer branches of the entire tree, Malagasy still remains near the Finnish-Estonian pair. A particular misplacement can be observed in the outer connection of the Semitic language pair into the IE family, paired closely with the Celtic language pair. These imperfections of the PCM-0J tree yield room for improvements in the syntactic approach by bettering language inclusion, parameter modification, or both.

3.3 Significance testing: Syntactic

The goal of significance testing in the syntactic approach is to establish that the results obtained from the approach are significantly different from a random approach. While certain insights into linguistic relatedness that the matrices above present adhere to previous research on language families, it certainly would not be enough to rely on this intuition alone for the integrity of the research. Thus, these matrices also require a degree of significance testing to support their claimed accuracy.

For significance testing of a syntactic approach, the researcher ought to show that the parametrically corresponding value matches are not chance resemblances, and actually signify a linguistic connection between the languages. In this regard, when Language A is compared against Language B, each parameter’s value for Language A is compared to a random parameter value on the list of Language B, as opposed to the corresponding value. These random matches are done between all language pairs, and the resulting random table is compared against the matrix where values are matched parametrically.

Practically speaking, this means that instead of matching the values with each other for a Jaccard metric, the value for a random parameter is chosen instead for comparison. For example, when comparing two languages on the parameter FSN – “Number spread to N”, instead of taking the value for FSN for both languages, the value for a random parameter for L_1 is compared against the value of FSN for L_2.

This random value comparison is done for every parameter in the list, for 500 iterations. Using a 99% confidence level, the number of instances where the random matching of the syntactic parameters outperforms the benchmark should be limited to five at most.

One issue does come up in this random matching, considering the distances between the most distant languages. In particular Mandarin, Cantonese, Japanese, and Korean have a distance of roughly around 0.75-0.80 from a majority of the other languages. For these languages, even in 50 iterations, it was possible to see the random matching outperform the benchmark in more than 30 iterations. However, this does not necessarily mean the benchmark distance of these languages is faulty. Consider the distribution of the values of the parameters. If two languages L_1 and L_2 exhibit parameters in a mutually exclusive manner, then random matching is certain to outperform the expected distance between these two languages. This means that the random matches outperform the benchmark for Mandarin, Cantonese, Japanese, and Korean, simply due to the fact that these languages more exclusively exhibit parameters compared to other languages. In other words, often when languages necessarily display a parameter, these four languages do not, and vice versa.

In order to account for this, the outperformances were limited to the cases where the benchmark distance was less than 0.50, inclusive. Most relevant to the present thesis is whether there were any outliers in Turkish. Some languages crossed the five iteration threshold for outperforming the structured PCM-0J in a randomized matching in Turkish, and are considered outliers. Notably these pairs include Turkish-Dutch, Turkish-Romanian, Turkish-Japanese, Turkish-Finnish, and Turkish-Estonian among others at 9/500 iterations, 5/500 iterations, 15/500 iterations, 20/500 iterations, and 14/100 iterations respectively.

The existence of outliers from Finnic languages (Finnish, Estonian, Mari) is notable considering the historically controversial relationship between Turkic-Finnic languages. Despite a 99% confidence level being relatively strict, it nevertheless signals that PCM's current parameter selection could be improved to better capture the relationship between Turkic-Finnic languages. The same effect can also be observed in yet another controversial language pair, Turkish-Japanese. Since Finnic languages and Japanese do not pass the significance testing when paired with Turkish, among the others mentioned above, they were omitted from the qualitative study.

Once the qualitative consideration of the PCM-0J tree is supported by the statistical confidence in the underlying model, the final ordered language list can be created. The language list based on linguistic distance from Turkish can be found in Appendix C, and the general linguistic distance matrix of PCM-0J can be found in Appendix D.

CHAPTER 4

METHODOLOGY

4.1 Dependent and independent variables

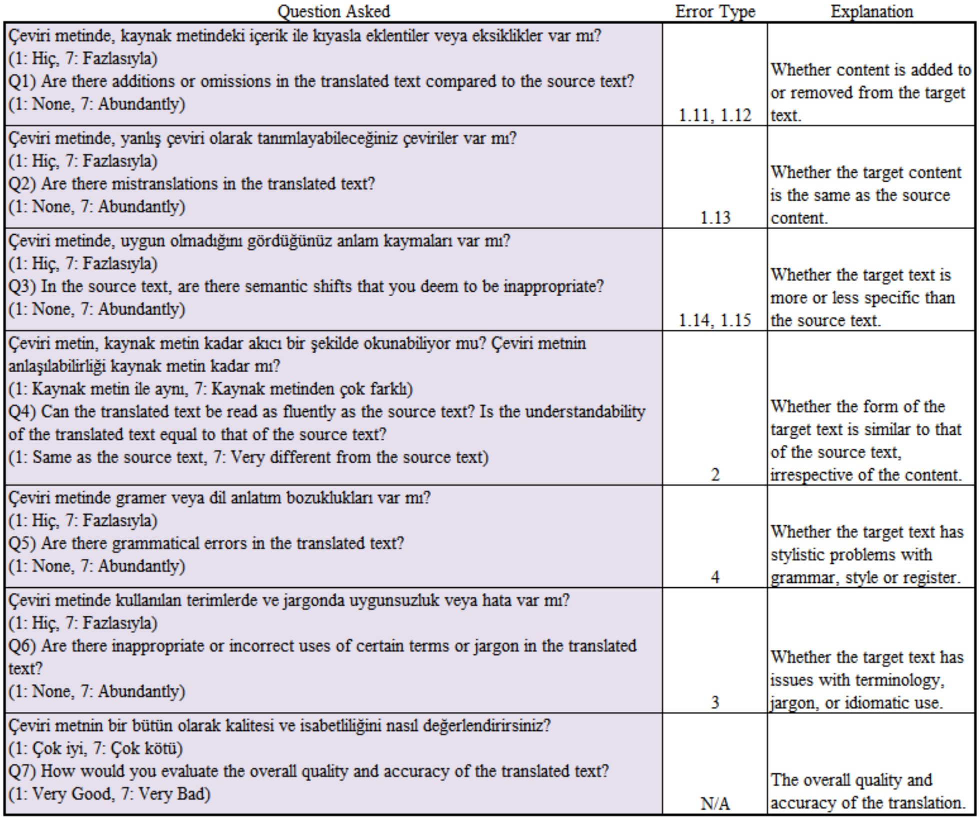
The question posed by this thesis is “How does neural machine translation quality change based on the linguistic distance between the source and target languages?” In order to quantify machine translation quality, it is represented by a dependent variable hereinafter referred to as Translation Quality (TQ). TQ score is based on the evaluation of translators and language professionals – hereinafter also referred to as “participants” – give to the presented text. In desiring to structure this evaluation, guiding questions were created which indicate to the participants how to evaluate a machine translated piece of text. These questions are formed to correspond to a high-level error type described in the DQF-MQM Error Typology framework (Lommel et al., 2015) and are scored on a 7-point Likert scale, inspired by the creativity assessment questions from Guerberof-Arenas and Toral (Guerberof-Arenas & Toral, 2020). The DQF-MQM Error Typology was developed as part of the project Quality Translation 21 (QT21), which is a machine translation project funded by the EU’s Horizon 2020 research and innovation program (Lommel et al., 2015). It is a framework that standardizes translation error types and sub-types with definitions and examples, and according to Guerberof-Arenas and Toral, “ . . . unifies evaluation practices from academia and industry” (Guerberof-Arenas & Toral, 2020, p. 11). The relevant error types and their sub-types are given in Appendix E.

The questions as they are asked in Turkish and their English translations are given in Appendix F. For example, the second question (Q2) asked in the survey is:

> Q2: Çeviri metinde, kaynak metindeki içerik ile kıyasla eklentiler veya eksiklikler var mı? (1: *Hiç*, 7: *Fazlasıyla*)
>
> Q2: Are there additions or omissions in the translated text compared to the source text? (1: *None,* 7: *Abundantly*)

This question directly corresponds to error sub-types 1.11 “Additions” and 1.12 “Omissions” of the DQF-MQM framework. Not all high-level error types and their subcategories are represented in the questions of this thesis, mainly due to a dispreference towards extending participants’ workload. The likelihood of getting cooperation and willingness to join from participants drop significantly in proportion to the length of the study, which puts limitations on both the number and types of questions asked, as well as the length of the texts themselves. Furthermore, it is unclear whether a researcher ought to weigh all error types equally. Whether or not error type 4.41 “Awkward Style” has the same significance as 2.23 “Fluency – Grammar” for the text is largely left to the deliberation of the researcher. Thus, the error types inquired in this thesis are subjectively restricted to: Error sub-types 1.11-1.12 in Question 1, sub-type 1.13 in Question 2, sub-type 1.14-1.15 in Question 3, type 2 in Question 4, type 4 in Question 5, and type 3 in Question 6 – see Appendix E. Questions exploring high-level error types 2, 3, and 4 are deliberately phrased broadly in order to best accommodate the respective error subcategories. The questions, corresponding error types, and their measured values are summarized in Table 4 below.

Table 4. Survey Questions and Corresponding Error Types



Question Asked	Error Type	Explanation
Çeviri metinde, kaynak metindeki içerik ile kıyasla eklentiler veya eksiklikler var mı? (1: Hiç, 7: Fazlasıyla) Q1) Are there additions or omissions in the translated text compared to the source text? (1: None, 7: Abundantly)	1.11, 1.12	Whether content is added to or removed from the target text.
Çeviri metinde, yanlış çeviri olarak tanımlayabileceğiniz çeviriler var mı? (1: Hiç, 7: Fazlasıyla) Q2) Are there mistranslations in the translated text? (1: None, 7: Abundantly)	1.13	Whether the target content is the same as the source content.
Çeviri metinde, uygun olmadığını gördüğünüz anlam kaymaları var mı? (1: Hiç, 7: Fazlasıyla) Q3) In the source text, are there semantic shifts that you deem to be inappropriate? (1: None, 7: Abundantly)	1.14, 1.15	Whether the target text is more or less specific than the source text.
Çeviri metin, kaynak metin kadar akıcı bir şekilde okunabiliyor mu? Çeviri metnin anlaşılabilirliği kaynak metin kadar mı? (1: Kaynak metin ile aynı, 7: Kaynak metinden çok farklı) Q4) Can the translated text be read as fluently as the source text? Is the understandability of the translated text equal to that of the source text? (1: Same as the source text, 7: Very different from the source text)	2	Whether the form of the target text is similar to that of the source text, irrespective of the content.
Çeviri metinde gramer veya dil anlatım bozuklukları var mı? (1: Hiç, 7: Fazlasıyla) Q5) Are there grammatical errors in the translated text? (1: None, 7: Abundantly)	4	Whether the target text has stylistic problems with grammar, style or register.
Çeviri metinde kullanılan terimlerde ve jargonda uygunsuzluk veya hata var mı? (1: Hiç, 7: Fazlasıyla) Q6) Are there inappropriate or incorrect uses of certain terms or jargon in the translated text? (1: None, 7: Abundantly)	3	Whether the target text has issues with terminology, jargon, or idiomatic use.
Çeviri metnin bir bütün olarak kalitesi ve isabetliliğini nasıl değerlendirirsiniz? (1: Çok iyi, 7: Çok kötü) Q7) How would you evaluate the overall quality and accuracy of the translated text? (1: Very Good, 7: Very Bad)	N/A	The overall quality and accuracy of the translation.

The final question on the list asks the participants to rate the overall quality (TQ) of the text to the best of their ability, which was compared against the composite evaluation arising from the responses to the previous questions. The dependent variable of machine translation quality, quantified and obtained in this manner, was then measured against the independent variable of linguistic distance.

The independent variable of the thesis is the linguistic distance between languages of the source text (L_1) and target text (L_2). Linguistic distance can be varied by changing the language of the translated text while the source language stays the same.

4.2 Source language selection

For a standardized scale of linguistic distance, it is most germane to use a single source language for all language pairs involved in the study. In principle, the particular language chosen as the source language does not matter, as long as the scope of selected language pairs offers a wide range of similarities and dissimilarities when converted into numerical data. In other words, when selecting a source language, it is important to make sure to have representation of target languages that are known to be related to the source, as well as target languages that are known to not be related.

The selection of Turkish offers a unique and apt selection of target languages. It belongs to a well-defined primary language family– the Turkic language family. It also remains outside of the large IE language family, allowing for a wide array of potentially linguistically distant languages. In addition, Turkish also enjoys common vocabulary with regional languages it does not otherwise share linguistic relationship with, such as Arabic and Persian.

As mentioned previously, it is critical for the researcher to avoid committing lexical mistakes arising from working with languages they are unfamiliar with. In line with this, Turkish also presents itself under a unique light, being the language that the present researcher is most familiar with, and their native tongue. It is with this familiarity that the mishaps of previous literature can be identified, which is especially worthwhile for a language that is less attentively studied than IE languages.

4.3 Text types and text generation

Alongside language selection, there were a number of other factors that needed to be controlled for in this thesis. Translation quality could have a propensity to vary depending on the type of text being translated. One could expect that more abstract, literary texts are more prone to machine translation errors. One of the reasons for this could be the prevalence of technical terms in various languages. Terms which are specific to certain fields, such as law, medicine, or even some scientific terms, can show similarities that allow machine learning algorithms to easily replicate its performance in one language pair in another when translating technical texts. One of the implications from Şahin and Gürses's research was that machine translation could be more suitable for certain types of texts, and they questioned whether their singular choice of Charles Dickens could be a contributing factor to their results (Şahin and Gürses, 2021). This reasoning makes intuitive sense: suppose a text employs heavy use of figures of speech, it could be the case that neural machine translation produces a more literal, and thereby a less desirable, result. On the other hand, if the textual material is already translated and is part of the machine learning training set of the translation engine, then it could be expected that the machine translation output would be identical to human translation. Regardless of which of these cases is true, it remains possible that the text chosen has an impact on the machine translation output. In order to control for this, the ideal choice would be to use a variety of text types. These texts would encompass different genres, authors, periods and length in order to account for the various qualities of a text. However, one has to recognize the practical drawbacks of this ambitious attempt. Realistically, participants partaking in this thesis might not desire to comb through numerous texts sifted through a variety of translation engines, resulting in an hours-long work of quality assurance. Therefore, an alternative way of text selection is required.

Another quality desired in the texts to be translated is having them be previously un-translated. This is needed to avoid any instances where one of the texts happen to be a part of the training set of one of the machine translation tools, resulting in an unexpectedly high quality translation. The most direct way to ensure that texts are un-translated is to present original texts to the neural machine translation tools. This way, the content of the texts can also be controlled, and different text types can be represented. It also provides another opportunity for this particular thesis. Using syntactic parameters to create linguistic distance matrices also allows the use of syntactic parameters in text origination. It is appropriate to try to exhibit phrases in the translated texts which correspond to the syntactic parameters in PCM-0J, since the syntactic parameters are also determined by phrases in each language that exclusively portray the parameter in question –the aforementioned p-expressions (Crisma et al., 2020). Even though the exact p-expressions in Crisma's Restricted List (Crisma et al., 2020) are not publicly available, one can still work backwards from the parameters and form sentences that display the desired parameter. An example of this type of text origination is given in Table 5 below, representing the first text (T_1) of this thesis:

Table 5. Sample Text Origination and Parameters Involved

Text 1	Legend
Kendisini sevmeseler de onların arasına katılmak istiyordu Elif. O üst mahalle çocukları dünyaya başka bir gözle bakıyorlardı sanki. Onun yeri ise yokuşun aşağısındaki alt mahalledeydi. Elif aralarındaki yakınlığa imrenmişti en çok. Hepsi birbirini tanıyordu! Alt mahallede yakınlık, ihtiyaçtan doğan bir şeydi.	• grammaticalized morphology • grammaticalized gender • collective number • grammaticalized agreement grammaticalized number • number spread to N • adjectival possessives

Here it is important to elaborate Crisma's frequent use of "grammaticalized" as an adjective for parameters, relating specifically to its corresponding phrase being included in the Restricted List. "Grammaticalized" means that the feature in question necessarily places a grammatical constraint on possible phrases in the language (Crisma et al., 2020). In the sample text in Table 5 above, the first parameter "grammaticalized morphology" would refer to a language necessarily having to modify nouns in order to express morphological qualities – such as by conjugating. A language like Mandarin Chinese would not be regarded as having "grammaticalized morphology", as it does not conjugate nouns and does not necessarily express morphology. The burden of a feature being "grammaticalized", and therefore receiving a positive "+" value in syntactic comparison, is on the necessity of its expression.

When determining which parameters to include p-expressions for, some filtering needed to be done on Crisma's parameters list. As it stood, 94 parameters were infeasible to be included in separate p-expressions within short texts of a paragraph each. This would most likely require either longer texts, or a larger amount of shorter texts, both of which creates a higher burden on the participants and reduces willingness to participate. Thus, a narrower selection of parameters is required.

Filtering the parameters was done on the principle that the most "competitive" parameters would be chosen. In other words, parameters that show the largest amount of divergence within PCM would be selected. This way, the most delineation of languages by using the least amount of parameters could be achieved, thereby making it possible to fit these parameters into short texts. Parameters that had no language with a "-" value, such as "FGN – Grammaticalized Number" (referring to a grammatical necessity to express the number modifier of the noun – for example, pluralization), no language with a "+" value, such as "FPC – Grammaticalized Perception" (referring to a grammatically necessary constraint to express perception), or with an overwhelming amount of implications with "0" value "FGC – Grammaticalized Classifier" (referring to a grammatically necessary constraint to include a classifier word, such as measure words in Chinese) are ignored. Optimal parameters are ones similar to "FGG – Grammaticalized Gender" (41 "+" values and 24 "-" values, referring to a grammatically necessary constraint to express gender of nouns, absent in genderless languages like Turkish) or "ARR – Free Reduced Relatives" (38 "+" values and 31 "-" values, referring to free positioning of relative clauses). Some fundamental parameters like "FSN – Number Spread to N" (number of a sentence being expressed on a noun phrase) or "FGP – Grammaticalized Person" (constraint of expressing person noun or noun modifier) are represented mainly due to them necessarily appearing in comprehensible texts.

The four texts prepared for this thesis (T_1, T_2, T_3, T_4), are given in Appendix G, alongside a sample English translation. They are marked with which syntactic parameters are denoted by which fragment or p-expression.

4.4 Machine translation tools

Machine translation tools needed to be controlled for as well, in order to be able to deduce more general conclusions about neural machine translation as a whole, as opposed to one particular piece of translation software. In order to achieve this, source texts were translated using different machine translation tools. Each participant acquired various text excerpts in paragraph form, from different genres of text, each ran through different machine translation software, which participants remained blind to. The four translation tools chosen for this thesis were Google Translate, Yandex Translate, LibreTranslate, and Windows Translator, respectively as MT_1, MT_2, MT_3, and MT_4. Translations of the prepared texts were obtained in April 2022, using the most recently available, public version of each tool.

The selected machine translation tools all employ neural networks in some capacity. Google Translate had used a popular “Long short-term memory” (LSTM) neural network with 8 layers of nodes (Wu, Schuster, Chen, Le & Norouzi, 2016) up until 2020. LSTM is one particular construction of a neural network that has forward flow of information as well as feedback, or a backwards flow (Hochreiter & Schmidhuber, 1997). Since 2020, Google now uses a proprietary neural network model dubbed as “Transformer” (Vaswani et al., 2017). Transformer is a different network architecture that uses a metric called “attention” to provide the context for each of the semantic units in a given source text (Vaswani et al., 2017). Suggested translations in the target language are then made not based on a single vector value derived from the source word, but a matrix of values derived from all of the other words in the source text, weighted by their relevance (or "attention") to the particular semantic unit being translated at each step. Google's own documentation of their translation performance is highly relevant for the present thesis as well. In one blog post, Google mentions one potential performance-impacting mechanism outside of linguistic distance:

> Nevertheless, state-of-the-art systems lag significantly behind human performance in all but the most specific translation tasks. And while the research community has developed techniques that are successful for high-resource languages like Spanish and German, for which there exist copious amounts of training data, performance on low-resource languages, like Yoruba or Malayalam, still leaves much to be desired. (Caswell & Liang, 2020, para. 1)

The software engineers from the quote above also highlight the lacking quality of even the latest machine translation software in performance compared to human translation. They focus on the quality of machine translation in regard to resource availability, referring to amount of translated material available in both the target and source languages, and do not mention possible inference of linguistic distance.

Yandex Translate operates on a hybrid model of both a neural network and a statistical machine translation model. Statistical machine translation is another process that feeds on previously translated material from two languages. In this case, instead of letting the software "learn" on its own and imprint its sub-processes on layers of nodes, statistical machine translation has a more rigid mode of operation. The software makes an index of all the words and phrase structures it observes in its training material, and calculates how often certain words and phrases seem to be paired up together. When a new, full text is presented, the algorithm devises numerous potential translations and selects the best one on a statistical, probabilistic model (Yandex). In each translation case, Yandex Translate mentions that they use an open-source algorithm called CatBoost to select which method's translation is preferable, statistical model or neural network.

Microsoft Translator also operates on a neural network, although they do offer the choice to translate based on an older statistical translation model as well. Microsoft shares that they employ an LSTM structure similar to that of Google Translate up to 2020 (Microsoft). In their case, their neural network assigns values to each word on a 500-dimension vector space, based on the word's semantic and lexical qualities (Microsoft). As it is described on their page, "[these layers] could encode simple concepts like gender (feminine, masculine, neutral), politeness level (slang, casual, written, formal, etc.), type of word (verb, noun, etc.), but also any other non-obvious characteristics as derived from the training data" (Microsoft, para. 22). Each vectored representation of the word is then passed onto a second layer which encodes the data further into a 1000-dimension vector space. The process is then repeated for fine tuning. Alongside an attention layer that sequences which words are to be translated, and a decoder layer that produces a translation from the vector space representation, the Microsoft algorithm can be said to employ a 4 layered structure.

The pieces of translation software above were picked due to their widespread use. Being supported by large technological companies, it can safely be assumed that these tools provide some of the most current machine translation technologies. An evaluation of translation performance by employing these tools ensures that the thesis is relevant to the contemporary machine translation industry. Other machine translation software offered by smaller enterprises might even actually be using programs – APIs – that directly send the requested translation through the servers of the tools above. Despite all these, one more piece of translation software was used, particularly because of the proprietary nature of these powerful tools. Unique and more importantly open-source software would provide for more opportunities to discuss the internal working process of neural networks should its performance be an outlier. For this purpose, LibreTranslate was chosen, a web interface tool of Argos Translate software. Argos Translate is dependent on a tool called Stanza for sentence detection, and uses a Python-based, open-source piece of machine translation software called OpenNMT (Klein, Kim, Deng, Senellart & Rush, 2017). Argos Translate works on a sentence level, and breaks down sentences into "tokens" in a process dubbed "tokenization" (Argos Open Technologies). Tokens might be a word itself or a part of a word. The tokens within a sentence are sequenced and translated using a pre-trained model of the "CTranslate2" process under OpenNMT (Argos Open Technologies).

For most these tools above, while it has been argued that an intermediary language such as English could have been used especially with older neural models (Benjamin, 2019), and could still be used when training newer tools, there is no official, public disclosure on whether intermediary languages are used when translating between Turkish and other particular languages. The exception to this is LibreTranslate, where it can be seen that Turkish-English direct translation is supported in its open-source documentation, but all other translations that include Turkish as one leg use English as an intermediary language in between (Argos Open Technologies, 2020).

Evaluating the machine translation outputs would indicate the relative strengths of the neural networks used by each translation engine. LibreTranslate's reliance on sentence based translation could also result in comprehension issues on a paragraph level, or its reliance on English as an intermediary language in translations involving Turkish might be causing issues in quality. When the results from the participants' evaluation are collated against the linguistic distance based on the source language (L_1 = Turkish), attention was paid to how participants' scores shift between languages. Scoring worse on the results could implicate a particular piece of software as underperforming, a particular text type as difficult, or a particular participant as unreliable. Linguistic distance can only be said to affect machine translation quality when scores are consistent despite these, or in other words, when these variables are controlled.

4.5 Participants

Finally, the assessing participants needed to be controlled for. While it would perhaps seem ideal to include both translations from L_1 to L_2 as well as from L_2 to L_1, and therefore translators from both languages evaluating each output, it would give rise to two significant issues. First issue is the operational scope of the research becoming excessively broad. There are 58 languages present in PCM-0J, and to try to include a group of translators for each of these groups would increase the participant count beyond what can be feasibly conducted. The second issue arises from the inherent differences between the translators themselves. It is not possible to control for attributes such as attitude towards translation, or expectations from machine translations, when the groups of translators assessing these translations are unique human beings. The only way to control these attributes is to have the same group of translators assess outputs in each language, and such levels of polyglotism in translators is unfortunately absent in the status quo. These limits precluded the study from being conducted in both ways, from L_1 to L_2 and from L_2 to L_1. Therefore, for practicality, translations of only one way – from L_1 to L_2 – were considered in this thesis, and the individual qualities of participants were checked by having a plurality of participants for each language pair. Language pairs used in this thesis were determined by the language distance list presented in Appendix C.

4.6 Qualitative survey

The final list of linguistic distance was broken down into six tiers with increasing language dissimilarity to Turkish, the source language. When constructing these tiers, special attention was also paid to the ease of finding potential survey participants. These tiers and the languages in each tier are represented below, listed in order from closest to farthest, in Table 6:

Table 6. Ordered Tiers of Language Pairs Based on Increasing Linguistic Distance to Turkish

Tier 1	Turkish	Kazakh Kyrgyz Uzbek
Tier 2	Turkish	Spanish Greek
Tier 3	Turkish	Russian Polish
Tier 4	Turkish	Italian Portuguese English Arabic
Tier 5	Turkish	German French
Tier 6	Turkish	Mandarin Korean

Two to four participants were found for each tier, with a median and mean participant count of three, for a total of 21 participants. When selecting the participants, it was sought that each participant would be someone that can prove their proficiency in their respective language. Keeping with this theme, most of the participants were working or have worked as professional translators, or were advanced students of Translation Studies. The select remaining few had demonstrated their language proficiency, for example via a language proficiency test (B2 equivalent or above). A multiplicity of the participants was needed to control for the biases of any one particular participant.

Unique packages for each language were constructed in which the four texts given in Appendix G were translated into the language of the respective package by the machine translation tools mentioned in section 4.4. The names of the translation tools were hidden as to alleviate any preconceived notion in regard to the quality of a particular translation engine that the participants might have. The constructed packages were then distributed to each participant, who completed their evaluations on their own means. In these packages, they would find the questions from Appendix F to evaluate. A sample from the Turkish-English package for one text can be found in Appendix H.

The average return time of each package was a little over a week, skewed by a couple of participants that had taken upwards of a month to complete their package. The texts being previously untranslated meant that the participants could not have taken outside assistance by consulting to any other readily available translation.

Challenges surfaced especially on the discovery and selection of participants. It proved to be a challenge to find qualified participants in some languages, which was the main reason why not every language in the tier list was represented in the study. Out of the languages that were represented, Turkish-Mandarin Chinese, Turkish-Uzbek, and Turkish-Kazakh were the most difficult to find qualified participants for. While it is relatively common to find speakers of both Turkish and other Turkic languages (Uzbek and Kazakh), few of these speakers had reputable proof of their language proficiency.

The vast majority of the participants were found from university bodies, current masters' students, alumni, or teaching staff. The remaining few were professional contacts who worked as translators for their respective languages. The full list of anonymous participants and their scoring for each output can be found in Appendix I.

Once all the data was congregated from every participant, the dataset was inspected closely by using statistical analyses. Descriptive statistics – simple attributes of the numerical dataset such as mean, median, range, percentiles, and so on – can provide valuable, yet concise information about the values, their spread, and occurrence frequency in this numerical dataset.

Another, more sophisticated statistical analysis is the use of multivariate linear regressions. A regressive analysis aims to fit a linear expression to a numerical dataset. Using multiple variables when calculating the regression allows the observation of the effect of one particular variable when every other variable is held constant, isolating that one variable's effect. Regressive models are used in statistics for data modeling, and data prediction. A linear expression, if fitting well to the data, allows the researcher to input new values into independent variables and calculate what the result would be for the dependent variable. In the present thesis, it allows the calculation of TQ for custom values of linguistic distance, when keeping other variables such as text type or machine translation tool type constant.

How is it determined whether a linear expression is a good fit for the given numerical dataset? For this, it is important to turn to a value produced from each regression, named the *R-squared* (R^2). Put simply, R^2 is the correlation between the values of the dataset, and the dependent variable values that are derived from the model for the same independent variable values (Devore, 2011). Higher R^2 values mean the linear model fits the given dataset better.

While R^2 is a value that measures the overall fit of a particular regression, the significance of an individual variable, such as linguistic distance is determined by its *p-value* (or the corresponding *t-stat*). Only variables with p-values less than 0.05 (corresponding to a 95% confidence level), and equivalent t-stats above 2.00 would signal a statistically significant effect of the respective variable.

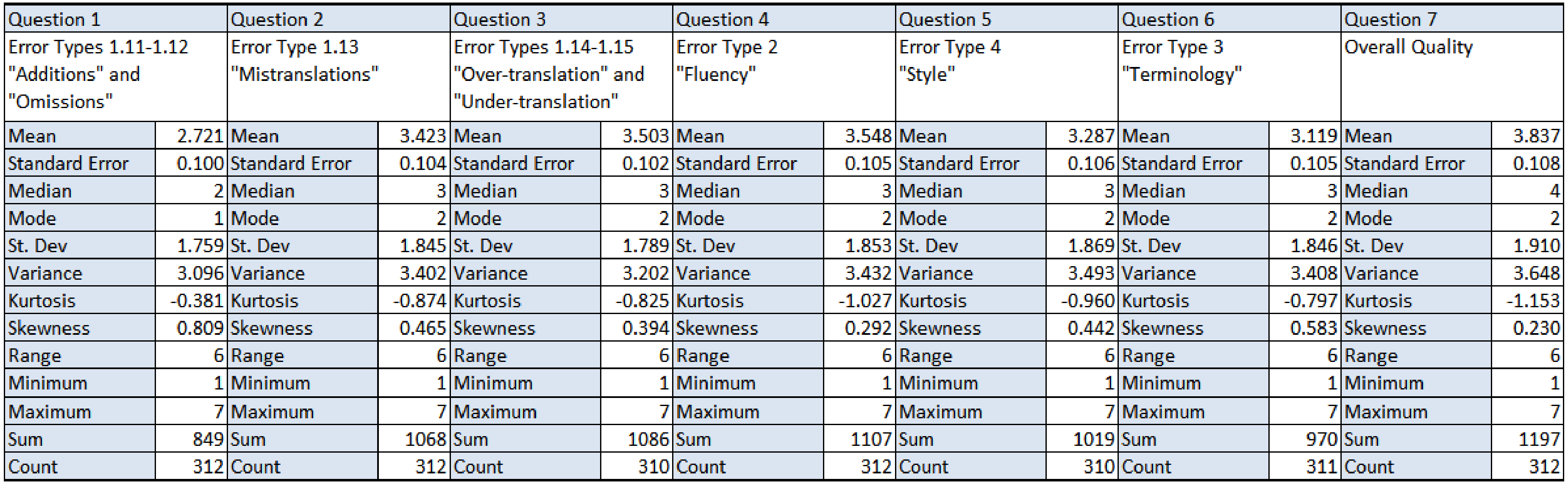
CHAPTER 5

ASSESSING MACHINE TRANSLATION QUALITY BY EXPERIMENTATION

5.1 Descriptive statistics

In order to establish an overview of the responses from the participants, a look at the descriptive statistics of each question is warranted at first. Slight deviance can be seen for Question 1 which asks about additions or omissions, corresponding to the error sub-types 1.11 and 1.12. Compared to other questions, results of Question 1 have a lower mean than others – 2.72 against roughly 3.5 – a lower median of 2 against 3 in others, and a lower mode of 1 against 2 in others. The lower and therefore better scoring of Question 1 implies that translation software might not make as many addition or omission errors as other error types. It is worthy to note the performance of Question 4 in these statistics. Question 4 broadly corresponds to the error type 4 “Style” in the error framework, and observing its mean and variance being close to every other error type was noteworthy. This denotes that at first glance, neural machine translation is not more or less likely to make stylistic errors than any other type. Question 7, which asks about TQ scores, observes the highest mean and median, implying that users tend to regard the translation quality worse than any one particular error type. This makes intuitive sense, as one would expect translations that have errors in different categories would score worse in aggregate than any one of those categories on their own. Standard deviations and variances of each question remain similar to each other. Further, more detailed descriptive statistics can be seen in Table 7 (a more readable version of the table is given in Appendix J).

Table 7. Descriptive Statistics of Survey Results



Question 1		Question 2		Question 3		Question 4		Question 5		Question 6		Question 7	
Error Types 1.11-1.12 "Additions" and "Omissions"		Error Type 1.13 "Mistranslations"		Error Types 1.14-1.15 "Over-translation" and "Under-translation"		Error Type 2 "Fluency"		Error Type 4 "Style"		Error Type 3 "Terminology"		Overall Quality	
Mean	2.721	Mean	3.423	Mean	3.503	Mean	3.548	Mean	3.287	Mean	3.119	Mean	3.837
Standard Error	0.100	Standard Error	0.104	Standard Error	0.102	Standard Error	0.105	Standard Error	0.106	Standard Error	0.105	Standard Error	0.108
Median	2	Median	3	Median	3	Median	3	Median	3	Median	3	Median	4
Mode	1	Mode	2	Mode	2	Mode	2	Mode	2	Mode	2	Mode	2
St. Dev	1.759	St. Dev	1.845	St. Dev	1.789	St. Dev	1.853	St. Dev	1.869	St. Dev	1.846	St. Dev	1.910
Variance	3.096	Variance	3.402	Variance	3.202	Variance	3.432	Variance	3.493	Variance	3.408	Variance	3.648
Kurtosis	-0.381	Kurtosis	-0.874	Kurtosis	-0.825	Kurtosis	-1.027	Kurtosis	-0.960	Kurtosis	-0.797	Kurtosis	-1.153
Skewness	0.809	Skewness	0.465	Skewness	0.394	Skewness	0.292	Skewness	0.442	Skewness	0.583	Skewness	0.230
Range	6	Range	6	Range	6	Range	6	Range	6	Range	6	Range	6
Minimum	1	Minimum	1	Minimum	1	Minimum	1	Minimum	1	Minimum	1	Minimum	1
Maximum	7	Maximum	7	Maximum	7	Maximum	7	Maximum	7	Maximum	7	Maximum	7
Sum	849	Sum	1068	Sum	1086	Sum	1107	Sum	1019	Sum	970	Sum	1197
Count	312	Count	312	Count	310	Count	312	Count	310	Count	311	Count	312

The various text types are compared to one another in how they have scored, given in Figure 4. All text types have scored within one point of each other in every question.

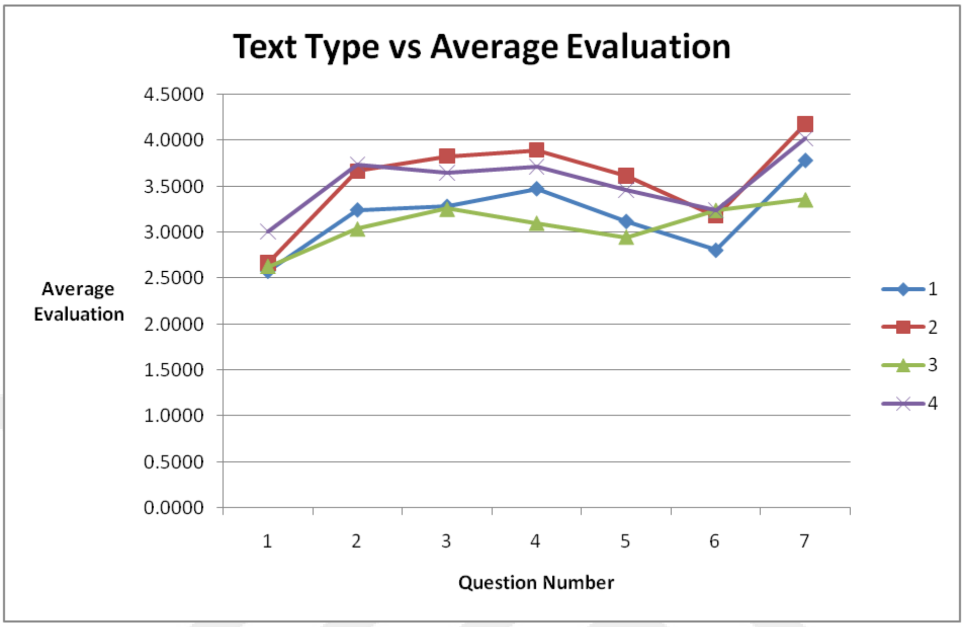


Figure 4. Average evaluation of the four different text types

It seems that Text 2 is the one that scores the highest on the questions as an average of all the machine translation software, suggesting its translations are of the poorest quality (1 being the best score and 7 being the worst score across all questions for the survey). On the other hand, Text 3 seems to score the lowest, initially suggesting that the highly technical, financial text was more easily translated by the software.

Figures 5 and 6 demonstrate in two ways how different machine translation software perform, holding the text type constant. Both of the figures below can be used to compare the performance of machine translation tools.

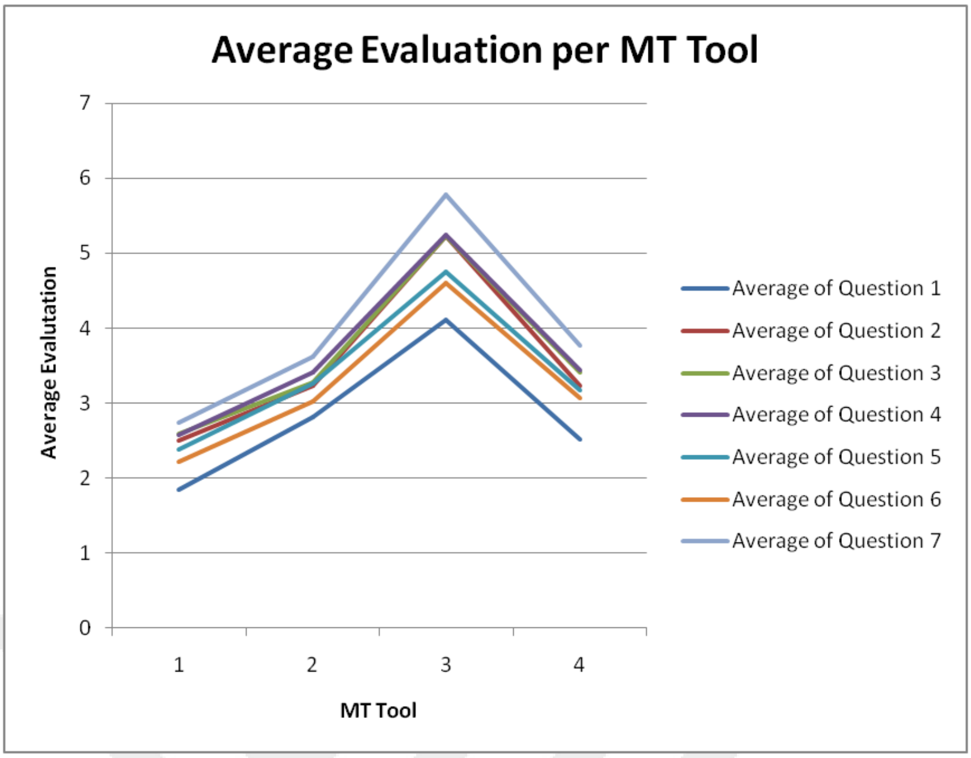


Figure 5. Average evaluation per MT tool

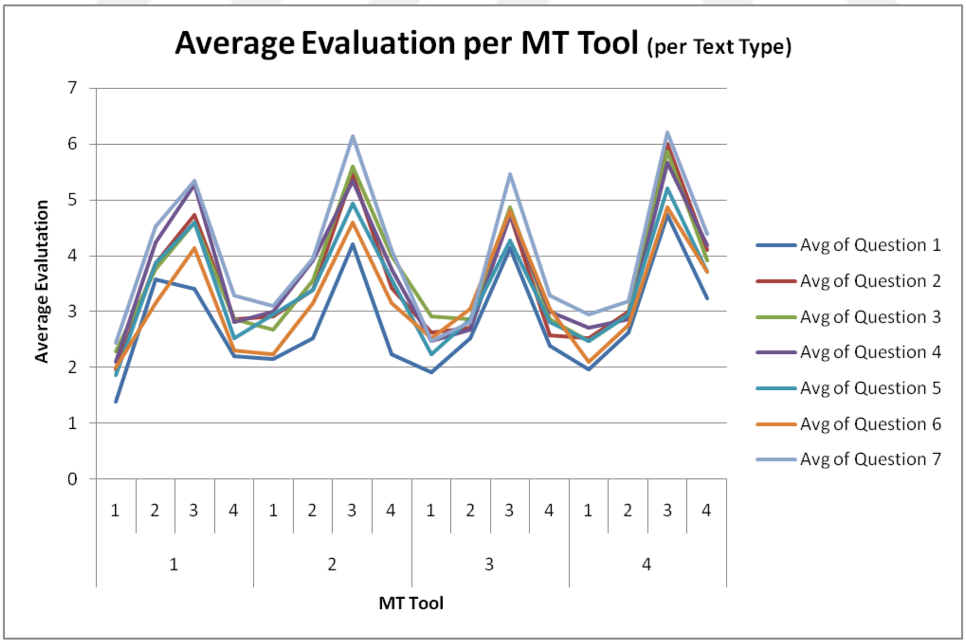


Figure 6. Average evaluation per MT, per text type

While Figure 5 only takes the broad average of machine translation tool performances, Figure 6 divides the tools for each Text Type, in order to see if the performances are consistent across all texts. In both of these graphs, MT_3 (LibreTranslate) stands out as the highest scoring, compared to all other software, average of all questions. In other words, Google Translate can be seen as performing marginally better than Yandex Translate and Windows Translator, while LibreTranslate is the worst performing of the cohort. Poor performance by LibreTranslate can be attributed to previous intuitions related to the use of English as an intermediary language, or use of sentence-based translation structure.

The below illustrations on Figures 7 and 8 show an initial look at the relationship between linguistic distance and the evaluations of questions. With the exception of Question 1, all the other questions seem to follow a trend. Small variances in the less distant languages – below a distance of 0.60 – yield themselves into an increasing, and thereby worsening, evaluations as the languages get more distant above a distance of 0.60. Figure 8 removes the evaluations of Question 1 in order to better illustrate this trend among other questions.

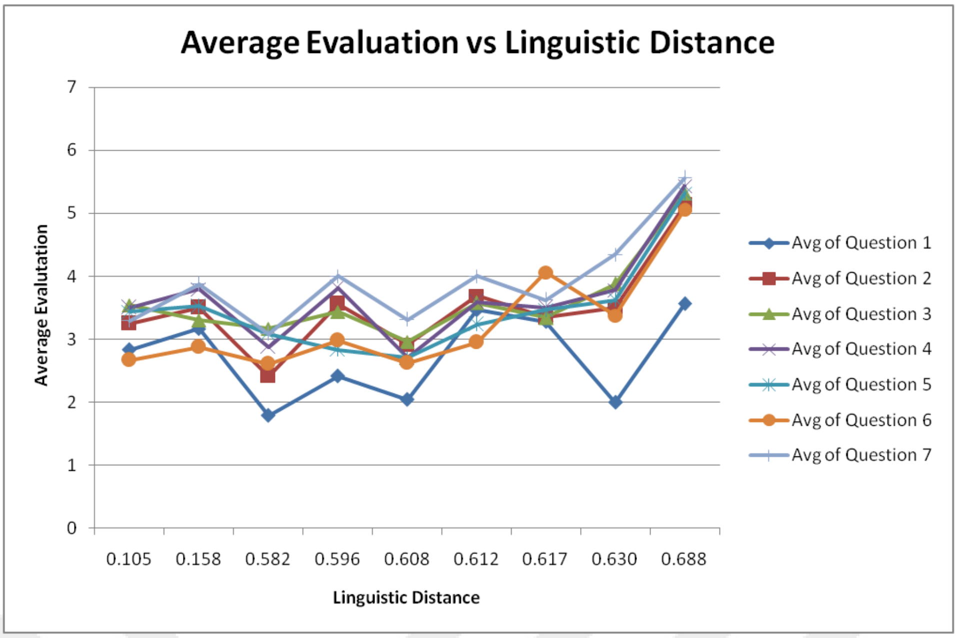


Figure 7. Average evaluation against linguistic distance: All questions

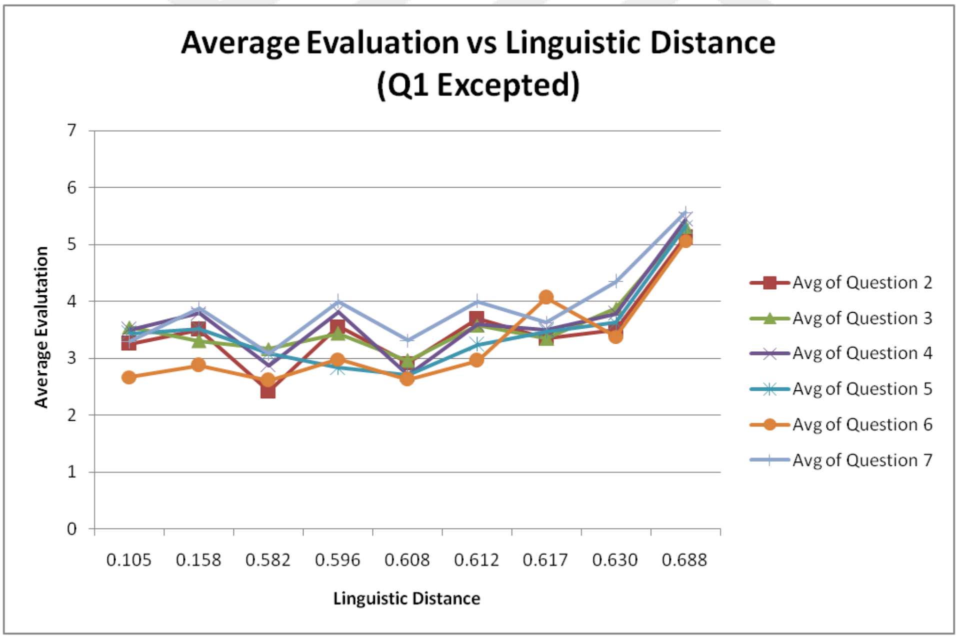


Figure 8. Average evaluation against linguistic distance: Question 1 exempted

This trend relates directly to this thesis's main question. It suggests that, while the existence of additions and omissions in translation might be distributed differently, other error types – as well as overall quality – worsen as languages get more distant, particularly on the latter end of the distance spectrum.

While graphical illustrations are a valuable starting point, they lack the concrete statistical data needed to arrive at trustworthy results. The more robust, and statistically precise, manner of identifying how evaluations change based on different factors, is to run multivariate regressions to elaborate the data.

5.2 Multivariate regressions

Seven regressions were run at first, each taking one question as the dependent variable. In each regression, linguistic distance was taken as the independent variable, alongside of text number and machine translation software as categorical variables, and participant number as a numerical variable. The results from the first set of regressions are given in Appendix K.

When observing the effects of linguistic distance while holding the other variables constant, it was found that for the majority of the questions it did not manage to produce a statistically significant effect. The only exception to this observation was the regression for Question 1, where linguistic distance had a statistically significant effect with a large negative coefficient. This exception was in concordance with the outlying behavior of the average evaluation plot for Question 1. In fact, Questions 1-5 all observed a linguistic distance variable with a negative coefficient, despite most of it being statistically insignificant. This implied that, as linguistic distance increases away from Turkish, the machine translation software were less prone to making errors, which was against intuitions about linguistic distance.

In order to better understand the inverse effect of Questions 1-5, the data itself was more closely inspected. It was found that one particular participant – participant 15 – of Turkish-Kazakh and Turkish-Uzbek language pairs often gave scores of all 1s or 7s for different translations. To see if the regressions were skewed by the evaluations from participant 15, a new set of regressions were run with participant 15 removed, reducing the total number of observations to 283. The resultant regression tables were noteworthy, with 3 out of 7 regressions now producing a statistically significant coefficient at a 95% confidence level and 5 out of 7 regressions at a 90% confidence level for the variable of linguistic distance. Questions 1 and 5 are the exceptions with statistically insignificant effects. Also worth noting is the positive coefficients of 6 out of 7 regressions on linguistic distance, which is in line with previous intuitions. The second set of regressions is given in Appendix L.

To understand which set of regressions suits the dataset better, and whether individual regressions within the sets are a good predictor of their respective question, R^2 values are consulted. Overall, the R^2 values for both sets of regressions remain on the low side for statistical standards, indicating a poor fit. In the first set, questions 1-6 have R^2 values below 0.30 with question 7 being only slightly higher at 0.314. In the second set R^2 values average slightly above 0.30 for the first six questions, while question 7 reaches 0.380. The regressions and exact R^2 values can be found in Appendix K for the first set and Appendix L for the second set.

It can be seen that the R^2 values for the second set are marginally higher than the first set. This would suggest that the regression models of the second set fit slightly better to the dataset. However, R^2 values ranging around 0.30-0.40 is not an indicator that the constructed model explains the data well. In other words, these regressions can predict scores for their respective question only with 30-40% accuracy. Yet, this does not mean the regressions do not provide useful information. When relationship between variables, especially the relationship between linguistic distance and the question result, are statistically significant, valuable conclusions can be drawn.

Focusing attention on the second set alone due to slightly better R^2 values, it can be seen that Questions 2, 6, and 7 showed a statistically significant effect of linguistic distance with magnitudes of 1.513, 2.945, and 1.998 points respectively. To interpret these coefficients suppose two hypothetical languages L_x and L_y; where L_x has perfect similarity to Turkish (a linguistic distance score of 0.00) and L_y has perfect dissimilarity to Turkish (a linguistic distance score of 1.00). A coefficient of 1.513 in Question 2 would mean that Turkish- L_y language pair would score 1.513 points worse on the evaluation on mistranslations compared to Turkish- L_x language pair, when all other factors remain constant.

The coefficient of Question 6 was more pronounced, with a statistically significant value of 2.945 – nearly 3 points out of the 7-point Likert scale. Lastly, the coefficient of Question 7 that indicates overall quality of machine translation with respect to linguistic distance, showed a 1.998 point difference in quality between a language pair of perfect similarity and another of perfect dissimilarity.

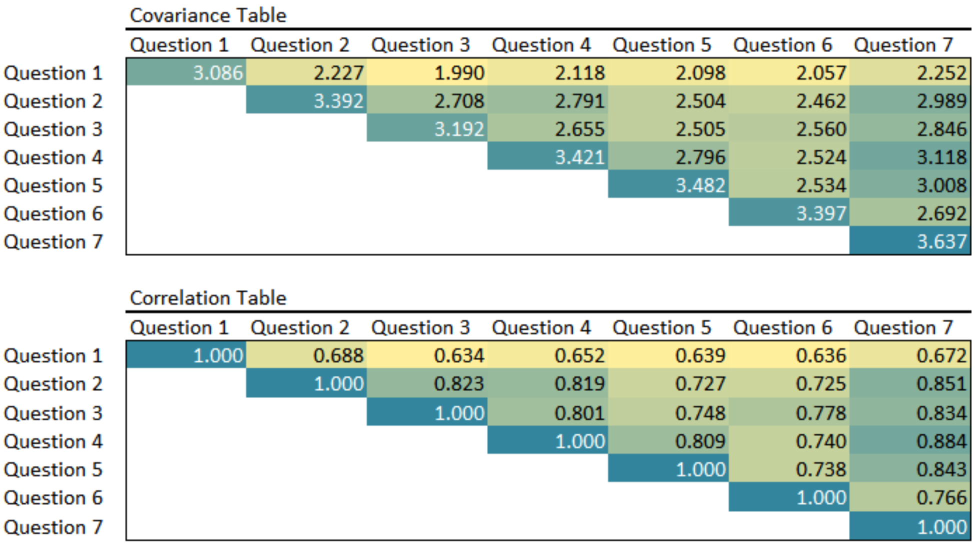
Looking at the other variables, the different text types were most often not statistically significant in the regressions. In regressions where text type did have a statistically significant effect, it was not the same text number that had this effect. For example, while in Question 2 Text 4 had a significant effect, in Questions 3 and

5 it was Text 2 that had the significant coefficient. Counter to the text type, machine translation type was almost always statistically significant, keeping other variables constant. Through all questions Windows Translator and Yandex Translate performed worse than Google Translate, with Windows Translator scoring an average of 0.857 points higher, and Yandex Translate scoring an average of 0.504 points higher. A significant outlier was LibreTranslate, performing an average of 2.488 points worse across Questions 1-6 and 2.975 points worse on Question 7. This solidifies the previous insight that LibreTranslate performed worse than the other translation engines irrespective of the language pair, while Google Translate performed best. The reason behind the poor performance of LibreTranslate might be related to the structure of its neural network – the amount of nodes, layers, training method, or others – or the amount of content the default engine is trained by. However, doing any comparison between LibreTranslate and other tools to see which quality is exactly lacking is infeasible, due to the fact that the other machine translation tools are proprietary and most information relating to their structure or trained content are not publicly available.

By looking at the statistically significant effects of text type, inferences can be made about the parameters that are associated with each error type. In Question 2, relating to error sub-type 1.13 “Mistranslation”, Text 4 observes the only statistically significant effect, and relatively largest coefficient at 0.583 points. Looking at the parameters specifically represented in Text 4, the unique parameters are “PSC - Plural Spread from Cardinal Quantifiers” and “OPK - Null Possessive Licensing Article with Kinship Nouns”. These parameters provide implications that there is a higher likelihood of mistranslations occurring (corresponding to the error type represented in Question 2) in neural machine translation when the values of these parameters differ between the two languages involved. For example, “PSC - Plural Spread from Cardinal Quantifiers” is observed when pluralization occurs with cardinal numbers. While in English there is a “plural spread” meaning that in a phrase such as “two cars” the noun “car” takes on a plural suffix due to the quantifier before it, in Turkish the equivalent phrase “iki araba” exhibits no plural suffix on the noun (equivalent to English “two car”). The other parameter “OPK - Null Possessive Licensing Article with Kinship Nouns” does not differ with the represented languages in this study. Similarly in Questions 3 and 5, with the statistically significant effect of Text 2, the parameter “DGR - Grammaticalized Specified Quantity” could be said to have resulted in a difference. Perhaps it is the case that over-translations or under-translations occur when translating between a language that necessarily has to specify noun quantities and another that does not share this necessity. Confounding factors could have also contributed to the statistical significance of texts. One particularly interesting observation is how Text 3 never seems to cross the significance threshold despite being a financial text, with an abundance of technical terms, jargon, and idiomatic speech. This lack of effect from Text 3 suggests the parameter representation might not be an exhaustive manner to assess text quality by.

From these regressions, the effect of linguistic distance can be observed in relation to participant, text, and machine translation types. Nevertheless, these regressions and the poor R^2 values can potentially be improved by considering relationships hitherto unconsidered: between the results of the questions. Put another way, the evaluation of one question may tend to occur concurrent with evaluations of another question. When variations in evaluations of two questions occur concurrently, the questions are said to have high covariance. When the values of the evaluations between two questions vary towards the same direction, whether negative or positive, the questions are said to have high correlation. To determine these relationships between the questions present in this thesis, covariance and correlation tables are created, and can be found below.

Table 8. Covariance and Correlation Tables between Questions of the Survey



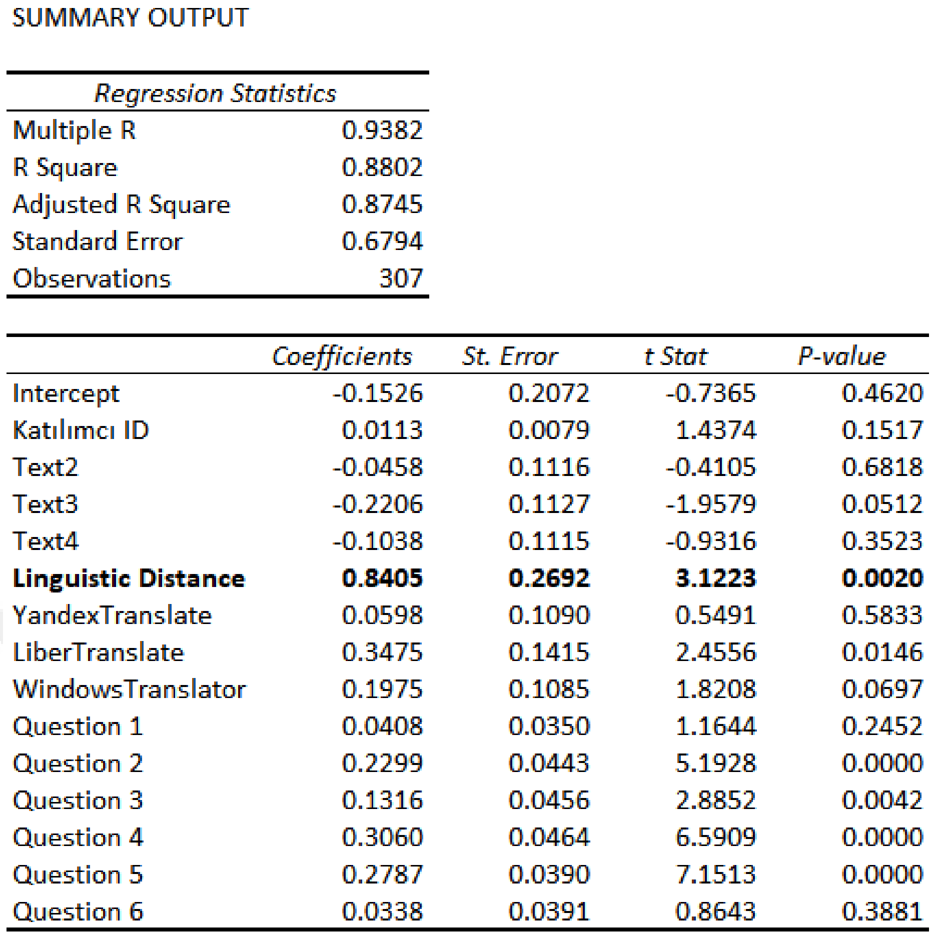
Covariance Table

	Question 1	Question 2	Question 3	Question 4	Question 5	Question 6	Question 7
Question 1	3.086	2.227	1.990	2.118	2.098	2.057	2.252
Question 2		3.392	2.708	2.791	2.504	2.462	2.989
Question 3			3.192	2.655	2.505	2.560	2.846
Question 4				3.421	2.796	2.524	3.118
Question 5					3.482	2.534	3.008
Question 6						3.397	2.692
Question 7							3.637

Correlation Table

	Question 1	Question 2	Question 3	Question 4	Question 5	Question 6	Question 7
Question 1	1.000	0.688	0.634	0.652	0.639	0.636	0.672
Question 2		1.000	0.823	0.819	0.727	0.725	0.851
Question 3			1.000	0.801	0.748	0.778	0.834
Question 4				1.000	0.809	0.740	0.884
Question 5					1.000	0.738	0.843
Question 6						1.000	0.766
Question 7							1.000

The positive covariance values attest to the concurrent variation of the evaluations between the questions. In particular, correlations above 0.5, and in many cases reaching above 0.8, confirm the close relationship between the questions. Therefore, one final regression was done – with all of the observations included – that aimed to comprehensively investigate the results of the survey question that directly corresponds to the present thesis' research question: Question 7 (TQ). In order to account for the ~3.0 point covariance and an average correlation of 0.808 between the other questions and Question 7, the results from the other six questions were included as variables in the final regression. This regression can be seen on Figure 9 below, as well as Appendix M.



SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.9382
R Square	0.8802
Adjusted R Square	0.8745
Standard Error	0.6794
Observations	307

	Coefficients	*St. Error*	*t Stat*	*P-value*
Intercept	-0.1526	0.2072	-0.7365	0.4620
Katılımcı ID	0.0113	0.0079	1.4374	0.1517
Text2	-0.0458	0.1116	-0.4105	0.6818
Text3	-0.2206	0.1127	-1.9579	0.0512
Text4	-0.1038	0.1115	-0.9316	0.3523
Linguistic Distance	**0.8405**	**0.2692**	**3.1223**	**0.0020**
YandexTranslate	0.0598	0.1090	0.5491	0.5833
LiberTranslate	0.3475	0.1415	2.4556	0.0146
WindowsTranslator	0.1975	0.1085	1.8208	0.0697
Question 1	0.0408	0.0350	1.1644	0.2452
Question 2	0.2299	0.0443	5.1928	0.0000
Question 3	0.1316	0.0456	2.8852	0.0042
Question 4	0.3060	0.0464	6.5909	0.0000
Question 5	0.2787	0.0390	7.1513	0.0000
Question 6	0.0338	0.0391	0.8643	0.3881

Figure 9. Comprehensive multivariate regression on linguistic distance versus translation quality

In this regression above, the statistical values are much more promising. The R^2 value is at 0.880, indicating a good fit. To put into words, the regression above would be able to predict the TQ score of a particular text with 88% accuracy, when given the variables above, which include the scores from the other questions. A statistically significant relationship can be observed for linguistic distance, with a t-stat above 2, and p-value below 0.05.

Residuals of predicted TQ scores from the regression in Figure 9 are nearly normally distributed, further supporting the integrity of the model. 34% of the distribution is represented by 0.862 standard deviations in the residuals in the increasing direction from the mean and 0.920 standard deviations in the decreasing direction, compared to 1.000 standard deviations in the normal distribution. Similarly 48% of the distribution is associated with 2.185 standard deviation in the positive direction and 1.988 standard deviation in the negative direction compared to the benchmark of 2.000. The kurtosis of the distribution is 0.7996, compared to the benchmark 3.000 in the perfect normal distribution, suggesting a platykurtic deviance in the residual distribution. A visual representation of residuals can be found in Appendix N.

CHAPTER 6

DISCUSSION

In relation to the research question of this thesis, “How does neural machine translation quality change based on the linguistic distance between the source and target languages?” the results show that there exists a statistically significant effect of linguistic distance on neural machine translation quality. The positive coefficients of the linguistic distance variable found in the second set of regressions align well with the coefficient obtained from the comprehensive regression for Question 7. This comprehensive regression, given in Figure 9 above, notably scores a satisfying R^2 value of 0.880. Broadly speaking, the final regression model can explain 88% of the variation in the overall quality of a neural machine translation output, when given the other variables. Specifically, on the variable of linguistic distance, the regressions serve to prove that, holding other variables such as text type, participant type, and machine translation software tool constant, an inverse relationship between linguistic distance and neural machine translation quality exists. Keeping in mind that positive coefficients imply poorer performance in this thesis, the regression above claims that a translation between two languages of perfect dissimilarity would perform 0.840 points worse than a translation between two languages of perfect similarity.

It is important to stress exactly what this result is, suggested by the way the final regression has been constructed. When interpreting the result for the variable of linguistic distance, other variables that were controlled for need to be interpreted as being held constant. For example, it can be said that the variable text type does not have a significant effect on machine translation quality, by looking at the corresponding t-stat for that variable. However, when interpreting the variable of linguistic distance, it must be said that the effect linguistic distance has is irrespective of the text type chosen. Linguistic distance a statistically significant effect when controlling for the text type, participant type, translation tool, and even frequency of different types of errors. This insight has to influence the interpretation of the overall regression because of the correlated questions involved.

Whereas previously the individual regression for Question 7 in the second set had a poor R^2 value of 0.397, the overall regression had an R^2 value of 0.880, due to the introduction of the results from other questions as variables in the final regression. It would seem that roughly 60% of the variation in the evaluations for Question 7 can be explained by the variations in other questions. The R^2 value changes by 2% when the variable of linguistic distance is removed, supporting the intuition that most of the variance in data is explained by the covariance with the evaluations of other questions. This makes intuitive sense. Whether or not a text has any errors at all has a profound effect on whether that text is regarded as having high or low quality. This means that the effect of linguistic distance has to be contextualized. Therefore, it must be said that linguistic distance still has an inverse and significant effect on overall quality of a machine translation output, irrespective of the amount errors that may exist in the translation – even if the amount of errors is zero. Put another way, a text that might be otherwise error-free might still be perceived as having poorer quality because of the underlying linguistic distance between the source and target languages.

While there is a statistically significant effect, interpreting the magnitude of this result is subjective. On one hand, it does affirm that as languages get more linguistically distant, the neural machine translation quality drops. On the other hand, it could be said that a 0.840 points difference between perfect similarity and dissimilarity is not impactful for practical purposes. Suppose we take two language pairs from this thesis directly: Turkish-Uzbek against Turkish-English. The linguistic distance scores are 0.056 and 0.500, respectively. According to the results of the final regression, holding other variables constant, the overall quality of a Turkish-English translation would only be 0.373 points worse than that of the Turkish-Uzbek translation on a 7-point Likert scale. This effect could also be affected by other confounding factors, such as network structure or breadth of training material. In future studies, these findings could be supported by research on translation time for texts of equivalent length. If machine translation does in fact take longer, which is likely to be within milliseconds, it could further imply that linguistic distance might be related to poorer performance, or higher computational resources required. Alternatively, future studies could employ multiple open-source translation tools, take the time to vary all of the higher-order parameters of the same neural network, or train the same open-source tool on differing sizes of corpora with a similar qualitative evaluation and compare them. By doing so, the question of whether the effect of linguistic distance can be caused by confounding variables could be answered.

The implications of these results depend on exactly how they are manifested. Machine translation process differences of a few milliseconds is not likely to matter in all but the most particular situations, such as high-frequency financial trading based on countries’ economic statements published in other languages. When the perceived quality deterioration manifests but the effect is minimal enough, it might not matter in most situations, except when the content is particularly sensitive, such as high-level political communication. Whether this deterioration based on linguistic distance exists for human translation as well can be confirmed by studying the process of multilingual translators with similar studies. Despite the existence of an effect, even when controlling for present errors, the magnitude of this effect limits the consequences on practical use of translation. While being a significant variable, the coefficient of linguistic distance in relation to the whole 7-point Likert scale remains small. Users of machine translation could well within reason regard this difference as professionally acceptable.

CHAPTER 7

CONCLUSION

The question posed by this thesis was “How does neural machine translation quality change based on the linguistic distance between the source and target languages?” It was hypothesized that the more distant languages are to one another, quality in translation would worsen, since larger distance would provide a greater challenge to translation. In order to assess this relationship, a survey was constructed, in which participants rated machine translations of four, short, original texts. To select the participants, a linguistic distance matrix was constructed between 33 languages first, using the latest development in linguistic distance research, the Parametric Comparison Method. Languages were then taken and divided up into six tiers based on their relative distance to the selected source language of Turkish. Two to four participants were found for each tier to participate in the survey. Seven different questions, six of them relating to an error type and one relating to the overall quality, were asked to each participant; each question was evaluated on a 7-point scale. Their results were gathered together, and multivariate regressions were run to assess them, alongside their descriptive statistics. After poor fits of the individual regressions, the questions were introduced as variables into a final regression that measured neural machine translation quality across all variables. Based on this final regression, linguistic distance was found to have a statistically significant effect on machine translation quality, when all other variables are held constant.

The present thesis confirms that there exists an inverse relationship between linguistic distance between languages and neural machine translation quality, thus proving the null hypothesis. The more linguistically distant languages are, the worse machine translation quality gets. This inverse relationship is found to exist across multiple text types and machine translation tools, and even when the number of other error types present in the translations is controlled for. Despite these insights, there are reasons to believe this inverse effect of linguistic distance may not be detrimental to daily use of neural machine translation. A worse performance of 0.840 points on a 7-point scale between a language pair of perfect similarity and another of perfect dissimilarity could be regarded as acceptable, especially when language pairs will realistically be closer to each other than a perfect similarity/dissimilarity duality.

There are a number of limitations in this study that can be improved upon. In relation to the underlying methodology employed to find linguistic distance, PCM has opportunities to be modified to achieve more robust results, since it is a relatively recent method. The significance testing of PCM-0 in this thesis serves to prove this fact by showing that parametric selection could be tuned further in order to accommodate different language pairings, such as Turkic-Finnic languages or Turkish-Japanese. The addition of more languages would necessarily introduce new challenges and opportunities to improve parameter selection. Likely due to the limited number of languages represented, the resultant tree from PCM-0J still deviated from expected placements with the connection of Semitic languages into the IE family before Celtic languages.

In regard to the surveying method, different aspects could be improved. The clearest improvement would be to increase the number of participants, texts, and neural machine translation tools. By doing so, the intended scale effect can be observed better, where subjectivity of one individual participant would have an even lesser effect on the overall study. This way, incidents observed in the present thesis, such as participant 15 skewing the results can be mitigated better. It would also be ideal to have participants for every single language of the linguistic distance table. This was especially apparent in the present thesis' lack of qualified participants who speak Uralic languages and Turkish, which would represent an additional node of the language trees. Perhaps financial incentives can be offered to participants to increase their willingness to join, and maintain that willingness through longer, more numerous texts. In addition, the study can be replicated using different source languages to see if the results remain consistent.

It also remains possible to have human evaluations compared against automated evaluation models. Using BLEU, FEMTI or an alternative automatic evaluation model provides for an interesting opportunity to see if results stay consistent with qualitative evaluation. Doing so would require reliable human translations to compare machine translation performance against, and it would be preferable to consult participants who are not included in qualitative evaluation, to avoid familiarity with presented machine translation outputs.

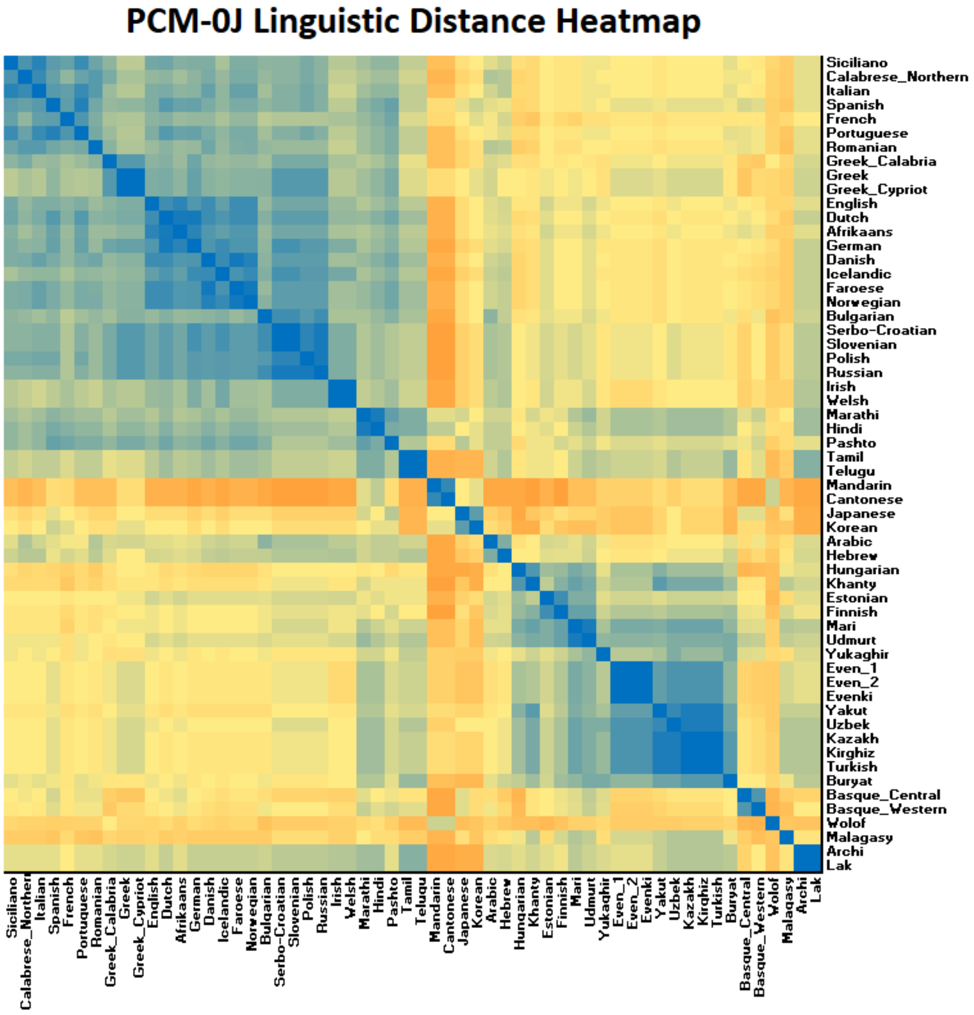
There are opportunities for further research in connecting the results of this thesis to other qualities of neural machine translation performance. In order to better understand the area of inquiry relating to machine translation accommodation of challenges presented by linguistic distance, further studies can be done to connect machine translation performance in process times and computer resources expended to linguistic distance.

The two key areas of interest of this thesis, under its research question were about difficulty of translation between distant languages, and potential professional concerns of users of machine translation with distant languages. According to the results of this thesis, there is a signal that there could be a greater effort required when an individual or software attempts to translate between languages of greater linguistic distance; since the inverse relationship manifested as statistically significant. Nevertheless, users of machine translation software need not have worries when using machine translation tools to translate between two distant languages, should they consider the effect of the inverse relationship to not be detrimental.

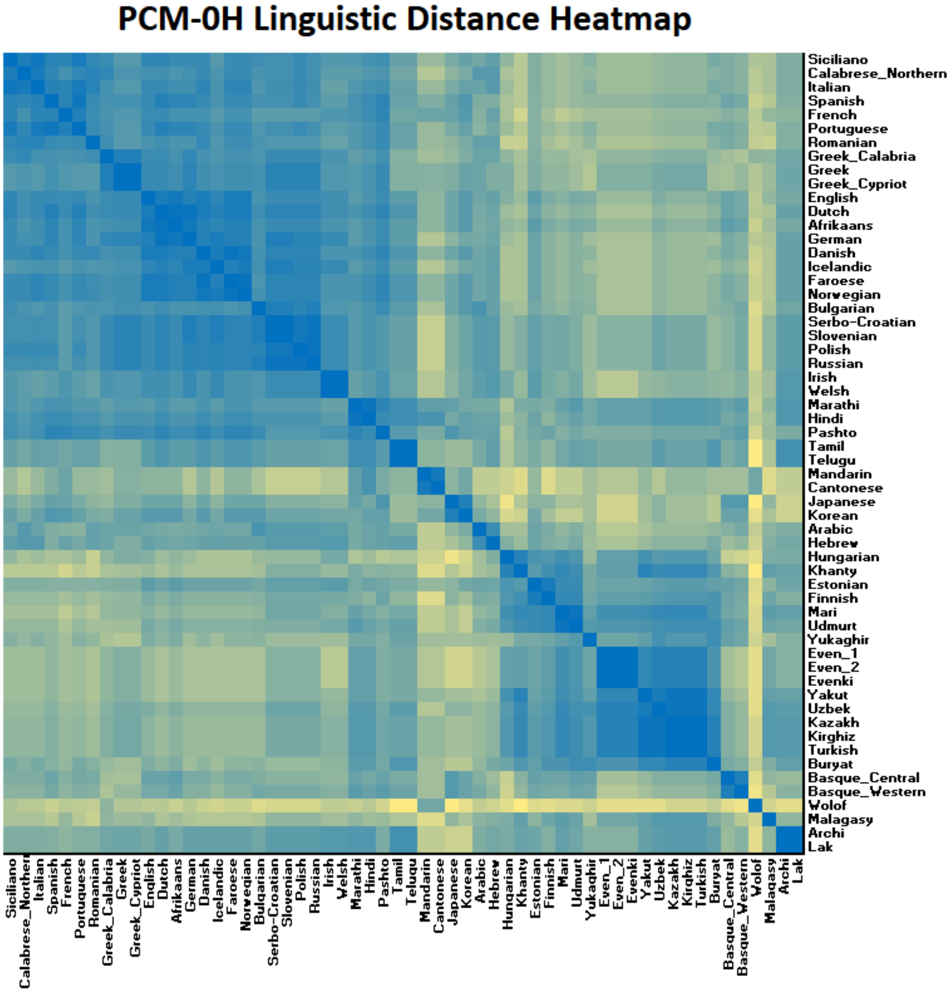
APPENDIX A

HEATMAPS OF PCM-0 AND PCM-1 SYNTACTIC METHODS

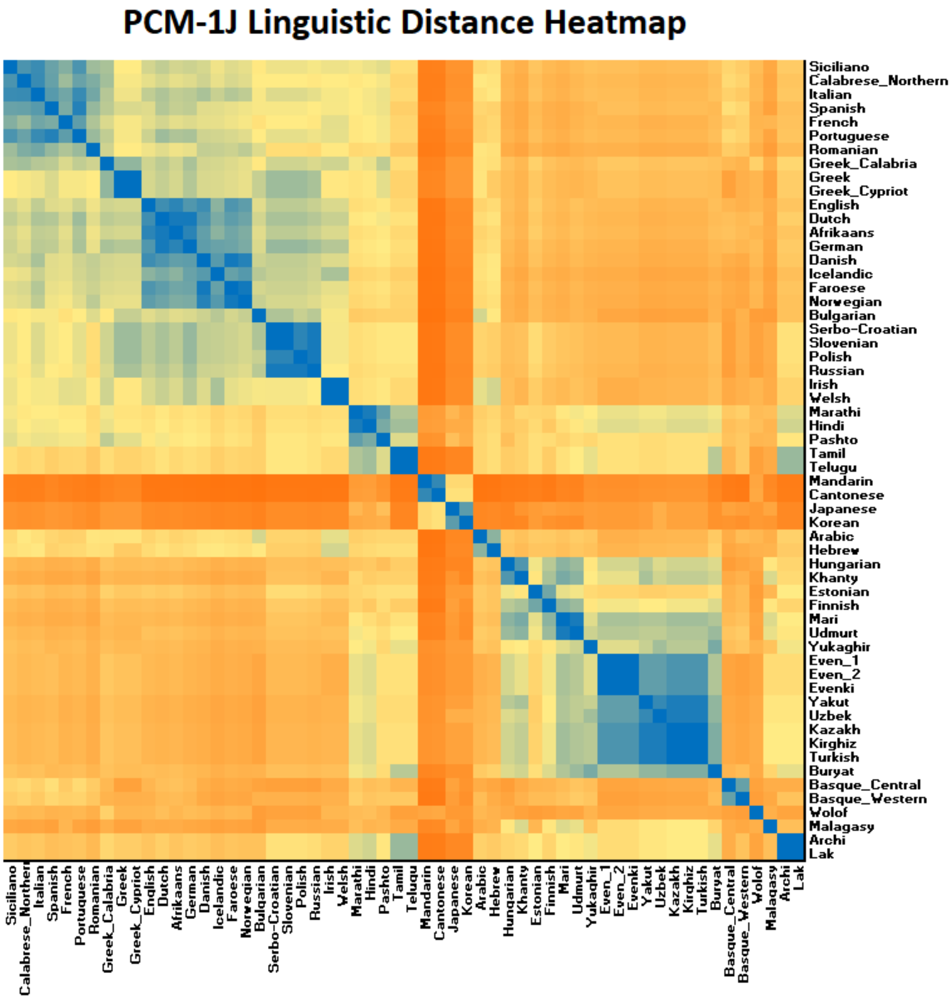
PCM-0J Implied Relationship Syntactic Linguistic Distance Heatmap by Jaccard Distance Metric



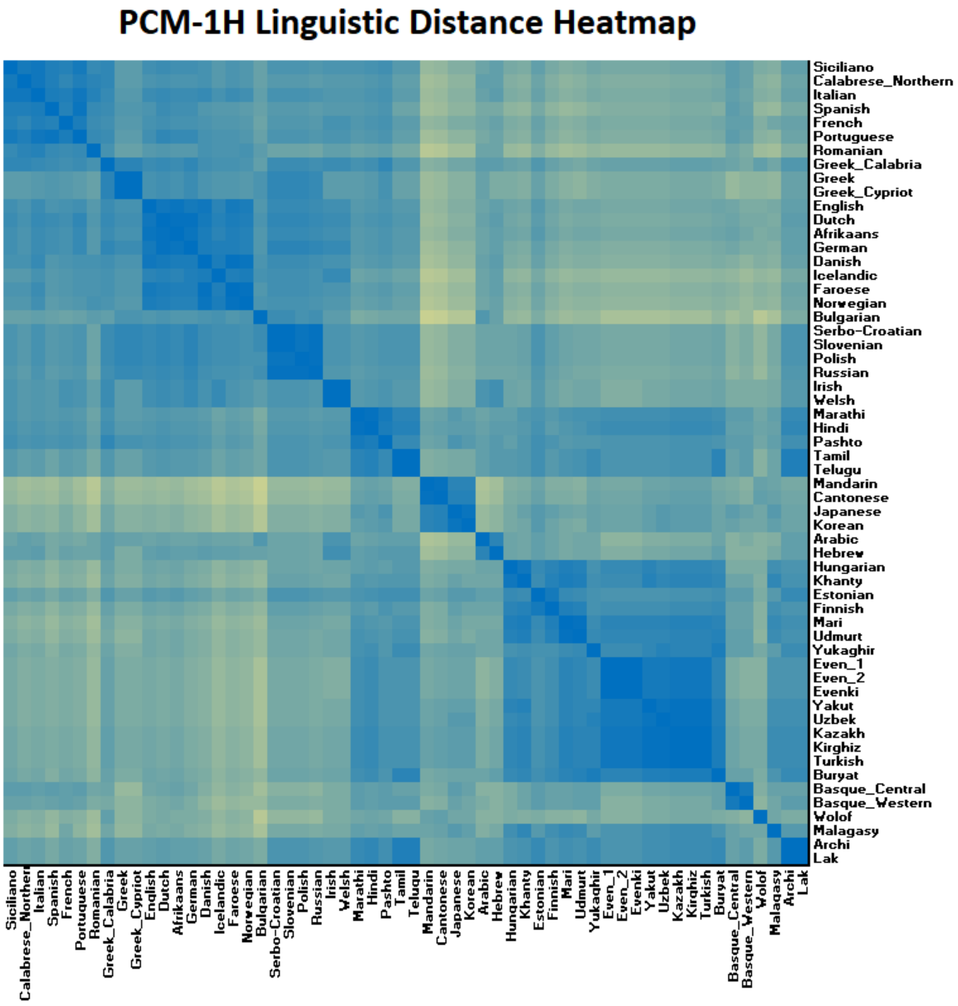
PCM-0H Implied Relationship Syntactic Linguistic Distance Heatmap by Hamming Distance Metric



PCM-1J Unimplied Relationship Syntactic Linguistic Distance Heatmap by Jaccard Distance Metric

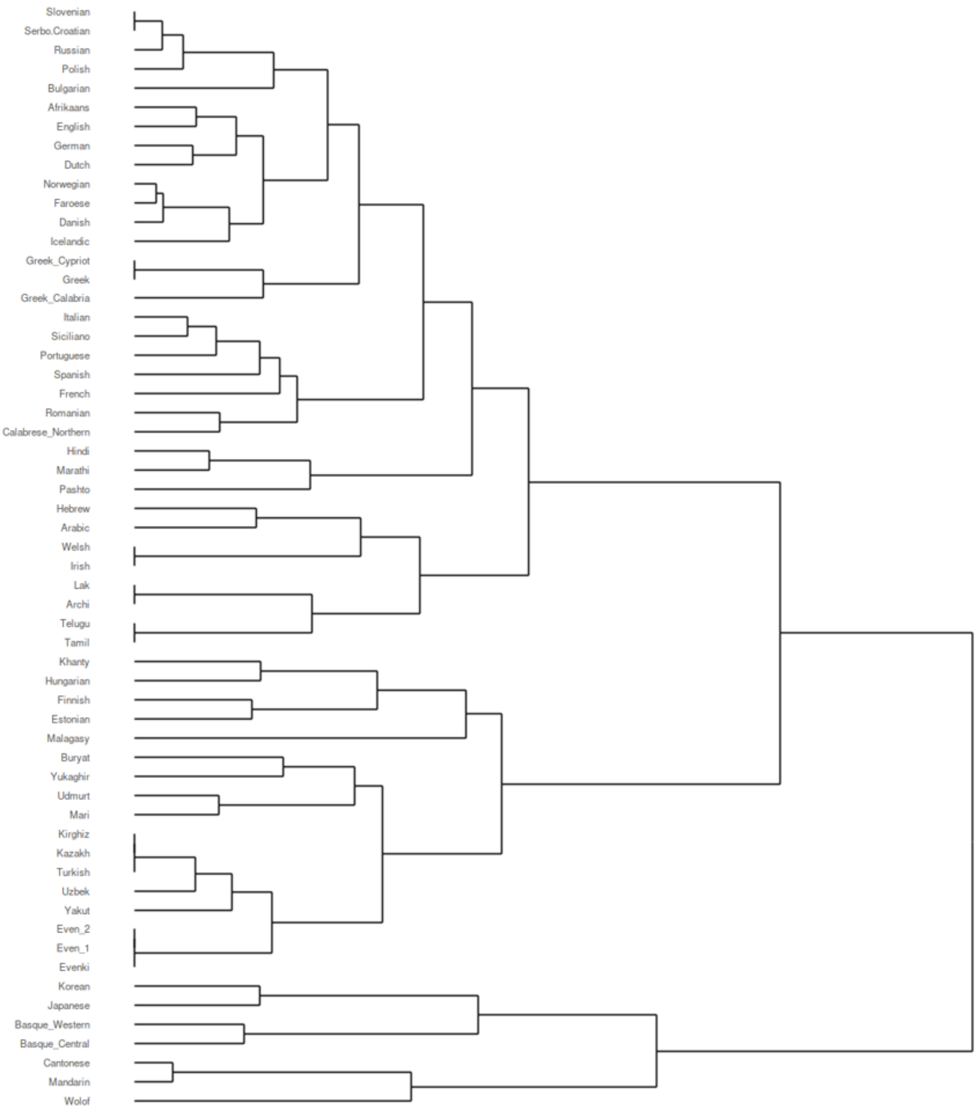


PCM-1H Unimplied Relationship Syntactic Linguistic Distance Heatmap by Hamming Distance Metric


PCM-1H Linguistic Distance Heatmap

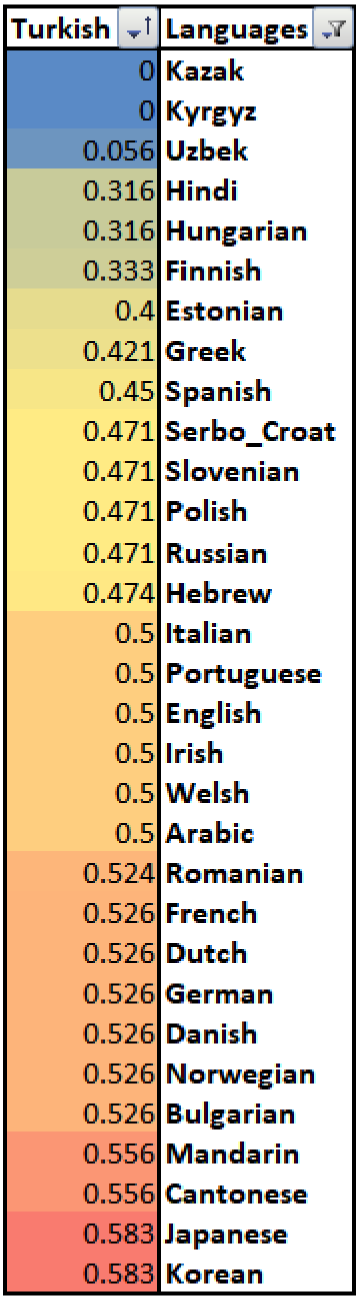
APPENDIX B

PCM-0J SYNTACTIC METHOD CLUSTERED LANGUAGE TREE



APPENDIX C

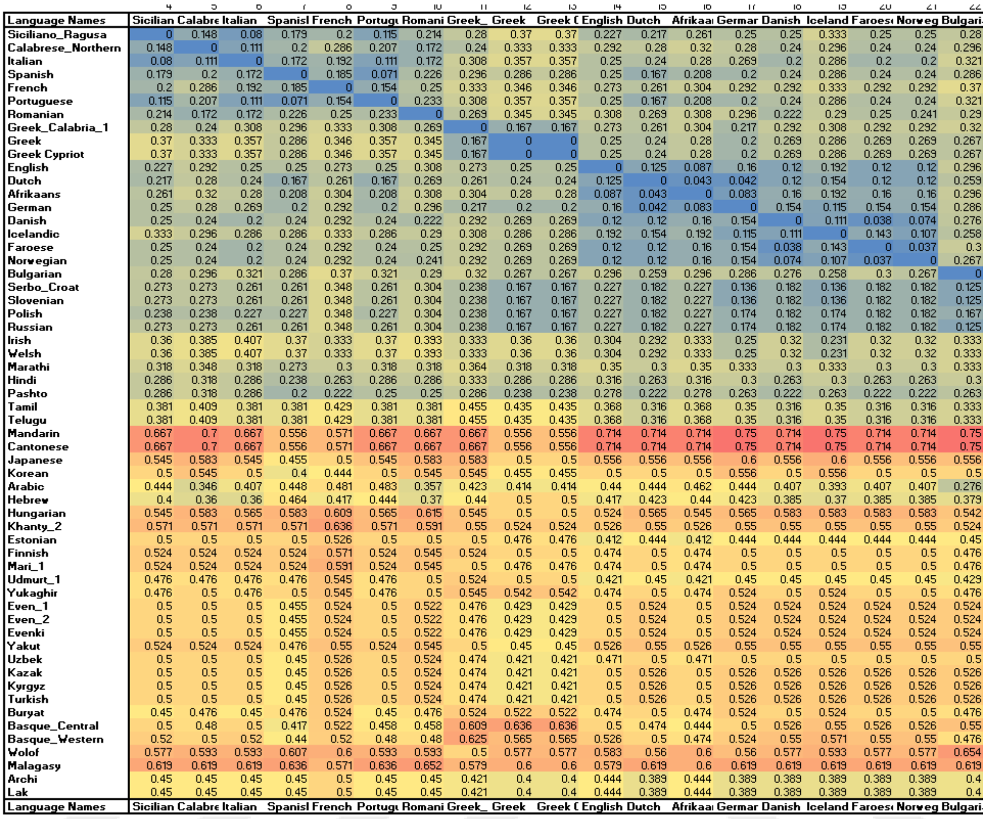
LINGUISTIC DISTANCES OF SELECTED LANGUAGES TO TURKISH



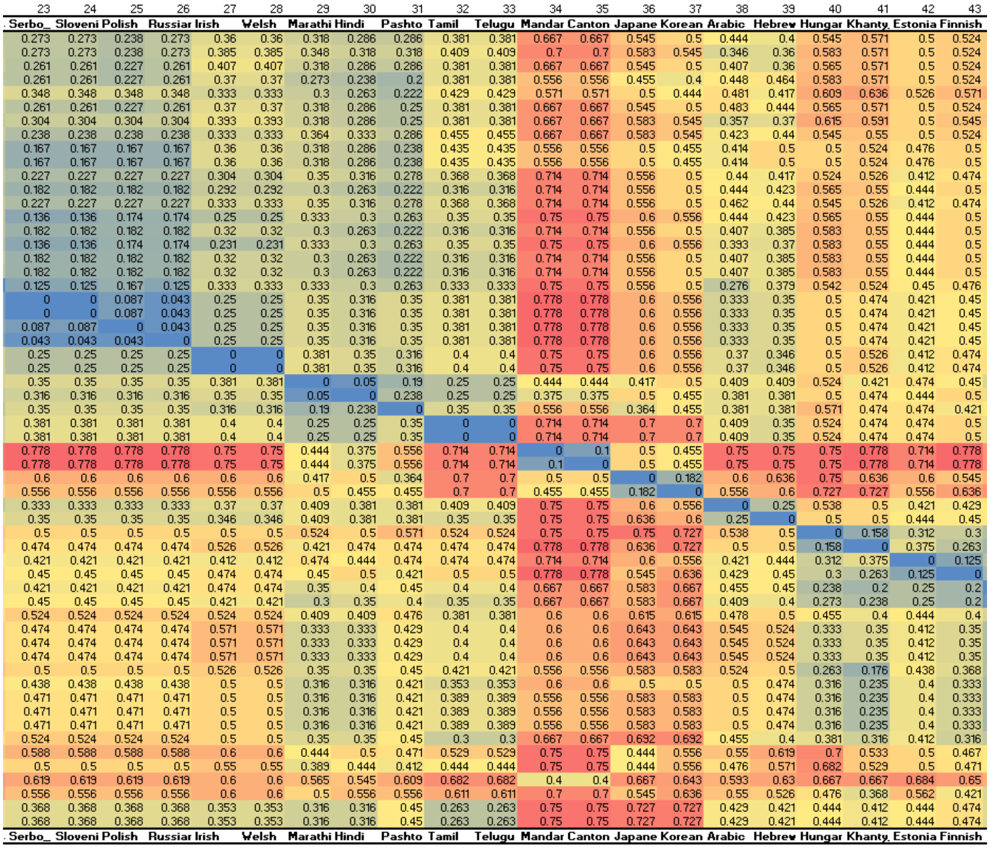
Turkish	Languages
0	Kazak
0	Kyrgyz
0.056	Uzbek
0.316	Hindi
0.316	Hungarian
0.333	Finnish
0.4	Estonian
0.421	Greek
0.45	Spanish
0.471	Serbo_Croat
0.471	Slovenian
0.471	Polish
0.471	Russian
0.474	Hebrew
0.5	Italian
0.5	Portuguese
0.5	English
0.5	Irish
0.5	Welsh
0.5	Arabic
0.524	Romanian
0.526	French
0.526	Dutch
0.526	German
0.526	Danish
0.526	Norwegian
0.526	Bulgarian
0.556	Mandarin
0.556	Cantonese
0.583	Japanese
0.583	Korean

APPENDIX D

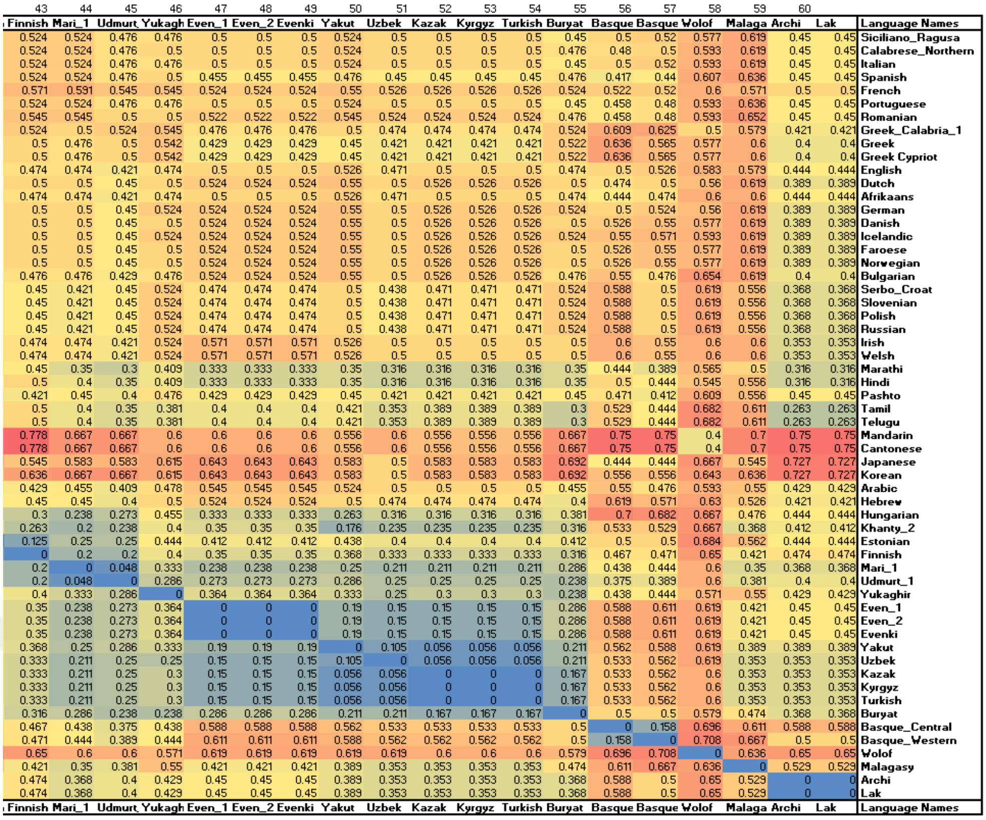
PCM-0J SYNTACTIC LINGUISTIC DISTANCE MATRIX



	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
Language Names	Sicilian	Calabre	Italian	Spanisl	French	Portugu	Romani	Greek_	Greek	Greek (	English	Dutch	Afrikaa	Germar	Danish	Iceland	Faroes	Norweg	Bulgari
Siciliano_Ragusa	0	0.148	0.08	0.179	0.2	0.115	0.214	0.28	0.37	0.37	0.227	0.217	0.261	0.25	0.25	0.333	0.25	0.25	0.28
Calabrese_Northern	0.148	0	0.111	0.2	0.286	0.207	0.172	0.24	0.333	0.333	0.292	0.28	0.32	0.28	0.24	0.296	0.24	0.24	0.296
Italian	0.08	0.111	0	0.172	0.192	0.111	0.172	0.308	0.357	0.357	0.25	0.24	0.28	0.269	0.2	0.286	0.2	0.2	0.321
Spanish	0.179	0.2	0.172	0	0.185	0.071	0.226	0.296	0.286	0.286	0.25	0.167	0.208	0.2	0.24	0.286	0.24	0.24	0.286
French	0.2	0.286	0.192	0.185	0	0.154	0.25	0.333	0.346	0.346	0.273	0.261	0.304	0.292	0.292	0.333	0.292	0.292	0.37
Portuguese	0.115	0.207	0.111	0.071	0.154	0	0.233	0.308	0.357	0.357	0.25	0.167	0.208	0.2	0.24	0.286	0.24	0.24	0.321
Romanian	0.214	0.172	0.172	0.226	0.25	0.233	0	0.269	0.345	0.345	0.308	0.269	0.308	0.296	0.222	0.29	0.25	0.241	0.29
Greek_Calabria_1	0.28	0.24	0.308	0.296	0.333	0.308	0.269	0	0.167	0.167	0.273	0.261	0.304	0.217	0.292	0.308	0.292	0.292	0.32
Greek	0.37	0.333	0.357	0.286	0.346	0.357	0.345	0.167	0	0	0.25	0.24	0.28	0.2	0.269	0.286	0.269	0.269	0.267
Greek Cypriot	0.37	0.333	0.357	0.286	0.346	0.357	0.345	0.167	0	0	0.25	0.24	0.28	0.2	0.269	0.286	0.269	0.269	0.267
English	0.227	0.292	0.25	0.25	0.273	0.25	0.308	0.273	0.25	0.25	0	0.125	0.087	0.16	0.12	0.192	0.12	0.12	0.296
Dutch	0.217	0.28	0.24	0.167	0.261	0.167	0.269	0.261	0.24	0.24	0.125	0	0.043	0.042	0.12	0.154	0.12	0.12	0.259
Afrikaans	0.261	0.32	0.28	0.208	0.304	0.208	0.308	0.304	0.28	0.28	0.087	0.043	0	0.083	0.16	0.192	0.16	0.16	0.296
German	0.25	0.28	0.269	0.2	0.292	0.2	0.296	0.217	0.2	0.2	0.16	0.042	0.083	0	0.154	0.115	0.154	0.154	0.286
Danish	0.25	0.24	0.2	0.24	0.292	0.24	0.222	0.292	0.269	0.269	0.12	0.12	0.16	0.154	0	0.111	0.038	0.074	0.276
Icelandic	0.333	0.296	0.286	0.286	0.333	0.286	0.29	0.308	0.286	0.286	0.192	0.154	0.192	0.115	0.111	0	0.143	0.107	0.258
Faroese	0.25	0.24	0.2	0.24	0.292	0.24	0.25	0.292	0.269	0.269	0.12	0.12	0.16	0.154	0.038	0.143	0	0.037	0.3
Norwegian	0.25	0.24	0.2	0.24	0.292	0.24	0.241	0.292	0.269	0.269	0.12	0.12	0.16	0.154	0.074	0.107	0.037	0	0.267
Bulgarian	0.28	0.296	0.321	0.286	0.37	0.321	0.29	0.32	0.267	0.267	0.296	0.259	0.296	0.286	0.276	0.258	0.3	0.267	0
Serbo_Croat	0.273	0.273	0.261	0.261	0.348	0.261	0.304	0.238	0.167	0.167	0.227	0.182	0.227	0.136	0.182	0.136	0.182	0.182	0.125
Slovenian	0.273	0.273	0.261	0.261	0.348	0.261	0.304	0.238	0.167	0.167	0.227	0.182	0.227	0.136	0.182	0.136	0.182	0.182	0.125
Polish	0.238	0.238	0.227	0.227	0.348	0.227	0.304	0.238	0.167	0.167	0.227	0.182	0.227	0.174	0.182	0.174	0.182	0.182	0.167
Russian	0.273	0.273	0.261	0.261	0.348	0.261	0.304	0.238	0.167	0.167	0.227	0.182	0.227	0.174	0.182	0.174	0.182	0.182	0.125
Irish	0.36	0.385	0.407	0.37	0.333	0.37	0.393	0.333	0.36	0.36	0.304	0.292	0.333	0.25	0.32	0.231	0.32	0.32	0.333
Welsh	0.36	0.385	0.407	0.37	0.333	0.37	0.393	0.333	0.36	0.36	0.304	0.292	0.333	0.25	0.32	0.231	0.32	0.32	0.333
Marathi	0.318	0.348	0.318	0.273	0.3	0.318	0.318	0.364	0.318	0.318	0.35	0.3	0.35	0.333	0.3	0.333	0.3	0.3	0.333
Hindi	0.286	0.318	0.286	0.238	0.263	0.286	0.286	0.333	0.286	0.286	0.316	0.263	0.316	0.3	0.263	0.3	0.263	0.263	0.3
Pashto	0.286	0.318	0.286	0.2	0.222	0.25	0.25	0.286	0.238	0.238	0.278	0.222	0.278	0.263	0.222	0.263	0.222	0.222	0.263
Tamil	0.381	0.409	0.381	0.381	0.429	0.381	0.381	0.455	0.435	0.435	0.368	0.316	0.368	0.35	0.316	0.35	0.316	0.316	0.333
Telugu	0.381	0.409	0.381	0.381	0.429	0.381	0.381	0.455	0.435	0.435	0.368	0.316	0.368	0.35	0.316	0.35	0.316	0.316	0.333
Mandarin	0.667	0.7	0.667	0.556	0.571	0.667	0.667	0.667	0.556	0.556	0.714	0.714	0.714	0.75	0.714	0.75	0.714	0.714	0.75
Cantonese	0.667	0.7	0.667	0.556	0.571	0.667	0.667	0.667	0.556	0.556	0.714	0.714	0.714	0.75	0.714	0.75	0.714	0.714	0.75
Japanese	0.545	0.583	0.545	0.455	0.5	0.545	0.583	0.583	0.5	0.5	0.556	0.556	0.556	0.6	0.556	0.6	0.556	0.556	0.556
Korean	0.5	0.545	0.5	0.4	0.444	0.5	0.545	0.545	0.455	0.455	0.5	0.5	0.5	0.556	0.5	0.556	0.5	0.5	0.5
Arabic	0.444	0.346	0.407	0.448	0.481	0.483	0.357	0.423	0.414	0.414	0.44	0.444	0.462	0.444	0.407	0.393	0.407	0.407	0.276
Hebrew	0.4	0.36	0.36	0.464	0.417	0.444	0.37	0.44	0.5	0.5	0.417	0.423	0.44	0.423	0.385	0.37	0.385	0.385	0.379
Hungarian	0.545	0.583	0.565	0.583	0.609	0.565	0.615	0.545	0.5	0.5	0.524	0.565	0.545	0.565	0.583	0.583	0.583	0.583	0.542
Khanty_2	0.571	0.571	0.571	0.571	0.636	0.571	0.591	0.55	0.524	0.524	0.526	0.55	0.526	0.55	0.55	0.55	0.55	0.55	0.524
Estonian	0.5	0.5	0.5	0.5	0.526	0.5	0.5	0.5	0.476	0.476	0.412	0.444	0.412	0.444	0.444	0.444	0.444	0.444	0.45
Finnish	0.524	0.524	0.524	0.524	0.571	0.524	0.545	0.524	0.5	0.5	0.474	0.5	0.474	0.5	0.5	0.5	0.5	0.5	0.476
Mari_1	0.524	0.524	0.524	0.524	0.591	0.524	0.545	0.5	0.476	0.476	0.474	0.5	0.474	0.5	0.5	0.5	0.5	0.5	0.476
Udmurt_1	0.476	0.476	0.476	0.476	0.545	0.476	0.5	0.524	0.5	0.5	0.421	0.45	0.421	0.45	0.45	0.45	0.45	0.45	0.429
Yukaghir	0.476	0.5	0.476	0.5	0.545	0.476	0.5	0.545	0.542	0.542	0.474	0.5	0.474	0.524	0.5	0.524	0.5	0.5	0.476
Even_1	0.5	0.5	0.5	0.455	0.524	0.5	0.522	0.476	0.429	0.429	0.5	0.524	0.5	0.524	0.524	0.524	0.524	0.524	0.524
Even_2	0.5	0.5	0.5	0.455	0.524	0.5	0.522	0.476	0.429	0.429	0.5	0.524	0.5	0.524	0.524	0.524	0.524	0.524	0.524
Evenki	0.5	0.5	0.5	0.455	0.524	0.5	0.522	0.476	0.429	0.429	0.5	0.524	0.5	0.524	0.524	0.524	0.524	0.524	0.524
Yakut	0.524	0.524	0.524	0.476	0.55	0.524	0.545	0.5	0.45	0.45	0.526	0.55	0.526	0.55	0.55	0.55	0.55	0.55	0.55
Uzbek	0.5	0.5	0.5	0.45	0.526	0.5	0.524	0.474	0.421	0.421	0.471	0.5	0.471	0.5	0.5	0.5	0.5	0.5	0.5
Kazak	0.5	0.5	0.5	0.45	0.526	0.5	0.524	0.474	0.421	0.421	0.5	0.526	0.5	0.526	0.526	0.526	0.526	0.526	0.526
Kyrgyz	0.5	0.5	0.5	0.45	0.526	0.5	0.524	0.474	0.421	0.421	0.5	0.526	0.5	0.526	0.526	0.526	0.526	0.526	0.526
Turkish	0.5	0.5	0.5	0.45	0.526	0.5	0.524	0.474	0.421	0.421	0.5	0.526	0.5	0.526	0.526	0.526	0.526	0.526	0.526
Buryat	0.45	0.476	0.45	0.476	0.524	0.45	0.476	0.524	0.522	0.522	0.474	0.5	0.474	0.524	0.5	0.524	0.5	0.5	0.476
Basque_Central	0.5	0.48	0.5	0.417	0.522	0.458	0.458	0.609	0.636	0.636	0.5	0.474	0.444	0.5	0.526	0.55	0.526	0.526	0.55
Basque_Western	0.52	0.5	0.52	0.44	0.52	0.48	0.48	0.625	0.565	0.565	0.526	0.5	0.474	0.524	0.55	0.571	0.55	0.55	0.476
Wolof	0.577	0.593	0.593	0.607	0.6	0.593	0.593	0.5	0.577	0.577	0.583	0.56	0.6	0.56	0.577	0.593	0.577	0.577	0.654
Malagasy	0.619	0.619	0.619	0.636	0.571	0.636	0.652	0.579	0.6	0.6	0.579	0.619	0.6	0.619	0.619	0.619	0.619	0.619	0.619
Archi	0.45	0.45	0.45	0.45	0.5	0.45	0.45	0.421	0.4	0.4	0.444	0.389	0.444	0.389	0.389	0.389	0.389	0.389	0.4
Lak	0.45	0.45	0.45	0.45	0.5	0.45	0.45	0.421	0.4	0.4	0.444	0.389	0.444	0.389	0.389	0.389	0.389	0.389	0.4
Language Names	Sicilian	Calabre	Italian	Spanisl	French	Portugu	Romani	Greek_	Greek	Greek (	English	Dutch	Afrikaa	Germar	Danish	Iceland	Faroes	Norweg	Bulgari



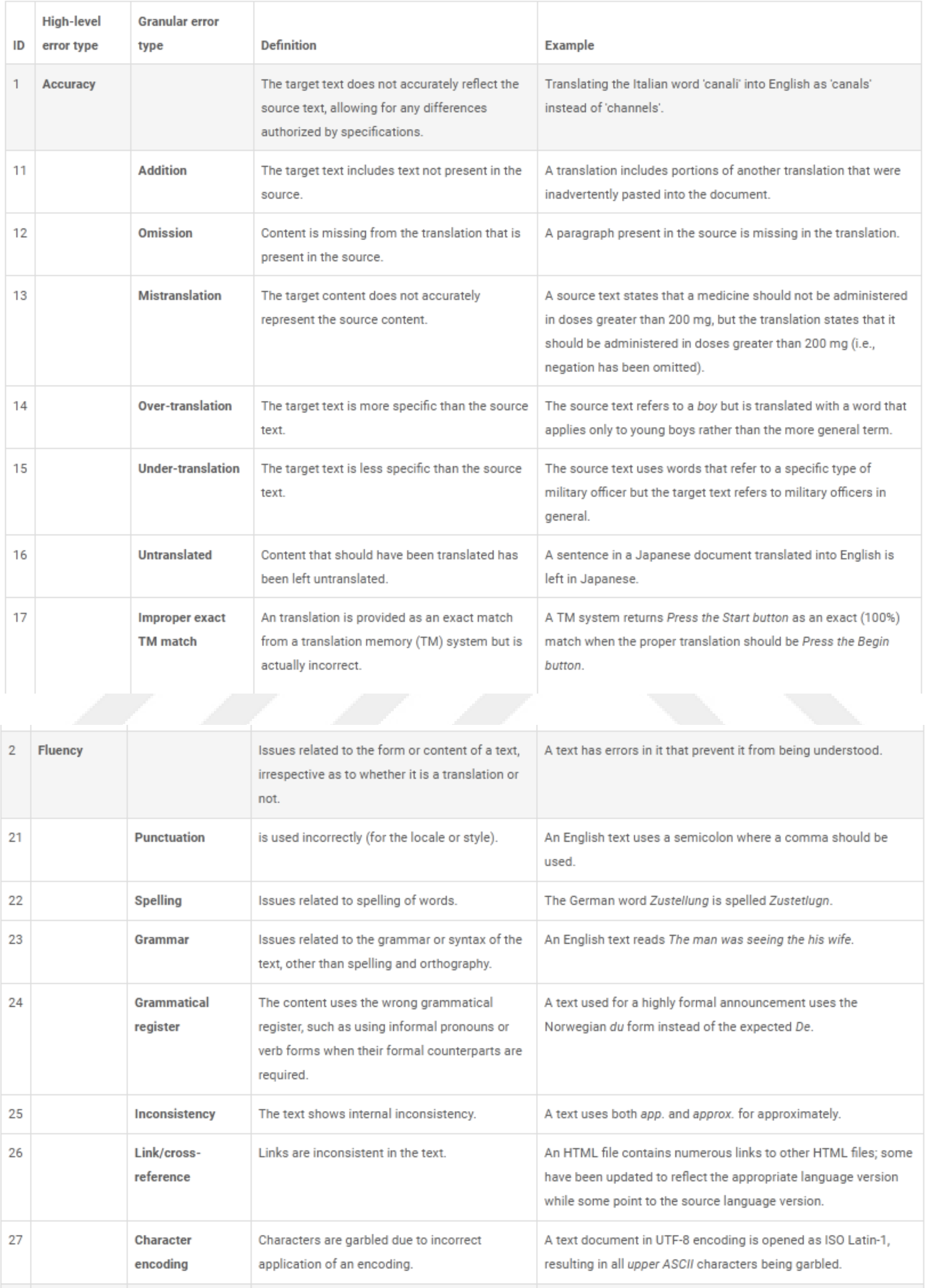
23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43
Serbo_	Sloveni	Polish	Russiar	Irish	Welsh	Marathi	Hindi	Pashto	Tamil	Telugu	Mandar	Canton	Japane	Korean	Arabic	Hebrew	Hungar	Khanty	Estonia	Finnish
0.273	0.273	0.238	0.273	0.36	0.36	0.318	0.286	0.286	0.381	0.381	0.667	0.667	0.545	0.5	0.444	0.4	0.545	0.571	0.5	0.524
0.273	0.273	0.238	0.273	0.385	0.385	0.348	0.318	0.318	0.409	0.409	0.7	0.7	0.583	0.545	0.346	0.36	0.583	0.571	0.5	0.524
0.261	0.261	0.227	0.261	0.407	0.407	0.318	0.286	0.286	0.381	0.381	0.667	0.667	0.545	0.5	0.407	0.36	0.565	0.571	0.5	0.524
0.261	0.261	0.227	0.261	0.37	0.37	0.273	0.238	0.2	0.381	0.381	0.556	0.556	0.455	0.4	0.448	0.464	0.583	0.571	0.5	0.524
0.348	0.348	0.348	0.348	0.333	0.333	0.3	0.263	0.222	0.429	0.429	0.571	0.571	0.5	0.444	0.481	0.417	0.609	0.636	0.526	0.571
0.261	0.261	0.227	0.261	0.37	0.37	0.318	0.286	0.25	0.381	0.381	0.667	0.667	0.545	0.5	0.483	0.444	0.565	0.571	0.5	0.524
0.304	0.304	0.304	0.304	0.393	0.393	0.318	0.286	0.25	0.381	0.381	0.667	0.667	0.583	0.545	0.357	0.37	0.615	0.591	0.5	0.545
0.238	0.238	0.238	0.238	0.333	0.333	0.364	0.333	0.286	0.455	0.455	0.667	0.667	0.583	0.545	0.423	0.44	0.545	0.55	0.5	0.524
0.167	0.167	0.167	0.167	0.36	0.36	0.318	0.286	0.238	0.435	0.435	0.556	0.556	0.5	0.455	0.414	0.5	0.5	0.524	0.476	0.5
0.167	0.167	0.167	0.167	0.36	0.36	0.318	0.286	0.238	0.435	0.435	0.556	0.556	0.5	0.455	0.414	0.5	0.5	0.524	0.476	0.5
0.227	0.227	0.227	0.227	0.304	0.304	0.35	0.316	0.278	0.368	0.368	0.714	0.714	0.556	0.5	0.44	0.417	0.524	0.526	0.412	0.474
0.182	0.182	0.182	0.182	0.292	0.292	0.3	0.263	0.222	0.316	0.316	0.714	0.714	0.556	0.5	0.444	0.423	0.565	0.55	0.444	0.5
0.227	0.227	0.227	0.227	0.333	0.333	0.35	0.316	0.278	0.368	0.368	0.714	0.714	0.556	0.5	0.462	0.44	0.545	0.526	0.412	0.474
0.136	0.136	0.174	0.174	0.25	0.25	0.333	0.3	0.263	0.35	0.35	0.75	0.75	0.6	0.556	0.444	0.423	0.565	0.55	0.444	0.5
0.182	0.182	0.182	0.182	0.32	0.32	0.3	0.263	0.222	0.316	0.316	0.714	0.714	0.556	0.5	0.407	0.385	0.583	0.55	0.444	0.5
0.136	0.136	0.174	0.174	0.231	0.231	0.333	0.3	0.263	0.35	0.35	0.75	0.75	0.6	0.556	0.393	0.37	0.583	0.55	0.444	0.5
0.182	0.182	0.182	0.182	0.32	0.32	0.3	0.263	0.222	0.316	0.316	0.714	0.714	0.556	0.5	0.407	0.385	0.583	0.55	0.444	0.5
0.182	0.182	0.182	0.182	0.32	0.32	0.3	0.263	0.222	0.316	0.316	0.714	0.714	0.556	0.5	0.407	0.385	0.583	0.55	0.444	0.5
0.125	0.125	0.167	0.125	0.333	0.333	0.333	0.3	0.263	0.333	0.333	0.75	0.75	0.556	0.5	0.276	0.379	0.542	0.524	0.45	0.476
0	0	0.087	0.043	0.25	0.25	0.35	0.316	0.35	0.381	0.381	0.778	0.778	0.6	0.556	0.333	0.35	0.5	0.474	0.421	0.45
0	0	0.087	0.043	0.25	0.25	0.35	0.316	0.35	0.381	0.381	0.778	0.778	0.6	0.556	0.333	0.35	0.5	0.474	0.421	0.45
0.087	0.087	0	0.043	0.25	0.25	0.35	0.316	0.35	0.381	0.381	0.778	0.778	0.6	0.556	0.333	0.35	0.5	0.474	0.421	0.45
0.043	0.043	0.043	0	0.25	0.25	0.35	0.316	0.35	0.381	0.381	0.778	0.778	0.6	0.556	0.333	0.35	0.5	0.474	0.421	0.45
0.25	0.25	0.25	0.25	0	0	0.381	0.35	0.316	0.4	0.4	0.75	0.75	0.6	0.556	0.37	0.346	0.5	0.526	0.412	0.474
0.25	0.25	0.25	0.25	0	0	0.381	0.35	0.316	0.4	0.4	0.75	0.75	0.6	0.556	0.37	0.346	0.5	0.526	0.412	0.474
0.35	0.35	0.35	0.35	0.381	0.381	0	0.05	0.19	0.25	0.25	0.444	0.444	0.417	0.5	0.409	0.409	0.524	0.421	0.474	0.45
0.316	0.316	0.316	0.316	0.35	0.35	0.05	0	0.238	0.25	0.25	0.375	0.375	0.5	0.455	0.381	0.381	0.5	0.474	0.444	0.5
0.35	0.35	0.35	0.35	0.316	0.316	0.19	0.238	0	0.35	0.35	0.556	0.556	0.364	0.455	0.381	0.381	0.571	0.474	0.474	0.421
0.381	0.381	0.381	0.381	0.4	0.4	0.25	0.25	0.35	0	0	0.714	0.714	0.7	0.7	0.409	0.35	0.524	0.474	0.474	0.5
0.381	0.381	0.381	0.381	0.4	0.4	0.25	0.25	0.35	0	0	0.714	0.714	0.7	0.7	0.409	0.35	0.524	0.474	0.474	0.5
0.778	0.778	0.778	0.778	0.75	0.75	0.444	0.375	0.556	0.714	0.714	0	0.1	0.5	0.455	0.75	0.75	0.75	0.778	0.714	0.778
0.778	0.778	0.778	0.778	0.75	0.75	0.444	0.375	0.556	0.714	0.714	0.1	0	0.5	0.455	0.75	0.75	0.75	0.778	0.714	0.778
0.6	0.6	0.6	0.6	0.6	0.6	0.417	0.5	0.364	0.7	0.7	0.5	0.5	0	0.182	0.6	0.636	0.75	0.636	0.6	0.545
0.556	0.556	0.556	0.556	0.556	0.556	0.5	0.455	0.455	0.7	0.7	0.455	0.455	0.182	0	0.556	0.6	0.727	0.727	0.556	0.636
0.333	0.333	0.333	0.333	0.37	0.37	0.409	0.381	0.381	0.409	0.409	0.75	0.75	0.6	0.556	0	0.25	0.538	0.5	0.421	0.429
0.35	0.35	0.35	0.35	0.346	0.346	0.409	0.381	0.381	0.35	0.35	0.75	0.75	0.636	0.6	0.25	0	0.5	0.5	0.444	0.45
0.5	0.5	0.5	0.5	0.5	0.5	0.524	0.5	0.571	0.524	0.524	0.75	0.75	0.75	0.727	0.538	0.5	0	0.158	0.312	0.3
0.474	0.474	0.474	0.474	0.526	0.526	0.421	0.474	0.474	0.474	0.474	0.778	0.778	0.636	0.727	0.5	0.5	0.158	0	0.375	0.263
0.421	0.421	0.421	0.421	0.412	0.412	0.474	0.444	0.474	0.474	0.474	0.714	0.714	0.6	0.556	0.421	0.444	0.312	0.375	0	0.125
0.45	0.45	0.45	0.45	0.474	0.474	0.45	0.5	0.421	0.5	0.5	0.778	0.778	0.545	0.636	0.429	0.45	0.3	0.263	0.125	0
0.421	0.421	0.421	0.421	0.474	0.474	0.35	0.4	0.45	0.4	0.4	0.667	0.667	0.583	0.667	0.455	0.45	0.238	0.2	0.25	0.2
0.45	0.45	0.45	0.45	0.421	0.421	0.3	0.35	0.4	0.35	0.35	0.667	0.667	0.583	0.667	0.409	0.4	0.273	0.238	0.25	0.2
0.524	0.524	0.524	0.524	0.524	0.524	0.409	0.409	0.476	0.381	0.381	0.6	0.6	0.615	0.615	0.478	0.5	0.455	0.4	0.444	0.4
0.474	0.474	0.474	0.474	0.571	0.571	0.333	0.333	0.429	0.4	0.4	0.6	0.6	0.643	0.643	0.545	0.524	0.333	0.35	0.412	0.35
0.474	0.474	0.474	0.474	0.571	0.571	0.333	0.333	0.429	0.4	0.4	0.6	0.6	0.643	0.643	0.545	0.524	0.333	0.35	0.412	0.35
0.474	0.474	0.474	0.474	0.571	0.571	0.333	0.333	0.429	0.4	0.4	0.6	0.6	0.643	0.643	0.545	0.524	0.333	0.35	0.412	0.35
0.5	0.5	0.5	0.5	0.526	0.526	0.35	0.35	0.45	0.421	0.421	0.556	0.556	0.583	0.583	0.524	0.5	0.263	0.176	0.438	0.368
0.438	0.438	0.438	0.438	0.5	0.5	0.316	0.316	0.421	0.353	0.353	0.6	0.6	0.5	0.5	0.5	0.474	0.316	0.235	0.4	0.333
0.471	0.471	0.471	0.471	0.5	0.5	0.316	0.316	0.421	0.389	0.389	0.556	0.556	0.583	0.583	0.5	0.474	0.316	0.235	0.4	0.333
0.471	0.471	0.471	0.471	0.5	0.5	0.316	0.316	0.421	0.389	0.389	0.556	0.556	0.583	0.583	0.5	0.474	0.316	0.235	0.4	0.333
0.471	0.471	0.471	0.471	0.5	0.5	0.316	0.316	0.421	0.389	0.389	0.556	0.556	0.583	0.583	0.5	0.474	0.316	0.235	0.4	0.333
0.524	0.524	0.524	0.524	0.5	0.5	0.35	0.35	0.45	0.3	0.3	0.667	0.667	0.692	0.692	0.455	0.4	0.381	0.316	0.412	0.316
0.588	0.588	0.588	0.588	0.6	0.6	0.444	0.5	0.471	0.529	0.529	0.75	0.75	0.444	0.556	0.55	0.619	0.7	0.533	0.5	0.467
0.5	0.5	0.5	0.5	0.55	0.55	0.389	0.444	0.412	0.444	0.444	0.75	0.75	0.444	0.556	0.476	0.571	0.682	0.529	0.5	0.471
0.619	0.619	0.619	0.619	0.6	0.6	0.565	0.545	0.609	0.682	0.682	0.4	0.4	0.667	0.643	0.593	0.63	0.667	0.667	0.684	0.65
0.556	0.556	0.556	0.556	0.6	0.6	0.5	0.556	0.556	0.611	0.611	0.7	0.7	0.545	0.636	0.55	0.526	0.476	0.368	0.562	0.421
0.368	0.368	0.368	0.368	0.353	0.353	0.316	0.316	0.45	0.263	0.263	0.75	0.75	0.727	0.727	0.429	0.421	0.444	0.412	0.444	0.474
0.368	0.368	0.368	0.368	0.353	0.353	0.316	0.316	0.45	0.263	0.263	0.75	0.75	0.727	0.727	0.429	0.421	0.444	0.412	0.444	0.474
Serbo_	Sloveni	Polish	Russiar	Irish	Welsh	Marathi	Hindi	Pashto	Tamil	Telugu	Mandar	Canton	Japane	Korean	Arabic	Hebrew	Hungar	Khanty	Estonia	Finnish



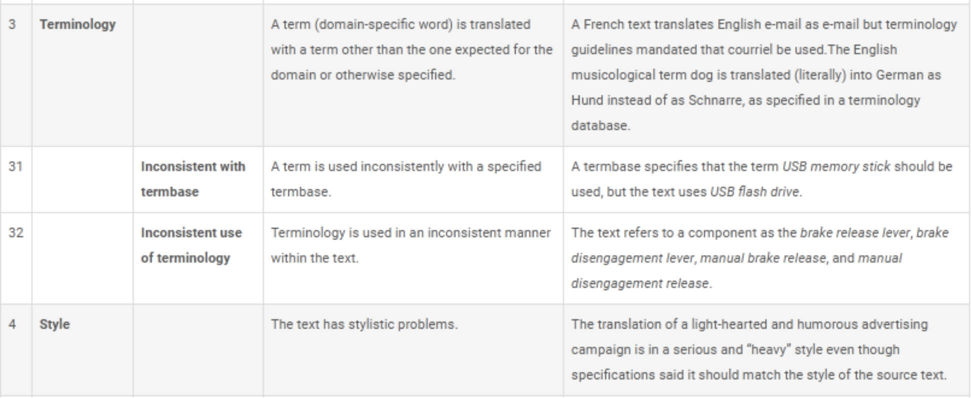
43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60		
Finnish	Mari_1	Udmurt	Yukagh	Even_1	Even_2	Evenki	Yakut	Uzbek	Kazak	Kyrgyz	Turkish	Buryat	Basque	Basque	Wolof	Malaga	Archi	Lak	Language Names
0.524	0.524	0.476	0.476	0.5	0.5	0.5	0.524	0.5	0.5	0.5	0.5	0.45	0.5	0.52	0.577	0.619	0.45	0.45	Siciliano_Ragusa
0.524	0.524	0.476	0.5	0.5	0.5	0.5	0.524	0.5	0.5	0.5	0.5	0.476	0.48	0.5	0.593	0.619	0.45	0.45	Calabrese_Northern
0.524	0.524	0.476	0.476	0.5	0.5	0.5	0.524	0.5	0.5	0.5	0.5	0.45	0.5	0.52	0.593	0.619	0.45	0.45	Italian
0.524	0.524	0.476	0.5	0.455	0.455	0.455	0.476	0.45	0.45	0.45	0.45	0.476	0.417	0.44	0.607	0.636	0.45	0.45	Spanish
0.571	0.591	0.545	0.545	0.524	0.524	0.524	0.55	0.526	0.526	0.526	0.526	0.524	0.522	0.52	0.6	0.571	0.5	0.5	French
0.524	0.524	0.476	0.476	0.5	0.5	0.5	0.524	0.5	0.5	0.5	0.5	0.45	0.458	0.48	0.593	0.636	0.45	0.45	Portuguese
0.545	0.545	0.5	0.5	0.522	0.522	0.522	0.545	0.524	0.524	0.524	0.524	0.476	0.458	0.48	0.593	0.652	0.45	0.45	Romanian
0.524	0.5	0.524	0.545	0.476	0.476	0.476	0.5	0.474	0.474	0.474	0.474	0.524	0.609	0.625	0.5	0.579	0.421	0.421	Greek_Calabria_1
0.5	0.476	0.5	0.542	0.429	0.429	0.429	0.45	0.421	0.421	0.421	0.421	0.522	0.636	0.565	0.577	0.6	0.4	0.4	Greek
0.5	0.476	0.5	0.542	0.429	0.429	0.429	0.45	0.421	0.421	0.421	0.421	0.522	0.636	0.565	0.577	0.6	0.4	0.4	Greek Cypriot
0.474	0.474	0.421	0.474	0.5	0.5	0.5	0.526	0.471	0.5	0.5	0.5	0.474	0.5	0.526	0.583	0.579	0.444	0.444	English
0.5	0.5	0.45	0.5	0.524	0.524	0.524	0.55	0.5	0.526	0.526	0.526	0.5	0.474	0.5	0.56	0.619	0.389	0.389	Dutch
0.474	0.474	0.421	0.474	0.5	0.5	0.5	0.526	0.471	0.5	0.5	0.5	0.474	0.444	0.474	0.6	0.6	0.444	0.444	Afrikaans
0.5	0.5	0.45	0.524	0.524	0.524	0.524	0.55	0.5	0.526	0.526	0.526	0.524	0.5	0.524	0.56	0.619	0.389	0.389	German
0.5	0.5	0.45	0.5	0.524	0.524	0.524	0.55	0.5	0.526	0.526	0.526	0.5	0.526	0.55	0.577	0.619	0.389	0.389	Danish
0.5	0.5	0.45	0.524	0.524	0.524	0.524	0.55	0.5	0.526	0.526	0.526	0.524	0.55	0.571	0.593	0.619	0.389	0.389	Icelandic
0.5	0.5	0.45	0.5	0.524	0.524	0.524	0.55	0.5	0.526	0.526	0.526	0.5	0.526	0.55	0.577	0.619	0.389	0.389	Faroese
0.5	0.5	0.45	0.5	0.524	0.524	0.524	0.55	0.5	0.526	0.526	0.526	0.5	0.526	0.55	0.577	0.619	0.389	0.389	Norwegian
0.476	0.476	0.429	0.476	0.524	0.524	0.524	0.55	0.5	0.526	0.526	0.526	0.476	0.55	0.476	0.654	0.619	0.4	0.4	Bulgarian
0.45	0.421	0.45	0.524	0.474	0.474	0.474	0.5	0.438	0.471	0.471	0.471	0.524	0.588	0.5	0.619	0.556	0.368	0.368	Serbo_Croat
0.45	0.421	0.45	0.524	0.474	0.474	0.474	0.5	0.438	0.471	0.471	0.471	0.524	0.588	0.5	0.619	0.556	0.368	0.368	Slovenian
0.45	0.421	0.45	0.524	0.474	0.474	0.474	0.5	0.438	0.471	0.471	0.471	0.524	0.588	0.5	0.619	0.556	0.368	0.368	Polish
0.45	0.421	0.45	0.524	0.474	0.474	0.474	0.5	0.438	0.471	0.471	0.471	0.524	0.588	0.5	0.619	0.556	0.368	0.368	Russian
0.474	0.474	0.421	0.524	0.571	0.571	0.571	0.526	0.5	0.5	0.5	0.5	0.5	0.6	0.55	0.6	0.6	0.353	0.353	Irish
0.474	0.474	0.421	0.524	0.571	0.571	0.571	0.526	0.5	0.5	0.5	0.5	0.5	0.6	0.55	0.6	0.6	0.353	0.353	Welsh
0.45	0.35	0.3	0.409	0.333	0.333	0.333	0.35	0.316	0.316	0.316	0.316	0.35	0.444	0.389	0.565	0.5	0.316	0.316	Marathi
0.5	0.4	0.35	0.409	0.333	0.333	0.333	0.35	0.316	0.316	0.316	0.316	0.35	0.5	0.444	0.545	0.556	0.316	0.316	Hindi
0.421	0.45	0.4	0.476	0.429	0.429	0.429	0.45	0.421	0.421	0.421	0.421	0.45	0.471	0.412	0.609	0.556	0.45	0.45	Pashto
0.5	0.4	0.35	0.381	0.4	0.4	0.4	0.421	0.353	0.389	0.389	0.389	0.3	0.529	0.444	0.682	0.611	0.263	0.263	Tamil
0.5	0.4	0.35	0.381	0.4	0.4	0.4	0.421	0.353	0.389	0.389	0.389	0.3	0.529	0.444	0.682	0.611	0.263	0.263	Telugu
0.778	0.667	0.667	0.6	0.6	0.6	0.6	0.556	0.6	0.556	0.556	0.556	0.667	0.75	0.75	0.4	0.7	0.75	0.75	Mandarin
0.778	0.667	0.667	0.6	0.6	0.6	0.6	0.556	0.6	0.556	0.556	0.556	0.667	0.75	0.75	0.4	0.7	0.75	0.75	Cantonese
0.545	0.583	0.583	0.615	0.643	0.643	0.643	0.583	0.5	0.583	0.583	0.583	0.692	0.444	0.444	0.667	0.545	0.727	0.727	Japanese
0.636	0.667	0.667	0.615	0.643	0.643	0.643	0.583	0.5	0.583	0.583	0.583	0.692	0.556	0.556	0.643	0.636	0.727	0.727	Korean
0.429	0.455	0.409	0.478	0.545	0.545	0.545	0.524	0.5	0.5	0.5	0.5	0.455	0.55	0.476	0.593	0.55	0.429	0.429	Arabic
0.45	0.45	0.4	0.5	0.524	0.524	0.524	0.5	0.474	0.474	0.474	0.474	0.4	0.619	0.571	0.63	0.526	0.421	0.421	Hebrew
0.3	0.238	0.273	0.455	0.333	0.333	0.333	0.263	0.316	0.316	0.316	0.316	0.381	0.7	0.682	0.667	0.476	0.444	0.444	Hungarian
0.263	0.2	0.238	0.4	0.35	0.35	0.35	0.176	0.235	0.235	0.235	0.235	0.316	0.533	0.529	0.667	0.368	0.412	0.412	Khanty_2
0.125	0.25	0.25	0.444	0.412	0.412	0.412	0.438	0.4	0.4	0.4	0.4	0.412	0.5	0.5	0.684	0.562	0.444	0.444	Estonian
0	0.2	0.2	0.4	0.35	0.35	0.35	0.368	0.333	0.333	0.333	0.333	0.316	0.467	0.471	0.65	0.421	0.474	0.474	Finnish
0.2	0	0.048	0.333	0.238	0.238	0.238	0.25	0.211	0.211	0.211	0.211	0.286	0.438	0.444	0.6	0.35	0.368	0.368	Mari_1
0.2	0.048	0	0.286	0.273	0.273	0.273	0.286	0.25	0.25	0.25	0.25	0.238	0.375	0.389	0.6	0.381	0.4	0.4	Udmurt_1
0.4	0.333	0.286	0	0.364	0.364	0.364	0.333	0.25	0.3	0.3	0.3	0.238	0.438	0.444	0.571	0.55	0.429	0.429	Yukaghir
0.35	0.238	0.273	0.364	0	0	0	0.19	0.15	0.15	0.15	0.15	0.286	0.588	0.611	0.619	0.421	0.45	0.45	Even_1
0.35	0.238	0.273	0.364	0	0	0	0.19	0.15	0.15	0.15	0.15	0.286	0.588	0.611	0.619	0.421	0.45	0.45	Even_2
0.35	0.238	0.273	0.364	0	0	0	0.19	0.15	0.15	0.15	0.15	0.286	0.588	0.611	0.619	0.421	0.45	0.45	Evenki
0.368	0.25	0.286	0.333	0.19	0.19	0.19	0	0.105	0.056	0.056	0.056	0.211	0.562	0.588	0.619	0.389	0.389	0.389	Yakut
0.333	0.211	0.25	0.25	0.15	0.15	0.15	0.105	0	0.056	0.056	0.056	0.211	0.533	0.562	0.619	0.353	0.353	0.353	Uzbek
0.333	0.211	0.25	0.3	0.15	0.15	0.15	0.056	0.056	0	0	0	0.167	0.533	0.562	0.6	0.353	0.353	0.353	Kazak
0.333	0.211	0.25	0.3	0.15	0.15	0.15	0.056	0.056	0	0	0	0.167	0.533	0.562	0.6	0.353	0.353	0.353	Kyrgyz
0.333	0.211	0.25	0.3	0.15	0.15	0.15	0.056	0.056	0	0	0	0.167	0.533	0.562	0.6	0.353	0.353	0.353	Turkish
0.316	0.286	0.238	0.238	0.286	0.286	0.286	0.211	0.211	0.167	0.167	0.167	0	0.5	0.5	0.579	0.474	0.368	0.368	Buryat
0.467	0.438	0.375	0.438	0.588	0.588	0.588	0.562	0.533	0.533	0.533	0.533	0.5	0	0.158	0.696	0.611	0.588	0.588	Basque_Central
0.471	0.444	0.389	0.444	0.611	0.611	0.611	0.588	0.562	0.562	0.562	0.562	0.5	0.158	0	0.708	0.667	0.5	0.5	Basque_Western
0.65	0.6	0.6	0.571	0.619	0.619	0.619	0.619	0.619	0.6	0.6	0.6	0.579	0.696	0.708	0	0.636	0.65	0.65	Wolof
0.421	0.35	0.381	0.55	0.421	0.421	0.421	0.389	0.353	0.353	0.353	0.353	0.474	0.611	0.667	0.636	0	0.529	0.529	Malagasy
0.474	0.368	0.4	0.429	0.45	0.45	0.45	0.389	0.353	0.353	0.353	0.353	0.368	0.588	0.5	0.65	0.529	0	0	Archi
0.474	0.368	0.4	0.429	0.45	0.45	0.45	0.389	0.353	0.353	0.353	0.353	0.368	0.588	0.5	0.65	0.529	0	0	Lak
Finnish	Mari_1	Udmurt	Yukagh	Even_1	Even_2	Evenki	Yakut	Uzbek	Kazak	Kyrgyz	Turkish	Buryat	Basque	Basque	Wolof	Malaga	Archi	Lak	Language Names

APPENDIX E

DQF-MQM ERROR TYPOLOGY FRAMEWORK



ID	High-level error type	Granular error type	Definition	Example
1	**Accuracy**		The target text does not accurately reflect the source text, allowing for any differences authorized by specifications.	Translating the Italian word 'canali' into English as 'canals' instead of 'channels'.
11		**Addition**	The target text includes text not present in the source.	A translation includes portions of another translation that were inadvertently pasted into the document.
12		**Omission**	Content is missing from the translation that is present in the source.	A paragraph present in the source is missing in the translation.
13		**Mistranslation**	The target content does not accurately represent the source content.	A source text states that a medicine should not be administered in doses greater than 200 mg, but the translation states that it should be administered in doses greater than 200 mg (i.e., negation has been omitted).
14		**Over-translation**	The target text is more specific than the source text.	The source text refers to a *boy* but is translated with a word that applies only to young boys rather than the more general term.
15		**Under-translation**	The target text is less specific than the source text.	The source text uses words that refer to a specific type of military officer but the target text refers to military officers in general.
16		**Untranslated**	Content that should have been translated has been left untranslated.	A sentence in a Japanese document translated into English is left in Japanese.
17		**Improper exact TM match**	An translation is provided as an exact match from a translation memory (TM) system but is actually incorrect.	A TM system returns *Press the Start button* as an exact (100%) match when the proper translation should be *Press the Begin button.*
2	**Fluency**		Issues related to the form or content of a text, irrespective as to whether it is a translation or not.	A text has errors in it that prevent it from being understood.
21		**Punctuation**	is used incorrectly (for the locale or style).	An English text uses a semicolon where a comma should be used.
22		**Spelling**	Issues related to spelling of words.	The German word *Zustellung* is spelled *Zustetlugn*.
23		**Grammar**	Issues related to the grammar or syntax of the text, other than spelling and orthography.	An English text reads *The man was seeing the his wife.*
24		**Grammatical register**	The content uses the wrong grammatical register, such as using informal pronouns or verb forms when their formal counterparts are required.	A text used for a highly formal announcement uses the Norwegian *du* form instead of the expected *De*.
25		**Inconsistency**	The text shows internal inconsistency.	A text uses both *app.* and *approx.* for approximately.
26		**Link/cross-reference**	Links are inconsistent in the text.	An HTML file contains numerous links to other HTML files; some have been updated to reflect the appropriate language version while some point to the source language version.
27		**Character encoding**	Characters are garbled due to incorrect application of an encoding.	A text document in UTF-8 encoding is opened as ISO Latin-1, resulting in all *upper ASCII* characters being garbled.



3	**Terminology**		A term (domain-specific word) is translated with a term other than the one expected for the domain or otherwise specified.	A French text translates English e-mail as e-mail but terminology guidelines mandated that courriel be used.The English musicological term dog is translated (literally) into German as Hund instead of as Schnarre, as specified in a terminology database.
31		**Inconsistent with termbase**	A term is used inconsistently with a specified termbase.	A termbase specifies that the term *USB memory stick* should be used, but the text uses *USB flash drive*.
32		**Inconsistent use of terminology**	Terminology is used in an inconsistent manner within the text.	The text refers to a component as the *brake release lever*, *brake disengagement lever*, *manual brake release*, and *manual disengagement release*.
4	**Style**		The text has stylistic problems.	The translation of a light-hearted and humorous advertising campaign is in a serious and "heavy" style even though specifications said it should match the style of the source text.

APPENDIX F

QUESTIONS IN THE QUALITATIVE SURVEY AND ENGLISH TRANSLATIONS

Q1) Çeviri metinde, kaynak metindeki içerik ile kıyasla eklentiler veya eksiklikler var mı?

(Q1) Are there additions or omissions in the translated text compared to the source text?)

(1: Hiç, 7: Fazlasıyla) (1: None, 7: Abundantly)

Q2) Çeviri metinde, yanlış çeviri olarak tanımlayabileceğiniz çeviriler var mı?

(Q2) Are there mistranslations in the translated text?)

(1: Hiç, 7: Fazlasıyla) (1: None, 7: Abundantly)

Q3) Çeviri metinde, uygun olmadığını gördüğünüz anlam kaymaları var mı?

(Q3) In the source text, are there semantic shifts that you deem to be inappropriate?)

(1: Hiç, 7: Fazlasıyla) (1: None, 7: Abundantly)

Q4) Çeviri metin, kaynak metin kadar akıcı bir şekilde okunabiliyor mu? Çeviri metnin anlaşılabilirliği kaynak metin kadar mı?

(Q4) Can the translated text be read as fluently as the source text? Is the understandability of the translated text equal to that of the source text?)

(1: Kaynak metin ile aynı, 7: Kaynak metinden çok farklı) (1: Same as the source text, 7: Very different from the source text)

Q5) Çeviri metinde gramer veya dil anlatım bozuklukları var mı?

(Q5) Are there grammatical errors in the translated text?)

(1: Hiç, 7: Fazlasıyla) (1: None, 7: Abundantly)

Q6) Çeviri metinde kullanılan terimlerde ve jargonda uygunsuzluk veya hata var mı?

(Q6) Are there inappropriate or incorrect uses of certain terms or jargon in the translated text?)

(1: Hiç, 7: Fazlasıyla) (1: None, 7: Abundantly)

Q7) Çeviri metnin bir bütün olarak kalitesi ve isabetliliğini nasıl değerlendirirsiniz?

(Q7) How would you evaluate the overall quality and accuracy of the translated text?)

(1: Çok iyi, 7: Çok kötü) (1: Very good, 7: Very bad)

APPENDIX G

GENERATED TEXTS FOR THE THESIS, EXEMPLIFIED SYNTACTIC CATEGORIES, AND ENGLISH TRANSLATIONS

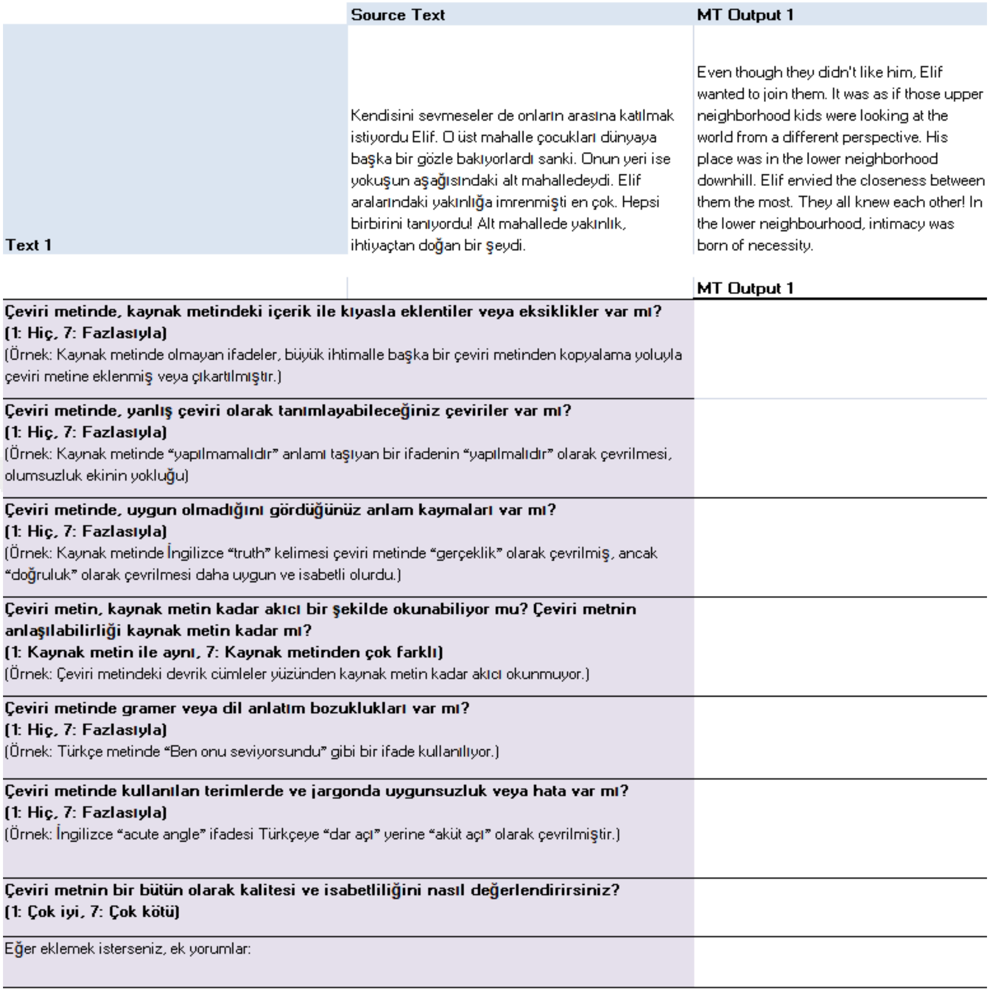
Turkish Text	Legend	Sample English Translation
Kendisini sevmeseler de onların arasına katılmak istiyordu Elif. O üst mahalle çocukları dünyaya başka bir gözle bakıyorlardı sanki. Onun yeri ise yokuşun aşağısındaki alt mahalledeydi. Elif aralarındaki yakınlığa imrenmişti en çok. Hepsi birbirini tanıyordu! Alt mahallede yakınlık, ihtiyaçtan doğan bir şeydi.	• grammaticalized morphology • grammaticalized gender • collective number • grammaticalized agreement - grammaticalized number • number spread to N • adjectival possessives	Even though they didn't like him, Elif wanted to join them. It was as if those upper neighborhood kids were looking at the world from a different perspective. His place was in the lower neighborhood downhill. Elif envied the closeness between them the most. They all knew each other! In the lower neighborhood, intimacy was born of necessity.

Ben de herhangi biri sayılırım dostlarım. Et yiyen, çorba içen, fazla düşünen bir vatansever. Sesimin yüksek çıkmasının sebebi içimdeki yazarlık tutkusudur sadece. Hatta sesim zamanla yükseldi diyebiliriz. Şiirler yazdım, makale yazıyorum, kitaplar yazacağım. Bir gün de herhangi biri benim de sonum gelecek.	• linkers • grammaticalized number agreement • relative clauses • grammaticalized Specified Quantity • grammaticalized person	I'm just another person, my friends. A patriot who eats meat, drinks soup, thinks too much. The reason why my voice is so loud is only my passion for writing. I can even say that my voice rose over time. I wrote poems, I write articles, I will write books. One day, any of them will come to an end for me.
Amerikan **borsa**larında büyük kayıplar yaşandı. NASDAQ **borsa**sında yaklaşık %11’lik bir düşüş ile, giderek artan **jeopolitik** belirsizlikten dolayı artık **yatırımcı**larda bir satış tepkisinin tetiklendiği haber verilmekte. **Endeks**teki bazı şirketler %20’lere kadar değer kaybı görürken, en ciddi kayıplarda teknoloji	• free reduced relatives • number spread to N • relative clauses • grammaticalized number agreement • linkers • idiomatic speech • **jargon / terms**	There were huge losses in the American stock markets. With the NASDAQ stock market down nearly 11%, it is reported that a sell-off reaction has now been triggered by investors due to the growing geopolitical uncertainty. While some companies in the index saw a loss of up to 20%,

hisseleri başı çekti. Küresel yelpazede yaşanan **yarı iletken** kıtlığı karşısında iyi konumlanan şirketler bugünkü satış furyasında değer kaybetmeye en dirençli **hisse**ler oldular.		technology stocks led the way in the most serious losses. Companies that are well positioned in the face of the global semiconductor shortage have become the stocks most resistant to depreciation in today's sales frenzy.
“o demin gelen kimdi?” diye sordu Maria. “Peter bana iki çanta bırakmaya gelmiş” diye cevapladım. “Hani Peter’le artık görüşmüyordunuz” diye sorgulaması rahatsız etti beni. Bir elini kapının demir kulbu üstünde tutuyordu, ilişkimizden de evden de her an çıkmaya hazırdı. “Kardeşini kaybetmiş,” dedim “bu eşyalar onu hatırlatıyormuş kendisine.“	• grammaticalized gender • relative clauses • plural spread from cardinal quantifiers • grammaticalized person • null possessive with kinship nouns	“Who was that who just arrived?” she asked. “Peter came to drop me two bags,” I replied. His questioning, "You weren't seeing Peter anymore," bothered me. He was holding one hand on the iron handle of the door, ready to leave our relationship and the house at any moment. “He lost his brother,” I said, “these items reminded him of him.”

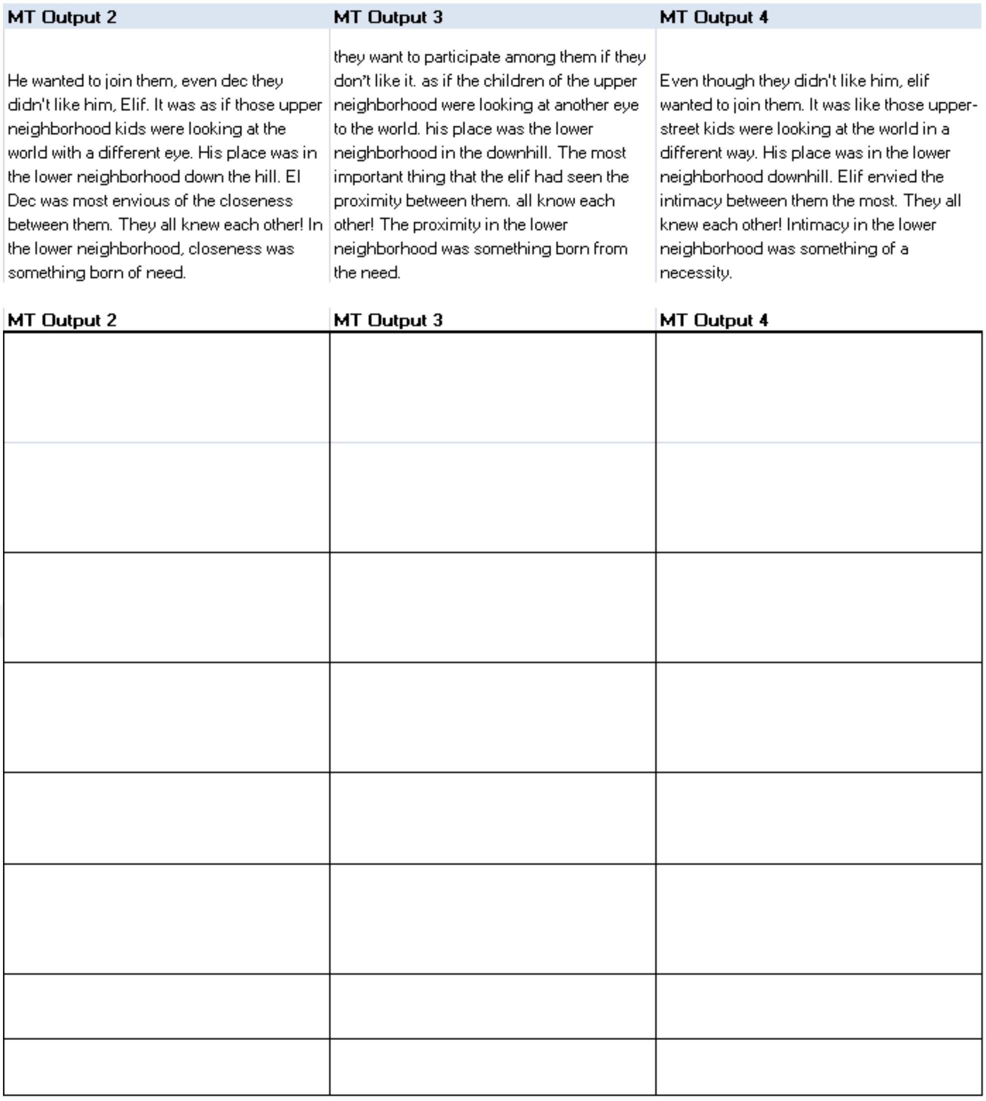
APPENDIX H

SAMPLE ENGLISH SURVEY PRESENTED TO A PARTICIPANT



	Source Text	MT Output 1
Text 1	Kendisini sevmeseler de onların arasına katılmak istiyordu Elif. O üst mahalle çocukları dünyaya başka bir gözle bakıyorlardı sanki. Onun yeri ise yokuşun aşağısındaki alt mahalledeydi. Elif aralarındaki yakınlığa imrenmişti en çok. Hepsi birbirini tanıyordu! Alt mahallede yakınlık, ihtiyaçtan doğan bir şeydi.	Even though they didn't like him, Elif wanted to join them. It was as if those upper neighborhood kids were looking at the world from a different perspective. His place was in the lower neighborhood downhill. Elif envied the closeness between them the most. They all knew each other! In the lower neighbourhood, intimacy was born of necessity.

	MT Output 1
Çeviri metinde, kaynak metindeki içerik ile kıyasla eklentiler veya eksiklikler var mı? **(1: Hiç, 7: Fazlasıyla)** (Örnek: Kaynak metinde olmayan ifadeler, büyük ihtimalle başka bir çeviri metinden kopyalama yoluyla çeviri metine eklenmiş veya çıkartılmıştır.)	
Çeviri metinde, yanlış çeviri olarak tanımlayabileceğiniz çeviriler var mı? **(1: Hiç, 7: Fazlasıyla)** (Örnek: Kaynak metinde "yapılmamalıdır" anlamı taşıyan bir ifadenin "yapılmalıdır" olarak çevrilmesi, olumsuzluk ekinin yokluğu)	
Çeviri metinde, uygun olmadığını gördüğünüz anlam kaymaları var mı? **(1: Hiç, 7: Fazlasıyla)** (Örnek: Kaynak metinde İngilizce "truth" kelimesi çeviri metinde "gerçeklik" olarak çevrilmiş, ancak "doğruluk" olarak çevrilmesi daha uygun ve isabetli olurdu.)	
Çeviri metin, kaynak metin kadar akıcı bir şekilde okunabiliyor mu? Çeviri metnin anlaşılabilirliği kaynak metin kadar mı? **(1: Kaynak metin ile aynı, 7: Kaynak metinden çok farklı)** (Örnek: Çeviri metindeki devrik cümleler yüzünden kaynak metin kadar akıcı okunmuyor.)	
Çeviri metinde gramer veya dil anlatım bozuklukları var mı? **(1: Hiç, 7: Fazlasıyla)** (Örnek: Türkçe metinde "Ben onu seviyorsundu" gibi bir ifade kullanılıyor.)	
Çeviri metinde kullanılan terimlerde ve jargonda uygunsuzluk veya hata var mı? **(1: Hiç, 7: Fazlasıyla)** (Örnek: İngilizce "acute angle" ifadesi Türkçeye "dar açı" yerine "akút açı" olarak çevrilmiştir.)	
Çeviri metnin bir bütün olarak kalitesi ve isabetliliğini nasıl değerlendirirsiniz? **(1: Çok iyi, 7: Çok kötü)**	
Eğer eklemek isterseniz, ek yorumlar:	

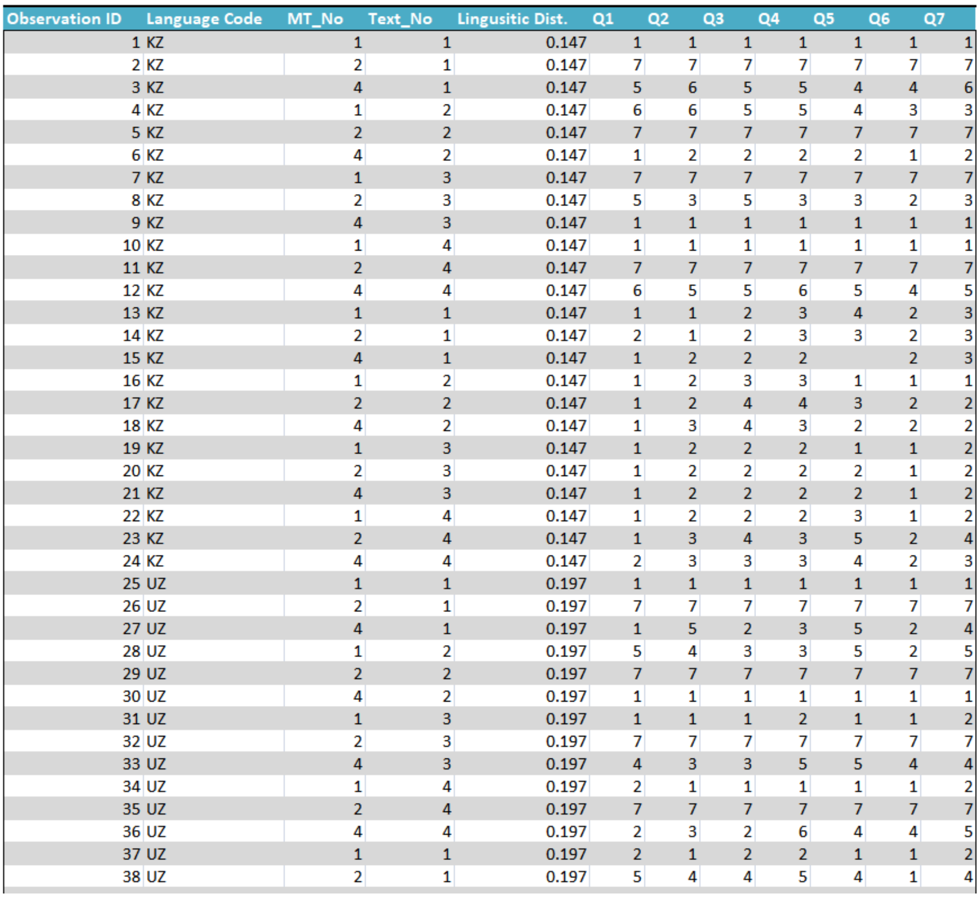


MT Output 2	MT Output 3	MT Output 4
He wanted to join them, even dec they didn't like him, Elif. It was as if those upper neighborhood kids were looking at the world with a different eye. His place was in the lower neighborhood down the hill. El Dec was most envious of the closeness between them. They all knew each other! In the lower neighborhood, closeness was something born of need.	they want to participate among them if they don't like it. as if the children of the upper neighborhood were looking at another eye to the world. his place was the lower neighborhood in the downhill. The most important thing that the elif had seen the proximity between them. all know each other! The proximity in the lower neighborhood was something born from the need.	Even though they didn't like him, elif wanted to join them. It was like those upper-street kids were looking at the world in a different way. His place was in the lower neighborhood downhill. Elif envied the intimacy between them the most. They all knew each other! Intimacy in the lower neighborhood was something of a necessity.

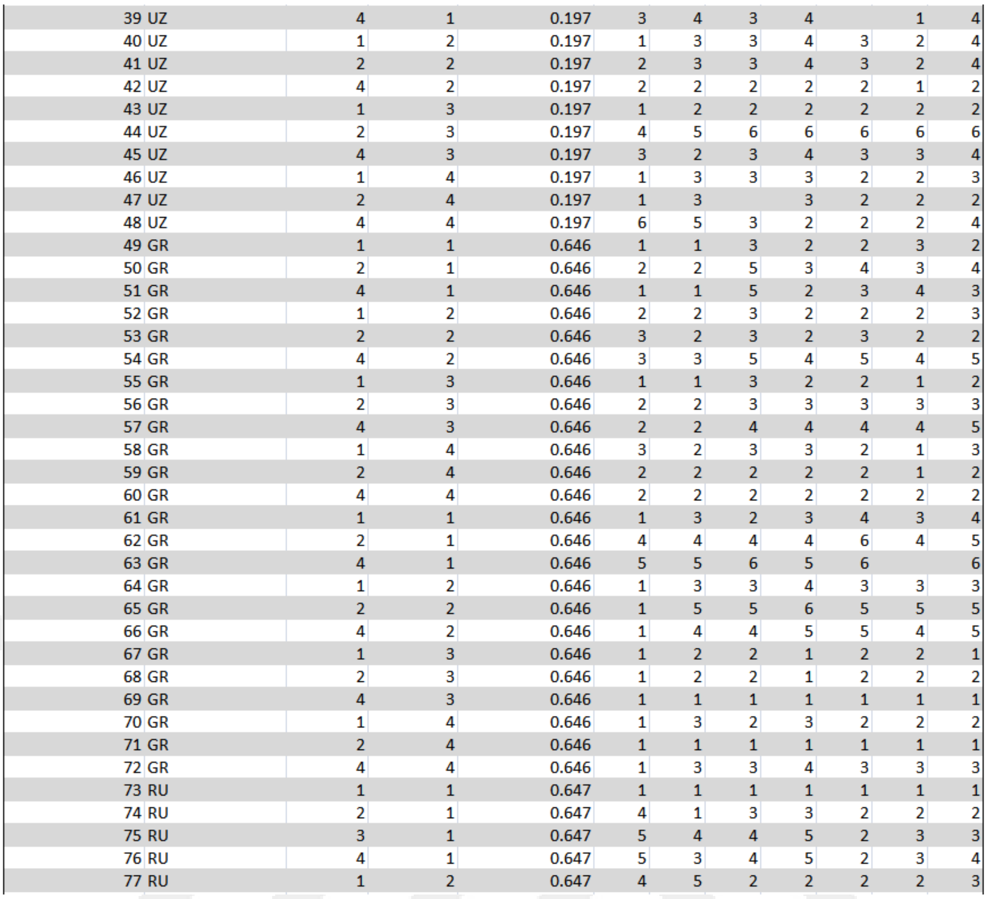
MT Output 2	MT Output 3	MT Output 4

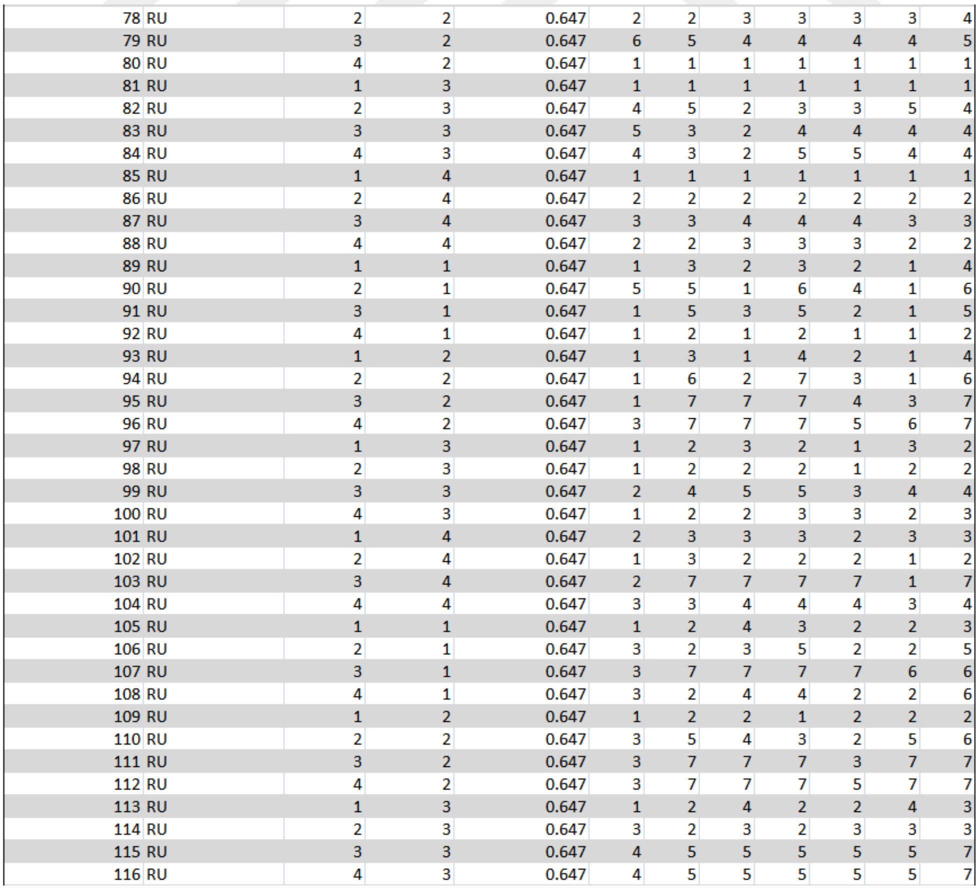
APPENDIX I

PARTICIPANT EVALUATIONS DATA

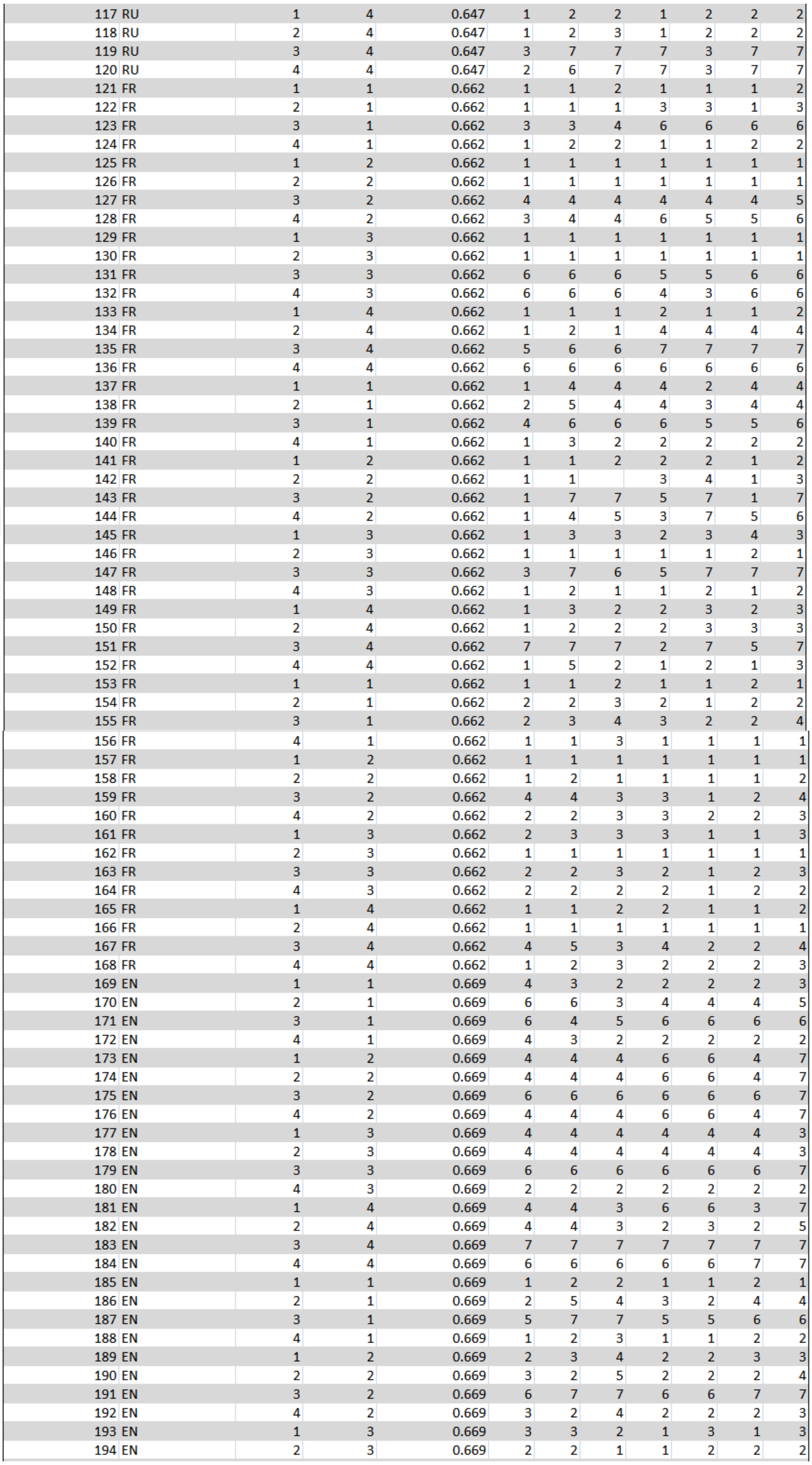


Observation ID	Language Code	MT_No	Text_No	Lingusitic Dist.	Q1	Q2	Q3	Q4	Q5	Q6	Q7
1	KZ	1	1	0.147	1	1	1	1	1	1	1
2	KZ	2	1	0.147	7	7	7	7	7	7	7
3	KZ	4	1	0.147	5	6	5	5	4	4	6
4	KZ	1	2	0.147	6	6	5	5	4	3	3
5	KZ	2	2	0.147	7	7	7	7	7	7	7
6	KZ	4	2	0.147	1	2	2	2	2	1	2
7	KZ	1	3	0.147	7	7	7	7	7	7	7
8	KZ	2	3	0.147	5	3	5	3	3	2	3
9	KZ	4	3	0.147	1	1	1	1	1	1	1
10	KZ	1	4	0.147	1	1	1	1	1	1	1
11	KZ	2	4	0.147	7	7	7	7	7	7	7
12	KZ	4	4	0.147	6	5	5	6	5	4	5
13	KZ	1	1	0.147	1	1	2	3	4	2	3
14	KZ	2	1	0.147	2	1	2	3	3	2	3
15	KZ	4	1	0.147	1	2	2	2		2	3
16	KZ	1	2	0.147	1	2	3	3	1	1	1
17	KZ	2	2	0.147	1	2	4	4	3	2	2
18	KZ	4	2	0.147	1	3	4	3	2	2	2
19	KZ	1	3	0.147	1	2	2	2	1	1	2
20	KZ	2	3	0.147	1	2	2	2	2	1	2
21	KZ	4	3	0.147	1	2	2	2	2	1	2
22	KZ	1	4	0.147	1	2	2	2	3	1	2
23	KZ	2	4	0.147	1	3	4	3	5	2	4
24	KZ	4	4	0.147	2	3	3	3	4	2	3
25	UZ	1	1	0.197	1	1	1	1	1	1	1
26	UZ	2	1	0.197	7	7	7	7	7	7	7
27	UZ	4	1	0.197	1	5	2	3	5	2	4
28	UZ	1	2	0.197	5	4	3	3	5	2	5
29	UZ	2	2	0.197	7	7	7	7	7	7	7
30	UZ	4	2	0.197	1	1	1	1	1	1	1
31	UZ	1	3	0.197	1	1	1	2	1	1	2
32	UZ	2	3	0.197	7	7	7	7	7	7	7
33	UZ	4	3	0.197	4	3	3	5	5	4	4
34	UZ	1	4	0.197	2	1	1	1	1	1	2
35	UZ	2	4	0.197	7	7	7	7	7	7	7
36	UZ	4	4	0.197	2	3	2	6	4	4	5
37	UZ	1	1	0.197	2	1	2	2	1	1	2
38	UZ	2	1	0.197	5	4	4	5	4	1	4

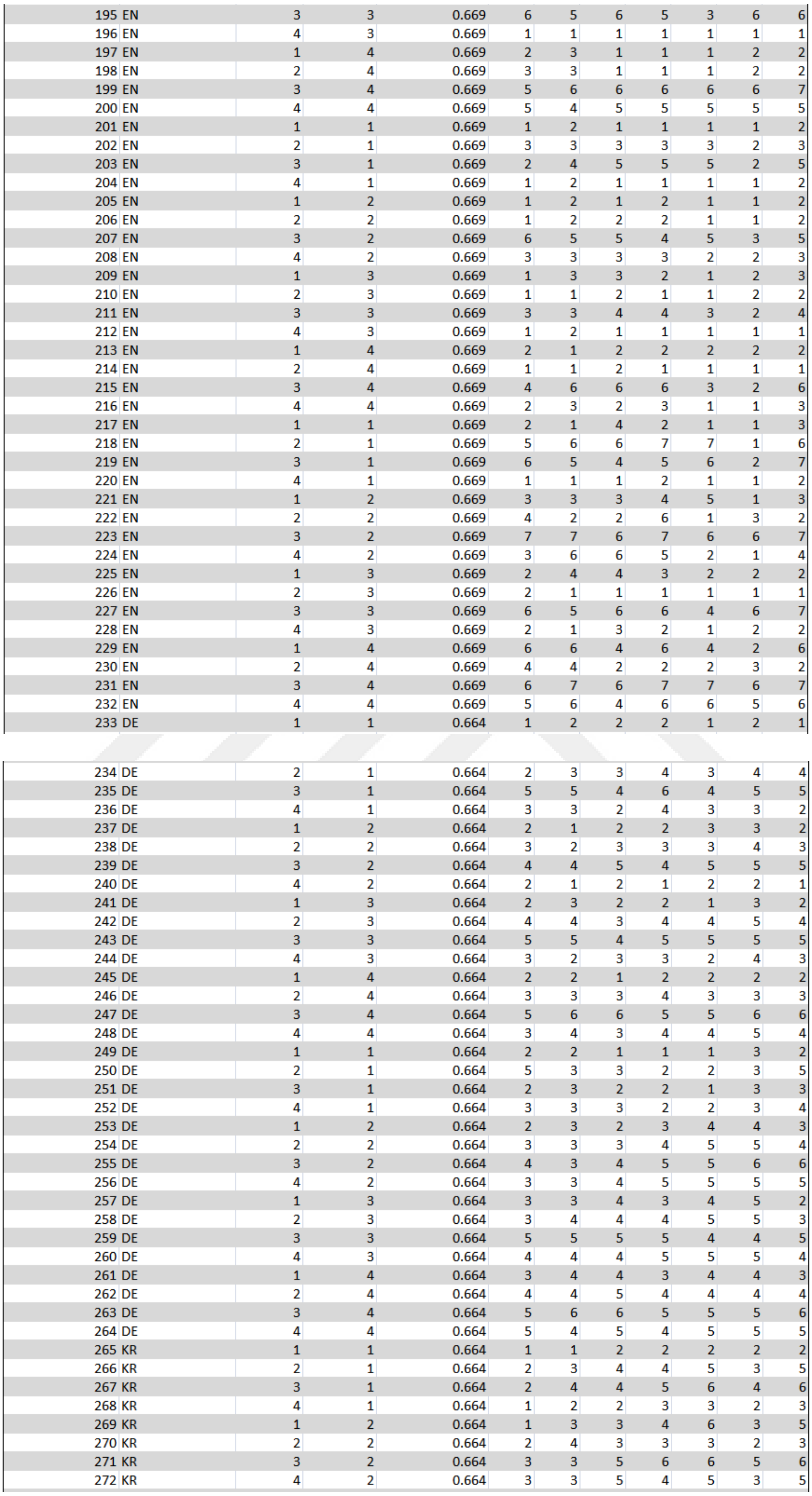




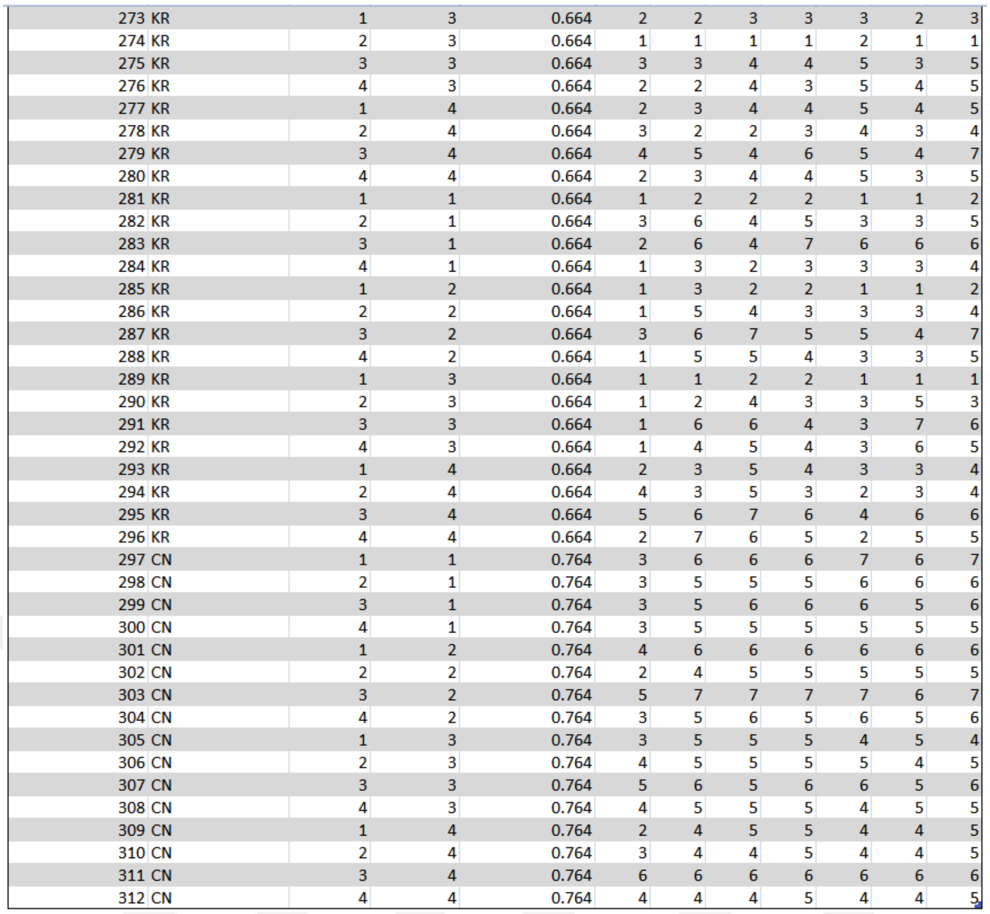
39	UZ	4	1	0.197	3	4	3	4		1	4
40	UZ	1	2	0.197	1	3	3	4	3	2	4
41	UZ	2	2	0.197	2	3	3	4	3	2	4
42	UZ	4	2	0.197	2	2	2	2	2	1	2
43	UZ	1	3	0.197	1	2	2	2	2	2	2
44	UZ	2	3	0.197	4	5	6	6	6	6	6
45	UZ	4	3	0.197	3	2	3	4	3	3	4
46	UZ	1	4	0.197	1	3	3	3	2	2	3
47	UZ	2	4	0.197	1	3		3	2	2	2
48	UZ	4	4	0.197	6	5	3	2	2	2	4
49	GR	1	1	0.646	1	1	3	2	2	3	2
50	GR	2	1	0.646	2	2	5	3	4	3	4
51	GR	4	1	0.646	1	1	5	2	3	4	3
52	GR	1	2	0.646	2	2	3	2	2	2	3
53	GR	2	2	0.646	3	2	3	2	3	2	2
54	GR	4	2	0.646	3	3	5	4	5	4	5
55	GR	1	3	0.646	1	1	3	2	2	1	2
56	GR	2	3	0.646	2	2	3	3	3	3	3
57	GR	4	3	0.646	2	2	4	4	4	4	5
58	GR	1	4	0.646	3	2	3	3	2	1	3
59	GR	2	4	0.646	2	2	2	2	2	1	2
60	GR	4	4	0.646	2	2	2	2	2	2	2
61	GR	1	1	0.646	1	3	2	3	4	3	4
62	GR	2	1	0.646	4	4	4	4	6	4	5
63	GR	4	1	0.646	5	5	6	5	6		6
64	GR	1	2	0.646	1	3	3	4	3	3	3
65	GR	2	2	0.646	1	5	5	6	5	5	5
66	GR	4	2	0.646	1	4	4	5	5	4	5
67	GR	1	3	0.646	1	2	2	1	2	2	1
68	GR	2	3	0.646	1	2	2	1	2	2	2
69	GR	4	3	0.646	1	1	1	1	1	1	1
70	GR	1	4	0.646	1	3	2	3	2	2	2
71	GR	2	4	0.646	1	1	1	1	1	1	1
72	GR	4	4	0.646	1	3	3	4	3	3	3
73	RU	1	1	0.647	1	1	1	1	1	1	1
74	RU	2	1	0.647	4	1	3	3	2	2	2
75	RU	3	1	0.647	5	4	4	5	2	3	3
76	RU	4	1	0.647	5	3	4	5	2	3	4
77	RU	1	2	0.647	4	5	2	2	2	2	3
78	RU	2	2	0.647	2	2	3	3	3	3	4
79	RU	3	2	0.647	6	5	4	4	4	4	5
80	RU	4	2	0.647	1	1	1	1	1	1	1
81	RU	1	3	0.647	1	1	1	1	1	1	1
82	RU	2	3	0.647	4	5	2	3	3	5	4
83	RU	3	3	0.647	5	3	2	4	4	4	4
84	RU	4	3	0.647	4	3	2	5	5	4	4
85	RU	1	4	0.647	1	1	1	1	1	1	1
86	RU	2	4	0.647	2	2	2	2	2	2	2
87	RU	3	4	0.647	3	3	4	4	4	3	3
88	RU	4	4	0.647	2	2	3	3	3	2	2
89	RU	1	1	0.647	1	3	2	3	2	1	4
90	RU	2	1	0.647	5	5	1	6	4	1	6
91	RU	3	1	0.647	1	5	3	5	2	1	5
92	RU	4	1	0.647	1	2	1	2	1	1	2
93	RU	1	2	0.647	1	3	1	4	2	1	4
94	RU	2	2	0.647	1	6	2	7	3	1	6
95	RU	3	2	0.647	1	7	7	7	4	3	7
96	RU	4	2	0.647	3	7	7	7	5	6	7
97	RU	1	3	0.647	1	2	3	2	1	3	2
98	RU	2	3	0.647	1	2	2	2	1	2	2
99	RU	3	3	0.647	2	4	5	5	3	4	4
100	RU	4	3	0.647	1	2	2	3	3	2	3
101	RU	1	4	0.647	2	3	3	3	2	3	3
102	RU	2	4	0.647	1	3	2	2	2	1	2
103	RU	3	4	0.647	2	7	7	7	7	1	7
104	RU	4	4	0.647	3	3	4	4	4	3	4
105	RU	1	1	0.647	1	2	4	3	2	2	3
106	RU	2	1	0.647	3	2	3	5	2	2	5
107	RU	3	1	0.647	3	7	7	7	7	6	6
108	RU	4	1	0.647	3	2	4	4	2	2	6
109	RU	1	2	0.647	1	2	2	1	2	2	2
110	RU	2	2	0.647	3	5	4	3	2	5	6
111	RU	3	2	0.647	3	7	7	7	3	7	7
112	RU	4	2	0.647	3	7	7	7	5	7	7
113	RU	1	3	0.647	1	2	4	2	2	4	3
114	RU	2	3	0.647	3	2	3	2	3	3	3
115	RU	3	3	0.647	4	5	5	5	5	5	7
116	RU	4	3	0.647	4	5	5	5	5	5	7



117	RU	1	4	0.647	1	2	2	1	2	2	2
118	RU	2	4	0.647	1	2	3	1	2	2	2
119	RU	3	4	0.647	3	7	7	7	3	7	7
120	RU	4	4	0.647	2	6	7	7	3	7	7
121	FR	1	1	0.662	1	1	2	1	1	1	2
122	FR	2	1	0.662	1	1	1	3	3	1	3
123	FR	3	1	0.662	3	3	4	6	6	6	6
124	FR	4	1	0.662	1	2	2	1	1	2	2
125	FR	1	2	0.662	1	1	1	1	1	1	1
126	FR	2	2	0.662	1	1	1	1	1	1	1
127	FR	3	2	0.662	4	4	4	4	4	4	5
128	FR	4	2	0.662	3	4	4	6	5	5	6
129	FR	1	3	0.662	1	1	1	1	1	1	1
130	FR	2	3	0.662	1	1	1	1	1	1	1
131	FR	3	3	0.662	6	6	6	5	5	6	6
132	FR	4	3	0.662	6	6	6	4	3	6	6
133	FR	1	4	0.662	1	1	1	2	1	1	2
134	FR	2	4	0.662	1	2	1	4	4	4	4
135	FR	3	4	0.662	5	6	6	7	7	7	7
136	FR	4	4	0.662	6	6	6	6	6	6	6
137	FR	1	1	0.662	1	4	4	4	2	4	4
138	FR	2	1	0.662	2	5	4	4	3	4	4
139	FR	3	1	0.662	4	6	6	6	5	5	6
140	FR	4	1	0.662	1	3	2	2	2	2	2
141	FR	1	2	0.662	1	1	2	2	2	1	2
142	FR	2	2	0.662	1	1		3	4	1	3
143	FR	3	2	0.662	1	7	7	5	7	1	7
144	FR	4	2	0.662	1	4	5	3	7	5	6
145	FR	1	3	0.662	1	3	3	2	3	4	3
146	FR	2	3	0.662	1	1	1	1	1	2	1
147	FR	3	3	0.662	3	7	6	5	7	7	7
148	FR	4	3	0.662	1	2	1	1	2	1	2
149	FR	1	4	0.662	1	3	2	2	3	2	3
150	FR	2	4	0.662	1	2	2	2	3	3	3
151	FR	3	4	0.662	7	7	7	2	7	5	7
152	FR	4	4	0.662	1	5	2	1	2	1	3
153	FR	1	1	0.662	1	1	2	1	1	2	1
154	FR	2	1	0.662	2	2	3	2	1	2	2
155	FR	3	1	0.662	2	3	4	3	2	2	4
156	FR	4	1	0.662	1	1	3	1	1	1	1
157	FR	1	2	0.662	1	1	1	1	1	1	1
158	FR	2	2	0.662	1	2	1	1	1	1	2
159	FR	3	2	0.662	4	4	3	3	1	2	4
160	FR	4	2	0.662	2	2	3	3	2	2	3
161	FR	1	3	0.662	2	3	3	3	1	1	3
162	FR	2	3	0.662	1	1	1	1	1	1	1
163	FR	3	3	0.662	2	2	3	2	1	2	3
164	FR	4	3	0.662	2	2	2	2	1	2	2
165	FR	1	4	0.662	1	1	2	2	1	1	2
166	FR	2	4	0.662	1	1	1	1	1	1	1
167	FR	3	4	0.662	4	5	3	4	2	2	4
168	FR	4	4	0.662	1	2	3	2	2	2	3
169	EN	1	1	0.669	4	3	2	2	2	2	3
170	EN	2	1	0.669	6	6	3	4	4	4	5
171	EN	3	1	0.669	6	4	5	6	6	6	6
172	EN	4	1	0.669	4	3	2	2	2	2	2
173	EN	1	2	0.669	4	4	4	6	6	4	7
174	EN	2	2	0.669	4	4	4	6	6	4	7
175	EN	3	2	0.669	6	6	6	6	6	6	7
176	EN	4	2	0.669	4	4	4	6	6	4	7
177	EN	1	3	0.669	4	4	4	4	4	4	3
178	EN	2	3	0.669	4	4	4	4	4	4	3
179	EN	3	3	0.669	6	6	6	6	6	6	7
180	EN	4	3	0.669	2	2	2	2	2	2	2
181	EN	1	4	0.669	4	4	3	6	6	3	7
182	EN	2	4	0.669	4	4	3	2	3	2	5
183	EN	3	4	0.669	7	7	7	7	7	7	7
184	EN	4	4	0.669	6	6	6	6	6	7	7
185	EN	1	1	0.669	1	2	2	1	1	2	1
186	EN	2	1	0.669	2	5	4	3	2	4	4
187	EN	3	1	0.669	5	7	7	5	5	6	6
188	EN	4	1	0.669	1	2	3	1	1	2	2
189	EN	1	2	0.669	2	3	4	2	2	3	3
190	EN	2	2	0.669	3	2	5	2	2	2	4
191	EN	3	2	0.669	6	7	7	6	6	7	7
192	EN	4	2	0.669	3	2	4	2	2	2	3
193	EN	1	3	0.669	3	3	2	1	3	1	3
194	EN	2	3	0.669	2	2	1	1	2	2	2



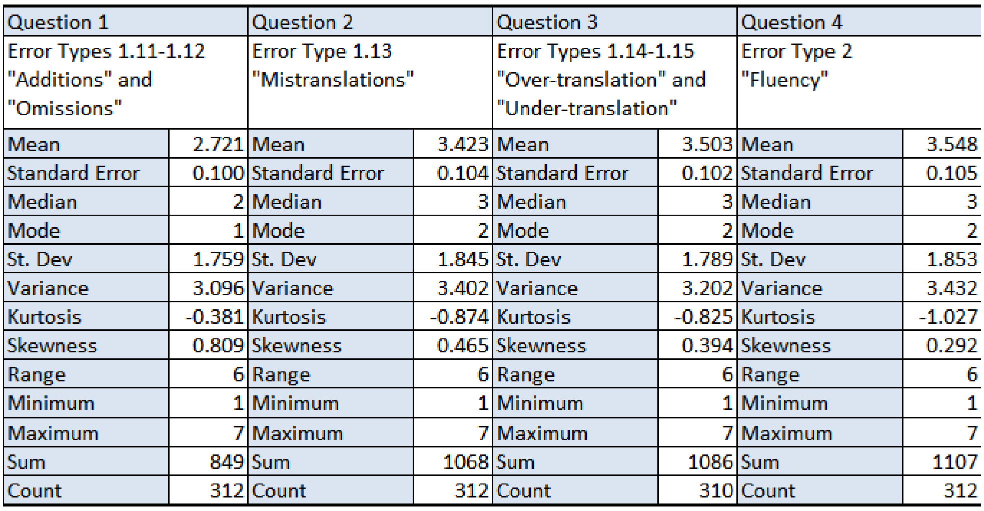
195	EN	3	3	0.669	6	5	6	5	3	6	6
196	EN	4	3	0.669	1	1	1	1	1	1	1
197	EN	1	4	0.669	2	3	1	1	1	2	2
198	EN	2	4	0.669	3	3	1	1	1	2	2
199	EN	3	4	0.669	5	6	6	6	6	6	7
200	EN	4	4	0.669	5	4	5	5	5	5	5
201	EN	1	1	0.669	1	2	1	1	1	1	2
202	EN	2	1	0.669	3	3	3	3	3	2	3
203	EN	3	1	0.669	2	4	5	5	5	2	5
204	EN	4	1	0.669	1	2	1	1	1	1	2
205	EN	1	2	0.669	1	2	1	2	1	1	2
206	EN	2	2	0.669	1	2	2	2	1	1	2
207	EN	3	2	0.669	6	5	5	4	5	3	5
208	EN	4	2	0.669	3	3	3	3	2	2	3
209	EN	1	3	0.669	1	3	3	2	1	2	3
210	EN	2	3	0.669	1	1	2	1	1	2	2
211	EN	3	3	0.669	3	3	4	4	3	2	4
212	EN	4	3	0.669	1	2	1	1	1	1	1
213	EN	1	4	0.669	2	1	2	2	2	2	2
214	EN	2	4	0.669	1	1	2	1	1	1	1
215	EN	3	4	0.669	4	6	6	6	3	2	6
216	EN	4	4	0.669	2	3	2	3	1	1	3
217	EN	1	1	0.669	2	1	4	2	1	1	3
218	EN	2	1	0.669	5	6	6	7	7	1	6
219	EN	3	1	0.669	6	5	4	5	6	2	7
220	EN	4	1	0.669	1	1	1	2	1	1	2
221	EN	1	2	0.669	3	3	3	4	5	1	3
222	EN	2	2	0.669	4	2	2	6	1	3	2
223	EN	3	2	0.669	7	7	6	7	6	6	7
224	EN	4	2	0.669	3	6	6	5	2	1	4
225	EN	1	3	0.669	2	4	4	3	2	2	2
226	EN	2	3	0.669	2	1	1	1	1	1	1
227	EN	3	3	0.669	6	5	6	6	4	6	7
228	EN	4	3	0.669	2	1	3	2	1	2	2
229	EN	1	4	0.669	6	6	4	6	4	2	6
230	EN	2	4	0.669	4	4	2	2	2	3	2
231	EN	3	4	0.669	6	7	6	7	7	6	7
232	EN	4	4	0.669	5	6	4	6	6	5	6
233	DE	1	1	0.664	1	2	2	2	1	2	1
234	DE	2	1	0.664	2	3	3	4	3	4	4
235	DE	3	1	0.664	5	5	4	6	4	5	5
236	DE	4	1	0.664	3	3	2	4	3	3	2
237	DE	1	2	0.664	2	1	2	2	3	3	2
238	DE	2	2	0.664	3	2	3	3	3	4	3
239	DE	3	2	0.664	4	4	5	4	5	5	5
240	DE	4	2	0.664	2	1	2	1	2	2	1
241	DE	1	3	0.664	2	3	2	2	1	3	2
242	DE	2	3	0.664	4	4	3	4	4	5	4
243	DE	3	3	0.664	5	5	4	5	5	5	5
244	DE	4	3	0.664	3	2	3	3	2	4	3
245	DE	1	4	0.664	2	2	1	2	2	2	2
246	DE	2	4	0.664	3	3	3	4	3	3	3
247	DE	3	4	0.664	5	6	6	5	5	6	6
248	DE	4	4	0.664	3	4	3	4	4	5	4
249	DE	1	1	0.664	2	2	1	1	1	3	2
250	DE	2	1	0.664	5	3	3	2	2	3	5
251	DE	3	1	0.664	2	3	2	2	1	3	3
252	DE	4	1	0.664	3	3	3	2	2	3	4
253	DE	1	2	0.664	2	3	2	3	4	4	3
254	DE	2	2	0.664	3	3	3	4	5	5	4
255	DE	3	2	0.664	4	3	4	5	5	6	6
256	DE	4	2	0.664	3	3	4	5	5	5	5
257	DE	1	3	0.664	3	3	4	3	4	5	2
258	DE	2	3	0.664	3	4	4	4	5	5	3
259	DE	3	3	0.664	5	5	5	5	4	4	5
260	DE	4	3	0.664	4	4	4	5	5	5	4
261	DE	1	4	0.664	3	4	4	3	4	4	3
262	DE	2	4	0.664	4	4	5	4	4	4	4
263	DE	3	4	0.664	5	6	6	5	5	5	6
264	DE	4	4	0.664	5	4	5	4	5	5	5
265	KR	1	1	0.664	1	1	2	2	2	2	2
266	KR	2	1	0.664	2	3	4	4	5	3	5
267	KR	3	1	0.664	2	4	4	5	6	4	6
268	KR	4	1	0.664	1	2	2	3	3	2	3
269	KR	1	2	0.664	1	3	3	4	6	3	5
270	KR	2	2	0.664	2	4	3	3	3	2	3
271	KR	3	2	0.664	3	3	5	6	6	5	6
272	KR	4	2	0.664	3	3	5	4	5	3	5



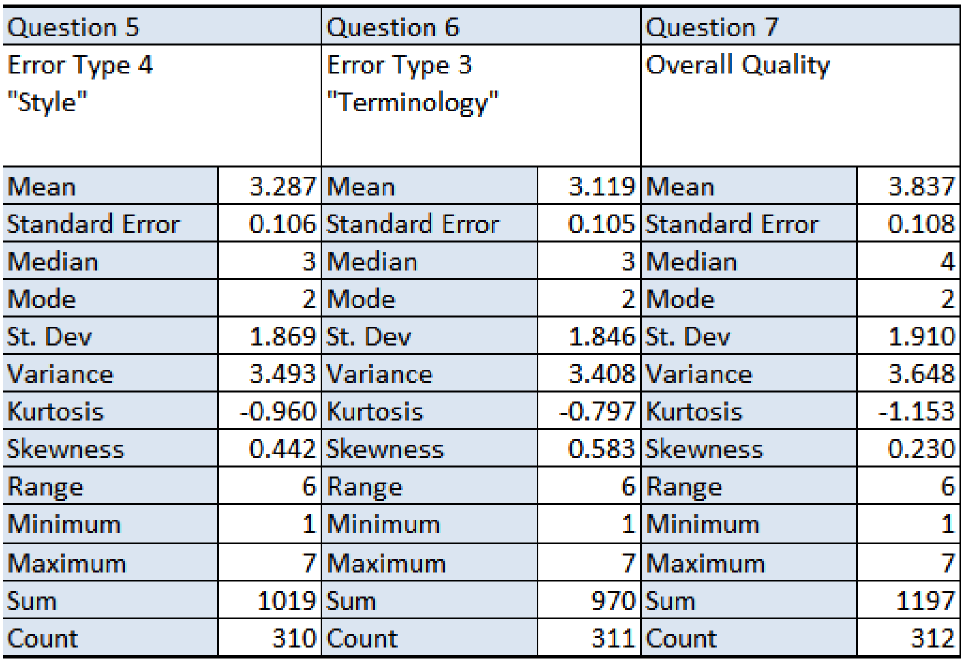
273	KR	1	3	0.664	2	2	3	3	3	2	3
274	KR	2	3	0.664	1	1	1	1	2	1	1
275	KR	3	3	0.664	3	3	4	4	5	3	5
276	KR	4	3	0.664	2	2	4	3	5	4	5
277	KR	1	4	0.664	2	3	4	4	5	4	5
278	KR	2	4	0.664	3	2	2	3	4	3	4
279	KR	3	4	0.664	4	5	4	6	5	4	7
280	KR	4	4	0.664	2	3	4	4	5	3	5
281	KR	1	1	0.664	1	2	2	2	1	1	2
282	KR	2	1	0.664	3	6	4	5	3	3	5
283	KR	3	1	0.664	2	6	4	7	6	6	6
284	KR	4	1	0.664	1	3	2	3	3	3	4
285	KR	1	2	0.664	1	3	2	2	1	1	2
286	KR	2	2	0.664	1	5	4	3	3	3	4
287	KR	3	2	0.664	3	6	7	5	5	4	7
288	KR	4	2	0.664	1	5	5	4	3	3	5
289	KR	1	3	0.664	1	1	2	2	1	1	1
290	KR	2	3	0.664	1	2	4	3	3	5	3
291	KR	3	3	0.664	1	6	6	4	3	7	6
292	KR	4	3	0.664	1	4	5	4	3	6	5
293	KR	1	4	0.664	2	3	5	4	3	3	4
294	KR	2	4	0.664	4	3	5	3	2	3	4
295	KR	3	4	0.664	5	6	7	6	4	6	6
296	KR	4	4	0.664	2	7	6	5	2	5	5
297	CN	1	1	0.764	3	6	6	6	7	6	7
298	CN	2	1	0.764	3	5	5	5	6	6	6
299	CN	3	1	0.764	3	5	6	6	6	5	6
300	CN	4	1	0.764	3	5	5	5	5	5	5
301	CN	1	2	0.764	4	6	6	6	6	6	6
302	CN	2	2	0.764	2	4	5	5	5	5	5
303	CN	3	2	0.764	5	7	7	7	7	6	7
304	CN	4	2	0.764	3	5	6	5	6	5	6
305	CN	1	3	0.764	3	5	5	5	4	5	4
306	CN	2	3	0.764	4	5	5	5	5	4	5
307	CN	3	3	0.764	5	6	5	6	6	5	6
308	CN	4	3	0.764	4	5	5	5	4	5	5
309	CN	1	4	0.764	2	4	5	5	4	4	5
310	CN	2	4	0.764	3	4	4	5	4	4	5
311	CN	3	4	0.764	6	6	6	6	6	6	6
312	CN	4	4	0.764	4	4	4	5	4	4	5

APPENDIX J

DESCRIPTIVE STATISTICS OF SURVEY RESULTS



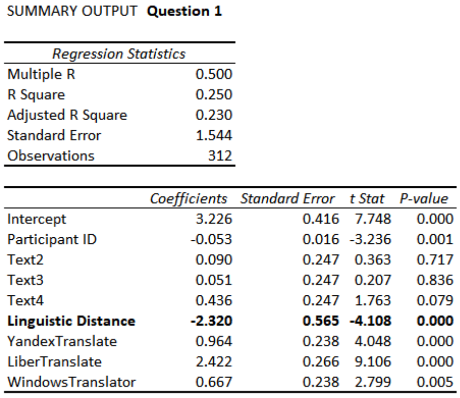
Question 1		Question 2		Question 3		Question 4	
Error Types 1.11-1.12 "Additions" and "Omissions"		Error Type 1.13 "Mistranslations"		Error Types 1.14-1.15 "Over-translation" and "Under-translation"		Error Type 2 "Fluency"	
Mean	2.721	Mean	3.423	Mean	3.503	Mean	3.548
Standard Error	0.100	Standard Error	0.104	Standard Error	0.102	Standard Error	0.105
Median	2	Median	3	Median	3	Median	3
Mode	1	Mode	2	Mode	2	Mode	2
St. Dev	1.759	St. Dev	1.845	St. Dev	1.789	St. Dev	1.853
Variance	3.096	Variance	3.402	Variance	3.202	Variance	3.432
Kurtosis	-0.381	Kurtosis	-0.874	Kurtosis	-0.825	Kurtosis	-1.027
Skewness	0.809	Skewness	0.465	Skewness	0.394	Skewness	0.292
Range	6	Range	6	Range	6	Range	6
Minimum	1	Minimum	1	Minimum	1	Minimum	1
Maximum	7	Maximum	7	Maximum	7	Maximum	7
Sum	849	Sum	1068	Sum	1086	Sum	1107
Count	312	Count	312	Count	310	Count	312



Question 5		Question 6		Question 7	
Error Type 4 "Style"		Error Type 3 "Terminology"		Overall Quality	
Mean	3.287	Mean	3.119	Mean	3.837
Standard Error	0.106	Standard Error	0.105	Standard Error	0.108
Median	3	Median	3	Median	4
Mode	2	Mode	2	Mode	2
St. Dev	1.869	St. Dev	1.846	St. Dev	1.910
Variance	3.493	Variance	3.408	Variance	3.648
Kurtosis	-0.960	Kurtosis	-0.797	Kurtosis	-1.153
Skewness	0.442	Skewness	0.583	Skewness	0.230
Range	6	Range	6	Range	6
Minimum	1	Minimum	1	Minimum	1
Maximum	7	Maximum	7	Maximum	7
Sum	1019	Sum	970	Sum	1197
Count	310	Count	311	Count	312

APPENDIX K

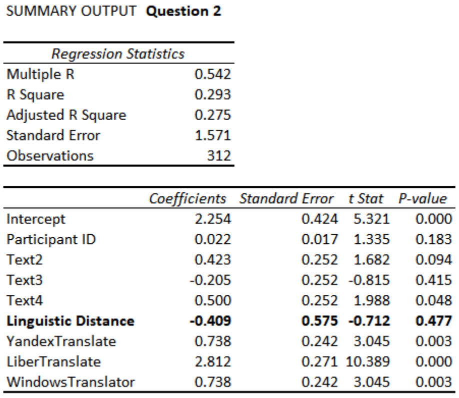
MULTIVARIATE REGRESSIONS – FIRST SET



SUMMARY OUTPUT **Question 1**

Regression Statistics	
Multiple R	0.500
R Square	0.250
Adjusted R Square	0.230
Standard Error	1.544
Observations	312

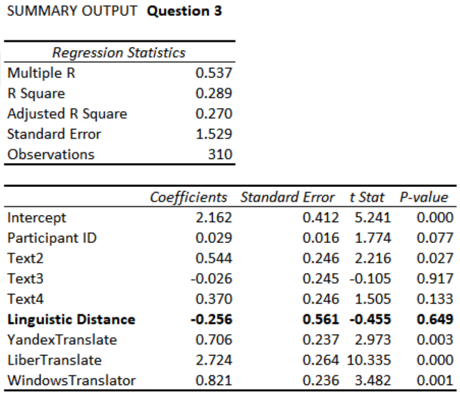
	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	3.226	0.416	7.748	0.000
Participant ID	-0.053	0.016	-3.236	0.001
Text2	0.090	0.247	0.363	0.717
Text3	0.051	0.247	0.207	0.836
Text4	0.436	0.247	1.763	0.079
Linguistic Distance	**-2.320**	**0.565**	**-4.108**	**0.000**
YandexTranslate	0.964	0.238	4.048	0.000
LiberTranslate	2.422	0.266	9.106	0.000
WindowsTranslator	0.667	0.238	2.799	0.005



SUMMARY OUTPUT **Question 2**

Regression Statistics	
Multiple R	0.542
R Square	0.293
Adjusted R Square	0.275
Standard Error	1.571
Observations	312

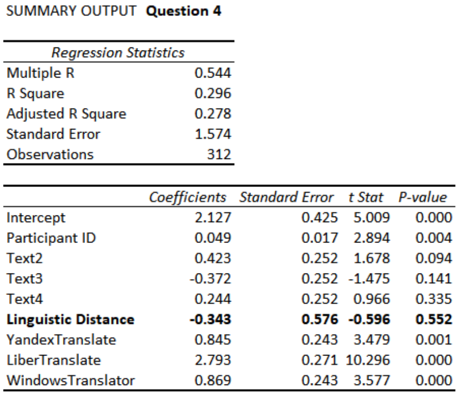
	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	2.254	0.424	5.321	0.000
Participant ID	0.022	0.017	1.335	0.183
Text2	0.423	0.252	1.682	0.094
Text3	-0.205	0.252	-0.815	0.415
Text4	0.500	0.252	1.988	0.048
Linguistic Distance	**-0.409**	**0.575**	**-0.712**	**0.477**
YandexTranslate	0.738	0.242	3.045	0.003
LiberTranslate	2.812	0.271	10.389	0.000
WindowsTranslator	0.738	0.242	3.045	0.003



SUMMARY OUTPUT **Question 3**

Regression Statistics	
Multiple R	0.537
R Square	0.289
Adjusted R Square	0.270
Standard Error	1.529
Observations	310

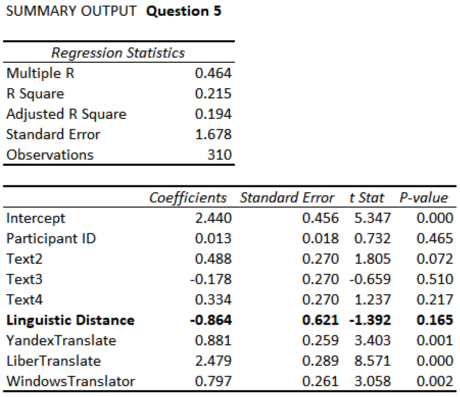
	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	2.162	0.412	5.241	0.000
Participant ID	0.029	0.016	1.774	0.077
Text2	0.544	0.246	2.216	0.027
Text3	-0.026	0.245	-0.105	0.917
Text4	0.370	0.246	1.505	0.133
Linguistic Distance	**-0.256**	**0.561**	**-0.455**	**0.649**
YandexTranslate	0.706	0.237	2.973	0.003
LiberTranslate	2.724	0.264	10.335	0.000
WindowsTranslator	0.821	0.236	3.482	0.001



SUMMARY OUTPUT **Question 4**

Regression Statistics	
Multiple R	0.544
R Square	0.296
Adjusted R Square	0.278
Standard Error	1.574
Observations	312

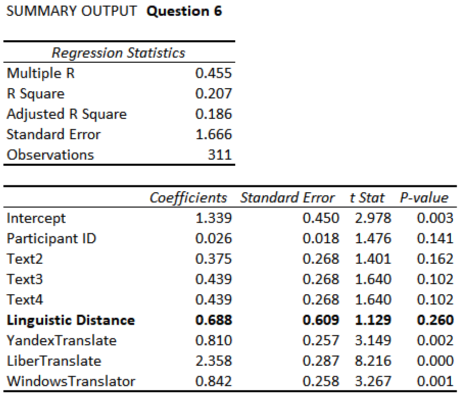
	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	2.127	0.425	5.009	0.000
Participant ID	0.049	0.017	2.894	0.004
Text2	0.423	0.252	1.678	0.094
Text3	-0.372	0.252	-1.475	0.141
Text4	0.244	0.252	0.966	0.335
Linguistic Distance	**-0.343**	**0.576**	**-0.596**	**0.552**
YandexTranslate	0.845	0.243	3.479	0.001
LiberTranslate	2.793	0.271	10.296	0.000
WindowsTranslator	0.869	0.243	3.577	0.000



SUMMARY OUTPUT **Question 5**

Regression Statistics	
Multiple R	0.464
R Square	0.215
Adjusted R Square	0.194
Standard Error	1.678
Observations	310

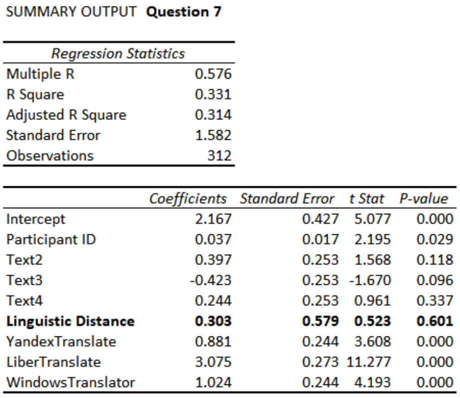
	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	2.440	0.456	5.347	0.000
Participant ID	0.013	0.018	0.732	0.465
Text2	0.488	0.270	1.805	0.072
Text3	-0.178	0.270	-0.659	0.510
Text4	0.334	0.270	1.237	0.217
Linguistic Distance	**-0.864**	**0.621**	**-1.392**	**0.165**
YandexTranslate	0.881	0.259	3.403	0.001
LiberTranslate	2.479	0.289	8.571	0.000
WindowsTranslator	0.797	0.261	3.058	0.002



SUMMARY OUTPUT **Question 6**

Regression Statistics	
Multiple R	0.455
R Square	0.207
Adjusted R Square	0.186
Standard Error	1.666
Observations	311

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	1.339	0.450	2.978	0.003
Participant ID	0.026	0.018	1.476	0.141
Text2	0.375	0.268	1.401	0.162
Text3	0.439	0.268	1.640	0.102
Text4	0.439	0.268	1.640	0.102
Linguistic Distance	**0.688**	**0.609**	**1.129**	**0.260**
YandexTranslate	0.810	0.257	3.149	0.002
LiberTranslate	2.358	0.287	8.216	0.000
WindowsTranslator	0.842	0.258	3.267	0.001



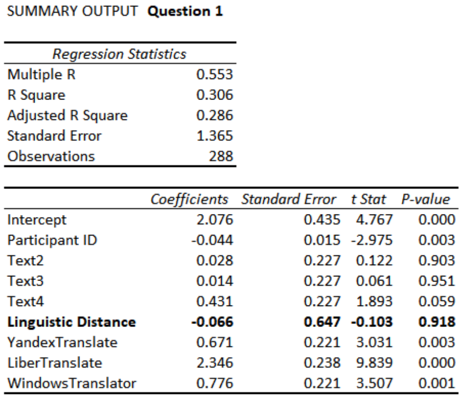
SUMMARY OUTPUT **Question 7**

Regression Statistics	
Multiple R	0.576
R Square	0.331
Adjusted R Square	0.314
Standard Error	1.582
Observations	312

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	2.167	0.427	5.077	0.000
Participant ID	0.037	0.017	2.195	0.029
Text2	0.397	0.253	1.568	0.118
Text3	-0.423	0.253	-1.670	0.096
Text4	0.244	0.253	0.961	0.337
Linguistic Distance	**0.303**	**0.579**	**0.523**	**0.601**
YandexTranslate	0.881	0.244	3.608	0.000
LiberTranslate	3.075	0.273	11.277	0.000
WindowsTranslator	1.024	0.244	4.193	0.000

APPENDIX L

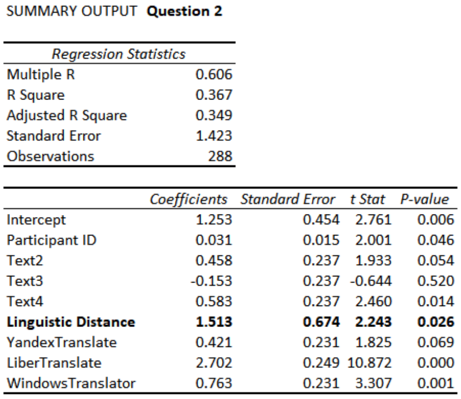
MULTIVARIATE REGRESSIONS – SECOND SET



SUMMARY OUTPUT **Question 1**

Regression Statistics	
Multiple R	0.553
R Square	0.306
Adjusted R Square	0.286
Standard Error	1.365
Observations	288

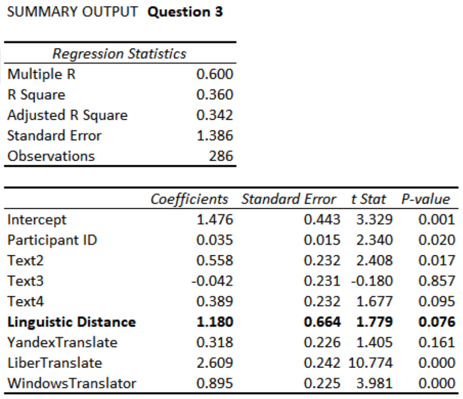
	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	2.076	0.435	4.767	0.000
Participant ID	-0.044	0.015	-2.975	0.003
Text2	0.028	0.227	0.122	0.903
Text3	0.014	0.227	0.061	0.951
Text4	0.431	0.227	1.893	0.059
Linguistic Distance	**-0.066**	**0.647**	**-0.103**	**0.918**
YandexTranslate	0.671	0.221	3.031	0.003
LiberTranslate	2.346	0.238	9.839	0.000
WindowsTranslator	0.776	0.221	3.507	0.001



SUMMARY OUTPUT **Question 2**

Regression Statistics	
Multiple R	0.606
R Square	0.367
Adjusted R Square	0.349
Standard Error	1.423
Observations	288

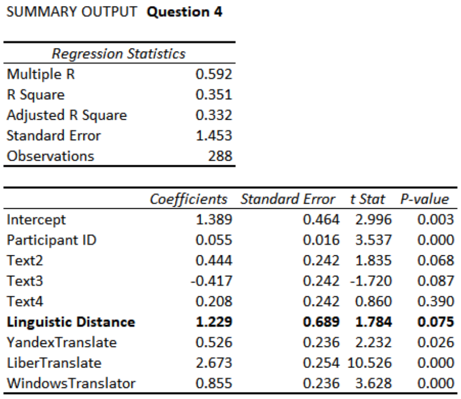
	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	1.253	0.454	2.761	0.006
Participant ID	0.031	0.015	2.001	0.046
Text2	0.458	0.237	1.933	0.054
Text3	-0.153	0.237	-0.644	0.520
Text4	0.583	0.237	2.460	0.014
Linguistic Distance	**1.513**	**0.674**	**2.243**	**0.026**
YandexTranslate	0.421	0.231	1.825	0.069
LiberTranslate	2.702	0.249	10.872	0.000
WindowsTranslator	0.763	0.231	3.307	0.001



SUMMARY OUTPUT **Question 3**

Regression Statistics	
Multiple R	0.600
R Square	0.360
Adjusted R Square	0.342
Standard Error	1.386
Observations	286

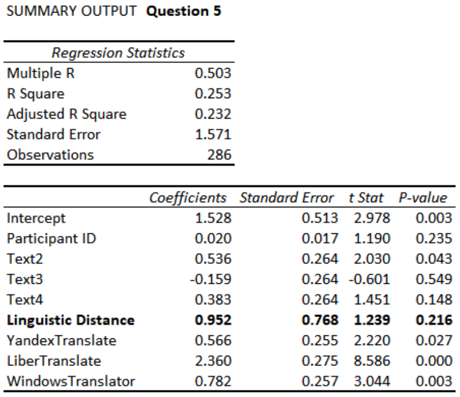
	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	1.476	0.443	3.329	0.001
Participant ID	0.035	0.015	2.340	0.020
Text2	0.558	0.232	2.408	0.017
Text3	-0.042	0.231	-0.180	0.857
Text4	0.389	0.232	1.677	0.095
Linguistic Distance	**1.180**	**0.664**	**1.779**	**0.076**
YandexTranslate	0.318	0.226	1.405	0.161
LiberTranslate	2.609	0.242	10.774	0.000
WindowsTranslator	0.895	0.225	3.981	0.000



SUMMARY OUTPUT **Question 4**

Regression Statistics	
Multiple R	0.592
R Square	0.351
Adjusted R Square	0.332
Standard Error	1.453
Observations	288

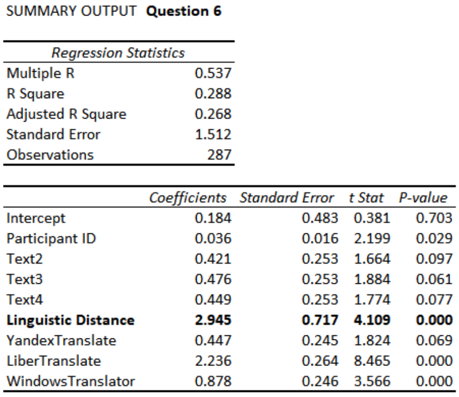
	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	1.389	0.464	2.996	0.003
Participant ID	0.055	0.016	3.537	0.000
Text2	0.444	0.242	1.835	0.068
Text3	-0.417	0.242	-1.720	0.087
Text4	0.208	0.242	0.860	0.390
Linguistic Distance	**1.229**	**0.689**	**1.784**	**0.075**
YandexTranslate	0.526	0.236	2.232	0.026
LiberTranslate	2.673	0.254	10.526	0.000
WindowsTranslator	0.855	0.236	3.628	0.000



SUMMARY OUTPUT **Question 5**

Regression Statistics	
Multiple R	0.503
R Square	0.253
Adjusted R Square	0.232
Standard Error	1.571
Observations	286

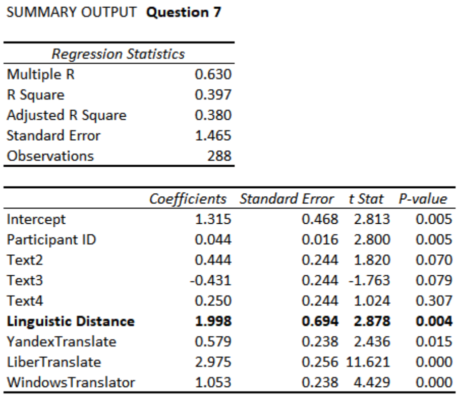
	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	1.528	0.513	2.978	0.003
Participant ID	0.020	0.017	1.190	0.235
Text2	0.536	0.264	2.030	0.043
Text3	-0.159	0.264	-0.601	0.549
Text4	0.383	0.264	1.451	0.148
Linguistic Distance	**0.952**	**0.768**	**1.239**	**0.216**
YandexTranslate	0.566	0.255	2.220	0.027
LiberTranslate	2.360	0.275	8.586	0.000
WindowsTranslator	0.782	0.257	3.044	0.003



SUMMARY OUTPUT **Question 6**

Regression Statistics	
Multiple R	0.537
R Square	0.288
Adjusted R Square	0.268
Standard Error	1.512
Observations	287

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	0.184	0.483	0.381	0.703
Participant ID	0.036	0.016	2.199	0.029
Text2	0.421	0.253	1.664	0.097
Text3	0.476	0.253	1.884	0.061
Text4	0.449	0.253	1.774	0.077
Linguistic Distance	**2.945**	**0.717**	**4.109**	**0.000**
YandexTranslate	0.447	0.245	1.824	0.069
LiberTranslate	2.236	0.264	8.465	0.000
WindowsTranslator	0.878	0.246	3.566	0.000



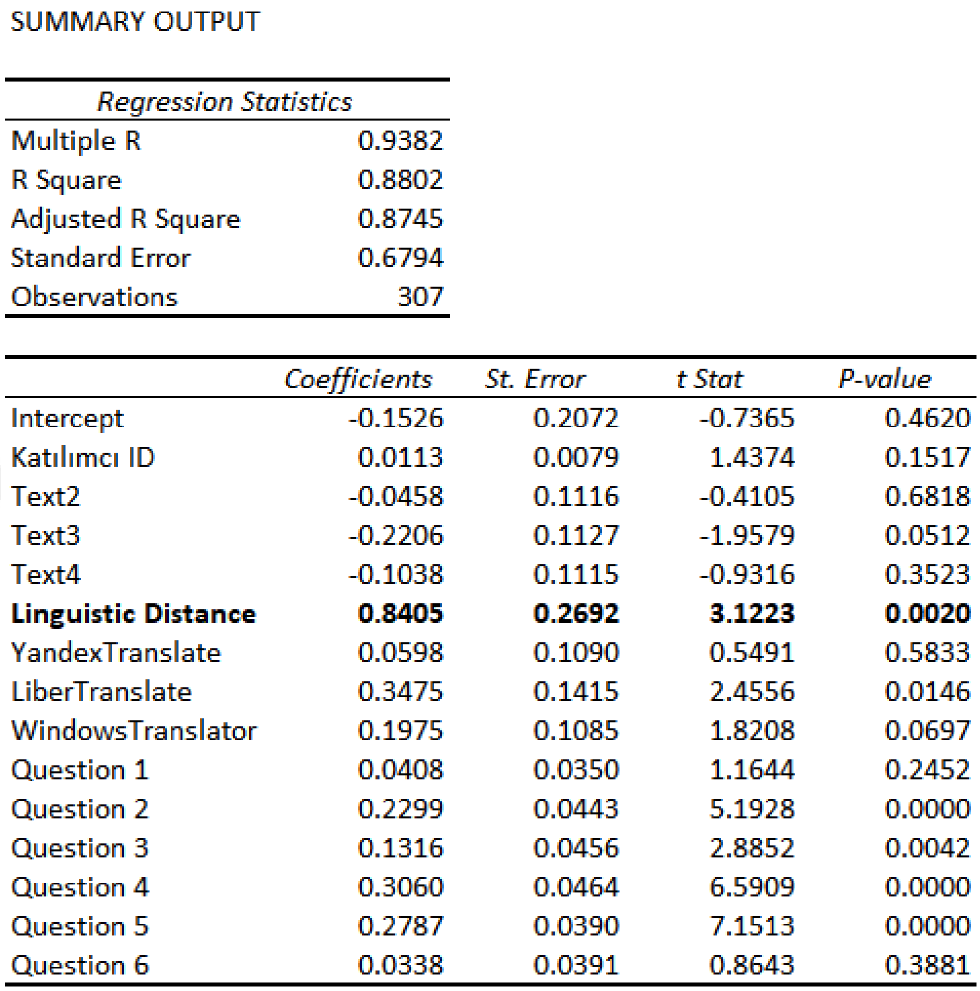
SUMMARY OUTPUT **Question 7**

Regression Statistics	
Multiple R	0.630
R Square	0.397
Adjusted R Square	0.380
Standard Error	1.465
Observations	288

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	1.315	0.468	2.813	0.005
Participant ID	0.044	0.016	2.800	0.005
Text2	0.444	0.244	1.820	0.070
Text3	-0.431	0.244	-1.763	0.079
Text4	0.250	0.244	1.024	0.307
Linguistic Distance	**1.998**	**0.694**	**2.878**	**0.004**
YandexTranslate	0.579	0.238	2.436	0.015
LiberTranslate	2.975	0.256	11.621	0.000
WindowsTranslator	1.053	0.238	4.429	0.000

APPENDIX M

MULTIVARIATE REGRESSION – COMPREHENSIVE ON TQ



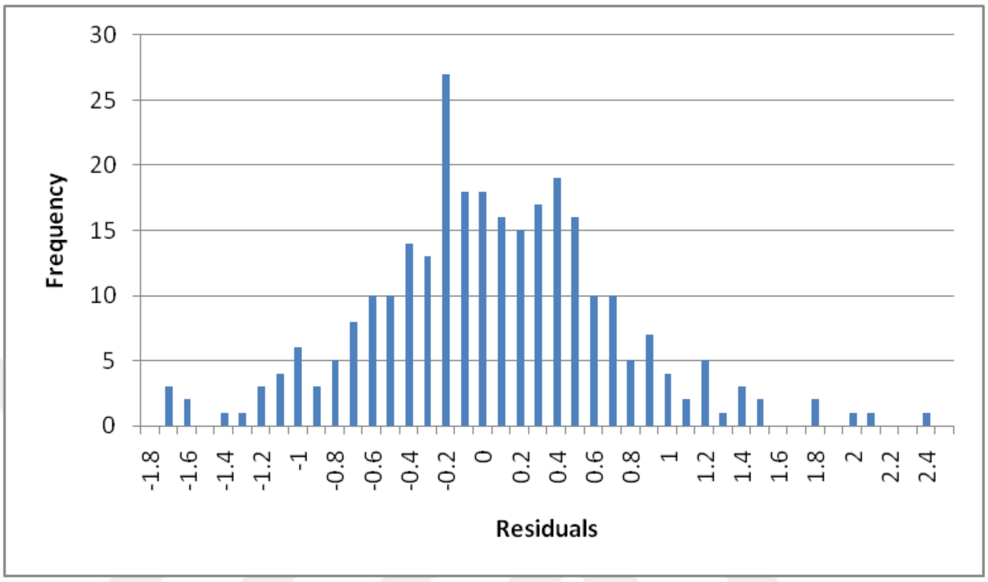
SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.9382
R Square	0.8802
Adjusted R Square	0.8745
Standard Error	0.6794
Observations	307

	Coefficients	*St. Error*	*t Stat*	*P-value*
Intercept	-0.1526	0.2072	-0.7365	0.4620
Katılımcı ID	0.0113	0.0079	1.4374	0.1517
Text2	-0.0458	0.1116	-0.4105	0.6818
Text3	-0.2206	0.1127	-1.9579	0.0512
Text4	-0.1038	0.1115	-0.9316	0.3523
Linguistic Distance	**0.8405**	**0.2692**	**3.1223**	**0.0020**
YandexTranslate	0.0598	0.1090	0.5491	0.5833
LiberTranslate	0.3475	0.1415	2.4556	0.0146
WindowsTranslator	0.1975	0.1085	1.8208	0.0697
Question 1	0.0408	0.0350	1.1644	0.2452
Question 2	0.2299	0.0443	5.1928	0.0000
Question 3	0.1316	0.0456	2.8852	0.0042
Question 4	0.3060	0.0464	6.5909	0.0000
Question 5	0.2787	0.0390	7.1513	0.0000
Question 6	0.0338	0.0391	0.8643	0.3881

APPENDIX N

VISUAL REPRESENTATION OF THE DISTRIBUTION OF THE RESIDUALS FOR THE FINAL COMPREHENSIVE REGRESSION



APPENDIX O

ETHICS COMMITTEE APPROVAL

Evrak Tarih ve Sayısı: 26.03.2022-59477

T.C.
BOĞAZİÇİ ÜNİVERSİTESİ
SOSYAL VE BEŞERİ BİLİMLER YÜKSEK LİSANS VE DOKTORA TEZLERİ ETİK İNCELEME KOMİSYONU
TOPLANTI KARAR TUTANAĞI

Toplantı Sayısı	: 29
Toplantı Tarihi	: 24.03.2022
Toplantı Saati	: 10:00
Toplantı Yeri	: Zoom Sanal Toplantı
Bulunanlar	: Prof. Dr. Ebru Kaya, Dr. Öğr. Üyesi Yasemin Sohtorik İlkmen
Bulunmayanlar	:

Ata Leblebici
Çeviribilim

Sayın Araştırmacı,
"Makine Çevirisi İsabetliliği ve Dil Mesafesi Arasındaki İlişki" başlıklı projeniz ile ilgili olarak yaptığınız SBBEAK 2022/15 sayılı başvuru komisyonumuz tarafından 24 Mart 2022 tarihli toplantıda incelenmiş ve uygun bulunmuştur.

Bu karar tüm üyelerin toplantıya çevrimiçi olarak katılımı ve oybirliği ile alınmıştır. COVID-19 önlemleri kapsamında kurul üyelerinden ıslak imza alınamadığı için bu onay mektubu üye ve raportör olarak Yasemin Sohtorik İlkmen tarafından bütün üyeler adına e-imzalanmıştır.

Saygılarımızla, bilgilerinizi rica ederiz.

Dr. Öğr. Üyesi Yasemin
SOHTORİK İLKMEN
ÜYE

e-imzalıdır
Dr. Öğr. Üyesi Yasemin Sohtorik
İlkmen
Öğretim Üyesi
Raportör

SOBETİK 29 24.03.2022

Bu belge, güvenli elektronik imza ile imzalanmıştır.

REFERENCES

Aksoy, B. (2005). Translation activities in the Ottoman Empire. *Erudit*. doi:doi.org/10.7202/011606ar

Argos Open Technologies, LLC. (n.d.). *Argos Translate Documentation*. Retrieved August 26, 2022, from: https://argos-translate.readthedocs.io/en/latest/

Baker, M. (Eds.). (2005). *Routledge Encyclopedia of Translation Studies*. New York, NY, United States of America: Routledge.

Benjamin, M. (2019). How GT pivots through English. In *Teach You Backwards: An In-Depth Study of Google Translate for 108 Languages*. Immediate sharing [Web log post]. Retrieved January 12, 2023, from: https://www.teachyoubackwards.com/extras/pivot/

Besacier, L., & Schwartz, L. (2015). Automated translation of a literary work: A pilot study. In *Proceedings of the Fourth Workshop on Computational Linguistics for Literature*, 114–122. Denver, CO, United States of America: Association for Computational Linguistics. doi:doi.org/10.3115/v1/W15-0713

Blatz, J., Fitzgerald, E., Foster, G., Gandrabur, S., Goutte, C., Kulesza, A., . . . Ueffing, N. (2004). Confidence estimation for machine translation. In *COLING 2004: Proceedings of the 20th international conference on Computational Linguistics*. doi:10.3115/1220355.1220401

Brown, C., Holman, E., Wichmann, S., & Velupillai, V. (2008). Automated classification of the world's languages: A description of the method and preliminary results. In *STUF - Language Typology and Universals*, *61*, 285-308.

Caswell, I., & Liang, B. (2020). *Recent Advances in Google Translate*. Immediate sharing [Web log post]. Retrieved August 26, 2022, from: https://ai.googleblog.com/2020/06/recent-advances-in-google-translate.html

Ceolin, A. (2019). Significance testing of the Altaic family. In *Diachronica*, *36*(3), 299–336. doi: doi.org/10.1075/dia.17007.ceo

Ceolin, A., Longobardi, G., Guardiano, C., & Irimia, M. A. (2020). Formal syntax and deep history. In *Frontiers in Psychology*, *11*. doi:doi.org/10.3389/fpsyg.2020.488871

Ceolin, A., Guardiano, C., Longobardi, G., Irimia, M. A., Bortolussi L., & Sgarro A. (2021). At the boundaries of syntactic prehistory. In *Philosophical Transactions of the Royal Society B*, *376*. doi:10.1098/rstb.2020.0197

Chiswick, B. R., & Miller, P. W. (2004). Linguistic distance: A quantitative measure of the distance between English and other languages. In *Journal of Multilingual and Multicultural Development*, *26*, 1–11. doi:10.1080/14790710508668395

Clauson, G. (2005). Altayca teorisinin leksikoistatistiksel bir değerlendirmesi. In *Journal of Turkish World Studies*, *5*(2), 311-323. İzmir: Türkiye.

Crisma, P., Guaridano, C., & Longobardi, G. (2020). Syntactic diversity and language learnability. In *Studi e Saggi Linguistici*, *58*(2), 99-130. York, United Kingdom: University of York. doi:doi.org/10.4454/ssl.v58i2.265

Devore, J. L. (2011). *Probability and Statistics for Engineering and the Sciences* (8th ed.). Boston, MA, United Stated of America: Cengage Learning. 508–510.

Dryer, M. S., & Haspelmath, M. (Eds.). (2013). The world atlas of language structures online. In *Max Planck Institute for Evolutionary Anthropology. Online Database*. Retrieved on May 2, 2022, from: http://wals.info

Even-Zohar, I. (1990). The position of translated literature within the literary polysystem. In *Poetics Today*, *11*(1), 45–51. Durham, NC, United States of America: Duke University Press.

Google Translate. (n.d.). Retrieved from: http://translate.google.com/

Guerberof-Arenas, A., & Toral, A. (2020). The impact of post-editing and machine translation on creativity and reading experience. In *Translation Spaces*, *9*(2), 255-282. doi:doi.org/10.1075/ts.20035.gue

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. In *Neural Computation*, *9*(8), 1735-1780.

Holmes, J. S. (1988). The name and nature of translation studies. In *Translated!: Papers on Literary Translation and Translation Studies*, 66-80. Netherlands: Rodopi.

Hovy, E., King, M., & Popescu-Belis, A. (2002). Principles of context-based machine translation evaluation. In *Machine Translation*, *17*(1), 43-75.

Hutchins, J. (2007). Machine translation: A concise history. In *Computer aided translation: Theory and Practice*, *13*(29-70), 11. Spain: Universitat Oberta de Catalunya.

Isphording, I. E., & Otten, S. (2011). Linguistic distance and the language fluency of immigrants. In *Ruhr Economic Paper No. 274,* doi:dx.doi.org/10.2139/ssrn.1919474

Karaca, O. S. (2011). Çağdaş Türk lehçelerinin söz varlığındaki ortaklığa karşılaştırmalı bir bakış. In *International Periodical For The Languages, Literature and History of Turkish or Turkic*, *6*(1), 1340-1352. Türkiye.

Kenny, D. (Eds.). (2022). *Machine Translation for Everyone: Empowering Users in the Age of Artificial Intelligence*. Berlin, Germany: Language Science Press.

Kessler, B. (2007). Word similarity metrics and multilateral comparison. In *Proceedings of Ninth Meeting of the ACL Special Interest Group in Computational Morphology and Phonology*, 6-14. Association for Computational Linguistics.

Klein, G., Kim, Y., Deng, Y., Senellart, J., & Rush, A. (2017). The OpenNMT neural machine translation toolkit: 2020 edition. In *Proceedings of ACL 2017, System Demonstrations*. Vancouver, Canada: Association for Computational Linguistics.

LibreTranslate. (n.d.). Argos Open Technologies, LLC. Retrieved from: https://www.argosopentech.com/

Lommel, A., Görög, A., Melby, A., Uszkoreit, H., Burchardt, A., & Popović, M. (2015). D.3.1 harmonized metric. In *Quality Translation 21*. Retrieved from: https://www.taus.net/qt21-project#dqf-qt21

Longobardi, G., & Guardiano, C. (2009) Evidence for syntax as a signal of historical relatedness. In *Lingua*, *119*(11), 1679-1706. Italy.

Longobardi, G, Guardiano, C., & Crisma, P. (2020). Syntactic parameters and language learnability. In *Studi e Saggi Linguistici*, 99-130.

Marcolli, M. (2016). Syntactic parameters and a coding theory perspective on entropy and complexity of language families. In *Entropy, 18*(4), 110. doi:doi.org/10.3390/e18040110

Moorkens, J., & Rocchi, M. (2021). *Ethics in the Translation Industry*. In K. Koskinen, & N. K. Pokorn (Eds.), *The Routledge Handbook of Translation and Ethics*. Abingdon: Routledge. doi:10.4324/9781003127970-24

Microsoft Translator. (n.d.). Retrieved from: https://www.microsoft.com/en-us/translator/personal/

Microsoft. (n.d.). *Neural Machine Translation*. Retrieved from Microsoft website: https://www.microsoft.com/en-us/translator/business/machine-translation/

Nida, E. (1964). *Towards a Science of Translating*. Leiden, Netherlands: Brill.

Papineni, K., Roukos, S., Ward, T., & Zhu, W. (2002). BLEU: A method for automatic evaluation of machine translation. In *40th Annual Meeting of the Association for Computational Linguistics*, 311–318, Philadelphia, PA, United States of America.

Piazzalunga, D., Strøm, S., Venturini, A., & Villosio, C. (2018). Wage assimilation of immigrants and internal migrants: the role of linguistic distance. In *Regional Studies*, *52*(10), 1423-1434. doi:10.1080/00343404.2017.1395003.

Poibeau, T. (2017). *Machine Translation*. Cambridge, MA, United States of America: The MIT Press.

Pool, J. (2022). *PanLex Swadesh Lists*. Retrieved from: https://dev.panlex.org/

Sokal, R. R., & Michener, C. D. (1958). A statistical method for evaluating systematic relationships. In *University of Kansas Science Bulletin*. *38*, 1409–1438.

Sokal, R. R. (1988). Genetic, geographic, and linguistic distances in Europe. In *Proceedings of the National Academy of Sciences,* 1722-1726. doi:10.1073/pnas.85.5.1722

Somers, H. L. (2005). Machine translation: History. In M. Baker (Eds.) *Routledge Encyclopedia of Translation Studies*. 140-143. New York, NY, United States of America: Routledge.

Sørensen, T. (1948). A method of establishing group of equal amplitude in plant sociobiology based on similarity of species content and its application to analyses of the vegetation on Danish commons. In *Kongelige Danske Videnskabernes Selskab, Biologiske Skrifter*, *5*, 1-34.

Specia, L., Hajlaoui, N., Hallett, C., & Aziz, W. (2011). Predicting machine translation adequacy. In *Machine Translation Summit,* 13(2011), 19-23.

Stolze, R. (2020). *Çeviri Kuramları, Bir Giriş* (E. Büyüknisan, Trans.). İstanbul, Türkiye: Runik Kitap.

Swadesh, M. (1950). Salish internal relationships. In *International Journal of American Linguistics*, *16*, 157–167.

Swadesh, M. (1971). *The Origin and Diversification of Language*. Chicago, IL, United States of America: Aldine.

Şahin, M., & Duman, D. (2013). Multilingual chat through machine translation: A case of English-Russian. In *Meta*, *58*(2), 397-410. doi:doi.org/10.7202/1024180ar

Şahin, M., & Gürses, S. (2021). English-Turkish literary translation through human-machine interaction. In *Revista Tradumàtica: tecnologies de la traducció*. Spain: Universitat Autònoma de Barcelona. doi:doi.org/10.5565/rev/tradumatica.284

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., . . . Polosukhin, I. (2017). Attention is all you need. In *Advances in neural information processing systems*, *30*. Ithaca, NY, United States of America: Cornell University. doi:doi.org/10.48550/arXiv.1706.03762

Wichmann, S., Holmanc, E. W., Bakker, D., & Brown, C. H. (2010). Evaluating linguistic distance measures. In *Physica A: Statistical Mechanics and its Applications*, *389*(17), 3632-3639. doi:doi.org/10.1016/j.physa.2010.05.011

Wichmann, S., Holman, E. W., & Brown, C. H. (Eds.). (2020). *The ASJP Database*. Retrieved May 4, 2022 from: https://asjp.clld.org/

Wittgenstein, L. (2005). *Tractatus Logico-Philosophicus* (C. K. Ogden, Trans.). London, United Kingdom: Routledge. (Original work published 1921).

Wu, Y., Schuster, M., Chen, Z., Le, Q. V., & Norouzi, M. (2016). *Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation*. Ithaca, NY, United States of America: Cornell University. doi:doi.org/10.48550/arXiv.1609.08144

Yakhontov, S. (1991). *Concept Lists*. Retrieved May 4, 2022, from: https://concepticon.clld.org/contributions/Yakhontov-1991-100

Yandex Translate. (n.d.). Retrieved from: https://translate.yandex.com/

Yandex. *About Machine Translation*. Immediate sharing [Web log post]. Retrieved August 28, 2022, from: https://yandex.com/dev/translate/doc/dg/concepts/how-works-machine-translation.html