National and Kapodistrian University of Athens

Department of English Language and Literature

MA Programme "Linguistics: Theory and Applications"

Differences between Human and Machine-generated Institutional Translations: A comparative analysis using quantitative methods

Maria Bourou

217018

February 2019

Supervisors

George Mikros, Professor, Department of Italian Language and Literature, National and Kapodistrian University of Athens

Elly Ifantidou, Professor, Department of English Language and Literature, National and Kapodistrian University of Athens

Vasiliki Nikiforidou, Professor, Department of English Language and Literature, National and Kapodistrian University of Athens

Declaration

This submission is my own work. Any quotation from, or description of, the work of others is acknowledged herein by reference to the sources, whether published or unpublished.

Maria Bourou

Abstract

Machine translation, commonly referred to as MT, has gained popularity over the recent years; however, it has not yet reached the quality and naturalness of human writing. The present thesis aims to explore how human and automatic English translations of Greek institutional texts differ by comparing quantitative characteristics of the two translation types. Statistical analysis using independent samples t-tests revealed that the two corpora differed in a range of linguistic features including descriptive characteristics (e.g. word length), word information (e.g. parts of speech, word frequency), lexical diversity, syntax and cohesion; however, the degree of variation was not striking. In a follow-up examination, using Multilayer Perceptron neural network, the machine was able to classify correctly almost 82% of the texts as automatic or human-produced. These results suggest that the differences between HT and MT regarding the subgenre in question are detectable using machine learning techniques, but the distinction is not as clear-cut as expected. Further research is needed to determine whether the text properties that differ most in the two corpora can be used effectively as predictors of translation quality.

Keywords: machine translation (MT); human translation (HT); translation quality assessment; institutional texts; Greek-English language pair

Table of Contents

Chapter 1: Introduction

1. Scope and Aim

The need for intercultural communication and instant access to information in a variety of languages calls for the development of automatic translation systems that break down language barriers. The European Institutions have introduced automatic tools since 1970s to meet the increasing demands for translation of political, legal and administrative documents in all EU official languages (Eisele and Lavecchia 2011: 3). A growing interest in machine translation, also called 'MT', is also observed in areas, like international commerce and administration (Puchała-Ladzińska 2016: 90-91) as well as in the field of software localization, i.e. adaption of computer programs for foreign target audiences (Puchała-Ladzińska 2016: 91) to name but a few.

Currently, Google Translate is probably the most popular machine translation platform with an average of 200 million users and a billion translations per day, reported in 2013 (Shankland, 2013 ctd in Li, Greasser and Cai 2014). However convenient, instant translations have not yet reached the quality standards of professional human translations and it is still questionable whether they will ever perform as good (Puchała-Ladzińska 2016: 92). No software so far has been able to imitate human creativity and intellect in preserving the aesthetic value and function of the source text (Puchała-Ladzińska 2016: 97). Machine translations suffice only for limited purposes – to cater for basic communicative needs (Puchała-Ladzińska 2016: 95). So, when MT systems are used, human involvement at least in the form of post-editing is necessary (Ahrenberg 2017: 28).

The present study was stimulated by the need to assess the strengths and weaknesses of machine translation. Efficient evaluation is the first step towards optimizing translation quality and creating accurate post-editing systems. Using quantitative methods, I compare corpora of human translations (HT) and machine translations (MT) in terms of language and text characteristics. The corpora consist of formal announcements and statements published on the website of the Hellenic Ministry of Foreign Affairs and translated into English either by experts or Google Translate. Based on the premise that expert translations are more reliable than automatic, this comparative analysis attempts to bring to our attention some limitations of MT and pave the way to their resolution in the future. As in most automatic evaluation systems, closeness to human-generated translations, that are deemed accurate, is used as a quality metric (Finch, Hwang and Sumita 2005: 17). Proximity to the reference text or "gold standard" is an indicator of accuracy, while deviations point to potentially problematic areas that need investigation and improvement.

The criteria used in the comparison include both shallow aspects (e.g. number of tokens and types, average word/ sentence/ paragraph length, parts of speech, first/second/third person pronouns etc) and deeper linguistic. The indices concerning the latter operate at sentence and whole paragraph level and assess syntactic complexity, coherence, cohesion (surface/ referential/ deep), lexical diversity, readability and discourse structure. All sets of characteristics were selected as indicative of (a) linguistic complexity (b) register and (c) comprehensibility.

In a follow up experiment, I test the accuracy of a neural network, trained on the same sample, in predicting the category of texts (human or machine-translated). This test is performed in order to discover the extent to which the two translation types are distinguishable and detect which linguistic features contribute more to the classification. The standout features that distinguish the two samples are expected to be good predictors of translation quality. This assumption is based on the surmise that closeness and divergence to the reference translations expose some strong and weak aspects of MT, respectively. Likewise, the features that provide this information are likely to be useful in evaluation.

The specific research questions concerning this thesis are the following: (a) in what respects do human and machine-produced translations differ? (b) Are these differences detectable by machine learning algorithms and which variables included in the classifier are the most influential?

In other words, I intent to discover whether human and machine translations truly differ in terms of lexical, syntactic, discourse and readability properties. To date, the most widely used automatic evaluation metrics have focused on lexical similarity and n-gram overlap between reference and candidate translations (for instance, BLEU; Papineni et al. 2002 or NIST; Doddington 2002). However, previous research findings support that the integration of deeper linguistic knowledge in evaluation confers many advantages (Amigo et al. 2009; Comelles et al. 2012; Gimenez and Marquez 2007; Joty et al 2017; Pado et al 2009; Scarton and Specia 2015). If this is the case, a hybrid, comprehensive approach that combines micro and macro levels of analysis may be the key to effective evaluation. Examining surface similarities is not adequate to capture underlying problems related to discourse structure. Issues of cohesion and coherence, comprehensibility, improper tone and inconsistent stylistic or lexical choices can only be identified if thorough linguistic analysis is conducted at multiple levels. For example, only a linguistically-informed metric would be efficient to assess the degree of formality within a text. The present study aims at identifying some text properties that could potentially be used successfully as predictors of translation quality in future metrics.

It is worth to mention that quality assessments need to be discussed with reference to the intended use of translation and any specifications required by the user (Dorr et al 2011: 745; Melby 2005: 6-7). As Nord (2003) states, the task of translation involves the transfer of a text into a new semiotic system; the produced output needs to fulfil pragmatic, cultural and formal conventions (ctd in Vela, Schumann and Wurm 2014). Also, expectations vary across contexts; tone, stylistics and accuracy may be crucial in some applications but irrelevant in other cases (Dorr et al 2011: 745) when the purpose of translation is to convey the text's general message (Aiken et al 2009: 67). Therefore, contextual factors, like text type and expected audience, determine which parametres will be assigned most weight in the evaluation (Hovy, King and Popescu-Belis 2002a: 3).

The language of civil texts is expected to be impersonal, rigid, and structurally complex, elements associated with increased formality (Heylighen and Dewaele 1999). So, in the present sample, appropriacy and high register are as important to maintain as content. This explains why stylistic aspects are thoroughly addressed in the analysis alongside surface and deep linguistic features. Finally, the selected methodology was tailored to describe the specific subgenre and type of data. As a result, this evaluation concerns exclusively the language of institutional translations and the given language pair.

2. Structure of the thesis

The first chapter introduced the scope and aim of this study. Chapter 2 moves on to some fundamentals in machine translation evaluation and provides an elaborate review of relevant research. Specifically, it presents some of the most common MT evaluation metrics along with their criticism: BLEU, NIST, WER, TER and METEOR. Then, it reports some important findings on the integration of linguistic features in evaluation and brings up some advantages of these approaches. Finally, it describes some differences of human and machine translated output as evidenced in relevant case studies. These studies adopt both quantitative and descriptive methods and concern a number of different language pairs, for instance Chinese-English (Li et al 2014), English-Swedish (Ahrenberg 2017), English-Spanish (Chen et al 2016) and other. To my knowledge, there is no relevant research based on Greek-English data so far. Elaborating on the HT/ MT distinction, I also present some research on automatic text classification using machine learning techniques.

Chapter 3 describes the selected methodology. First, all choices made at the stage of corpus design are presented and justified. Certain alternatives are also discussed. The first section mostly contains descriptive information about the corpora: source, length, language and text domain. The next section offers a detailed description of the tools and materials used in the analysis, specifically QUITA and

Coh-Metrix. All indices are listed and explained along with the criteria for their selection.

Results are presented in Chapter 4. Their interpretation and implications are discussed in Chapter 5. Chapter 6 summarizes the findings, presents some limitations and suggests certain alternatives. Towards the end, I also discuss the usefulness of this study and propose ideas for further research. An appendix with additional information about the data can be found on the last pages, after the reference list.

Chapter 2: Literature Review

2.1 Machine Translation Evaluation

The problem of MT evaluation has been concerning researchers for a long time. A research focus to evaluation was proposed as early as in 1966 by the Automatic Language Processing Advisory Committee (ALPAC 1966) that recommended, among others, further work in new evaluation methods and machine-aided translation (White 2003: 212). To date, MT evaluation has been a complex and challenging task (Mauser et al 2008: 3089). As White (2003: 213) phrases it, “correct translation” is an elusive target – it is hard to measure quality in the absence of a prototype or standard. First, there can be disagreement as to what constitutes a perfect translation and, second, the rich variability of language and the remarkable creativity that goes into the act of translating allows for a set of different translations to be valid (White 2003: 213).

Before the shift to statistical approaches, machine translations were evaluated manually (Dorr, 2010: 805). Early traditional methods were exclusively dependent on intuitive, subjective judgments of evaluators, who rated sentences based on error point scales (White 2003). As Hovy et al. (2002a: 1) report in their overview, the notions of quality and fidelity have stood out in evaluation. Quality or fluency, as it is often called, is evidenced by two indicators: proper lexical use and existence of wellformed, grammatically accurate structures (Hovy et al. 2002a: 1). Fidelity, on the other side, refers to the semantics of the output – the meaning conveyed in the original should not be twisted or distorted (Han et al 2016: Hovy et al 2002a: 1). Additional qualities of automatic translation that still concern users and developers are comprehensibility, adequacy and informativeness (Dorr et al 2011; Han et al 2016; White 1995). Other criteria pertain to the translation system itself, such as extensibility (whether it allows for addition of grammar, words, structures etc) and cost (White 1995). Customized multi-point scales, designed to assign numerical scores to the system output, have been developed for many of the above criteria and are still used today (Dorr et al 2011: 748).

Recent developments in computational and corpus linguistics have opened up new possibilities in MT evaluation. Automatic measures address the shortcomings of traditional evaluation techniques that are expensive, time-consuming, labor-intensive, subjective and, sometimes, inconsistent (Chatzitheodorou and Chatzistamatis 2013: 83; Koehn 2004: 389; Lin and Och 2004; Papineni et al. 2002: 311; Scarton and Specia: 2015: 3; White 2003: 241). While manual evaluation requires much work from trained evaluators (Koehn 2004: 389; Lin and Och 2004; Papineni et al. 2002: 311; White 2003: 241) automatic systems can provide instant results based on objective measures. They usually compare the system output against a set of reference translations, produced by experts (Dorr et al 2011: 759; Graham et al. 2014: 266; Sun 2018: 1). The core principle of this method, summarized by Papineni et al (2002: 311) is that "[t]he closer a machine translation is to a professional human translation, the better it is". According to Mauser et al. (2008: 3090), most of these metrics show "a reasonable correlation with human judgments".

Proximity to the reference can be assessed on different grounds. The first automatic evaluation tools, BLEU (Papineni et al. 2002) and NIST (Doddington 2002) measure n-gram co-occurrence between a candidate and a reference translation and penalize translations that are too short. METEOR (Banerjee and Lavie 2005; Lavie and Denkowski 2009; Denkowski and Lavie 2011) also allows for morphological variants and synonyms to be matched and incorporates recall as an extra component. Other measures, such as WER (Nieβen et al. 2000), PER (Tillmann et al. 1997) and TER (Snover et al. 2006) calculate the edit distance between a hypothesis and a set of reference translations. A more elaborate description of these measures is presented in the next subsection.

2.2 Automatic Evaluation Metrics

2.2.1 BLEU (BiLingual Evaluation Understudy)

BLEU, created by Papineni et al. (2002), has been the most frequently used automatic metric in the last few years (Federmann 2014 para 10; Lavie and Denkowski 2009: 106; Vela et al. 2014). As specified by its creators, it is intended as an "understudy to skilled human judges" and captures fluency and adequacy (Papineni et al. 2002: 311, 313). BLEU employs a numerical 'translation closeness' metric (Papineni et al. 2002). It assesses similarity between a candidate and a corpus of reference translations in terms of n-gram matching. The greater the overlap, the better the translation is judged to be.

The following example is borrowed by Papineni (2002: 312). Given a hypothesis, BLEU assesses overall n-gram overlap with three reference translations. The underlined constituents in Candidate 1 are those that occur in at least one of the references.

> Candidate 1: It is a guide to action which ensures that the military always obeys the commands of the party.
>
> Candidate 2: It is to insure the troops forever hearing the activity guidebook that party direct.
>
> Reference 1: It is a guide to action that ensures that the military will forever heed Party commands.
> Reference 2: It is the guiding principle which guarantees the military forces always being under the command of the Party.
> Reference 3: It is the practical guide for the army always to heed the directions of the party.

Similarity is assessed on a scale from 0 to 1. Proximity to 1 indicates high quality. Candidate 1, which is both accurate and fluent, will receive a high score as it exhibits many matches. Candidate 2 is clearly inferior in quality, as co-occurrences are scarce and short. Therefore, this alternative will fail to get an optimal score.

Regarding sentence length, candidate translations should not deviate too much from the original. To this end, BLEU features a brevity penalty for short translations. Moreover, it penalizes words that are repeated more frequently than in reference texts as well as spurious content (e.g. Reference: The cat is on the mat. Candidate: the the the the the the the [Papineni et al 2002: 312]). In this way, inadequate candidates do not achieve high scores.

BLEU is reported to correlate well with human judgments at corpus level (Papineni et al 2002: 317).

2.2.2 NIST (National Institute of Standards and Technology)

Along the same lines as BLEU, the NIST metric (Doddington 2002) is based on n-gram co-occurrence between a hypothesis and the references. Rather than precision, NIST measures n-gram informativeness by assigning more weight to components that are rare (Doddington 2002: 141). In other words, frequency is used as an indicator of importance: n-grams that occur more often are treated as less informative.

Moreover, its brevity penalty is revised, so that the impact on the overall score is reduced (Federmann 2014: para 10; Mauser et al. 2008: 3091). NIST, similarly to BLEU fails to achieve a good correlation score at the sentence level (Federmann 2014: para 10).

2.2.3 METEOR

METEOR, developed in 2004, assesses sentence-level similarity between a candidate and a set of reference translations based on word-to-word matching (Lavie and Denkowski 2009). What differentiates METEOR from previous systems, such as BLEU or NIST, is that it also allows for morphological variants and synonyms to be matched, thus, addressing the issue of lexical variability. (Lavie and Denkowski 2009: 106). It also places emphasis on unigram-recall rather than precision, which has been found to correlate better with human judgments (Lavie et al. 2004). Unlike BLEU, METEOR was designed to achieve optimal correlation with evaluators at the sentence level (Lavie and Denkowski 2009).

2.2.4 WER (Word Error Rate)

WER is derived from the Levenshtein distance (Levenshtein 1966), that is the "minimum number of word insertions, substitutions and deletions necessary to transform the candidate translation into the reference translation" (Mauser et al. 2008: 3090). WER has been described as the standard metric used in Automatic Speech Recognition but applications extend to machine translation as well (Dorr et al. 2011: 759). Some proposals regarding WER have been made by Nießen et al. (2000) and Su et al. (1992). The assumption behind these metrics, as well as TER presented below, is that translation quality can be estimated based on the minimum number of post-editing steps required for the given output to match a valid translation.

2.2.5 TER (Translation Error Rate)

Translation Edit Rate (Snover et al. 2006: 561) calculates the amount of post-editing required by humans so that the system output is fixed into a fluent and semantically accurate translation (Snover et al. 2006: 561-562). Therefore, it is essentially an estimate of the amount of human labour needed for a candidate to meet the expected quality standards. Distance is assessed by reference to the closest possible correct translation (Snover et al. 2006). All edits are weighted equally and may involve insertion, deletion, substitution, modifications in word order and re-capitualization (Snover et al. 2006: 563). TER is computed as the number of edits divided by the average number of reference words (Snover et al. 2006: 563). Since it is essentially an error rate, lower scores indicate better performance (Snover et al. 2006).

According to its creators, TER correlates reasonably well with human judgments (Snover et al. 2006). However, as they comment, performance can be further improved by using targeted references created by human annotators to approximate a particular system output, as happens in its semi-automatic variant, HTER (Human-targeted Translation Edit Rate). They propose that if human work is to be used, the evaluators' role should be to create target references or calculate errors instead of making subjective judgments (Snover et al. 2006: 569).

2.3 Criticism of Lexical Similarity metrics

Automatic techniques have not been without criticism. Although they are fast and convenient, their interpretation is a complex task (Federmann 2014 para. 11). Given that some of the metrics, for instance BLEU, depend on lexical similarity, it is likely that candidates are acknowledged only if they share the exact same lexicon as the reference document (Zhou et al. 2008: 1121). However, owing to the enormous variability and flexibility of natural languages, the same message can be conveyed by varying paraphrased versions that are impossible to predict when only a limited set of references is at hand (Culy and Riehemann 2003: 6; Gimenez 2008: 30). Therefore, a valid translation may not get an optimal score if the reference words are paraphrased and n-grams do not overlap (Federmann, 2014 para 11). This could partly explain why metrics like BLEU have reportedly failed to meet the expectations for robust correlations with human judgments (Callison-Burch et al. 2006). Along the same lines, a greater number of reference translations has been found to yield higher scores in evaluations that employ n-gram matching (Culy and Riehemann 2003: 6). This brings up another disadvantage of these methods: their strong dependence on the existence of reference translations, which are not readily available in most environments (Culy and Riehemann 2003: 2; Melby 2005: 6).

As mentioned, lexical similarity is neither a sufficient nor a necessary condition for semantic equivalence (Gimenez and Marquez 2007: 256). Evaluation is further complicated when it comes to free-word order languages with morphological richness and many inflectional forms (Tripathi and Kansal 2017). The inadequacy of n-gram-based metrics to capture variation in word order and lexical choice has been widely criticized in literature (Callison-Burch, Osborne and Koehn 2006; Culy and Riehemann 2003; Zhou et al 2008). Proximity to the reference should probably be assessed upon broader criteria, more closely associated with MT quality standards.

Furthermore, extant n-gram-based metrics are largely dependent on the type of translation used as reference; Strict and relatively literal reference translations result in better scores for machine translation systems, while professional translations of superior quality and creativity may get lower scores if they deviate remarkably from the source text (Culy and Riehemann 2003: 6). To address this shortcoming, the creators of BLEU suggest using multiple human translations with different styles as references (Papineni et al. 2002: 313). Metrics that calculate edit distance from reference translations also depend fundamentally on the choice of sample, as Nießen (2000) comments. This means that reliability of results inevitably relies on translation availability that often falls out of the researcher’s control.

Different to the classic methods, Machine Translation Quality Estimation (MTQE) approaches have been proposed that forego the need for reference translations (Bechara 2016: 256; Scarton and Specia: 2015: 4). Quality estimation is a challenging task and predicts translation quality based on data extracted from the source text and, sometimes, the corresponding MT system (Felice and Specia 2012: 96; Scarton and Specia 2015: 4). These measures are used to train unsupervised Machine Learning models and predict scores for unseen translations (Scarton and Specia 2015: 4). Research in this area has given rise to interesting findings to be reported in the next section.

Another shortcoming of many automatic measures lies on their inefficiency to weight the importance of linguistic information (Callison-Burch et al. 2006; Koehn 2010 ctd in Han et al. 2006). As Han et al. (2006) explain, punctuation marks and functional words are less meaningful than name entities and core concepts; however, they are often weighted equally, and issues of relevance are overlooked. In BLEU, for instance, omission of content-bearing material is not penalized more than less informative strings (Callison-Burch 2006: 252). However, two sentences differing by a minor detail may convey completely opposite meanings, for instance, “There is no vase on the table” vs “There is a vase on the table” (Ulitkin 2013: Introduction, para 3). This failure to identify the most meaningful components reflects the current limitation to distinguish between plausible and unacceptable translations.

The same holds true for errors. According to Schiaffino and Zearo (2005), errors should be evaluated considering their consequences (ctd in Maney et al 2012: 1). Those that only contribute to grammatical correctness may be minor, but others are likely to distort the meaning altogether (Vilar et al. 2006 ctd in Maney et al. 2012: 2). Along these lines, Maney et al. (2012) investigated how common errors, such as omissions (deleted verbs, nouns and other POS classes) modifications of prepositions (e.g. in → at) and alterations in word order, affect the comprehensibility of machine translations. Modified word order did not lead to major comprehension errors (Maney et al. 2012). Deleted verbs and adjectives were found to hinder understanding to a greater extent than did nouns, pronouns and prepositions. Given their higher frequency over adjectives, this study highlights the importance of reducing errors, first and foremost, in the translation of verbs (Maney et al. 2012). This study may also implicate that more weight is to be assigned to errors that involve verbs.

Another topic of criticism is that automatic metrics usually produce an abstract score that is difficult to qualify (Han et al 2016). On many occasions, results are neither informative nor meaningful. For instance, a BLEU score of 28.47 or a METEOR score of 33.03 say nothing by themselves (Han et al 2016), unless individual scores of different texts are compared for classification purposes. Moreover, as Aiken and Balan (2011) observe, BLEU does not provide information on adequacy. For instance, different standards apply to legal or medical texts than less important material (Aiken and Balan 2011). In the first case, gathering the overall idea is not sufficient while, on other occasions, this could be more than enough. Therefore, a BLEU score of 50.0 is evaluated differently in all contexts.

Regarding methodology, Felice and Specia (2012: 97) and also Sun (2018: 1) convincingly state that non-linguistic features, employed to compute sentence similarity are limited in scope. In essence, these metrics aim to assess the extent of semantic equivalence between machine and expert translations (Finch et al 2005: 17). Few studies, however, have assessed discourse features such as coherence, cohesion and intertextuality adopting a macro-point of view (Sun and Zhou 2016; Sun 2018). Elements like sentence length and n-gram overlap convey “no notion of meaning, grammar or content” and “as a result they could be very biased towards describing only superficial aspects”, as Felice and Specia point out (2012: 97). In line with this view, Vela et al. (2014) illustrate that BLEU and METEOR are insufficient to provide meaningful evaluations that satisfy the standards of translation studies. They are also ineffective in dealing with lexical variation and dissimilarities that arise from the source text and the process of semantic transfer (Vela et al 2014).

According to Federmann (2014, para.23-24), hybrid approaches need to be employed that ideally bring in to evaluation linguistic information and human knowledge. As expected, the criticism of conventional metrics, which are constructed upon the idea of lexical similarity, paved the way to more linguistically-informed types of measures.

2.4 Automatic evaluation using linguistic features

As Amigo et al (2009: 306) assert, n-gram based approaches to MT evaluation have been dominant because the advantages of employing deep linguistic information lack clarification. Linguistic processing of syntactic, semantic and discourse features has shown considerable potential as an alternative measure to tackle the issue of language variability, which interferes with n-gram-based metrics. The following studies provide important supporting evidence.

Amigo et al. (2009: 311-313) showed that linguistically motivated metrics achieve better correlation with human judgments at system level and they are more sensitive to poor translations with high word overlapping (e.g. “Bush praises NASA’s Mars Mission” vs “Bush praises nasa of Mars mission”). The reason for this is that they incorporate additional restrictions for assigning high scores, so they are harder to deceive (Amigo et al. 2009: 313). They also found that the combination of standard evaluation metrics with techniques that employ linguistic features improved meta-evaluation performance (Amigo et al 2009: 313).

Gimenez and Marquez (2007) conducted a comparative study on the behavior of metrics that depend on lexical matching (e.g. BLEU, NIST) and others that employ deeper linguistic information (e.g. syntactic, shallow-semantic), under single and multiple-reference scenarios. Through evaluating MT systems of different nature (rule-based and statistical), they found that linguistic features provide more reliable system rankings than those that are limited to the lexical dimension, especially when the systems under evaluation are based on different paradigms (Gimenez and Marquez 2007). The researchers suggest that in order to provide a global measure of quality, instead of addressing partial aspects, metrics that operate on different linguistic levels should be incorporated into a single measure (Gimenez and Marquez 2007: 263). Pado et al. (2009) also advocate the development of a model that comprehensively assesses meaning equivalence independent from wording. To this end, they propose a strategy that brings in a rich set of features, including lexical, syntactic and compositional aspects (Pado et al. 2009).

In the same spirit, Comelles et al. (2012) propose a linguistically-motivated metric, VERTa that employs linguistic knowledge to provide a more comprehensive coverage compared to other specialized metrics. VERTa works at different layers. It integrates a lexical similarity module similar to METEOR that, apart from synonyms, stemming and paraphrasing, considers words of related semantic field: hyperonyms (barrel-keg), hyponyms (keg-barrel) and lemmas (danger-dangerous, is-are). It also contains morphological, syntactic and n-gram similarity metrics. Correlation with human assessments showed that the integration of linguistic features at different levels seems promising as it improves performance, compared to other well-known evaluation metrics (Comelles et al. 2012: 3949).

Discourse phenomena in the context of machine translation have been rather under-investigated. As Joty et al. (2017) claim, MT evaluation has been mostly focusing on assessing quality of individual sentences. The “coherent structure” of the text (Hobbs 1979) that connects clauses and sentences guiding the reader’s inferential processes has not yet been seriously addressed, although there has been agreement on the need to integrate discourse-related knowledge in evaluation metrics (Joty et al. 2017: 712). Suitability to the intended context of use has been set as a condition for quality assessment of machine-generated output for a long time, as well as readability, comprehensibility, coherence and cohesion (Hovy, King and Popescu-Belis 2002b ctd in Joty et al 2017: 712).

According to Joty et al. (2017: 685), modelling discourse using RST techniques (Rhetorical Structure Theory; Mann and Thompson 1988) encapsulates those internal semantic relations that make up the coherence structure of the text. Research findings following these methods suggest that sentence-level discourse information should be considered in future metrics as it is quite independent from other features, such as syntactic (Joty et al 2017: 698). They seem to provide extra information that overall result in better correlations with human judgments and when incorporated to existing metrics, they can improve performance (Joty et al 2017: 707). Another interesting finding was that good translations are closer to the reference translations than the poor ones, as evidenced by the respective discourse trees, indicating that discourse similarities are good predictors of translation quality (Joty et al 2017: 706-7). All these findings confirm the importance of taking discourse elements into account when evaluating machine translation.

Scarton and Specia (2015) also investigated the role of discourse phenomena in state-of-the-art Machine Translation. As they comment, extracting linguistic information has been a challenge in NLP, especially when it comes to modeling discourse rather than shallow properties such as syntax (Scarton and Specia 2015: 5). Discourse properties go beyond sentence-level to whole paragraph organization, anaphoric reference, genre and other specifications that are disregarded by state-of-the-art MT (SMT) systems (Scarton and Specia 2015: 5). In this respect, the output of SMT is expected to be flawed (Scarton and Specia 2015: 5). The researchers correlated sentence-level Quality Estimation features (number of tokens in the target/source sentence, punctuation marks in the source/target sentence etc.) and discourse features (lexical cohesion, LSA cohesion, connectives, pronouns etc.) against HTER scores (Human-Targeted Translation Error Rate, the semi-automatic variant of TER [Snover et al. 2006]). Correlation analysis using Pearson's *r* and Spearman's *rho* suggest that discourse information has the potential to improve QE models, as it correlated even higher with HTER than basic, sentence-level features.

The ambitious aim of human translation to satisfy the linguistic norms of the target language (Ahrenberg 2017: 21) and preserve features of discourse, e.g. tone, style, literality and register, further increases the difficulty of evaluation (Dorr et al., 2010: 801). The need to assess translations at a deeper level using linguistically-motivated methods that incorporate lexical, semantic, syntactic and discourse information underlies the studies reported in this part. The following section elaborates more on the idea of multi-dimensional analysis and presents the differences of human and machine-generated translations at multiple linguistic levels.

2.5 Comparison of Human and Automatic translations

In line with Joty et al. (2017), Li et al. (2014: 190) observe that previous empirical studies investigated syntactic, lexical and intelligibility features but none was concerned with discourse. According to them, a comprehensive evaluation of quality requires multilevel linguistic analysis that considers discourse, words, syntax, semantics and pragmatics (Li et al. 2014: 191).

At discourse level, Li et al. (2014) evaluated the accuracy of Google Translate using a corpus of Chinese-English articles. Specifically, they investigated formality and cohesion in expert and machine translations. Formality was assessed based on measures of narrativity, cohesion, embodiment and space-time. In order to obtain a broader picture of discourse, they used additional features extracted from Coh-Metrix (Graesser et al. 2004; McNamara et al. 2014) and Linguistic Inquiry and Word Count (LIWC; Pennebaker, Booth, and Francis 2007). Pearson correlations showed that both the machine and the professional English translations were associated with the Chinese with respect to formality, implicating that the two versions were similar (Li et al. 2014: 193). Interestingly, measures of cohesion, showed that the Chinese text correlated more positively with the automatic than the human translation. As the researchers explain, this finding implicates that Google translate is less flexible and creative in the choice of words and expressions and avoids reordering of sentences, while humans interfere more during the translation process (Li et al 2014: 193). Considering the differences between Chinese and English regarding syntactic structure, it is implied that human translators choose to split sentences and add connectives to facilitate comprehension and adapt the text to the new norms (Li et al 2014). In contrast, the almost perfect correlation regarding sentence length between

MT and the original suggests that Google translate is too inflexible. Although it produces decipherable output, grammatical errors occur especially when dealing with complex sentences (Li et al 2014: 194).

In a more descriptive case study, Ahrenberg (2017) compared two Swedish translations of a British opinion article. The first was human-generated while the second was obtained using Google translation. The researcher found that the human translation was longer with respect to sentence number, words and characters. Similar to Li et al (2014), the human translator preferred to split a considerable number of sentences into two or three parts. Regarding word order, the human translator was twice more intrusive, even though they did not deviate from the source text in general (Ahrenberg 2017).

Apart from modifying the grammar, the translator also improved the style of the text and tried to adapt it to the new audience (Ahrenberg 2017). Some shifts that the MT system did not do, apart from sentence splitting, were explicitation (adding information that is not in the source text to explain a referent), change of perspective (e.g. 'here' is substituted by 'Great Britain' in the target), paraphrasing and various functional/grammatical shifts (adverbs translated by adjectives, clauses reduced by ellipsis etc). Using subjective post-editing review, the study also reported the number of edits required for the MT output to achieve publication quality, estimated to about three per sentence. Most problems were associated with improper word use (Ahrenberg 2017). Ahrenberg (2017) claims that the identification of common weaknesses of MT can contribute to its optimization. However, he is skeptical or 'counting errors' as an evaluation method.

Freitas and Liu (2017) attempted to quantify how human and automatic translations differ using sentences collected from Chinese websites and broadcast.

Apart from key differences, they aimed at identifying what types of errors occur in MT. On par with most studies in this review, the researchers used Google's service to obtain machine translations. The language pair they were interested in was Chinese-English. Their selected criteria focused on two elements: the position of words in the sentences and their underlying structure. For the first part, they calculated edit distance metrics to identify inserts, omissions and substitutions, inflection errors (e.g. happy-happier) and flaws in word order. It is important to note that observed differences in the two samples do not necessarily allude to errors. The use of two synonyms, for instance, is counted as a difference between the texts, although they may produce equivalent effect. The two most differing components between translations were, first, the omission of some words and, second, the insertion of extra ones. The parts of speech that were mostly affected were noun-based, adjective-based and verb-based phrases as well as in-constructions. According to Freitas and Liu (2017) improvement of these constructions should be prioritized, because they occur more frequently over more obscure categories. With reference to inner structure, the most impacted dependency relationships involve among others, parataxis and multi-word expression.

Other studies on machine translation accuracy include Chen, Acosta and Barry (2016) who tested whether Google Translate can be effectively used in the translation of health education material in two language pairs: English-Spanish, English-Chinese. Google was more accurate for English-Spanish than Chinese, however, translation errors could potentially affect patient behaviour leading to undesirable outcomes. This study also provides evidence on the extent to which performance of Google Translate varies across languages. Low accuracy has also been reported for Indonesian (Nadhianti 2016).

Elaborating on cross-linguistic differences, Aiken and Balan (2011) tested the accuracy of Google Translate using 50 text samples in 2.550 language-pair combinations. Most translations were comprehensible according to BLEU scores. However, the system performed better when it translated between Western language pairs than Asian ones (Aiken and Balan 2011). It is also important to note that, due to the great number of languages investigated, only one reference translation was available for each sample. Moreover, this study tested an older version of the service before Google introduced neural machine translation, which reportedly increased accuracy and comprehensibility (Turovsky 2017).

Carter and Inkpen (2012) performed a series of experiments to investigate whether a support vector machine (SVM) could correctly classify machine translations and human-produced texts. The sample they used to train their model contained English and French texts. The researchers concluded that the traits associated with machine translation are actually machine-learnable and detectable using SVM, as the classification was successful for two out of three data sets. A similar approach had been adopted by Baroni and Bernardini (2006) who employed SVMs to recognize translated and original Italian articles. The model reached 86.7% accuracy, outperforming human subjects. Kurokawa, Goutte and Isabelle (2009) confirmed that SVM classifiers are able to detect original and translated texts, this time in English and French samples, with 90% accuracy.

The present study, similar to previous work reviewed in this chapter, attempts to identify some key differences between machine and human translations and to explore whether machine learning techniques are successful in detecting them. Special emphasis is given to the selection of features so that they cover multiple linguistic levels: lexical, syntactic, semantic and discourse. Neural Network analysis aims at identifying which features play the most defining role in the classification task.

In the next chapter, aspects of methodology are discussed before proceeding with the results.

Chapter 3: Research Methodology

3.1 Corpus design and material selection

The aim of this study is to explore differences between human and automatic translations in written form, so appropriate language resources were necessary for qualitative research. As far as I know, there are no open machine translation corpora for the Greek language. Nevertheless, there are quite a few multilingual parallel corpora which could have been used to generate automatic translations. The EuroParl corpus of parliamentary speeches (Koehn, 2005), for instance, covers 21 languages and consists of about 60 million words (Steinberger et al 2014: 684), offering many possibilities to scholars. This large-scale corpus is a valuable resource for Statistical Machine Translation, lexicology and comparative language studies (Steinberger et al 2014: 689). These translations are produced by professional translators, interpreters, linguistic administrators and lawyers to ensure that they meet the highest standards of linguistic quality and accuracy (European Parliament 2008).

One of its limitations, however, is directionality. As Steinberger (2014: 691) explicitly states, “the **source language** for most documents produced by the EU institutions is no longer known” (emphasis in the original). According to the same source, “[i]t is likely that at least some documents were translated via an intermediate language, i.e. that there are translations of translations” (Steinberger et al 2014: 691).

Directionality is an important aspect in translation studies, though. The impact of translation on readability has been investigated cross-linguistically (Ciobanu and Dinu 2014). Findings suggest that there are sufficient and detectable differences between original and translated texts that should be taken into account when training Machine Translation systems (Kurokawa et al 2009). Especially, in a study in which readability is used as an indicator for translation quality, the language of the original should be acknowledged. Therefore, the use of any European corpus was rendered unsuitable.

For the purpose of the current study, three distinct corpora had to be compiled anew; a source corpus, and two English translation corpora. In order to exclude the possibility that internal disparities within the source texts interfered with the results, I decided to focus on a specific genre and collect texts from a single source, namely the website of the Hellenic Ministry of Foreign Affairs. The sample consisted of formal announcements, statements and speeches so, texts did not differ much in terms of style and register. The Greek corpus was used to generate machine translations in English. The corresponding human translations, against which they were compared, were collected from the English website of the Ministry. Given that the government is the official source of information, translation quality was ensured.

The comparative advantages of the specific sample also pertain to practical considerations. First, governmental texts are publicly available, easily accessible and machine-readable. Second, the website of the Hellenic Ministry of Foreign Affairs features a Greek, English and French version.[1] Therefore, the Greek archive of announcements and speeches, which I used as source text, was already available in two translated versions which I could use. Given the existence of multiple text

[1] All Greek texts can be accessed online at https://www.mfa.gr. For the English and French archives, the interested reader can browse https://www.mfa.gr/en/ and https://www.mfa.gr/fr/ respectively.

analysis tools for English input, I preferred English as the corpus' target language. All texts can be retrieved from the website by selecting *Current Affairs> Announcements–Statements–Speeches* and *Meetings–Events* in the English navigation bar and for the Greek version, respectively, *Επικαιρότητα> Ανακοινώσεις–Δηλώσεις–Ομιλίες* and *Συναντήσεις–Δράσεις–Συμμετοχές*.

An important factor to be controlled, to the greatest possible extent, was text length. As Popescu (2009: 70) states, text length "usually exerts influence on many quantitative characteristics". Text length interferes with measurements of lexical diversity such as Type-token ratio (McCarthy and Jarvis 2010: 382), Repeat rate, Entropy and Lambda (Cech and Kubat 2016: 6), some of which I intended to use as indicators of vocabulary richness. So, selecting texts of comparable length was the preferred option. All translated texts had to be juxtaposed to the originals to make sure that they consisted the same amount of information and, of course, the same content. If part of the original was not included in the translated product, it was deleted. This instance occurred only once in the data. Statistical analysis confirmed that average number of tokens did not differ significantly between human and machine translations.

Texts that appeared in the Greek version of the website but were missing in English were discarded in both samples. I also chose to omit some speeches that could have been composed and delivered in English and backtranslated to Greek. As mentioned above, translated texts have been found to deviate from the originals (Kurokawa et al 2009) therefore obtaining information about the texts and filtering them is vital at the stage of corpus design to avoid biased results. Other than these exceptions, all texts were included, from the beginning of 2018 till the first week of

October, same year. As a next step, the corpus was cleaned of captions to avoid the undesirable impact of noise.

The number of texts was 428 in total. The Greek corpus consisted of 42.994 tokens and 7.424 types (N=214) and its English translation 45.468 tokens and 4.688 types (N=214), as measured using Anconc.[2] A table listing all text files and further details about selection can be found in the Appendix.

In order to compare human and automatic translations, I needed a machine-translated corpus to juxtapose to the English translations provided by the Ministry. The Greek texts were exclusively used for this purpose. I selected the freely available machine translation service offered by Google Inc.[3] Google translate has been used in similar comparative (Ahrenberg 2017; Freitas and Liu 2017; Li et al. 2014) and evaluative (Chen et al 2016) studies in the past. It is generally considered accurate at least for Western languages (Aiken and Balan 2011) or compared to other MT services (Aiken 2009). It currently employs Neural Machine Translation systems reaching greater overall accuracy (Wu et al. 2016).

The machine-translated corpus consisted of 214 texts, 43.578 tokens and 4.743 types. As a next step, the two translation corpora were subjected to quantitative analysis. In the following section, I provide a description of the tools used in the study.

---

[2] Anthony (2014) Available from http://www.laurenceanthony.net/software

[3] Available online at https://translate.google.com/

3.2 Text Analysis tools

In order to obtain a record of characteristics for the human and the machine-generated corpora, I made use of two text analysis tools, Coh-Metrix 3.0 (McNamara, Graesser, McCarthy and Cai 2014) and QUITA - Quantitative Indicator Text Analyzer (Kubát et al 2014).

QUITA is a user-friendly tool designed by Miroslav Kubát, Vladimír Matlach and Radek Čech to facilitate quantitative analysis (Kubát et al 2014). The program is freely distributed and provides many functions including text processing. For the purpose of this study, I used it as an indicator computing tool to obtain my data. QUITA mostly computes frequency structure indicators: h-point (h), Entropy (H), Lamda (Λ), Adjusted Modulus (A), Type-Token Ratio (TTR), Vocabulary Richness (R4), Repeat Rate (RR), Relative Repeat Rate of McIntosh (RR mc), Hapax Legomenon Percentage (HL), Gini Coefficient (G), Vocabulary Richness (R1), Curve length (L) and Curve length Indicator (R). Other features include thematic concentration indicators (Thematic Concentration, Secondary Thematic Concentration), Activity (Q) and Descriptivity (D), Writer's View (α), Average Tokens length (ATL) and Verb Distances (VD) (Kubát et al 2014).

Coh-Metrix, developed by Arthur C. Graesser, Danielle S. McNamara, Max M. Louwerse and Zhiqiang Cai, incorporates a wide variety of modules used in computational linguistics (Greasser et al 2004: 194). The user can enter an English text and automatically compute several measures, pertinent to cohesion, readability and language with little effort (Greasser et al 2014: 201). This tool also computes Coh-Metrix L2 Readability formula. The comparative advantage of this metric is that, unlike traditional readability formulas that rely exclusively on shallow metrics, Coh-

Metrix is sensitive to a wide range of linguistic features that affect comprehension such as cohesion, world knowledge and discourse characteristics (Graesser et al 2014: 194).

In this research, I used the most recent version of the tool, Coh-Metrix 3.0.[4] It currently incorporates 108 indices that can be classified under 11 categories: (1) Descriptive indices, that calculate simple metrics such as average length of words, sentences and paragraphs, (2) Text Easability Principle Component Scores, aligned with well-grounded theories of comprehension (e.g., Graesser, Singer, and Trabasso, 1994; Graesser & McNamara, 2011; Kintsch, 1998; McNamara & Magliano, 2009 ctd in McNamara et al 2014), (3) Referential cohesion, evaluated on the basis of content word overlap across sentences (4) Latent Semantic Analysis, (5) Lexical Diversity, (6) Connectives, (7) Situation Model, that involves text elements used in the construction of mental representations for a given text (8) Syntactic Complexity and (9) Syntactic Pattern Density, (10) Word Information, such as frequency scores for syntactic parts-of-speech and psychological ratings (e.g. familiarity, age of acquisition, concreteness, meaningfulness) and (11) Readability consisting of three formulas, Flesch Reading Ease, Flesch Kincaid Grade Level and Coh-Metrix L2 Readability.

The data extracted from both tools were subjected to statistical analysis, the results of which are presented in the next chapter. First, a classification of all indices is presented, along with their usability.

---

[4] Available online at http://tool.cohmetrix.com/

3.3 Overview of Indices

Quality assessment of translated output is done upon specific criteria: faithfulness to the source text, grammar, syntax, spelling, punctuation, style, coherence, fluency and other (Vandepitte 2017: 22). The domain of the text, the expected audience and the purpose of translation are some important aspects to consider when deciding which criteria will be given most weight in evaluation (Hovy et al 2002a: 3).

As mentioned in the Methodology, the corpora of this study contain translations of announcements, responses and statements composed by the Greek Ministry of Foreign Affairs. This sample falls into the category of 'institutional translations', characterized by standardized form and consistency in vocabulary, syntax and style (Schäffner, Tcaciuc and Tesseur 2014: 2). In view of existing literature on discourse analysis, I decided to prioritize appropriacy, comprehensibility, and fidelity as the expected specifications for this genre. The last criterion refers to the accuracy of information transferred to the target from the source. Therefore, this investigation necessitated close examination of the Greek text as well. Given the limited availability of text analysis tools for Greek resources, fidelity was not addressed in this thesis. The reasons that justify the choice of the other two criteria are presented below.

Translation of political texts plays a substantial role in international diplomatic relations (Schäffner et al. 2014: 4). As Schäffner et al (1997: 120) assert, political discourse addresses the public, so it concerns a wide audience. In view of these facts, these types of texts are regulated by rules (Schäffner et al. 2014). According to Biel (2017: 35), the notion of translation quality in EU institutions has been redefined. The focus on 'faithfulness' to the source text has been succeeded by a new narrative that brings communicative aspects such as ‘fitness for purpose’ and ‘clarity’ to the forefront (Biel 2017: 35). Texts are expected to be informative and perfectly clear, while ambiguity should be avoided. This brings up an additional challenge for translators of political texts: apart from meaning equivalence, they should make sure that readability and comprehensibility are not afflicted during the process of information transfer.

Context-dependency and vagueness can be minimized with the use of *formal style* (Heylighen and Dewaele 1999). Heylighen and Dewaele (1999) associate the avoidance of ambiguity and fuzziness with deep formality. They explain that the “attention to form” (Labov 1972), as the term suggests, stems from the desire of the speaker to prevent misunderstanding. Political texts are expected to be accurate and precise to convey the intended message effectively but also for the sake of the convention itself. This is what Heylighen and Dewaele (1999) refer to as ‘surface formality’.

Expectations of appropriacy also stem from the writer’s status. Formal announcements and statements mirror political action. The messages communicated on behalf of the Government need to inspire the required certainty, responsibility and integrity. As Berkenkotter and Huckin (1995) note, genres are informed by their user’s sociocognitive needs (ctd in Trosborg 1997: 149). Conventions of appropriate language use enacted in professional, institutional and other contexts reflect and reproduce existent social structures (Berkenkotter and Huckin 1995: 17 ctd in Trosborg 1997: 149). Stylistic choices are meaningful, thus important to preserve.

Relative to the communicative purpose of political discourse, Schäffner et al. (2014) studied the strategies followed in the translation service of the German Foreign Office. The translators reported that they adopted reader-oriented approaches and *comprehensibility* was their guiding principle (Schäffner et al. 2014: 8). This is expected because the number of translations published online for information purposes is increasing and their audience consists of foreign politicians and journalists (Schäffner et al. 2014: 8). With those readers in mind, professionals prefer to translate texts literally, avoiding insertions, omissions and restructuring to prevent distortion of meaning (Schäffner et al. 2014: 8). Their approach also involved careful consideration of stylistic conventions and aspects of cohesion (Schäffner et al. 2014: 8).

Considering the above, appropriacy and comprehensibility were selected as the expected standards that need to be addressed for the specific sample in question. Appropriacy is mostly satisfied by formality. Formal style is characterized by "detachment, precision, and 'objectivity', but also rigidity and cognitive load" (Heylighen and Dewaele 1999). These attributes have been previously quantified based on part-of-speech distribution: Nouns, adjectives, articles and prepositions occur more often in formal styles whereas pronouns, adverbs, verbs and interjections in informal texts (Heylighen and Dewaele 1999).

Regarding readability, numerous studies have been concerned with the development of linguistically-informed formulas and the exploration of the most influential features that affect ease or difficulty of comprehension, mostly in Foreign Language Teaching (Francois and Fairon, 2012; Francois and Miltsakaki, 2012; Liontou, 2013; Xia et al., 2016) but also for medical applications, to assist people with intellectual disabilities (Carroll et al. 1998; Feng et al, 2009) or to design appropriate information leaflets (Alotaibi et al, 2016). Some of the most prevalent features that have been found to foster readability are high word frequency, limited word and sentence length, lexical cohesion, simplified syntax and low conceptual density (Francois and Fairon, 2012; Liontou, 2013; Xia et al., 2016). Nevertheless, the most widely used metrics have been Flesch Reading Ease (Flesch 1948) and Flesch-Kincaid (Kincaid et al. 1975), focusing only on surface text characteristics, word and sentence length.

Based on this theoretical background, I decided to compare human and machine-generated translations with respect to six sets of linguistic features: Descriptive Information, Word Information, Lexical Richness, Syntax, Cohesion and Readability. These characteristics were selective as indicative of lexical diversity, structural complexity and discourse. The variability of features makes sure that deep linguistic knowledge is employed for a more comprehensive evaluation that confers many advantages as clarified in Chapter two. At the same time, these features are representative of criteria that concern evaluators of real-life applications and are carefully selected with the specific text sample in mind.

3.3.1 Descriptive information

Although descriptive information provides general statistics of the text, it is informative. The length of words, sentences and paragraphs is suggestive of lexical and syntactic complexity; Long sentences and paragraphs are considered more difficult to process because they involve many constituents imposing additional cognitive load (McNamara et al 2014). According to Zipf’s law (1935) long words are usually more difficult than short ones. The latter are easier to read because they tend to be more frequent and familiar to the reader (McNamara et al 2014).

Descriptive indices include: Types, Tokens, Average Tokens Length (ATL), Token Length Frequency Spectrum, Number of paragraphs (DESPC), Number of sentences (DESSC), Number of words (DESWC), Mean length of paragraphs calculated in sentences (DESPL), Mean length of sentences counted in words (DESSL) and Mean length of words, measured either in syllables (DESWLsy) or letters (DESWLlt). These features were collected from QUITA and Coh-Metrix 3.0.

Coh-Metrix also calculates Standard Deviations for most of the mentioned values. These scores indicate the degree of variation within the text. For instance, large standard deviations regarding the Mean length of sentences (DESSLd) indicate that both long and short sentences exist in the sample (McNamara et al 2014). Similar measures provided by this tool include Standard deviation of mean length of paragraphs (DESPLd) and Standard deviation of word length with respect to syllables (DESWLsyd) or letters (DESWLitd).

3.3.2 Word Information

This category includes information on Parts of Speech (Noun incidence, WRDNOUN; Verb incidence, WRDVERB; Adjective incidence, WRDADJ; Adverb incidence, WRDADV; Pronoun incidence, WRDPRO) and Perspective (First person singular pronouns, $WRDPRP_1s$; First person plural pronouns, $WRDPRP_1p$; Second person pronouns, $WRDPRP_2$; Third person singular pronouns, $WRDPRP_3s$; Third person plural pronouns, $WRDPRP_3s$).

Regarding psycholinguistic properties of words, the analysis covers Average CELEX word frequencies, Age of acquisition for content words (WRDAOAc), Familiarity for content words (WRDFAMc), Concreteness for content words (WRDCNCc), Imageability for content words (WRDIMGc), Meaningfulness (Colorado norms) for content words (WRDMEAc), Polysemy for content words (WRDPOLc), Hypernymy for nouns (WRDHYPn), Hypernymy for verbs

(WRDHYPv) and, lastly, Hypernymy for nouns and verbs (WRDHYPnv). All features in this category are selected from Coh-Metrix 3.0.

Word frequency is based on word ratings included in the CELEX database. This resource is offered from the Dutch Centre for Lexical Information and includes information extracted from analysis of 17.9 million words (Baayen, Piepenbrock, and Gulikers 1995 ctd in McNamara et al 2014: 73). CohMetrix only assesses the values of those words that are contained in the database. The rest are not taken into account. Therefore, the existence of nonsensical words that may occur in the automatic translations does not affect these results.

Estimates of polysemy and hypernymy are provided using WordNet (Fellbaum, 1998; Miller et al., 1990 ctd in McNamara 2014: 74). Polysemous words, though ambiguous, tend to be frequent (McNamara et al 2014: 43). Hypernymy refers to the number of conceptual taxonomic levels that precede a word e.g. table has seven hypernym levels (seat→furniture→furnishings→ instrumentality→ artifact→object→ →entity) (McNamara et al 2014: 43). Hypernymy is positively correlated with concreteness, that is concrete words have a greater number of hypernyms than abstract words (McNamara et al 2014: 44).

Age of acquisition, familiarity, concreteness, imagability and meaningfulness are based on information encoded in the MRC Psycholinguistic Database (Coltheart, 1981 ctd in McNamara et al 2014: 74). Again, it is important to note that these measures are estimates of the words included in the database. Non-existent words that may occur in the sample as a result of system failure do not affect the means of psycholinguistic indices, such as meaningfulness because they are not found in MRC. These measures only reflect how words are represented in the mind based on extant human ratings, e.g. the word "ball" is considered to be more concrete than "difference"; "reason" is low in imagability compared to "bracelet" that immediately evokes a mental image; "people" is considered more meaningful that "abbess" in the sense that is highly associated with other words etc (McNamara et al 2014: 75).

3.3.3 Lexical Richness

Lexical richness was estimated at multiple levels. First, standard and more sophisticated metrics were computed such as Type-Token Ratio (TTR) for all and content words, Measure of Textual Lexical Diversity (MTLD) and VOCD. While TTR is sensitive to variations in text length, MTLD was designed to overcome such disparities (McCarthy and Jarvis 2000 ctd in McNamara et al. 2014: 51). As McCarthy and Jarvis (2010: 382) explain, as a text becomes longer, the rate of increase for the number of new types slows and tokens become more repetitive. This balance is necessary so that the text is coherent and meaningful (McCarthy and Jarvis 2010: 382). Therefore, the gradual decrease in lexical diversity does not affect the reader's view of the text, but it may lead to misleading quantitative representations made by researchers (McCarthy and Jarvis 2010: 382). What further differentiates MTLD from VOCD and TTR is that, instead of approaching the text as a whole, it considers the sequence of the wording (McCarthy and Jarvis 2010). This approach may be preferable on the grounds that it preserves the integrity of the text (McCarthy and Jarvis 2010: 382). McCarthy and Jarvis (2010: 391) who investigated the validity of lexical diversity metrics advise researchers to use multiple measures, sequential and non-sequential, rather than a single index, noting that "lexical diversity can be assessed in many ways and each approach may be informative as to the construct under investigation".

Second, I calculated additional metrics, frequently used as indicators of lexical richness in cross-linguistic studies, comparison of genres and authorship attribution. Those were selected from QUITA: Vocabulary Richness (R1), Vocabulary Richness (R4), Hapax Legomena Percentage[5] (HL), Entropy (H), h-point (h), Lambda (Λ), Repeat Rate (RR), Relative Repeat Rate of McIntosh (RRmc), and Gini Coefficient[6] (G).

As Popescu (2009: 165) state, Entropy and Repeat Rate have gained particular attention in linguistic research. They are both used as measures of diversity, but entropy is mostly associated with dispersion and uncertainty while Repeat rate is a measure of concentration (Popescu 2009: 166; Popescu, Čech and Altmann 2011: 3). Smaller H indicates that the vocabulary is concentrated to a few words while greater values suggest that the distribution is more even, i.e. most of the words occur only once in the data (Popescu 2009: 173). Therefore, greater relativised Entropy means greater vocabulary richness (Popescu et al. 2011: 3). Repeat rate is also concerned with vocabulary concentration and can be interpreted by reference to spectrum and rank frequency (for a detailed review, see Popescu 2009: 166). The smaller the relativised Repeat Rate, the richer the vocabulary (Popescu 2009: 181; Popescu et al. 2011: 3).

Lambda and h-point are used to detect the degree of analytism and synthetism within a text (see Poiret and Liu 2017 for an investigation of Lambda's capacity). Greater values of Lambda generally mean greater vocabulary richness (Popescu, Čech and Altmann 2011: 9) H-point separates the vocabulary in two classes, systematics

---

[5] The ratio between the number of tokens and the number of words that occur only once in the text (Mandravickaite and Krilavičius 2018:62)

[6] This index shows the position of the text between maximal and minimal vocabulary richness (Popescu 2009: 57).

and autosemantics. Synsemantics (prepositions, conjuctions, pronouns, articles, particles etc) are more frequent in language than autosemantics, which constitute the core vocabulary of the text. Greater h is a sign of analytism; it indicates that there are less word forms and greatest prevalence of auxiliaries within a text (Popescu 2009: 23).

3.3.4 Syntax

Regarding Syntax, I measured indicators of (a) *Syntactic complexity* (Left embeddedness[7], SYNLE; Number of modifiers per noun phrase, SYNNP; Minimal Edit Distance, part of speech SYNMEDpos; Minimal Edit Distance all words SYNMEDwrd; Minimal Edit Distance lemmas, SYNMEDlem; Sentence syntax similarity of adjacent sentences, SYNSTRUTa; and Sentence syntax similarity all combinations across paragraphs SYNSTRUTt). The three Minimal Edit Distance indices assess semantic and syntactic dissimilarity by measuring the extent to which consecutive sentences within a text are structurally and lexically close (McNamara et al 2014: 70), e.g. “She took her stuff. She left her job”. Adjacent pair sentences may be syntactically similar but semantically different, so the indices work complementary (McNamara et al 2014: 70). SYNMEDpos calculates how many edits are needed to make the syntax of the first sentence match the second considering agreement in parts of speech. The rest two indices, SYNMEDwrd and SYNMEDlem consider word choice instead. As expected, SYNMEDpos tends to correlate with syntactic complexity while SYNMEDwrd and SYNMEDlem correlate stronger with measures of referential and semantic cohesion. (McNamara et al 2014: 70). The last two indices

[7] Refers to “the mean number of words before the main verb” (McNamara et al 2014: 70)

referring to syntax similarity also assess uniformity between consecutive sentences and across paragraphs (McNamara et al 2014: 71).

Coh-Metrix also computes measures of (b) *Syntactic pattern density* (Noun phrase density, DRNP; Verb phrase density, DRVP; Adverbial phrase density, DRAP and Preposition phrase density, DRPP; Agentless passive voice, DRPVAL; Negation density, DRNEG; Gerund density, DRGERUND and Infinitive density, DRINF). Texts with frequent noun and verb phrases are likely to be informationally dense therefore more complex (McNamara et al 2014: 72).

As McNamara et al (2014: 48, 70) mention, short sentences that follow a simple agent-action-object structure are easier to process than sentences with embedded clauses and passive voice that increase the cognitive load on working memory. Structurally complex sentences can be dense, ambiguous, even ungrammatical (Graesser 2004: 198) and are mostly associated with increased formality (Heylighen and Dewaele 1999).

3.3.5 Cohesion

Text cohesion is a complex property that needs to be assessed at multiple levels. Using Coh-Metrix, I selected to compute: *(a) Connectives* measured overall and by category (All connectives incidence, CNCall; Causal, CNCCaus; Logical, CNCLogic; Adversative-Contrastive, CNCADC; Temporal, CNCTemp; and Additive) *(b) Referential Cohesion* (Noun, Argument and Stem overlap within adjacent sentences and within all sentences in a paragraph; Content word overlap in all sentences and adjacent sentences, Means and Standard deviations) *(c) Latent Semantic Analysis* (LSA overlap between all sentences in a paragraph, adjacent sentences and adjacent paragraphs, Means and Standard deviations; LSA given/new sentences mean, LSAGN and standard deviation, LSAGNd) *(d) Situation model* (Causal verb incidence, SMCAUSv; Causal verb and particles incidence, SMCAUSvp; Intentional verbs incidence, SMINTEp; Ratio of casual particles to causal verbs, SMCAUSr; Ratio of intentional particles to intentional verbs, SMINTEr; LSA verb overlap, SMCAUSlsa; WordNet verb overlap, SMCAUSwn; and Temporal cohesion: tense and aspect repetition, mean, SMTEMP).

The use of *Connectives* guides the reader to the intended meaning, facilitating comprehension. *Referential cohesion* is slightly more implicit. It refers to the proportion of content words, mostly nouns, which co-occur between and across sentences. (McNamara et al. 2014: 64). *Latent Semantic Analysis* (Landaeuer et al. 2007) also measures cohesion based on semantic overlap (McNamara et al. 2014: 66). However, overlap is not limited in exact word forms or morphological variants, but extends to semantic information e.g. “home” exhibits high LSA overlap with “table”. The LSA given/new indices (Hempelmann et al. 2005; McCarthy, Dufty et al. 2012) assess the extent to which text information is given or introduced for the first time (McNamara et al 2014). Givenness is estimated by co-reference. New information increases the reader’s cognitive load, because it requires that they bridge any semantic gap inferentially (McNamara et al. 2014).

Relative to inference processes, *Situation model* refers to “the reader’s mental representation of the deeper underlying meaning of the text (Kintsch 1998 ctd in McNamara et al. 2014). Meaning is partly constructed with the use of causal and intentional verbs that describe actions (e.g. break, hit, impact) and goals of animate agents (e.g. walk, talk). If a great number of events is expressed in the text, the necessity for connectives to guide the reader inferential processes increases (McNamara 2014).

3.3.6 Readability

Readability was assessed using two traditional formulas, Flesch Reading Ease and Flesch-Kincaid Grade Level and a third index designed for second-language texts, Coh-Metrix L2 Readability. What differentiates the latter is that it goes beyond word level and considers aspects of cohesion (McNamara et al. 2014: 81). While traditional readability formulas have been widely criticized by cognitive researchers for their limited scope and lack of theoretical grounding, Coh-Metrix was designed to address these shortcomings (Crossley, Dufty, McCarthy and McNamara 2007: 197)

Coh-Metrix also computes eight Easability Scores derived from principal component analyses of 54 indices (Graesser, McNamara and Kulikowich 2011). According to previous research, these eight components accounted for a considerable amount of variability among texts of several domains and were thus, selected as indicative of text complexity (McNamara et al. 2014). These components consider Narrativity (affiliated with everyday conversation and familiar topics), Syntactic Simplicity, Word Concreteness (meaningful words evoke mental images juxtaposed to abstract words that are hard to process), Referential Cohesion (ideas are connected through overlap), Deep Cohesion (causal and logical relationships are made explicit with the use of connectives that guide comprehension), Verb Cohesion (repetition of verbs and events), Connectivity (use of adversative, additive and comparative connectives to explicate logical connections), and Temporality (consistency in the use of tenses and aspect facilitates understanding of events) (McNamara et al. 2014).

Chapter 4: Results

4.1 Statistical Analysis

The rationale of this research is that the identification of differences between machine and human-generated corpora may provide us with those text characteristics that discriminate the two translation types. These features are likely to be good predictors of translation quality, I assume, but that remains to be investigated.

Independent-samples t-test were conducted to examine whether machine and human translations differed in terms of six linguistic features: (1) Descriptive characteristics, e.g. number of types/tokens, word and sentence length, (2) Word Information, e.g. part-of-speech and psychological ratings, (3) Lexical Richness measured by standard metrics and stylometry indices, (4) Syntax, (5) Cohesion and (6) Readability. Then, I tested the extent to which the Multilayer Perceptron Neural Network could successfully predict the category of texts and which features contributed more to this classification.

Analysis was performed using IBM SPSS 25.0, the Statistical Package for Social Sciences. An alpha level of .05 was used for all statistical tests.

4.2 Differences between HT and MT

4.2.1 Descriptive Information

As mentioned in the Methodology section, the two corpora were balanced in terms of content in the sense that the MT system was provided with the exact amount of input that had been translated in the expert text. Therefore, any variations in the two corpora with regard to word count are due to different wording and translation strategies, for instance, insertions or omissions.

The mean number of tokens was slightly higher in human translations (M=212, SD=149) but the difference from the machine-generated output was not significant (M=204, SD=143); t (426) =.63, p=.532. Regarding *tokens* length, human translations contained longer words (M=5.33, SD=0.32) than MT on average (M=5.25, SD=.31); t (426) =2.89, p=.004. The existence of shorter words in MT is partly confirmed when looking at *word* length measured in syllables, t(426)=1.446, p=.149 or letters, t(426)=1.708, p=.088, however, these results may be random.

TABLE 1: Descriptive Statistics Means (Standard deviations)[8]

	HT	*MT*
Tokens, mean	212 (149)	204 (143)
Types, mean	109 (65.5)	109 (64.8)
Average Tokens Length	5.33* (0.32)	5.25 (0.31)
Word length, mean number of syllables	1.83 (0.13)	1.81 (0.12)
Word length, mean number of letters	5.30 (0.30)	5.25 (0.30)
Sentence count, number of sentences	10.43* (7.49)	8.78 (5.31)
Sentence length, mean number of words	21.07* (5.77)	22.90 (6.25)
Paragraph count, number of paragraphs	6.15 (2.95)	6.15 (2.93)
Paragraph length, mean number of sentences	1.61* (0.56)	1.39 (0.38)
Paragraph length, standard deviation	0.87* (0.65)	0.62 (0.49)

[8] The asterisks indicate statistical significance at .05

As can be seen in Table 1, human translations contain more sentences that are shorter in length while MT is characterised by fewer but more elaborate sentences. The differences regarding sentence count [t(383)=2.636, p=.009], and paragraph length [t(376)=4.737, p<.001] were found to be significant. This finding could be linked to the comparative prevalence of connectives, especially temporal, in automatic translations (see section 4.2.5). Interestingly, human translations were also found to be more heterogenous than automatic in terms of paragraph length as evidenced by greater standard deviations, t(396)=4.482, p<.001. The comparatively greater variation in HT is also indicated from the higher Standard deviations that correspond to the Mean values for 9 out of the 10 descriptive features presented in Table 1.

4.2.2 Word Information

Part-of-speech frequency did not differ significantly in human and machine translations, with the exception of Verbs, whose number was greater in HT (M=75.24, SD=25.5) than MT (M=66.72, SD=25.4); t (426)=3.52, p<.001.

TABLE 2: Frequency, Parts-of-speech (Standard Deviations)

	HT	*MT*
Noun incidence	399.5 (66.46)	396.0 (69.07)
Verb incidence	75.24* (24.5)	66.72 (25.4)
Adjective incidence	73.91 (29.8)	76.39 (30.3)
Adverb incidence	18.25 (14.05)	20.60 (14.91)
Pronoun incidence	22.77 (22.60)	24.24 (24.54)

Regarding psycholinguistic features, the two corpora differ with respect to CELEX Log frequency for all words (WRDFRQa) and CELEX Log minimum frequency for content words (WRDFRQmc).

Machine-produced output seems to contain more frequent words (M=3.02, SD=.13) compared to human translations (M=2.98, SD=.12), t(426)=-2.95, p=.003. This finding is partly confirmed when looking specifically at content words, but in this case, the higher word frequency observed in MT is not statistically significant (p=.116). As far as minimum word frequency for content words is concerned, the two samples differ significantly. Human translations exhibit higher values (M=.59, SD=.66) than machine translations (M=.40, SD=.58), t (420) =3.21, p=.001. All these findings show that HT contains less frequent words than MT, indicating that the vocabulary of expert translations may be more technical and domain-specific.

Following from the previous findings, mean familiarity for content words was higher in MT, but the difference was insignificant (p=.327). Important variations between the two corpora were not observed regarding age of acquisition for content words (p=.807), concreteness (p=.912), imageability (p=.287), meaningfulness (p=.176) and polysemy (p=.378). In contrast, hypernymy for verbs exhibited remarkable variation in HT (M=1.72, SD=.31) and MT samples (M=1.83, SD=.34), t(426)= -3.48, p=.001. This finding comes in stark contrast to the corresponding HT (M=5.48, SD=.58) and MT values (M= 5.46, SD=.62); t(426)=.351, p=.726 regarding hypernymy for nouns. These results suggest that automatic translations contain more concrete verbs, but this cannot be generalised to all word categories for two reasons. First, the results concerning hypernymy for nouns are conflicting and second, the two corpora did not exhibit significant variation with respect to word concreteness. Apart from that, the Text Easability index for word concreteness also provides contradictory information to the view of MT as less concrete, though not validated by significance levels, t(426)=-1.870, p=.062.

4.2.3 Lexical Richness

Lexical diversity seems to differ in the two corpora. TTR *for all words*, as measured by QUITA, was 54.3% for the human-translated corpus and 57.3% for machine translations (the same measure calculated by Coh-Metrix yielded similar results with only a slight increase in the two samples: 54.7% for HT and 57.7% for MT). The difference between the two corpora is statistically significant, $t(426)=-3.976$, $p<.001$. However, it should be noted that the mean number of Types was slightly higher in HT ($M=109.39$, $SD=65.50$) than MT ($M=109.26$, $SD=64.77$). In light of this, the previous results seem conflicting; they can be explained though if we consider the number of tokens in the two samples. HT contained slightly more tokens ($M=212$, $SD=148$) than MT ($M=203$, $SD=143$). Therefore, the results of TTR do not necessarily imply that automatic translations would be perceived as more lexically rich by the reader.

Regarding TTR for *content word lemmas*, the variation between HUM and MAC translations was also significant. Once more, machine translations exhibited a significantly higher TTR ($M=.78$, $SD=.07$) than human translations ($M=.73$, $SD=.81$); $t(426)=-6.729$, $p<.001$. The scores of MTLD for all words were consistent with those of TTR. Machine translations were found to be more diverse ($M=60.4$, $SD=18.66$) than human ($M=59.67$, $SD=15.83$), however, this difference was insignificant, $t(426)=-.441$, $p=.660$. Contrary to the previous results, lexical diversity as assessed by VOCD was higher for the HUM corpus ($M=50.23$, $SD=28.71$) compared to MAC ($M=49.59$, $SD=28.73$), but again, results may be random as evidenced by low significance levels $t(426)=.231$, $p=.818$.

Lexical richness was also assessed using a set of different indicators. In line with the above measures, the mean number of Hapax legomena was greater in automatic translations ($M=.41$, $SD=.087$) compared to HT ($M=.36$, $SD=.086$);

$t(426)=-5.625$, $p<.001$. Lambda was also significantly higher in the machine-generated sample ($M=1.38$, $SD=.15$) than in expert translations ($M=1.32$, $SD=.17$); $t(426)=-4.169$, $p<.001$. The rest stylometric measures did not reveal any significant variations between the two corpora.

Although the differences observed were not conclusive, some of these results stimulated scepticism regarding the plausibility of the computed metrics for the analysis of MT output. Certainly, lexical diversity as represented in this analysis should not be confounded with superior translation quality. The reason is twofold; first, regression analysis needs to be performed in order to examine the exact relationship between lexical diversity and quality in this data, whether the former can be used as a predictor of the latter. Second and most important, various factors might have affected the metrics, involving the existence of multiple ill-formed name entities in the MT sample. These possibilities are explained and discussed in more detail in the next Chapter.

4.2.4 Syntax

Machine translations contained more modifiers per noun phrase, $t(426)=-4.427$, $p<.001$, an indication of syntactic complexity. Other indicators such as Left embeddedness, Negation density and Passive voice partly confirm the existence of more complex structures in MT but are not validated by significance levels ($p=.219$, $p=.197$ and $p= 2.17$ respectively)

The two corpora differed with respect to the extent of semantic and syntactic dissimilarity between consecutive sentences. The three indices that employ Minimal Edit Distance showed that consecutive sentences were more similar in MT. All exhibited statistically significant results (SYNMEDpos: t(421)=2.641, p=.009, SYNMEDwrd: t(417)=2.351, p=.019, SYNMEDlem: t(418)=2.394, p=.017). SYNSTRUTt also approximated the level of significance. The results concerning all sentence combinations across paragraphs tell a different story. In this case, the syntactic composition in HT appears to be more uniform (M=.064, SD=.38) than MT (M=.058, SD=.01); t(306)=1.94, p=.053.

With regards to Syntactic patterns, Noun and Verb phrase incidences were higher in MT, but not at significant extent (p=.165 and p=.085 respectively). Only Gerund density seems to differentiate the two corpora. The expert translations contained a considerably greater number of gerund structures (M=15.10, SD=10.22) compared to machine translations (M=9.92, SD=10.09); t(426)=5.279, p<.001.

4.2.5 Cohesion

Cohesion was measured with respect to (a) Connectives, (b) Referential Cohesion, (c) Latent Semantic Analysis and (d) Situation Model.

The total number of *Connectives* did not differ significantly in the two corpora. Out of the 5 subcategories specified in section 3.4.5, only Temporal connectives varied significantly in the two samples. Specifically, the mean number of temporal connectives in MT (M=12.63, SD=10.83) exceeded that of HT (M=9.09, SD=8.68); t(406)=-3.736, p<.001. Although the results for the rest categories may be random, it is worth to mention that, overall, more connectives were found in MT (M=65.85, SD= 26.93) than HT (M=61.69, SD=25.90);t(426)=-1.63, p=.104 but also in 5 out of 6 subcategories: causal, logical, adversative-contrastive, temporal and additive connectives.

Regarding *Referential cohesion*, human translations contained more overlapping components. The categories that varied were Content word overlap across all sentences (CRFCWOa), $t(380)=4.585$, $p=.002$ and Content word overlap between adjacent sentences, $t(424)=2.099$, $p=.036$. Another interesting aspect was the notable difference in the standard deviations of CRFCWOa that were greater in HT ($M=.15$, $SD=.08$) than MT ($M=.12$, $SD=.05$); $t(388)=4.585$, $p<.001$.

Latent Semantic Analysis showed that text cohesion as estimated by the proportion of given/new information was higher in the HT ($M=.28$, $SD=.044$) than MT ($M=.26$, $SD=.055$);$t(409)=4.474$, $p<.001$. Therefore, the HT sample would probably require less inferencing on the part of the reader in bridging any semantic gaps at least regarding newly introduced information in comparison to the MT text.

Situation Model did not reveal remarkable differences, except for causal verbs that were slightly more in HT ($M=18.90$, $SD=8.09$) than MT ($M=17.03$, $SD=8.01$); $t(426)=2.407$, $p=.017$ and Causal Verbs and particles incidence in HT ($M=21.98$, $SD=9.99$) that again outnumbered MT ($M=19.86$, $SD=8.96$); $t(426)=2.316$, $p=.021$. However, given that the text mostly describes static and not action events, this variation was not given much weight in the evaluation.

These results, though inconclusive, indicate that machine translations contain more explicit cohesive cues, such as connectives but human translations exhibit greater cohesion at deeper level as evidenced by referential cohesion and situation model.

4.2.6 Readability

Traditional readability indices, Flesch Reading Ease and Flesch-Kincaid Grade level as well as Coh-Metrix L2 Readability evaluated the Human translated texts as slightly more readable; however, the difference was insignificant. The results of the measures are summarized in Table 4.

TABLE 3: Mean Readability Scores (Standard Deviation)

	HT	*MT*
Flesch Reading Ease	31.07 (13.26)	30.65 (13.03)
Flesch-Kincaid Grade level	14.17 (2.97)	14.67 (3.03)
Coh-Metrix L2 Readability	5.94 (5.37)	5.89 (4.80)

Flesch Reading Ease assigns a readability score from 0 to 100. Higher numbers suggest that the text is easier to understand. As mentioned on the website of CohMetrix[9], “an average document has a Flesch Reading Ease score between 6 and 70”. Input that is assigned scores between 0 and 30 is classified as ‘difficult’ and is better understood by college graduates (Readability Formulas n.d.). Therefore, both samples are relatively hard to comprehend. Machine translations are slightly more confusing in this respect but the difference is certainly insignificant, t(426)=.332, p=.740. Based on the same premises, the Flesch-Kincaid Grade converts the Flesch Reading Ease score to U.S. years of education required in order to comprehend the text. Grade levels range from 0 to 12. Flesch-Kincaid assigned almost 14 years of education to both samples. The results of these formulas are essentially the same, but elsehow represented. According to Coh-Metric L2 readability, HT was slightly more readable, but the difference was not at all significant, t(426)=.097, p=.923. In general

[9] Available at https://archive.is/20121214205404/http://cohmetrix.memphis.edu/CohMetrixWeb2/HelpFile2.htm

terms, readability as expressed globally in these three metrics does not seem to differentiate the two samples.

Text Easability scores focusing on specific linguistic areas revealed more differences. Human translations were found to be more readable with respect to Narrativity (z score), $t(426)=2.233$, $p=.026$, Syntactic simplicity (z score), $t(426)= 5.002$, $p<.001$ and Syntactic simplicity (percentile), $t(401)=5.507$, $p<.001$. Machine translations exhibited more instances of Verb cohesion (Text Easability component) ($M=44.80$, $SD=32.82$) than HT ($M=33.24$, $SD=27.78$); $t(414)=-3.935$, $p<.001$.

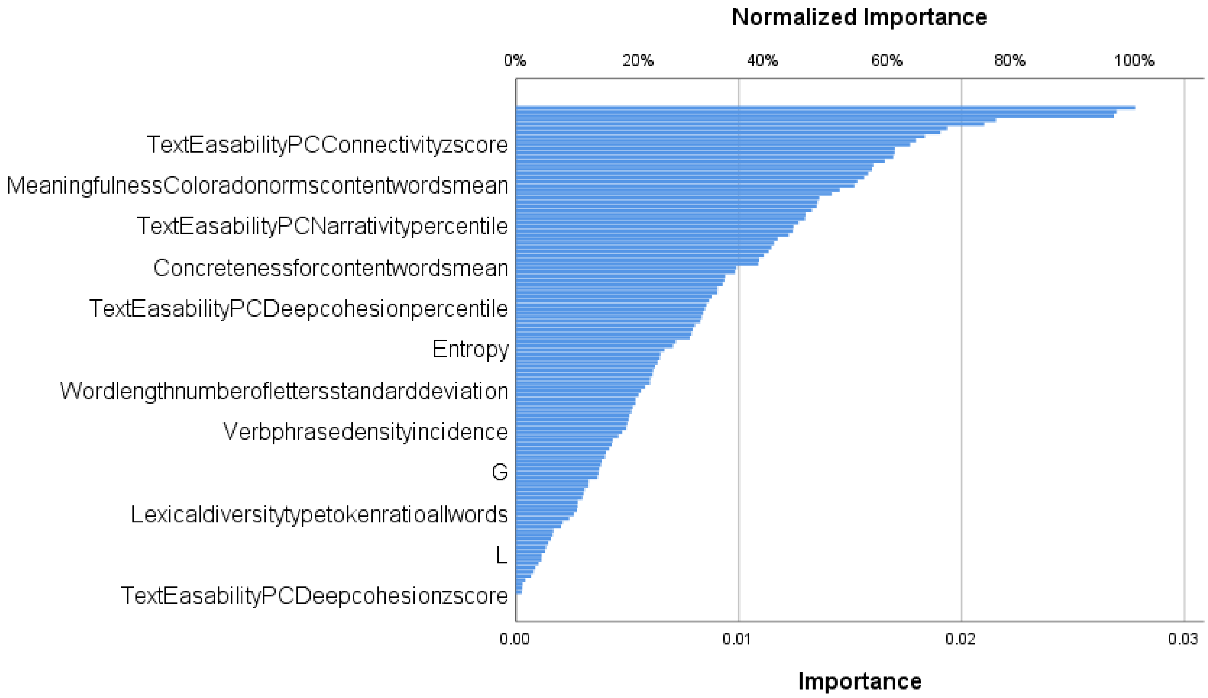
4.3 Text Classification and Sensitivity of variables

Having identified some detectable differences between machine and professional translations, I attempted to explore the predictive value of the variables in a text classification task. To this end, I performed artificial neural network analysis using Multilayer Perceptron classifiers. Covariates were rescaled using normalization in SPSS.

Of 428 texts in total, 312 were used for training (72.9%) and 116 for testing (27.1%). The model reached 83% accuracy in the training sample, while, 81.9% of the texts were grouped correctly in the testing sample. These results support that the differences between HT and MT are actually learnable and detectable; however, the model did not achieve optimal performance. 10 machine-generated texts were incorrectly classified as human-produced and 11 human translations were grouped as automatic. This indicates that first, the model can be improved and second, a respectable number of machine and human output, almost 18% of the sample, is not that easily classifiable. This probably implicates that the boundaries between human and machine translations are not as clear-cut as expected; some machine translations come very close to human writing.

Sensitivity analysis was also performed to estimate the effect of individual variables on how the network grouped the samples. These features are only indicative, since there was great variation in the results on each run based on the randomized selection of the training and testing samples and the assigned variable weights. Therefore, results should be regarded with caution.

Chart 1: Independent Variables Importance (SPSS)



The ten most important predictors of the network, sorted by normalized importance were: Text Easability PC Connectivity, z score (64.6%), Meaningfulness for content words (55.2%), Text Easability PC Narrativity, percentile (45.7%), Concreteness for content words (39.1%), Text Easability PC Deep Cohesion, percentile (30.8%), Entropy (25.4%), Word length measured in letters, SD (20.9%), Verb phrase density (17.8%), Gini's coefficient (13.4%) and Lexical Diversity - Type-Token Ratio, all words (9.9%). Text Easability scores account for a great part of the model. However, results are not easily interpretable due to the fuzziness of features; for instance, Text Easability PC Connectivity is categorized as a Readability index, but also reflects cohesion and discourse structure. In other words, the boundaries between some categories are not clear. Second, the ten more influential variables represent a broad range of linguistic properties. At least one component of all six categories (descriptive, word information, lexical diversity, syntax, cohesion, readability) constitutes a significant predictor for the model. Although it can be supported that all categories are to a certain extent influential, a principal component analysis, as described in Li et al (2014) would probably provide a more specific view of the effect of grouped variables in the classification task.

In the next chapter, the implications of the results are discussed in relation to the research questions of this thesis, as specified in the Introduction.

Chapter 5: Discussion

The present study was stimulated by the need to discover some strengths and weaknesses of machine translation. This objective was fulfilled through the comparison of HT and MT samples.

Quality assessment needs to be discussed with reference to specific specifications that arise from the domain of the evaluated text and the purpose of translation. The present sample contained statements on current topics and announcements of upcoming events composed by the Hellenic Ministry of Foreign Affairs. These texts, which formally address the public, have important informative value and reflect the conduct of the Ministry. Formal style is preferred to inspire sovereignty, firmness and credibility but also to avoid ambiguity and make oneself perfectly clear (Heylinghen and Dewaele 1999). Therefore, both the content and the stylistics of the source need to be preserved in the translation as they constitute well-informed choices that serve specific purposes.

Formal style, according to Heylighen and Dewaele (1999) is associated with "detachment, accuracy, rigidity and heaviness". The distribution of words in the corpus has been proposed as a method to quantify style: nouns, adjectives, articles and prepositions are more common in formal texts while pronouns, adverbs, verbs and interjections mostly occur in informal styles (Heylighen and Dewaele 1999). The present analysis did not reveal important differences in the distribution of parts of speech or POS-phrases (verb phrases, noun phrases, adjective phrases etc), indicating that machine-translations did not generally deviate from the expected specifications. However, the greater number of verbs in human translations needs to be further investigated by reference to the source text in order to explain this variation. Previous research has not identified remarkable differences in HT and MT with respect to formality either (Li et al. 2014).

Regarding word information, professional translations contained longer and less frequent words. This is potentially indicative of the use of more technical and domain-specific vocabulary in the HT sample. Schäffner et al. (2014: 8) report that professional translators working in institutional contexts give special emphasis to stylistic conventions, aspects of cohesion and comprehensibility. The use of correct terminology is also prioritized in European institutions as evidenced by the wide use of IT resources such as specialized translation memories and terminological databases e.g. EurLex, IATE in order to reduce the risk of human error (European Parliament n.d.). Automatic translation systems are more likely to transfer the meaning without notice to the domain of the target text. As stated by Osmałek (2014), machines are not in the position to recognize linguistic nuances, and in a situation in which a professional finds the most appropriate synonym for the given context the machine simply translates the word (ctd in Puchała-Ladzińska 2014: 95).

Nevertheless, it is important to note that the introduction of statistical methods in machine translation (SMT) has enhanced translation quality (Wu et al. 2016), especially in applications for which parallel texts are available (Puchała-Ladzińska 2016). Google switched to SMT in 2007, after the use of rule-based systems (Puchała-Ladzińska 2016: 94). It currently compares and detects common patterns in millions of human-translated documents and makes intelligent guesses based on this information (Puchała-Ladzińska 2016: 94). This method is expected to be particularly effective in standardized subgenres like the one in question, characterized by consistency in style and existence of repetitive structures. The most recent introduction of neural machine translation used to translate entire sentences rather than isolated phrases further upgraded the service for a number of language pairs (Turovsky 2017).

An important finding that arose from the comparison was the tendency of machine translations to be significantly more elaborate at sentence level. This may impact comprehensibility, especially if the reader is not familiar with the content. Specifically, MT contained fewer but longer sentences, while HT contained a greater number of sentences that were shorter in length. This finding is consistent with Ahrenberg (2017) and Li et al (2014) who observed that human translators tend to split sentences and deviate from the structure of the original when necessary to facilitate comprehension. Machine translations are more inflexible in this respect. The greater flexibility in HT is also evidenced in this study, when the greater standard deviations in professional translations are considered.

The existence of more elaborate sentences is aligned with greater syntactic complexity. Although there are strong indicators that point to the existence of more complex structures in MT, mostly modifiers and left embeddedness, the findings are not enough to validate this point because statistical significance was not observed in two out of the three computed measures. Nevertheless, a general overview of the findings reinforces the initial statement. The analysis of cohesion also gave rise to interesting findings. Although machine translations seem to contain more explicit cohesive cues, such as connectives, human translations are more cohesive at deeper level as evidenced by Referential cohesion and Situation model. Moreover, Latent Semantic Analysis revealed that HT would probably require less inference processing from the reader in relation to MT due to the proportion of given/new information. Correlation analysis between the two versions and the source text would reveal whether the differences are to be attributed to interventions on the part of the professional translator that are absent in the automatic output. Li et al. (2014) support that variation in the cohesion structure of the human translated text is a sign of flexibility while machine translations do not deviate from the original because their approach is more ‘mechanic’.

Readability metrics did not provide any significant results. This was expected because the individual measures upon which they depend were conflicting. Some characteristics of readable texts were most prevalent in the HT sample and others in the MT sample. For instance, longer words were observed in human-produced translations while longer sentences in machine-generated output. Considering that Flesch Reading Ease and Flesch Kincaid depend on word and sentence length, it is not surprising that readability scores for the two samples were balanced. An implication of these findings is that traditional readability metrics are probably not sufficient to account for subtle differences.

Although some variation in the corpora was certainly observed, the distinction between the two translation types is not that striking, especially if one considers the total number of text characteristics examined in this research (N=121). Significant differences were found in approximately 30% of the indices. A quick look at the texts would confirm that machine and human translations in the subgenre in question are at least semantically close. This is an excerpt from a medium sized text included in the corpus.

HT	*MT*
The Slovenian Ambassador referred to the support her country showed for Greece during the economic crisis,	The Slovenian Ambassador referred to her country's support for Greece during the economic crisis, she stressed that

stressed that Slovenia stands at Greece's side in handling the refugee/migration crisis, and referred to individual sectors in which the two country's collaborate, including food safety and water resource management. The Deputy Minister referred to the cooperation between the two countries, which is founded on historical ties and the two countries' shared course as allies and partners. He also expressed interest in further development of the two countries' cooperation in the culture and education sector, mentioning the potential for capitalizing on the common heritage of Ioannis Kapodistrias, the first Governor of the independent Greek state, whose family lineage traces back to the city of Koper in Slovenia.	Slovenia is on the side of Greece to address the refugee / immigration crisis and referred to specific areas of cooperation between the two countries, such as food safety, water management, etc. For his part, the State Secretary referred to the cooperation of the two countries, which is based on historical ties and the common course of the two countries as allies and partners. He also expressed interest in the further development of the cooperation of the two countries in the cultural and cultural field, indicating the possibility of exploiting the common heritage of John Capodistrias, the first governor of the free Greek state, whose deeper family background is drawn from the city of Koper Slovenia.

Although the human-translated text is certainly more fluent, the two versions share a considerable number of n-grams. A lexical similarity metric would probably assign a high score to the automatic output; however, the features of translationese in the second sample would most probably be perceived by the reader, especially when the two versions are read contrastively. Differences in vocabulary and syntax greatly affect fluency in this example, but these elements are disregarded by non-linguistic metrics as subtle. That is the reason why they have been criticized for their limited scope and bias towards superficial aspects (Felice and Specia 2012).

As mentioned in the results section, the findings regarding greater lexical diversity in MT stimulated questions and were accepted with reluctance. This is the reason why I further investigated if the number of types including in the MT texts could be attributed to some unwanted effect of the source language, Greek, that happens to be morphologically rich. Follow-up analysis using Sketch Engine[10]

[10] Kilgarriff et al. (2014). Available online at https://www.sketchengine.eu/

revealed that the machine translator failed to transfer the Greek names in their correct form that involves dropping the inflectional suffixes existent in the source text. For instance, the surnames of Greek Ministers “Quick” and “Katrougalos” occur 248 times in HT in proper formulation, while in MT they appear correctly in only 47 of the instances. The rest occurrences exhibit irregularities, pertaining to improper transliteration (e.g.“Koutougallos”, “Katoigalos” “Chik”, “Kouik”), unnecessary matching to Latinized forms (“George” instead of “Giorgos”) and unneeded Greek inflections (“Kotzia, Katrougallou”). These non-words were probably classified as separate word types, while they did not actually contribute to lexical richness. This, along with the slight variation in word tokens, could explain the oddness of results in indexes such as TTR and Hapax Legomena. The validity of type and token-based metrics as indicators of lexical richness may be compromised for the application in question, unless non-words are filtered.

Table 4: SketchEngine, Word List: Comparison, correct forms: *The word list contains information on frequent proper names and their occurrence in the corpus e.g. the surname 'Quick' appears 104 times in the HT corpus in its correct form, but only once in MT.*

	Human translations		*Machine translations*		
lc	frequency	frequency/mill	frequency	frequency/mill	Score
quick	104	1968.9	1	19.7	17.3
giorgos	77	1457.8	3	59.1	9.8
terens	45	851.9	0	0.0	9.5
programme	20	378.6	0	0.0	4.8
n't	19	359.7	0	0.0	4.6
meets	32	605.8	3	59.1	4.4
luncheon	16	302.9	0	0.0	4.0
fyrom	27	511.2	3	59.1	3.8
greece's	41	776.2	7	137.9	3.7
spokesperson	18	340.8	1	19.7	3.7
underscored	13	246.1	0	0.0	3.5
stated	20	378.6	2	39.4	3.4
moreover	14	265.0	1	19.7	3.0
attends	14	265.0	1	19.7	3.0
trilateral	20	378.6	3	59.1	3.0
defence	10	189.3	0	0.0	2.9
affairs	650	12305.7	218	4294.1	2.8
katrougalos	144	2726.2	46	906.1	2.8
regarding	44	833.0	12	236.4	2.8
u.s.	9	170.4	0	0.0	2.7
countries'	9	170.4	0	0.0	2.7
ioannis	36	681.5	10	197.0	2.6
marking	11	208.3	1	19.7	2.6
being	40	757.3	12	236.4	2.5
yennimatas	8	151.5	0	0.0	2.5

Table 5: SketchEngine - Word List: Comparison, incorrect forms: *Ill-formed names (e.g. Kotzia, Katrougalou, Chik, Koutougallos, Bolari etc) exhibit high frequencies in the MT sample*

	Machine translations		*Human translations*		
lc	frequency	frequency/mill	frequency	frequency/mill	Score
kotzia	57	1122.8	0	0.0	12.2
fm	71	1398.5	3	56.8	9.6
co-operation	33	650.0	0	0.0	7.5
katrougalou	30	590.9	0	0.0	6.9
george	47	925.8	3	56.8	6.5
chik	20	394.0	0	0.0	4.9
shik	17	334.9	0	0.0	4.3
lunch	16	315.2	0	0.0	4.2
koutougallos	16	315.2	0	0.0	4.2
kouik	16	315.2	0	0.0	4.2
bolari	16	315.2	0	0.0	4.2
tripartite	15	295.5	0	0.0	4.0
g	45	886.4	8	151.5	3.9
terence	48	945.5	9	170.4	3.9
particular	36	709.1	8	151.5	3.2
participation	57	1122.8	15	284.0	3.2
spokesman	11	216.7	0	0.0	3.2
katoigalos	11	216.7	0	0.0	3.2
program	20	394.0	3	56.8	3.2
said	43	847.0	11	208.3	3.1
messrs.	13	256.1	1	18.9	3.0
sidelines	10	197.0	0	0.0	3.0
chick	10	197.0	0	0.0	3.0

The findings of the present research only concern the specific subgenre and the sample in question. The relatively good performance of Google translate may not be confirmed if a more heterogenous sample had been used. Institutional translations, as Schäffner et al. (2014: 2) state, follow a specific standardized form. Regarding the present sample, it was consistent in vocabulary, syntax and style; the content and subject-matter of most text samples was quite similar and the terminology, the formulation especially in the headings (see Appendix) and the jargon were repetitive. Moreover, in such contexts, the authors are constrained by institutional procedures and strict guidelines because "the 'voice' of the institution is the one to be heard" (Schäffner et al. 2014: 2). Google translate is trained on large-scale data sets and currently employs Neural Machine Translation systems (Wu et al. 2016), thus its good performance on literal and homogenous text samples is not surprising. However, it would be interesting to investigate whether Google's translation would also cope well with literary texts of multiple authors, figurative language and ambiguous content.

To address the second research question, classification techniques were used to investigate whether an Artificial Neural Network could correctly predict if a text was human or machine-translated. The model achieved an accuracy of about 82% for the testing sample. In line with Carter and Inkpen (2012), these results support that the traits of machine translation are machine-learnable and detectable. Although machine learning techniques seem to be effective in distinguishing between HT and MT, further investigation is needed to determine whether those features that determine the network are also good predictors of translation quality. The present results were not very conclusive as to which linguistic properties were the most influential in the classification task. All categories seem to contain at least some informative components.

Chapter 6: Conclusion

This study presented a quantitative comparison of human and machine translated institutional texts in terms of a broad range of text properties pertinent to lexical richness, syntactic and discourse structure, readability and shallow linguistic aspects. Linguistic features showed remarkable discriminative potential. First, independent-samples t-tests revealed that HT and MT writing differs in terms of descriptive characteristics, sentence length, syntactic structure, cohesion and vocabulary. Second, a text classification experiment showed that these differences are detectable using machine learning techniques; nonetheless the distinction between the two translation types was not always clear cut.

The aim of this comparative analysis was to bring up some limitations of existent MT systems. Google translate was selected as a widespread service that employs state-of-the-art statistical methods. The output generally met the specifications of formality, although the vocabulary of human translations seems to be more suited for the given context. Regarding comprehensibility, there are some indicators that potentially point to problematic areas, such as the greater syntactic complexity in MT and the comparatively low deep cohesion, which is likely to impose unnecessary cognitive demands on the reader. This effect is certainly undesirable taking into account the important informative value of the texts.

Future research plans involve the comparison of translated outputs extracted from different MT services, each one ideally representing a different paradigm, e.g.

statistical vs rule-based systems. Another significant addition that would definitely contribute to the discussion is the analysis of the Greek source text alongside the translated output. Unfortunately, the Greek text was not exploited properly because the corresponding text analysis tools used for the English sample were not available in Greek. Under different circumstances, correlation analyses between the original and the translations would allow for a more comprehensive interpretation of results. Another limitation of the study concerns the ambiguity of results obtained through sensitivity analysis of the variables used in the classification task. As mentioned in the results, a principal component analysis, similar to the one described in Li et al (2014) would probably provide more valuable insight on the effect of grouped variables in the classification model.

Despite the limitations, this study gave rise to interesting and at times, unexpected findings. As specified in the introduction, this comparative analysis aimed at identifying some areas of variation between HT and MT with a view to exposing some weaknesses of automatic translations. This objective was fulfilled using a set of linguistic features, thus validating their usefulness in evaluation tasks.

References

Ahrenberg, L. (2017). Comparing machine translation and human translation: A case study. *In Proceedings of The First Workshop on Human-Informed Translation and Interpreting Technology* (pp. 21–28). Wolverhampton, UK.

Aiken, M. and Balan, S. (2011). An analysis of Google Translate accuracy. *Translation Journal*, 16 (2) http://translationjournal.net/journal/56google.htm.

Aiken, M., Ghosh, K., Wee, J., and Vanjani, M. (2009). An evaluation of the accuracy of online translation systems. *Communications of the IIMA*: *9*(4) http://scholarworks.lib.csusb.edu/ciima/vol9/iss4/6

ALPAC. (1966). *Language and Machines: Computers* in *Translation and Linguistics,* Report by the Automatic Language Processing Advisory Committee (ALPAC), Publication 1416. National Academy of Sciences, Washington, D.C.

Alotaibi, S., Al-Yahya, M., A-Khalifa, H., Alageel, S., and Abanmy, N. (2016). Readability of Arabic medicine information leaflets: A Machine Learning approach. *Symposium on Data Mining Applications* (pp. 122-126). Riyadh, Saudi Arabia.

Amigó, E., Giménez, J., Gonzalo, J., and Verdejo, F. (2009). The contribution of linguistic features to automatic machine translation evaluation. *In Proceedings of the 47th Annual Meeting of the ACL and the 4th IJCNLP of the AFNLP* (pp. 306-314). Suntec, Singapore.

Anthony, L. (2014). AntConc (Version 3.4.4) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.laurenceanthony.net/software

Baayen, R. H., Piepenbrock, R., and Gulikers, L. (1995). The CELEX lexical database (CD-ROM). *Linguistic Data Consortium* Philadelphia: University of Pennsylvania.

Banerjee, S., and Lavie, A. (2005). METEOR: an automatic metric for MT Evaluation with improved correlation with human judgments. *In Proceedings of the ACL workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization* (pp. 65-72). Stroudsburg: Association for Computational Linguistics. http://aclweb.org/anthology-new/W/W05/W05-0909.pdf.

Baroni, M., Bernardini, S. (2006). A new approach to the study of translationese: Machine-learning the difference between original and translated text. *Literary and Linguistic Computing*, *21*(3): 259-274.

Béchara, H., Parra Escartín, C., Orăsan, C. and Specia, L. (2016). Semantic Textual Similarity in Quality Estimation. *Baltic J. Modern Computing*, *4*(2): 256-68.

Berkenkotter, C. and Huckin, T. N. (1995). *Genre Knowledge in Disciplinary Communication: Cognition/culture/power.* Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Biel, Ł. (2017). Quality in institutional EU translation: Parameters, policies and practices. In T. Svoboda, Ł. Biel & K. Łoboda (Eds.), *Quality aspects in institutional translation* (pp.31–57). Berlin: Language Science Press. DOI:10.5281/zenodo.1048183

Callison-Burch, C., Osborne, M. and Koehn, P. (2006). Re-evaluating the role of BLEU in machine translation research. *In Proceedings of 11th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 249-256) Trento.

Carroll, J., Minnen, G., Canning, Y., Devlin, S., and Tait, J. (1998). Practical Simplification of English Newspaper Text to Assist Aphasic Readers. *In Proceedings of AAAI-98 Workshop on Integrating Artificial Intelligence and Assistive Technology* (pp.7-10) Madison, Wisconsin.

Carter D., and Inkpen, D. (2012). Searching for poor quality machine translated text: Learning the difference between human writing and machine translations. In Proceedings of the 25th Canadian Conference on Advances in Artificial Intelligence (pp. 49-60) Toronto, Canada.

Čech, R., and Kubát, M. (2016). Text length and the thematic concentration of text. *Mathematical Linguistics 2*, 5-13.

Chatzitheodorou, K. and Chatzistamatis, S. (2013). COSTA MT Evaluation tool: An open toolkit for human machine translation evaluation. *The Prague Bulletin of Mathematical Linguistics*, 83-89. DOI: 10.2478/pralin-2013-0014

Chen, X., Acosta, S., Barry, A.E. (2016). Evaluating the Accuracy of Google Translate for Diabetes Education Material. *JMIR Diabetes*, 1(1): 1-11.

Ciobanu, A. M., and Dinu, L. P. (2014). A Quantitative Insight into the Impact of Translation on Readability. *In Proceedings of the 3rd Workshop on Predicting and Improving Text Readability for Target Reader Populations* (pp. 104–113). Gothenburg, Sweden.

Coleman, M., and Liau, T. L. (1975). A computer readability formula designed for machine scoring. *Journal of Applied Psychology, 60*(2): 283-284.

Coltheart, M. (1981). The MRC Psycholinguistic Database. *Quarterly Journal of Experimental Psychology*, 33, 497–505.

Comelles, E., Atserias, J., Arranz, V., and Castellón, I. (2012). VERTa: Linguistic features in MT evaluation. *LREC*, 3944-3950.

Culy, C., Riehemann, S.Z. (2003). The Limits of N-gram Translation Evaluation Metrics. In *MT Summit IX* (pp. 1-7). New Orleans.

Denkowski, M. and Lavie, A. (2011). Meteor 1.3: Automatic metric for reliable optimization and evaluation of machine translation systems. In *Proceedings of the Sixth Workshop on Statistical Machine Translation* (pp. 85–91). Edinburgh.

Doddington, G. (2002). Automatic evaluation of machine translation quality using n-gram co-occurrence statistics. In *Proceedings of Human Language Technology Conference*. (pp.138-145). San Diego, California.

Dorr, B., Olive, J., McCary, J., and Christianson, C. (2011). Machine translation Evaluation (Chapter 5). In J., Olive, C., Christianson, and J., McCary (Eds.), *Handbook of Natural Language Processing and Machine Translation: DARPA Global Autonomous Language Exploitation,* (pp. 745-818). New York: Springer.

Eisele, A., and Lavecchia, C. (2011). Using statistical machine translation for computer-aided translation at the European Commission. In *Proceedings of the Third Joint EM+/CNGL Workshop "Bringing MT to the Research Meets Translators" (JEC '11)* (pp. 3-12). Luxembourg.

European Parliament. (n.d.) The profession of translator in the European Parliament Retrieved December 27, 2018 from http://www.europarl.europa.eu/multilingualism/trade_of_translator_en.htm

European Parliament. (2008). European Parliament - never lost in translation. Retrieved December 26, 2018 from http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT%20IM-PRESS%2020071017FCS11816%200%20DOC%20XML%20V0//EN

Federmann, C. (2014). How can we measure machine translation quality? *Tralogy* [En ligne], Session 5 - Quality in Translation / La qualité en traduction, Tralogy I, mis à jour le http://lodel.irevues.inist.fr/tralogy/index.php?id=76

Felice, M. and Specia, L. (2012). Linguistic Features for Quality Estimation. In *Proceedings of the 7th Workshop on Statistical Machine Translatio* (pp. 96-103) Stroudsburg: Association for Computational Linguistics. http://www.aclweb.org/anthology/W/W12/W12-3110.pdf.

Fellbaum, C. (Ed.). (1998). *WordNet: An Electronic Lexical Database* [CD-ROM]. Cambridge: MIT Press.

Feng, L., Elhadad, N., Huenerfauth, M. (2009). Cognitively motivated features for readability assessment. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, (pp. 229-237) Stroudsburg, PA, USA.

Finch, A., Hwang, Y.S. and Sumita, E. (2005). Using machine translation evaluation techniques to determine sentence-level semantic equivalence. In *Proceedings of the Third International Workshop on Paraphrasing (IWP2005)* (pp. 17-24) http://aclweb.org/anthology/I05-5003

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, 32 (3): 221-233.

François, T. and Fairon, C. (2012). An AI readability formula for French as a foreign language. *In Proceedings of the 2012 Conference on Empirical Methods in Natural Language Processing*, (pp. 466-477) Jeju.

François, T. and Miltsakaki, E. (2012). Do NLP and machine learning improve traditional readability formulas? *In Proceedings of the First Workshop on Predicting and improving text readability for target reader populations*, (pp. 49-57) Montréal.

Freitas, C. and Liu, Y. (2017). Exploring the Differences between Human and Machine Translation *WWU Honors Program Senior Projects*. 61 https://cedar.wwu.edu/wwu_honors/61

Gimenez, J. (2008). Empirical machine translation and its evaluation. PhD thesis, Departament de Llenguatges i Sistemes Informàtics, Universitat Politècnica de Catalunya, Barcelona.

Giménez, J. and Màrquez, L. (2007). Linguistic Features for Automatic Evaluation of Heterogenous MT Systems. In *Proceedings of the Second Workshop on Statistical Machine Translation (WMT'07),* (pp. 256-264) Prague, Czech Republic. http://aclweb.org/anthology/W07-0738

Graesser, A. C., McNamara, D. S., and Kulikowich, J. (2011). Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher 40*, 223–234.

Graesser, A. C., McNamara D. S., Louwerse, M. and Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers 2004*, *36* (2), 193-202.

Graham Y., Mathur, N. and Baldwin, T. (2014). Randomized significance tests in machine translation. In *Proceedings of the Ninth Workshop on Statistical Machine Translation,* (pp. 266–274) Baltimore, Maryland USA, Association for Computational Linguistics.

Hempelmann, C. F., Dufty, D., McCarthy, P., Graesser, A. C., Cai, Z., and McNamara, D. S. (2005). Using LSA to automatically identify givenness and newness of noun-phrases in written discourse. In *Proceedings of the 27th Annual Meeting of the Cognitive Science Society* (pp. 941–946). Mahwah, NJ.

Han, L., Wong, D. and Chao, L. (2016). Machine Translation Evaluation: A Survey. *arXiv*:1605.04515 https://arxiv.org/abs/1605.04515

Heylighen, F. and Dewaele J.M. (1999). Formality of Language: definition, measurement and behavioral determinants. Internal Report Center "Leo Apostel", Free University of Brussels.

Hobbs, J. (1979). Coherence and coreference. *Cognitive Science*, 3, 67–90.

Hovy, E., King, M. and Popescu-Belis A. (2002a). An introduction to MT evaluation. *In Proceedings of the Third International Conference on Language Resources and Evaluation LREC-2002,* (pp.1-7) Las Palmas, Canary Islands, Spain.

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

Joty, S., Guzman, F., Marquez, L. and Nakov, P. (2017). Discourse Structure in Machine Translation Evaluation. *Computational Linguistics 43*(4): 683-722.

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J. Rychlý, P. and Suchomel, V. (2014). The Sketch Engine: ten years on. *Lexicography*, 1, 7-36.

Kincaid, J. P., Fishburne, R. P. Jr., Rogers, R. L., and Chissom, B. S. (1975). Derivation Of New Readability Formulas (Automated Readability Index, Fog Count And Flesch Reading Ease Formula) For Navy Enlisted Personnel. CNTECHTRA Research Branch Report, 8-75.

Kintsch, W. (1998). *Comprehension: A Paradigm for Cognition.* Cambridge: Cambridge University Press.

Koehn. P. (2004). Statistical significance tests for machine translation evaluation. In *Proceedings of Empirical Methods in Natural Language Processing*, (pp. 388–395) Barcelona.

Koehn, P. (2005). Europarl: A Parallel Corpus for Statistical Machine Translation. *In Proceedings of the Tenth Machine Translation Summit,* (pp. 79-86), Phuket, Thailand: AAMT.

Kubát, M., Matlach, V., Čech, R. (2014). *QUITA - Quantitative Index text Analyzer.* Lüdensheid: RAM.

Kurokawa, D., Goutte C., Isabelle, P. (2009). Automatic detection of translated text and its impact on machine translation. In *Proceedings of the twelfth Machine Translation Summit,* (pp.81-88) Ottawa, Ontario, Canada.

Labov, W. (1972). *Sociolinguistic Patterns.* Philadelphia: University of Philadelphia Press.

Landauer, T., McNamara, D. S., Dennis, S., and Kintsch, W. (2007). *Handbook of Latent Semantic Analysis*. Mahwah, NJ: Erlbaum

Lavie, A., and Denkowski, M.J. (2009). The METEOR metric for automatic evaluation of machine translation. *Machine Translation*, 23, 105-115 DOI: 10.1007/s10590-009-9059-4

Lavie, A., Sagae, K., and Jayaraman, S. (2004). The significance of recall in automatic metrics for MT evaluation. In *Proceedings of the 6th conference of the association for machine translation in the Americas, (AMTA-2004)* (pp. 134–143) Washington, DC.

Levenshtein, V. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady,* 10(8):707-710.

Li, H., Graesser, A.C., and Cai, Z. (2014). Comparison of Google translation with human translation. *In Proceedings of the Twenty-Seventh International*

Florida Artificial Intelligence Research Society Conference, (pp. 190-195) Florida.

Lin, C., and Och, F.J. (2004). Automatic evaluation of machine translation quality using longest common subsequence and skip-bigram statistics. In *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics (ACL)*, (pp.605-612) Barcelona: Association for Computational Linguistics.

Liontou, T. (2013). The effect of text and reader variables on reading comprehension: the case of the Greek State Certificate of English Language Proficiency Exams (KPG) - A New Text Difficulty Index for Automatic Text Classification. PhD thesis. Faculty of English Language and Literature. National and Kapodistrian University of Athens.

Mandravickaite, J. and Krilavičius, T. (2018). Quantitative Analysis of Textual Genres: Comparison of English and Lithuanian. In *Proceedings of the International Conference on Information Technologies,* (pp. 61-67) Kaunas, Lithuania.

Maney, T., Sibert, L., Perzanowski, D., Gupta, K. and Schmidt-Nielsen, A. (2012). Toward determining the comprehensibility of machine translations. In *Proceedings of the First Workshop on Predicting and Improving Text Readability for target reader populations*. (pp.1-7) Montrèal, Canada: Association for Computational Linguistics.

Mann, W. and Thompson, S. (1988). Rhetorical structure theory: Toward a functional theory of text organization. *Text*, 8(3): 243–281.

Mauser, A., Hasan, S. and Ney, H. (2008). Automatic Evaluation Measures for Statistical Machine Translation system optimization. In *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC),* Marakech, Morocco.

McCarthy, P.M., Dufty, D., Hempelmann, C., Cai, Z., Graesser, A. C., and McNamara, D. S. (2012). Newness and givenness of information: Automated identification in written discourse. In P. M. McCarthy and C. Boonthum (Eds.), *Applied natural language processing and content analysis: Identification, investigation, and resolution* (pp. 457–478). Hershey, PA: IGI Global.

McCarthy, P. M., and Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, 42, 381–392.

McNamara, D., Graesser, A. C., McCarthy, P. M., and Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. New York: Cambridge University Press.

Melby, A. K. (2005). QT21: A New era for translators and the computer. In *Proceedings of the 37th Conference Translating and the Computer*, (pp. 1–11) London, UK.

Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. J. (1990). Introduction to WordNet: An on-line lexical database. *Journal of Lexicography*, 3, 235–244.

Nadhianti, M. (2016). An Analysis Of Accuracy Level Of Google Translate In English - Bahasa Indonesia And Bahasa Indonesia-English Translations. *Quill* 5(4): 296-303.

Nießen, S., Och, F. J., Leusch, G., & Ney, H. (2000). An Evaluation Tool for Machine Translation: Fast Evaluation for MT Research. In *Proceedings of the 2nd*

International Conference on Language Resources and Evaluation (LREC), (pp. 39-45). Athens, Greece.

Nord, C. (2003). Transparenz der Korrektur. In Handbuch Translation, pp. 384–387. Stauffenburg.

Osmałek A. (2014). „Dlaczego tłumaczenia maszynowe są nieprzydatne?”. Retrieved on: 9 January 2015 http://blog.supertresc.pl/dlaczego-tlumaczenia-maszynowe-sa-nieprzydatne#.VK_7HnuoMnI.

Padó, S., Galley, M. Jurafsky, D. and Manning, C. (2009). Robust Machine Translation Evaluation with Entailment Features. In *Proceedings of the 47th Annual Meeting of the ACL and the 4th IJCNLP of the AFNLP,* (pp. 297–305). Suntec, Singapore.

Papineni, K., Roukos, S., Ward, T., and Zhu, W. J. (2002). BLEU: A method for Automatic Evaluation of Machine Translation. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics,* (pp.311-318). Philadelphia, PA.

Pennebaker, J. W., Booth, R. J., and Francis, M. E. (2007). Linguistic Inquiry and Word Count: LIWC [Computer software]. TX: LIWC.net, Austin.

Poiret and Liu (2017). Mastering the measurement of text’s frequency structure: an investigation on Lambda’s reliability. *Glottometrics 37*: 82-100.

Popescu, I. (2009). *Word Frequency Studies.* Berlin/New York: Mouton de Gruyter.

Popescu, I. Čech, R. and Altmann, G. (2011). *The Lambda-structure of Texts.* Lüdenscheid: RAM-Verlag.

Puchała-Ladzińska, K. (2016). Machine translation: a threat or an opportunity for human translators? *Studia Anglica Resoviensia* 13: 89-98.

Scarton, C. and Specia, L. (2015). A Quantitative analysis of discourse phenomena in machine translation. *Discours* 16, 3-28.

Schäffner, C., Tcaciuc, L. S. and Tesseur, W. (2014). Translation practices in political institutions: a comparison of national, supranational, and nongovernmental organisations. *Perspectives: Studies in Translatology*, 22(4):1-27. http://centaur.reading.ac.uk/44097/

Schiaffino, R. and Zearo. F. (2005). Translation Quality Measurement in Practice. In *Proceedings of the 46th ATA Conference,* Seattle, Washington.

Shankland, S. (2013). Google Translate Now Serves 200 Million People Daily. *CNET*. Retrieved from http://news.cnet.com/8301-1023_3-57585143-93/googletranslate-now-serves-200-million-people-daily/

Snover, M. et al. (2006). A Study of Translation Edit Rate with Targeted Human Annotation. In *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas (AMTA06),* (pp. 223-231) Cambridge, Massachusetts. http://www.mt-archive.info/AMTA-2006-Snover.pdf

Steinberger, R., Ebrahim, M., Poulis, A., Carrasco-Benitez, M., Schluter, P., Przybyszewski, M. and Gilbro, S. (2014). An overview of the European Union’s highly multilingual parallel corpora *Lang Resources & Evaluation*. 48, 679–707. DOI 10.1007/s10579-014-9277-0

Su, K.Y., Wu, M. W. and Chang. J.S. (1992). A New Quantitative Quality Measure for Machine Translation Systems. In *Proceedings of COLING92* Nantes, France.

Sun, Y. (2018). Study on the comprehensive evaluation method of machine translation quality. *MATEC Web Conf*erence 173: 1-5.

Sun, Y., and Zhou, M. (2016). Study on evaluation method of machine translation quality based on questionnaires and data analysis. In *Proceedings of the 3rd International Conference on Materials Engineering, Manufacturing Technology and Control (ICMEMTC 2016)* (pp. 1565-1570) Taiyuan, China.

Readability Formulas (n.d.) The Flesch Reading Ease Readability Formula. Retrieved January 31, 2019 from http://www.readabilityformulas.com/flesch-reading-ease-readability-formula.php

Tillmann, C., Vogel, S., Ney, H., Zubiaga, A. and Sawaf, H. (1997). Accelerated DP based search for statistical translation. In *Proceedings of the Fifth European Conf. on Speech Communication and Technology,* (pp. 2667–2670) Rhodos, Greece.

Tripathi, S. and Kansal, V. (2017). Using linguistic knowledge for machine translation evaluation with Hindi as a target language. *Computación y Sistemas*, 21(4): 717–724.

Trosborg, A. (1997). Translating Hybrid Political Texts. In A. Trosborg (Ed.), *Text typology and translation* (pp. 145-158). Amsterdam: John Benjamins.

Turovsky, B. (2017). Higher quality neural translations for a bunch more languages. Retrieved January 23, 2019 from Google's blog https://www.blog.google/products/translate/higher-quality-neural-translations-bunch-more-languages/

Ulitkin, I. (2013). Human Translation vs. Machine Translation: Rise of the machines. *Translation Journal* 17.

Vandepitte, S. (2017). Translation product quality: A conceptual analysis. In T. Svoboda, Ł. Biel and K. Łoboda (Eds.). *Quality aspects in institutional translation* (pp. 15–29) Berlin: Language Science Press. DOI:10.5281/zenodo.1048180

Vela, M., Schumann, A. K. and Wurm, A. (2014). Human Translation Evaluation and its Coverage by Automatic Scores. In *Proceedings of the LREC Workshop on Automatic and Manual Metrics for Operational Translation Evaluation (MTE)* (pp. 20-30) Reykjavik, Iceland.

Vilar, D., Jia X., D'Haro, L.F., and Ney, H. (2006). Error Analysis of Statistical Machine Translation Output. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC'06),* (pp. 697–702) Genoa, Italy.

White, J.S. (1995). Approaches to black Box MT evaluation. In *Proceedings of Machine Translation Summit V,* Luxembourg.

White, J.S. (2003). How to evaluate machine translation. (Chapter 13) In H. Somers (Ed.). *Computers and Translation: A translator's guide* (pp. 211-244). Amsterdam and Philadelphia: John Benjamins.

Wu, Y., Schuster, M., Chen, Z., Le, Q.V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., Klingner, J., Shah, A., Johnson, M., Liu, X., Kaiser, L., Gouws, S., Kato, Y., Kudo, T., Kazawa, H., Stevens, K., Kurian, G., Patil, N., Wang, W., Young, C., Smith, J., Riesa, J., Rudnick, A., Vinyals, O., Corrado, G.S., Hughes, M., and Dean, J. (2016). Google's neural machine translation system: Bridging the gap between human and machine translation. *CoRR*, abs/1609.08144 https://arxiv.org/abs/1609.08144

Xia, M., Kochmar, E., and Briscoe, E. J. (2016). Text Readability Assessment for Second Language Learners. *Proceedings of the 11th Workshop on Innovative Use of NLP for Building Educational Applications*, (pp.12-22). San Diego, California.

Zhou, M., Wang, B., Liu, S., Li, M., Zhang, D. and Zhao, T. (2008). Diagnostic evaluation of Machine Translation systems using automatically constructed linguistic check-points. In *Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008)*, (pp. 1121–1128). Manchester, UK.

Zipf, G. K. (1949). *Human Behavior and the Principle of Least Effort*. Reading, MA: Addison-Wesley.

Appendix

Table 1: Texts in the Greek and Human Translated Corpora (Selected and Discarded)

	Original (GR)	Human Translation (EN)	Date	Discarded texts
1.	Συμμετοχή Υπουργού Εξωτερικών, Ν. Κοτζιά, στην 3η Υπουργική Σύνοδο του Περιφερειακού Φόρουμ της Ένωσης για τη Μεσόγειο (Βαρκελώνη, 8.10.2018)	Minister of Foreign Affairs Nikos Kotzias to participate in the 3rd Ministerial Meeting of the Union for the Mediterranean Regional Forum (Barcelona, 8 October 2018)	Sunday, 7 October 2018	
2.	Ανακοίνωση Υπουργείου Εξωτερικών για την απώλεια του Αριστείδη Ανδρουλάκη		Monday, 1 October 2018	No English translation
3.	Ανακοίνωση Υπουργείου Εξωτερικών για το αποτέλεσμα του δημοψηφίσματος στην πρώην Γιουγκοσλαβική Δημοκρατία της Μακεδονίας	Announcement by the Ministry of Foreign Affairs on the result of the referendum in the former Yugoslav Republic of Macedonia	Sunday, 30 September 2018	
4.	Συμμετοχή ΥΠΕΞ, Ν. Κοτζιά, στις εργασίες της 73ης Συνόδου της Γενικής Συνέλευσης του ΟΗΕ (Νέα Υόρκη, 23-28.09.2018)	Minister of Foreign Affairs, N. Kotzias, to participate in the proceedings of the 73rd Session of the UN General Assembly (New York City, 23-28 September 2018)	Sunday, 23 September 2018	
5.	Ανακοίνωση Υπουργείου Εξωτερικών σχετικά με την Τρίτη Συνάντηση Κορυφής των ηγετών της κορεατικής χερσονήσου (Πιονγκγιάνγκ, 18-20.09.2018)	Ministry of Foreign Affairs Announcement on the Third Summit of Korean peninsula leaders (Pyongyang, 18-20/09/2018)	Thursday, 20 September 2018	
6.	Συνάντηση Υπουργού Εξωτερικών, Ν. Κοτζιά, με τον Υπουργό Εξωτερικών της Γερμανίας, H. Maas (Αθήνα, 20.09.2018)	Minister of Foreign Affairs Nikos Kotzias to meet with the Minister of Foreign Affairs of Germany, Heiko Maas (Athens, 20 September 2018)	Wednesday 19 September 2018	
7.	Συνάντηση ΥΠΕΞ, Ν. Κοτζιά, με τον Αν. Πρωθυπουργό και ΥΠΕΞ του Κατάρ, Mohammed bin Abdulrahman bin Jassim Al Thani (Αθήνα, 19.9.2018)	Minister of Foreign Affairs Nikos Kotzias to meet with the Deputy Prime Minister and Foreign Minister of Qatar, Mohammed bin Abdulrahman bin Jassim Al Thani (Athens, 19 September 2018)	Tuesday September 18, 2018	
8.	Ανακοίνωση του Υπουργού Εξωτερικών Ν. Κοτζιά	Statement of the Minister of Foreign Affairs, N. Kotzias	Sunday, 16 September 2018	
9.	Δήλωση Υπουργού Εξωτερικών, Ν. Κοτζιά, μετά το πέρας της	Foreign Minister N. Kotzias' statements following his meeting	Friday, 14	

	συνάντησής του με τον Μακαριώτατο Αρχιεπίσκοπο Κύπρου κ.κ. Χρυσόστομο Β' (Λευκωσία, 14.09.2018)	with His Beatitude Archbishop Chrysostomos of Cyprus (Nicosia, 14/09/2018)	September 2018	
10.	Ανακοίνωση Υπουργείου Εξωτερικών για τη συμπλήρωση 19 ετών από την απώλεια του Γιάννου Κρανιδιώτη	Ministry of Foreign Affairs announcement on 19th anniversary of the passing of Giannos Kranidiotis	Friday, 14 September 2018	
11.	Επίσκεψη Υπουργού Εξωτερικών, Ν. Κοτζιά, σε Λευκωσία και Κάιρο (14.09.2018)	Minister of Foreign Affairs Nikos Kotzias to visit Nicosia and Cairo (14 September 2018)	Thursday, 13 September 2018	
12.	Συμμετοχή Υπουργού Εξωτερικών, Ν. Κοτζιά, στην τριμερή συνάντηση Υπουργών Εξωτερικών Ελλάδος-Κύπρου-Ισραήλ (Ιεροσόλυμα, 13.09.2018)	Minister of Foreign Affairs Nikos Kotzias to attend trilateral meeting of the Ministers of Foreign Affairs of Greece, Cyprus and Israel (Jerusalem, 13 September 2018)	Wednesday 12 September 2018	
13.	Συνάντηση Υπουργού Εξωτερικών, Ν. Κοτζιά, με τον Υπουργό Εξωτερικών της Κυπριακής Δημοκρατίας, Ν. Χριστοδουλίδη (ΥΠΕΞ, 12.09.2018)	Minister of Foreign Affairs Nikos Kotzias to meet with the Minister of Foreign Affairs of the Republic of Cyprus, Nikos Christodoulides (Foreign Ministry, 12 September 2018)	Tuesday, 11 September 2018	
14.	Σύγκληση Εθνικού Συμβουλίου Εξωτερικής Πολιτικής (ΥΠΕΞ, 12.09.2018)	National Council on Foreign Policy to convene (Ministry of Foreign Affairs, 12 September 2018)	Tuesday, 11 September 2018	
15.	Συνάντηση Υπουργού Εξωτερικών, Ν. Κοτζιά, με την ειδική απεσταλμένη του Γ.Γ. ΗΕ για το Κυπριακό, Jane Holl Lute (Αθήνα, 11.09.2018)	Minister of Foreign Affairs Nikos Kotzias to meet with the UN Secretary-General's special envoy on Cyprus, Jane Holl Lute (Athens, 11 September 2018)	Monday, 10 September 2018	
16.	*Ομιλία του Υπουργού Εξωτερικών, Νίκου Κοτζιά, στο Νορβηγικό Ινστιτούτο Διεθνών Σχέσεων (NUPI), (Όσλο, 10/09/2018)*	Speech of the Minister of Foreign Affairs, Nikos Kotzias, at the Norwegian Institute of International Affairs (NUPI), (Oslo, 10.09.2018)	Monday, 10 September 2018	The original language of the speech is unknown
17.	Επίσκεψη Υπουργού Εξωτερικών, Ν. Κοτζιά, στην Νορβηγία (Όσλο, 10.09.2018)	Minister of Foreign Affairs, N. Kotzias, to visit Norway (Oslo, 10 September 2018)	Sunday, 9 September 2018	
18.	Συνάντηση Υπουργού Εξωτερικών, Ν. Κοτζιά, με τον Υπουργό Εξωτερικών της Λετονίας, Edgars Rinkēvičs (Αθήνα, 07.09.2018)	Minister of Foreign Affairs Nikos Kotzias to meet with the Minister of Foreign Affairs of Latvia, Edgars Rinkēvičs (Athens, 7 September 2018)	Thursday, 6 September 2018	
19.	Συνάντηση Υπουργού Εξωτερικών, Ν. Κοτζιά, με τον Υπουργό Εξωτερικών και Ευρωπαϊκών Υποθέσεων της	Minister of Foreign Affairs Nikos Kotzias to meet with the French Minister for Europe and Foreign Affairs, Jean-Yves Le Drian	Wednesday 5 September 2018	

	Γαλλίας, Jean-Yves Le Drian (Αθήνα, 06.09.2018)	(Athens, 6 September 2018)		
20.	Ανακοίνωση του Υπουργού Εξωτερικών Ν. Κοτζιά	Announcement of the Minister of Foreign Affairs, Nikos Kotzias	Monday, 3 September 2018	
21.	Επίσκεψη Υπουργού Εξωτερικών, Ν. Κοτζιά, στη Σμύρνη (Σμύρνη, 04.09.2018)	Minister of Foreign Affairs, N. Kotzias, to visit Izmir (Izmir, 04.09.2018)	Monday, 3 September 2018	
22.	*Συνέντευξη Υπουργού Εξωτερικών, Ν. Κοτζιά, σε κινεζικό ειδησεογραφικό πρακτορείο XINHUA (Πεκίνο, 27.08.2018)*	Interview of Minister of Foreign Affairs, N. Kotzias, with XINHUA News Agency (Beijing, 27.8.2018)	Friday, 31 August 2018	The original language of the interview is unknown.
23.	Συμμετοχή Υπουργού Εξωτερικών, Ν. Κοτζιά, σε Dubrovnik Forum 2018 (Ντουμπρόβνικ, 31.08 - 01.09.2018)	Minister of Foreign Affairs, N. Kotzias, to attend Dubrovnik Forum 2018 (Dubrovnik, 31 August-1 September 2018)	Friday, 31 August 2018	
24.	*Δηλώσεις ΥΦΥΠΕΞ, Μ. Μπόλαρη, κατά την τελετή ανάληψης των καθηκόντων του (Αθήνα, 30.08.2018)*		Thursday, 30 August 2018	No English translation
25.	*Δηλώσεις απερχόμενου ΥΦΥΠΕΞ, Ι. Αμανατίδη, κατά τη τελετή παράδοσης καθηκόντων του (Αθήνα, 30.08.2018)*		Thursday, 30 August 2018	No English translation
26.	Παράδοση-ανάληψη καθηκόντων ΥΦΥΠΕΞ μεταξύ των κ.κ. Ι. Αμανατίδη και Μ. Μπόλαρη (ΥΠΕΞ, Βασιλίσσης Σοφίας 1, Μεγ. Αίθουσα Σοφιανοπούλου)	Ioannis Amanatidis to hand over Deputy Minister of Foreign Affairs portfolio to Markos Bolaris (MFA, 1 V. Sofias Ave., Sofianopoulos Hall)	Wednesday 29 August 2018	
27.	Συμμετοχή Υπουργού Εξωτερικών, Ν. Κοτζιά, στην Άτυπη Συνάντηση «Gymnich» των Υπουργών Εξωτερικών της ΕΕ (Βιέννη, 30-31.08.2018)	Minister of Foreign Affairs Nikos Kotzias to attend the Informal 'Gymnich' Meeting of EU Ministers of Foreign Affairs (Vienna, 30-31 August 2018)	Wednesday 29 August 2018	
28.	Επίσκεψη Υπουργού Εξωτερικών, Ν. Κοτζιά, στην Κίνα (Πεκίνο-Σαγκάη, 27-29.08.2018)	Minister of Foreign Affairs, N. Kotzias, to visit China (Beijing/Shanghai, 27-29.08.2018)	Sunday, 26 August 2018	
29.	Ανακοίνωση Υπουργείου Εξωτερικών για τον θάνατο του πρώην Γενικού Γραμματέα του ΟΗΕ, Κόφι Ανάν	Ministry of Foreign Affairs announcement on the death of former UN Secretary-General, Kofi Annan	Saturday, 18 August 2018	
30.	Δήλωση Υφυπουργού Εξωτερικών, Ι. Αμανατίδη, στην Παναγία Σουμελά Βερμίου (15.08.2018)	Statement of Deputy Minister of Foreign Affairs Ioannis Amanatidis at Panagia Soumela, Vermio (15 August 2018)	Wednesday 15 August 2018	

31.	Ανακοίνωση Υπουργείου Εξωτερικών για την απόφαση απελευθέρωσης των δύο Ελλήνων στρατιωτικών Α. Μητρετώδη και Δ. Κούκλατζη	Ministry of Foreign Affairs announcement on the release of the two members of the Greek Armed Forces, A. Mitretodis and D. Kouklatzis	Tuesday, 14 August 2018	
32.	Προτάσσοντας το εθνικό συμφέρον: Νηφάλια και σταθερά	Putting national interest first: Soberly and firmly	Friday, 10 August 2018	
33.	Ανακοίνωση Υπουργείου Εξωτερικών αναφορικά με τις πρόσφατες καταστρεπτικές πυρκαγιές στην Αττική	Ministry of Foreign Affairs announcement on the recent catastrophic wildfires in Attica	Wednesday 25 July 2018	
34.	Ανακοίνωση Υπουργείου Εξωτερικών για τις φονικές πυρκαγιές στην Αττική (24.07.2018)	Ministry of Foreign Affairs announcement on the deadly wildfires in Attica (24 July 2018)	Tuesday, 24 July 2018	
35.	*Συνέντευξη ΥΦΥΠΕΞ, Τέρενς Κουίκ, στον ρ/σ «Ραδιόφωνο 24/7» και το δημ/φο Βασίλη Σκουρή (23.07.2018)*		Monday, 23 July 2018	No English translation
36.	Ανακοίνωση Υπουργείου Εξωτερικών σχετικά με την απόφαση της Ι.Μ. Κύκκου να παραχωρήσει οικόπεδο έναντι της Μονής για ανέγερση Πρεσβείας και Πρεσβευτικής Κατοικίας της Ελλάδας	Ministry of Foreign Affairs announcement on the decision of Kykkos Monastery to grant a plot of land adjacent to the Monastery for construction of a Greek Embassy and Ambassadorial Residence	Monday, 23 July 2018	
37.	Δήλωση Υπουργού Εξωτερικών Ν. Κοτζιά για την 44η επέτειο της τουρκικής εισβολής στην Κύπρο	Statement of the Minister of Foreign Affairs, Nikos Kotzias, on the 44th anniversary of the Turkish invasion of Cyprus	Friday, 20 July 2018	
38.	Ανακοίνωση Υπουργείου Εξωτερικών αναφορικά με δηλώσεις της Εκπροσώπου του Ρωσικού Υπουργείου Εξωτερικών	Ministry of Foreign Affairs announcement on statements from the spokesperson of the Russian Ministry of Foreign Affairs	Wednesday 18 July 2018	
39.	Διοργάνωση επιστημονικής ημερίδας στο Υπουργείο Εξωτερικών με θέμα τις νομικές πτυχές της Συμφωνίας των Πρεσπών (ΥΠΕΞ, 19.07.2018)	Ministry of Foreign Affairs to host a scientific conference on the legal aspects of the Prespes Agreement (Ministry of Foreign Affairs, 19 July 2018)	Wednesday 18 July 2018	
40.	Ανακοίνωση Υπουργείου Εξωτερικών σχετικά με την τελετή ενταφιασμού οστών Ελλήνων στρατιωτών που σκοτώθηκαν στον Ελληνο-ιταλικό πόλεμο	Ministry of Foreign Affairs announcement on the burial ceremony for the remains of Greek soldiers who fell in the Greek-Italian war	Saturday, 14 July 2018	
41.	*Δήλωση Υπουργού Εξωτερικών, Ν. Κοτζιά, κατά την Συνέντευξη Τύπου μετά το πέρας της 2ης*	Statement of the Minister of Foreign Affairs, N. Kotzias, at the Press Conference following the	Saturday, 14 July 2018	Unknown source language

	Υπουργικής Διάσκεψης του Φόρουμ Αρχαίων Πολιτισμών (Λα Παζ, 13 Ιουλίου 2018)	2nd Ministerial Meeting of the Ancient Civilizations Forum (La Paz, 13.07.2018)		
42.	*Δηλώσεις Υπουργού Εξωτερικών, Ν. Κοτζιά, σε ΜΜΕ της Βολιβίας στο πλαίσιο της 2ης Υπουργικής Διάσκεψης του Φόρουμ των Αρχαίων Πολιτισμών (Λα Παζ, 13.07.2018)*	Statements of Minister of Foreign Affairs N. Kotzias to the Bolivian press at the 2nd Ministerial Conference of the Ancient Civilizations Forum (La Paz, 13 July 2018)	Friday, 13 July 2018	Unknown source language
43.	2η Υπουργική Σύνοδος του Φόρουμ των Αρχαίων Πολιτισμών (Λα Παζ, 13 Ιουλίου 2018)	2nd Ministerial Conference of the Ancient Civilizations Forum (La Paz, 13 July 2018)	Thursday, 12 July 2018	
44.	Χαιρετισμός Υφυπουργού Εξωτερικών, Ι. Αμανατίδη, στην 25η Επετειακή Γ.Σ. της Διακοινοβουλευτικής Συνέλευσης Ορθοδοξίας (Αθήνα, 25.06.2018)	Deputy Minister of Foreign Affairs I. Amanatidis' address to the 25th-Anniversary General Assembly of the Interparliamentary Assembly on Orthodoxy (Athens, 25 June 2018)	Monday, 25 June 2018	
45.	Συμμετοχή Υπουργού Εξωτερικών, Ν. Κοτζιά, στα Συμβούλια Εξωτερικών Υποθέσεων και Γενικών Υποθέσεων της Ευρωπαϊκής Ένωσης (Λουξεμβούργο, 25-26.06.2018)	Minister of Foreign Affairs Nikos Kotzias to attend meetings of the EU Foreign Affairs and General Affairs Councils (Luxembourg, 25-26 June 2018)	Sunday, 24 June 2018	
46.	Τρίτος γύρος συνομιλιών για την οριοθέτηση των θαλασσίων ζωνών μεταξύ Ελλάδας και Αλβανίας (Αθήνα, 22 Ιουνίου 2018)	Third round of talks on the delimitation of the maritime zones of Greece and Albania (Athens, 22 June 2018)	Saturday, 23 June 2018	
47.	Δήλωση Υπουργού Εξωτερικών, Ν. Κοτζιά, στο Αθηναϊκό-Μακεδονικό Πρακτορείο Ειδήσεων εν όψει της 3ης Υπουργικής Διάσκεψης της Ρόδου για την Ασφάλεια και τη Σταθερότητα	Statement of the Minister of Foreign Affairs, Nikos Kotzias, to the Athens-Macedonian News Agency ahead of the 3rd Ministerial Rhodes Conference for Security and Stability	Wednesday 20 June 2018	
48.	Τρίτη Διάσκεψη της Ρόδου για την Ασφάλεια και τη Σταθερότητα (Ρόδος, 21-22.06.2018)	Third Rhodes Conference for Security and Stability (Rhodes, 21-22 June 2018)	Wednesday 20 June 2018	
49.	Ανακοίνωση Υπουργείου Εξωτερικών για την υπογραφή συμφωνίας με την πρώην Γιουγκοσλαβική Δημοκρατία της Μακεδονίας (Πρέσπες,17.06.2018)	Ministry of Foreign Affairs announcement on the signing of the agreement with the former Yugoslav Republic of Macedonia (Lake Prespa, 17.06.2018)	Saturday, 16 June 2018	

50.	Ανακοίνωση Υπουργείου Εξωτερικών σχετικά με τη σύνοδο κορυφής των ηγετών των ΗΠΑ και Λ.Δ. της Κορέας (Σιγκαπούρη, 12.06.2018)	Ministry of Foreign Affairs announcement on the summit meeting between the leaders of the United States of America and the Democratic People’s Republic of Korea (Singapore, 12 June 2018)	Wednesday 13 June 2018	
51.	*Σύντομη Δήλωση του Υπουργού Εξωτερικών, Ν. Κοτζιά, κατά την έναρξη της συνάντησής του με τον Ομόλογό του της Ρωσικής Ομοσπονδίας, S. Lavrov (Μόσχα, 13.06.2018)*	Brief statement of the Minister of Foreign Affairs, Nikos Kotzias, at the opening of his meeting with his Russian Federation counterpart, Sergey Lavrov (Moscow, 13 June 2018)	Wednesday 13 June 2018	The original language of the statement is unknown.
52.	Επίσκεψη του Υπουργού Εξωτερικών, Ν. Κοτζιά, στη Ρωσική Ομοσπονδία (Μόσχα, 13.6.2018)	Minister of Foreign Affairs Nikos Kotzias to visit the Russian Federation (Moscow, 13 June 2018)	Tuesday, 12 June 2018	
53.	*Ανακοίνωση του Υπουργείου Εξωτερικών για την απώλεια της Βιργινίας Τσουδερού*		Monday, 11 June 2018	No English translation
54.	*Ανακοίνωση Υπουργείου Εξωτερικών αναφορικά με ανακοίνωση της Νέας Δημοκρατίας*		Thursday, 7 June 2018	No English translation
55.	Απάντηση Εκπροσώπου Υπουργείου Εξωτερικών, Αλέξανδρου Γεννηματά, σε ερώτηση δημοσιογράφου αναφορικά με τις δηλώσεις του Εκπροσώπου του τουρκικού Υπουργείου Εξωτερικών για το Ευρωπαϊκό Πρόγραμμα “Natura 2000”	Response of the spokesperson of the Ministry of Foreign Affairs, Alexandros Yennimatas, to a journalist’s question on statements from the spokesperson of the Turkish Ministry of Foreign Affairs regarding the European Programme “Natura 2000”	Wednesday 6 June 2018	
56.	Δήλωση Υπουργού Εξωτερικών, Ν. Κοτζιά, για τις εξελίξεις στη διαπραγμάτευση με την πΓΔΜ για το Ονοματολογικό (Βρυξέλλες, 28.05.2018)	Statement of the Minister of Foreign Affairs, N. Kotzias, regarding developments in the negotiations with fYROM on the name issue (Brussels, 28 May 2018)	Monday, 28 May 2018	
57.	Επίσκεψη του Υπουργού Εξωτερικών, Ν. Κοτζιά, στις ΗΠΑ (Ουάσιγκτων-Ν. Υόρκη, 21 - 25.5.2018)	Minister of Foreign Affairs Nikos Kotzias on visit to the United States of America (Washington, D.C.-New York City 21-25 May 2018)	Sunday, May 20, 2018	
58.	Δεύτερος γύρος συνομιλιών για την οριοθέτηση των θαλασσίων ζωνών Ελλάδος και Αλβανίας (Αθήνα, 15.05.2018)	Second round of talks on the delimitation of maritime zones between Greece and Albania (Athens, 15.05.2018)	Tuesday, 15 May 2018	
59.	Εκδήλωση για την παρουσίαση του στρατηγικού σχεδιασμού	Presentation of strategic planning and internationalization of the	Tuesday, 15 May	

	και την εξωστρέφεια της οικονομίας (ΥΠΕΞ, 16.05.2018)	economy (Ministry of Foreign Affairs, 16 May 2018)	2018	
60.	Ανακοίνωση Υπουργείου Εξωτερικών για την κλιμάκωση της κατάστασης στη Λωρίδα της Γάζας	Ministry of Foreign Affairs announcement on the escalation of the situation in the Gaza Strip	Monday, 14 May 2018	
61.	Συνάντηση Υπουργού Εξωτερικών, Ν. Κοτζιά, με τον ομόλογό του της πΓΔΜ, N. Dimitrov, παρουσία του Προσωπικού Απεσταλμένου του ΓΓ ΗΕ, Matthew Nimetz (Σούνιο, 12.05.2018)	Minister of Foreign Affairs, N. Kotzias, to meet with Minister of Foreign Affairs of fYROM, N. Dimitrov, in the presence of the UN Secretary General’s Personal Envoy, Matthew Nimetz (Sounion, 12.05.2018)	Friday, 11 May 2018	
62.	*Εκδήλωση για την ενίσχυση της εθελοντικής προσφοράς αιμοποιητικών κυττάρων (Υπουργείο Εξωτερικών, 14.05.2018)*		Friday, 11 May 2018	No English translation
63.	Δήλωση Υπουργού Εξωτερικών, Ν. Κοτζιά, κατά την άφιξή του στην 2η Υπουργική Συνάντηση των ΥΠΕΞ των χωρών Visegrad (Visegrad 4) και των Βαλκανικών κ-μ ΕΕ (Balkan 4) (V4+B4plus) στο Σούνιο (11.05.2018)	Statement of the Minister of Foreign Affairs, Nikos Kotzias, on arriving at the 2nd Ministerial Meeting of the Ministers of Foreign Affairs of the Visegrad countries (Visegrad 4) and the Balkan EU member states (Balkan 4) (V4+B4plus) in Sounion (11 May 2018)	Friday, 11 May 2018	The source language of the statement is unknown.
64.	2η Υπουργική Συνάντηση των Υπουργών Εξωτερικών της ομάδας Visegrad (Visegrad-4) και των Βαλκανικών κ-μ ΕΕ (Balkan-4) (Σούνιο, 11 Μαΐου, 2018)	2nd Meeting of the Ministers of Foreign Affairs of the Visegrad Group (Visegrad-4) and the Balkan EU member states (Balkan-4) (Sounion, 11 May 2018)	Thursday, 10 May 2018	
65.	Μήνυμα Υπουργού Εξωτερικών, Νίκου Κοτζιά, για την Ημέρα της Ευρώπης (Αθήνα, 09.05.2018)	Europe Day message from the Minister of Foreign Affairs, Nikos Kotzias (Athens, 9 May 2018)	Wednesday 9 May 2018	
66.	Επίσκεψη Υπουργού Εξωτερικών, Ν. Κοτζιά, στην Κύπρο (Λευκωσία, 07.05.2018)	Minister of Foreign Affairs Nikos Kotzias to visit Cyprus (Nicosia, 7 May 2018)	Sunday, 6 May 2018	
67.	3η Υπουργική Συνάντηση Ελλάδας, Αλβανίας, Βουλγαρίας και πρώην Γιουγκοσλαβικής Δημοκρατίας της Μακεδονίας (Θεσσαλονίκη, 3-4 Μαΐου,	3rd Ministerial Meeting between Greece, Albania, Bulgaria and the former Yugoslav Republic of Macedonia (Thessaloniki, 3-4 May 2018)	Thursday, 3 May 2018	

	2018)			
68.	Πρώτος γύρος συνομιλιών για την οριοθέτηση των θαλασσίων ζωνών Ελλάδος και Αλβανίας (Τίρανα, 30.04.2018)	First round of talks on the delimitation of maritime zones between Greece and Albania (Tirana, 30.04.2018)	Monday, 30 April 2018	
69.	Ανακοίνωση Υπουργείου Εξωτερικών αναφορικά με την σημερινή συνάντηση των ηγετών της Κορεατικής Χερσονήσου	Ministry of Foreign Affairs announcement on today's meeting between the leaders of the Korean Peninsula	Friday, 27 April 2018	
70.	Συμμετοχή Υπουργού Εξωτερικών, Ν. Κοτζιά, στη Σύνοδο Υπουργών Εξωτερικών του ΝΑΤΟ και συνάντησή του με τον Γενικό Γραμματέα του ΝΑΤΟ, Jens Stoltenberg (Βρυξέλλες, 26-27.4.2018)	Minister of Foreign Affairs, N. Kotzias, to attend the Meeting of NATO Foreign Ministers and meet with NATO Secretary General, J. Stoltenberg (Brussels, 26-27 April 2018)	Thursday, 26 April 2018	
71.	Συμμετοχή ΥΦΥΠΕΞ, Γ. Αμανατίδη, στη Σύνοδο Κορυφής Διαδικασίας Συνεργασίας ΝΑ Ευρώπης (SEECP SUMMIT) (Σλοβενία, 24.4.2018)	Deputy Minister of Foreign Affairs Ioannis Amanatidis participates in the South East European Cooperation Process (SEECP) Summit Meeting (Slovenia, 24 April 2018)	Tuesday, 24 April 2018	
72.	Ανακοίνωση Υπουργείου Εξωτερικών σχετικά με ατυχείς δηλώσεις Επιτρόπου Χαν	Ministry of Foreign Affairs announcement regarding regrettable statements from Commissioner Hahn	Tuesday, 24 April 2018	
73.	*Ανακοίνωση Υπουργείου Εξωτερικών για τη σημερινή ομόφωνη καταδικαστική απόφαση δικαστηρίου για την υπόθεση παράνομης κρατικής χρηματοδότησης ΜΚΟ για έργα απoναρκοθέτησης (24.4.2018)*		Tuesday, 24 April 2018	No English translation
74.	Συνάντηση Υπουργού Εξωτερικών, Ν. Κοτζιά, με τον ομόλογό του της πΓΔΜ, N. Dimitrov, παρουσία του Προσωπικού Απεσταλμένου του ΓΓ ΗΕ, Matthew Nimetz (Βιέννη, 25.04.2018)	Minister of Foreign Affairs, N. Kotzias, to meet with Minister of Foreign Affairs of fYROM, N. Dimitrov, in the presence of the UN Secretary General's Personal Envoy, Matthew Nimetz (Vienna, 25 April 2018)	Monday, April 23, 2018	
75.	Ο*μιλία Προέδρου της Δημοκρατίας, Π. Παυλόπουλου*, και Υπουργού Εξωτερικών, Ν. Κοτζιά, στην τελετή ορκωμοσίας της ΚΓ' Εκπαιδευτικής Σειράς Ακολούθων Πρεσβείας στο Υπουργείο Εξωτερικών (23.04.2018)	Statement of the Minister of Foreign Affairs, Nikos Kotzias, at the swearing-in ceremony of the 23rd Class of Ministry of Foreign Affairs Attachés (23 April 2018)	Monday, 23 April 2018	Deleted part – president's speech (omitted in the EN sample)

76.	Ανακοίνωση Υπουργείου Εξωτερικών για τη πρόκληση φθορών στην Πρεσβεία της Γαλλίας από μέλη αντιεξουσιαστικής ομάδας (Αθήνα, 22.4.2018)	Ministry of Foreign Affairs announcement on the damage caused to the French Embassy by members of an anarchist group (22 April 2018)	Sunday, 22 April 2018	
77.	*Ανακοίνωση Υπουργείου Εξωτερικών σχετικά με κακόηθες δημοσίευμα σε κυριακάτικη εφημερίδα*		Saturday, April 21, 2018	No English translation
78.	Τελετή ορκωμοσίας ΚΓ΄Εκπαιδευτικής Σειράς Ακολούθων στο Υπουργείο Εξωτερικών	Swearing-in ceremony of the 23rd Class of Attachés at the Ministry of Foreign Affairs	Friday, 20 April 2018	
79.	*Κλήση σε διαβούλευση για τα μέτρα συντονισμού και συνεργασίας προς διευκόλυνση της προξενικής προστασίας για μη αντιπροσωπευόμενους πολίτες της Ένωσης σε τρίτες χώρες*		Wednesday 18 April 2018	No English translation
80.	Ανακοίνωση Υπουργείου Εξωτερικών σχετικά με επανάληψη ανυπόστατων ισχυρισμών του τουρκικού Υπουργείου Εξωτερικών για τα Ίμια	Ministry of Foreign Affairs announcement on the Turkish Foreign Ministry's repeated groundless claims regarding Imia	Wednesday April 18, 2018	
81.	Ανακοίνωση Υπουργείου Εξωτερικών αναφορικά με το διευρυνσιακό πακέτο της Ευρωπαϊκής Επιτροπής	Ministry of Foreign Affairs announcement regarding the European Commission's enlargement package	Wednesday 18 April 2018	
82.	*Απάντηση Εκπροσώπου Υπουργείου Εξωτερικών, Α. Γεννηματά, σε ερώτηση δημ/φου ξένου ΜΜΕ σχετικά με το χθεσινό ανυπόστατο και συκοφαντικό δημοσίευμα της FAZ για την Ελλάδα*	Response of the spokesperson for the Ministry of Foreign Affairs, A. Yennimatas, to a question from a foreign journalist regarding yesterday's unsubstantiated and defamatory FAZ article on Greece	Monday, 16 April 2018	Unknown source language
83.	Συμμετοχή Υπουργού Εξωτερικών, Ν. Κοτζιά, στο Συμβούλιο Εξωτερικών Υποθέσεων της Ευρωπαϊκής Ένωσης (Λουξεμβούργο, 16.4.2018)	Minister of Foreign Affairs Nikos Kotzias to attend a meeting of the EU Foreign Affairs Council (Luxembourg, 16 April 2018)	Sunday, 15 April 2018	
84.	Ανακοίνωση Υπουργείου Εξωτερικών σχετικά με τις εξελίξεις στη Συρία	Ministry of Foreign Affairs announcement regarding developments in Syria	Saturday, April 14, 2018	
85.	Εφαρμογή και το 2018 της έκδοσης θεωρήσεων εισόδου μίας εβδομάδος για τουρίστες από την Τουρκία προς στα	Issuing of one-week entry visas for tourists visiting Greek islands from Turkey to continue in 2018	Wednesday 11 April 2018	

	ελληνικά νησιά			
86.	Επισκέψεις Υπουργού Εξωτερικών, Ν. Κοτζιά, σε πΓΔΜ και Κόσοβο (11-12.04.2018)	Minister of Foreign Affairs, Nikos Kotzias, to visit fYROM and Kosovo (11-12.04.2018)	Wednesday 11 April 2018	
87.	Επίσκεψη Υπουργού Εξωτερικών, Ν. Κοτζιά, στη Σερβία (Βελιγράδι, 10-11.04.2018)	Minister of Foreign Affairs, N. Kotzias, to visit Serbia (Belgrade, 10-11 April 2018)	Tuesday, 10 April 2018	
88.	Δήλωση Υφυπουργού Εξωτερικών, Ι. Αμανατίδη, στο αεροδρόμιο «Ελευθέριος Βενιζέλος» κατά την επιστροφή του στην Αθήνα με το Άγιο Φως (Αθήνα, 7.4.2018)	Deputy Minister of Foreign Affairs, I. Amantidis’ statement at ‘Eleftherios Venizelos’ airport, upon returning to Athens with the Holy Light (Athens, 7 April 2018)	Saturday, 7 April 2018	
89.	Δήλωση Υφυπουργού Εξωτερικών, Ι. Αμανατίδη, κατά την τελετή παραλαβής του Αγίου Φωτός από τον Πατριάρχη Ιεροσολύμων, κ.κ. Θεόφιλο Γ΄ (Ιεροσόλυμα, 7.4.2018)	Deputy Minister of Foreign Affairs, I. Amanatidis' statement at the ceremony for receiving the Holy Light from Patriarch Theophilos III of Jerusalem (Jerusalem, 7 April 2018)	Saturday, 7 April 2018	
90.	*Χαιρετισμός Υφυπουργού Εξωτερικών, Ι. Αμανατίδη, στο Διεθνές Συνέδριο με θέμα «την ενότητα στην πολυμορφία και τις βασικές αρχές ελευθερίας για Χριστιανούς και Μουσουλμάνους στην Μέση Ανατολή» (Βηρυτός, 3.4.2018)*	Deputy Minister of Foreign Affairs I. Amanatidis’ welcome speech at the International Conference on “Unity in diversity and the basic principles of freedom for Christians and Muslims in the Middle East" (Beirut, 03.04.2018)	Wednesday 4 April 2018	Unknown source language
91.	Χαιρετισμός Υφυπουργού Εξωτερικών, Τέρενς Κουίκ, στο Φόρουμ των απανταχού Νέων Ελλήνων, (Ζάππειο 31.03-01.04.2018)	Deputy Minister of Foreign Affairs T. Quick welcomes the Hellenic Youth in Action Forum (Zappeion, 31 March-1 April 2018)	Saturday, 31 March 2018	
92.	Ανακοίνωση Υπουργείου Εξωτερικών σχετικά με δήλωση εκπροσώπου τουρκικού Υπουργείου Εξωτερικών για τα Ίμια	Ministry of Foreign Affairs announcement on a statement regarding Imia made by the spokesperson of the Turkish Ministry of Foreign Affairs	Saturday, 31 March 2018	
93.	Διάψευση προβοκατόρικης πληροφορίας που διακινήθηκε σε ορισμένα Μ.Μ.Ε.	Rebuttal of provocative fake news disseminated by certain media	Friday, 30 March 2018	
94.	Συνάντηση των Υπουργών Εξωτερικών Ελλάδος και Αλβανίας (Τίρανα, 29.03.2018)	Meeting of the Ministers of Foreign Affairs of Greece and Albania (Tirana, 29.03.2018)	Thursday, 29 March 2018	
95.	Συνάντηση Υπουργού Εξωτερικών, Νίκου Κοτζιά, με τον Υπουργό Εξωτερικών της Αλβανίας, Ditmir Bushati	Minister of Foreign Affairs, Nikos Kotzias, to meet with the Minister of Foreign Affairs of Albania, Ditmir Bushati (Tirana,	Thursday, 29 March 2018	

	(Τίρανα, 29.03.2018)	29.03.2018)		
96.	*Δήλωση του Υφυπουργού Εξωτερικών κ. Ιωάννη Αμανατίδη για την 25η Μαρτίου στην παρέλαση της Θεσσαλονίκης*		Sunday, 25 March 2018	No English translation
97.	*Μήνυμα του Προέδρου της Δημοκρατίας κ. Προκόπιου Παυλόπουλου προς τον Απόδημο Ελληνισμό με την ευκαιρία της Εθνικής Εορτής της 25ης Μαρτίου*		Friday, 23 March 2018	No English translation
98.	Επίσκεψη Υπουργού Εξωτερικών, Ν. Κοτζιά, στα Σκόπια (22-23.3.2018) και συνάντησή του στη Βιέννη με τον ομόλογό του, N. Dimitrov, παρουσία του Προσωπικού Απεσταλμένου του ΓΓΗΕ, M. Nimetz (30.3.2018)	Minister of Foreign Affairs, Nikos Kotzias, to visit Skopje (22-23.3.2018) and to meet in Vienna with his counterpart, N. Dimitrov, in the presence of UNSG's Personal Envoy, M. Nimetz (30.3.2018)	Tuesday, 20 March 2018	
99.	Επίσκεψη Υπουργού Εξωτερικών, Ν. Κοτζιά, στο Κάιρο	Minister of Foreign Affairs Nikos Kotzias on visit to Cairo	Tuesday, 20 March 2018	
100.	Ομιλία ΑΝΥΠΕΞ, Γ. Κατρούγκαλου, στην επιχειρηματική συνάντηση Ελλάδος – Σερβίας (Θεσσαλονίκη 19.03.2018)	Alternate Minister of Foreign Affairs Giorgos Katrougalos' Speech at the business conference between Greece-Serbia (Thessaloniki, 19 March 2018)	Monday, 19 March 2018	
101.	Συμμετοχή Υπουργού Εξωτερικών, Ν. Κοτζιά, στο Συμβούλιο Εξωτερικών Υποθέσεων της Ευρωπαϊκής Ένωσης (Βρυξέλλες, 19.3.2018)	Minister of Foreign Affairs Nikos Kotzias to attend a meeting of the EU Foreign Affairs Council (Brussels, 19 March 2018)	Sunday, 18 March 2018	
102.	Ανακοίνωση Υπουργείου Εξωτερικών για την τρομοκρατική επίθεση σε βάρος του Πρωθυπουργού της Παλαιστινιακής Αρχής, Rami Hamdallah, στη Λωρίδα της Γάζας	Ministry of Foreign Affairs announcement on the terrorist attack on the Prime Minister of the Palestinian Authority, Rami Hamdalla, in the Gaza Strip	Wednesday 14 March 2018	
103.	Συνάντηση Υπουργού Εξωτερικών, Νίκου Κοτζιά, με τον νέο Υπουργό Εξωτερικών της Κυπριακής Δημοκρατίας, Νίκο Χριστοδουλίδη (ΥΠΕΞ, 5.3.2018)	Minister of Foreign Affairs Nikos Kotzias to meet with the new Minister of Foreign Affairs of the Republic of Cyprus, Nikos Christodoulides (Foreign Ministry, 5 March 2018)	Sunday, 4 March 2018	
104.	Χαιρετισμός Υφυπουργού Εξωτερικών, Ι. Αμανατίδη, στη λήξη των εργασιών του Συνεδρίου "Μοντέλο Ηνωμένων	Address by Deputy Minister of Foreign Affairs, Ioannis Amanatidis, at the closing Session of the "Model United Nations"	Saturday, 3 March 2018	

	Εθνών" (Αριστοτέλιο Κολλέγιο Θεσσαλονίκης, 3.3.2018)	Conference (Aristotle College of Thessaloniki, 3 March 2018)		
105.	*Απάντηση Υπουργού Εξωτερικών, Ν. Κοτζιά, σε επίκαιρη ερώτηση βουλευτή Αξιωματικής Αντιπολίτευσης, Ντ. Μπακογιάννη, με θέμα: «Η κυβέρνηση γκριζάρει το Ιόνιο. Κίνδυνοι στη νέα συμφωνία για τις θαλάσσιες ζώνες με την Αλβανία»*		Thursday, 1 March 2018	No English translation
106.	Συνέντευξη Υφυπουργού Εξωτερικών, Τέρενς Κουίκ, στον ρ/σ "Χρόνος FM" Κομοτηνής και το δημ/φο Δήμο Μπακιρτζάκη	Interview of Deputy Minister of Foreign Affairs, Terens Quick, on Komotini r/s "Chronos FM", with journalist Dimos Bakirtzakis	Sunday, 25 February 2018	
107.	Συνέντευξη Υφυπουργού Εξωτερικών, Γιάννη Αμανατίδη, στην εκπομπή της ΕΡΤ «Επτά» και τη δημ/φο Βάλια Πετούρη	Interview of Deputy Minister of Foreign Affairs, Ioannis Amanatidis, on ERT's 'Epta', with journalist Valia Petouri	Saturday, 24 February 2018	
108.	Επιστολή ΑΝΥΠΕΞ, Γ. Κατρούγκαλου, για υποψηφιότητα της ΠΓΔΜ στην EUSAIR	Alternate Minister of Foreign Affairs, G. Katrougalos, informs competent EU Commissioners on Greece's decision to support fYROM's EUSAIR candidacy	Thursday, 22 February 2018	
109.	Απάντηση Εκπροσώπου Υπουργείου Εξωτερικών, Α. Γεννηματά, σε ερώτηση δημ/φου σχετικά με σχόλια εκπροσώπου τ/ΥΠΕΞ για την συνέντευξη του Υπουργού Εξωτερικών, Ν. Κοτζιά, στον τ/σταθμό Alpha	Reply of the MFA's Spokesperson, Alexandros Yennimatas, to a journalist's question regarding comments made by the Spokesperson of the Turkish Ministry of Foreign Affairs on Minister of Foreign Affairs, Nikos Kotzias', interview on "Alpha" Channel	Saturday, 17 February 2018	
110.	Συμμετοχή Υπουργού Εξωτερικών, Ν. Κοτζιά, στην Άτυπη Συνάντηση «Gymnich» των Υπουργών Εξωτερικών της ΕΕ (Σόφια, 15-16.2.2018)	Minister of Foreign Affairs, Nikos Kotzias, to participate in the Informal "Gymnich" Meeting of EU Ministers of Foreign Affairs (Sofia, 15-16.2.2018)	Thursday, 15 February 2018	
111.	Δήλωση Εκπροσώπου Υπουργείου Εξωτερικών, Αλέξανδρου Γεννηματά, σε απάντηση ισχυρισμών Εκπροσώπου του τουρκικού Υπουργείου Εξωτερικών για το θέμα των Ιμίων	Statement of the Spokesperson of the Ministry of Foreign Affairs, Alexandros Yennimatas, in reply to allegations made by the Spokesperson of the Turkish Ministry of Foreign Affairs concerning Imia.	Tuesday, 13 February 2018	
112.	Ανακοίνωση για τη σημερινή συνάντηση του Προσωπικού Απεσταλμένου του Γενικού Γραμματέα του ΟΗΕ, κ.	Meeting of the Personal Envoy of the UN Secretary-General, Mr. Matthew Nimetz, with the Ministers for Foreign Affairs of	Tuesday, 13 February 2018	

	Matthew Nimetz, με τους Υπουργούς Εξωτερικών της Ελλάδος και της πρώην Γιουγκοσλαβικής Δημοκρατίας της Μακεδονίας	Greece and the former Yugoslav Republic of Macedonia		
113.	Ανακοίνωση Υπουργείου Εξωτερικών για πραγματοποίηση έντονου διαβήματος διαμαρτυρίας, σε επίπεδο Γενικού Γραμματέα, προς την Τουρκία, σχετικά με επικίνδυνο συμβάν στην περιοχή των Ιμίων	Announcement of the Ministry of Foreign Affairs on the strong demarche made to Turkey, at the Secretary General level, regarding a dangerous incident in the Imia area	Tuesday, 13 February 2018	
114.	Συνάντηση Υπουργού Εξωτερικών, Ν. Κοτζιά, με τον ομόλογό του της πΓΔΜ, N. Dimitrov και με τον Προσωπικό Απεσταλμένο του Γ.Γ. ΟΗΕ, Matthew Nimetz (Βιέννη, 12-13.02.2018)	Minister of Foreign Affairs, Nikos Kotzias, to meet with his fYROM counterpart, Nikola Dimitrov, and the UN Secretary General's Personal Envoy, Matthew Nimetz (Vienna, 12-13 February 2018)	Monday, 12 February 2018	
115.	Ανακοίνωση Υπουργείου Εξωτερικών σχετικά με τις παράνομες ενέργειες της Τουρκίας στην Κυπριακή ΑΟΖ	Ministry of Foreign Affairs announcement regarding Turkey's illegal actions in the Cypriot EEZ	Monday, 12 February 2018	
116.	7η συνάντηση για τα Μέτρα Οικοδόμησης Εμπιστοσύνης μεταξύ της Ελλάδας και της πρώην Γιουγκοσλαβικής Δημοκρατίας της Μακεδονίας (Σκόπια, 9.2.2018)	The 7th meeting on the confidence-building measures between Greece and the former Yugoslav Republic of Macedonia (Skopje, 09.02.2017)	Friday, 9 February 2018	
117.	*Μήνυμα Υφυπουργού Εξωτερικών, Τέρενς Κουίκ, για την Παγκόσμια Ημέρα Ελληνικής Γλώσσας*		Wednesday 7 February 2018	No English translation
118.	Ανακοίνωση Υπουργείου Εξωτερικών σχετικά με σημερινή απόφαση της κυβέρνησης της πρώην Γιουγκοσλαβικής Δημοκρατίας της Μακεδονίας να μετονομάσει το αεροδρόμιο των Σκοπίων και τον αυτοκινητόδρομο που συνδέει τα Σκόπια με τα ελληνικά σύνορα	Ministry of Foreign Affairs announcement regarding today's decision from the government of the former Yugoslav Republic of Macedonia to rename Skopje's airport and the highway linking Skopje and the Greek border	Tuesday, 6 February 2018	
119.	*Χαιρετισμός ΥΠΕΞ, Νίκου Κοτζιά, κατά την εκδήλωση κοπής της πρωτοχρονιάτικης πίτας του Υπουργείου Εξωτερικών (6.2.2018)*		Tuesday, 6 February 2018	No English translation
120.	Ανακοίνωση Υπουργείου	Ministry of Foreign Affairs	Sunday, 4	

	Εξωτερικών σχετικά με επαναλαμβανόμενη λανθασμένη αναφορά του Επιτρόπου Γιοχάνες Χαν στην ονομασία της πΓΔΜ στο γερμανικό περιοδικό Der Spiegel	announcement on Commissioner Johannes Hahn’s repeated erroneous reference to the name of fYROM in the German magazine Der Spiegel	February 2018	
121.	Ανακοίνωση Υπουργού Εξωτερικών, Ν. Κοτζιά, σχετικά με πρόσφατες εξελίξεις για την επίλυση του ονοματολογικού	Announcement of the Minister of Foreign Affairs, Nikos Kotzias, regarding recent developments on the resolution of the name issue (Athens, 2 February 2018)	Friday, 2 February 2018	
122.	Κύρια σημεία συνέντευξης Υπουργού Εξωτερικών Ν. Κοτζιά στο ειδησεογραφικό πρακτορείο Reuters (Αθήνα, 31 Ιανουαρίου 2018)	Key points of Minister of Foreign Affairs Nikos Kotzias’ interview with Reuters (Athens, 31 January 2018)	Wednesday 31 January 2018	
123.	Κοπή βασιλόπιτας Υπουργείου Εξωτερικών (Ακαδημίας 1, ισόγειο, 06.02.2018, 14:00)	Ministry of Foreign Affairs to cut New Year's Cake (14:00, 6 February 2018, 1 Akadimias St., ground floor)	Friday, 2 February 2018	
124.	Απάντηση Εκπροσώπου Υπουργείου Εξωτερικών Α. Γεννηματά σε ερώτηση δημοσιογράφου σχετικά με τις αχαρακτήριστες δηλώσεις συμβούλου του Τούρκου Προέδρου, αναφορικά με τα Ίμια	Response of the Spokesperson of the Ministry of Foreign Affairs, A. Yennimatas, to a journalist’s question concerning contemptible statements made, regarding Imia, by an advisor to the Turkish President	Thursday, 1 February 2018	
125.	Δηλώσεις Υπουργού Εξωτερικών, Νίκου Κοτζιά, με το πέρας της συνάντησής του με τον Προσωπικό Απεσταλμένο του Γ.Γ. ΟΗΕ, Matthew Nimetz (Αθήνα, 30.1.2018)	Statements of the Minister of Foreign Affairs, Nikos Kotzias, following his meeting with UN Secretary General'sPersonal Envoy, Matthew Nimetz (Athens, 30 January 2018)	Tuesday, 30 January 2018	
126.	Συνάντηση Υπουργού Εξωτερικών, Ν. Κοτζιά, με τον Προσωπικό Απεσταλμένο του Γ.Γ. ΟΗΕ, M. Nimetz (ΥΠΕΞ, 30.1.2018)	Minister of Foreign Affairs, Nikos Kotzias, to meet with the UN Secretary General’s Personal Envoy, Matthew Nimetz (Ministry of Foreign Affairs, 30 January 2018)	Monday, 29 January 2018	
127.	Ανακοίνωση Υπουργείου Εξωτερικών για δηλώσεις Αλβανών επισήμων περί συζήτησης του «τσαμικού» ζητήματος	Ministry of Foreign Affairs announcement on statements from Albanian officials regarding the “Cham” issue	Monday, 29 January 2018	
128.	Ανακοίνωση του Υπουργείου Εξωτερικών για την Ημέρα Μνήμης των Θυμάτων του Ολοκαυτώματος (27.01.2018)	Ministry of Foreign Affairs announcement on Holocaust Remembrance Day (27.1.2018)	Friday, 26 January 2018	
129.	*Χαιρετισμός Υφυπουργού*		Friday, 19	No English

	Εξωτερικών, Ι. Αμανατίδη, στην εκδήλωση για την παρουσίαση του βιβλίου του Γεώργιου Πηλιχού «Άουσβιτς: Έλληνες - Αριθμός Μελλοθανάτου»		January 2018	translation
130.	Ανακοίνωση Υπουργείου Εξωτερικών για έναρξη διαδικασίας αναζήτησης, εκταφής, ταυτοποίησης και ταφής οστών Ελλήνων πεσόντων στην Αλβανία	Ministry of Foreign Affairs announcement on the initiation of the process of searching for and disinterment, identification and burial of the remains of the Greek fallen in Albania	Monday, 22 January 2018	
131.	Κοινό Ανακοινωθέν των Υπουργείων Εξωτερικών Ελλάδας και Αλβανίας (Κορυτσά, 21.01.2018)	Joint announcement of the Ministries of Foreign Affairs of Greece and Albania (Korce, 21 January 2018)	Sunday, 21 January 2018	
132.	Διυπουργική σύσκεψη υπό τον Υπουργό Εξωτερικών, Ν. Κοτζιά, με αντικείμενο την προετοιμασία για το άνοιγμα της διασυνοριακής διάβασης των Πρεσπών (12.01.2018)	Minister of Foreign Affairs, N. Kotzias, chairs interministerial meeting on preparation for the opening of the Prespes border crossing (12 January 2018)	Friday, 12 January 2018	
133.	Συνάντηση Υπουργών Εξωτερικών, Νίκου Κοτζιά και Nikola Dimitrov (Θεσσαλονίκη, 11.1.2018)	Meeting between the Ministers of Foreign Affairs Nikos Kotzias and Nikola Dimitrov	Thursday, 11 January 2018	
134.	Ανακοίνωση Υπουργείου Εξωτερικών σχετικά με τη σημερινή ανακοίνωση του τουρκικού ΥΠΕΞ αναφορικά με την έγκριση του ν/σ για τη Σαρία από το ελληνικό Κοινοβούλιο	Ministry of Foreign Affairs announcement on today’s announcement from the Turkish Ministry of Foreign Affairs regarding the Hellenic Parliament’s approval of legislation on sharia	Thursday, 11 January 2018	