Efficiency of machine translation and post-editing

Vincelj, Melany

J. J. Strossmayer University of Osijek

Faculty of Humanities and Social Sciences

Study Programme: Double Major in English Language and Literature – Translation and Interpreting Studies and German Language and Literature – Translation and Interpreting Studies

Melany Vincelj

Efficiency of Machine Translation and Post-Editing

Master's Thesis

Supervisor: Dr. Marija Omazić, Full Professor

Osijek, 2021

J. J. Strossmayer University of Osijek

Faculty of Humanities and Social Sciences

Study Programme: Double Major in English Language and Literature – Translation and Interpreting Studies and German Language and Literature – Translation and Interpreting Studies

Melany Vincelj

Efficiency of Machine Translation and Post-Editing

Master's Thesis

Scientific area: Humanities

Scientific filed: Philology

Scientific branch: English Studies

Supervisor: Dr. Marija Omazić, Full Professor

Osijek, 2021

Sveučilište J. J. Strossmayera u Osijeku

Filozofski fakultet

Odsjek za engleski jezik i književnost

Studij: Dvopredmetni sveučilišni diplomski studij engleskog jezika i književnosti - prevoditeljski smjer i njemačkog jezika i književnosti – prevoditeljski smjer

Melany Vincelj

Učinkovitost strojnog prijevoda i redakture strojno prevedenih tekstova

Diplomski rad

Mentor: prof. dr. sc. Marija Omazić

Osijek, 2021.

Sveučilište J. J. Strossmayera u Osijeku
Filozofski fakultet
Odsjek za engleski jezik i književnost
Studij Dvopredmetni sveučilišni diplomski studij engleskog jezika i književnosti - prevoditeljski smjer i njemačkog jezika i književnosti – prevoditeljski smjer

Melany Vincelj

Učinkovitost strojnog prijevoda i redakture strojno prevedenih tekstova

Diplomski rad

Znanstveno područje: humanističke znanosti
Znanstveno polje: filologija
Znanstvena grana: anglistika
Mentor: prof. dr. sc. Marija Omazić
Osijek, 2021.

Prilog: Izjava o akademskoj čestitosti i o suglasnosti za javno objavljivanje

Obveza je studenta da donju Izjavu vlastoručno potpiše i umetne kao treću stranicu završnog odnosno diplomskog rada.

IZJAVA

Izjavljujem s punom materijalnom i moralnom odgovornošću da sam ovaj rad samostalno napravio te da u njemu nema kopiranih ili prepisanih dijelova teksta tuđih radova, a da nisu označeni kao citati s napisanim izvorom odakle su preneseni.
Svojim vlastoručnim potpisom potvrđujem da sam suglasan da Filozofski fakultet Osijek trajno pohrani i javno objavi ovaj moj rad u internetskoj bazi završnih i diplomskih radova knjižnice Filozofskog fakulteta Osijek, knjižnice Sveučilišta Josipa Jurja Strossmayera u Osijeku i Nacionalne i sveučilišne knjižnice u Zagrebu.

U Osijeku, datum
13. rujna 2021.

Julany Vuuf, 0122222245
ime i prezime studenta, JMBAG

Table of Contents

Abstract

Translation is a part of our everyday life even if we are not aware of it. Without it, we could not watch popular Hollywood movies, Netflix TV series or read bestseller books in our native tongue. It is also necessary for texts like legal documentation, different user manuals, newspaper articles etc. which are originally written in foreign language. The amount of the content that requires translating often exceeds the number of translators available and the task of translation is more often delegated to machines and the output then revised and its quality improved to make it appropriate for its purpose. This master thesis presents different studies on machine translation (MT) and deals with its development and different aspects of it – from different MT systems and the way they operate, their advantages and disadvantages, to the steps of the post-editing process and its efficiency in terms of the time, money and cognitive effort saved compared to human translation.

Key words: translation, MT, post-editing, efficiency of MT, cognitive effort

Sažetak

Prevođenje je dio svakodnevnog života iako toga nismo svjesni. Bez prevođenja ne bismo mogli gledati popularne holivudske filmove i serije na Netflixu niti čitati najprodavanije knjige na materinjem jeziku. Prevođenje je također neophodno za tekstove poput pravnih dokumenata, različitih priručnika za uporabu, novinskih članaka itd. koji su izvorno napisani na stranom jeziku. Opseg sadržaja koji je potrebno prevesti često premašuje broj dostupnih prevoditelja te se ovaj zadatak često povjerava računalima čiji se prijevod naknadno redigira, a kvaliteta popravlja kako bi tekst bio prikladan za uporabu. Ovaj diplomski rad prezentira različita istraživanja koja se bave strojnim prevođenjem (SP), kao i njegovim razvojem i različitim aspektima strojnog prevođenja – od različitih sustava i načina na koji funkcioniraju, njihovih prednosti i nedostataka, do koraka u procesu redakture te njezine učinkovitosti u pogledu uštede vremena, novca i kognitivnog napora u usporedbi s ljudskim prevođenjem.

Ključne riječi: prevođenje, strojno prevođenje, redaktura, učinkovitost strojnog prevođenja, kognitivni napor

Introduction

Even though it often gets overlooked, it is a fact that without the process of translation, communication between different languages and cultures or enjoying different digital content would simply not be possible. Various studies have shown the growing mismatch of users in the internet and the content available in their language. In the last two decades, the number of those users marks a constant growth, which also implies the increased amount of content that requires translation (Doherty, 2016:948-49). Back in 2016, at the time of publishing of Doherty's article, only 0.1% of digital content was being translated (Doherty, 2016:949). The interactionist approach suggests that CAT tools and machine translation indeed have changed the way we communicate (Doherty, 2016:950). Translation technologies have greatly impacted productivity and consistency, ensured better global language coverage and improved international communication (Doherty, 2016:950).

> "Indeed, the language services market as a whole has shown consistent year-on-year growth in recent years [article published in 2016] despite the global financial crisis, from US$23.50 billion in 2009 to US$34.78 billion in 2013 – an annual growth rate of 5.13%. Translation prices per word, however, have continued to decrease by up to 50% since 2008, a diminution that analysts attribute to budgetary pressures and increased acceptance of translation technologies." (Doherty, 2016:949)

Machine translation (MT) is still not accepted as equal to human translators and is often misunderstood as a consequence of two opposing viewpoints: there are those who are unable to understand the complexity of translation in general (the main point being a child's ability to learn language and the opinion that anyone knowing a foreign language is able to translate) and there are those who underestimate the contribution of machine translation (based on its inability to translate literary texts accurately). The aim of this master thesis is to define machine translation and its types, describe the process of post-editing (PE), point out the good and bad sides of using these two as translation strategies and compare different studies on the efficiency of PE.

The following chapters focus on different aspects of machine translation, listed as follows: The first chapter defines MT, presents its brief history and the development of different systems from its beginning till today and the principles they operate on. Chapter 2 compares disadvantages and advantages of using MT and presents different surveys on the quality of the raw MT output. Chapter 3 deals with the process of PE – its definition, different sets of guidelines, sets of competences required from post-editors and different surveys on the efficiency of such approach.

The last chapter outlines the research conducted based on the studies mentioned in this master thesis. It states the aim and the methods used in the research, provides an analysis of each text type separately and gives a general conclusion on the outcome of the research. The concluding part summarises the most important points of the thesis and states possible future directions.

1. Machine Translation – Definition, History and Types

Machine translation is briefly defined as "computerised systems responsible for production of translations from one natural language into another, with or without human assistance." (Hutchins and Somers 1992:3). It is also referred to as mechanic translation and automatic translation. It cannot be observed as a field in its own right since "it takes from linguistics, computer science, artificial intelligence, translation theory" (Hutchins and Somers 1992:3) and is dependent on their principles. The idea behind machine translation is the full automation of the translation process, but there is of course a significant difference between an ideal and the reality, that is, between what machine translation should be and what it really has to offer. Although the idea behind it was to produce high-quality outputs, there is no perfect machine translation. However, it has developed enough to produce outputs of fairly good quality that are comprehendible with little post-editing needed.

1.1. Brief History and the Development of MT

The interest in MT occurred in the United States in the 1950s, but gave away quickly due to the ALPAC's (The Automatic Language Processing Advisory Committee) report which criticized machine translation and investment in its further development because it was "slower, less accurate and twice as expensive as human translators" (Hutchins and Somers, 1992:7). The government was the only one still using this approach because of the need for quick gathering of information from foreign sources. The profession of a translator as such was not respected and the pay check was accordingly low. Anyone who had a major in languages was considered suitable for such tasks, especially if they were native speakers of a target foreign language (Slocum, 1985:109-110).

In other countries around the world, besides a major in languages and the theory of translation, translators were also required to have a practice in several specific fields they would take on as their domain and only after a few years of experience would they be considered as competent translators (Slocum, 1985:110). The main motif for the development of machine translation systems was the opinion that qualified human translators were "hard to find, expensive, and slow" (Slocum, 1985:111), an opinion that is still present in the industry.

The approach to translation can be divided into three categories, based on the level of dependency on a computerised system assisting in this process. According to Slocum (1985) there are Machine Translation (MT), Machine-Aided Translation (MAT), and Terminology Databanks (TD). MT translates without assistance of a human agent (pre-processing and post-editing not

accounted), MAT is divided into Human-Aided Machine Translation (HAMT) and Machine-Aided Human Translation (MAHT), where HAMT means that the machine performs the translation and interacts with a human agent when there is a problem, while MAHT means human translator assisted by a machine when a problem occurs. Lastly, TD offers translators the access to technical terminology, but not common words since they are already familiar. (Slocum, 1985:110-111). The following section will provide a further division of MT.

1.2. Types of MT

MT can further be divided to RBMT (Rule-Based Machine Translation, SMT (Statistical Machine Translation), hybrid systems, and the most recently developed NMT (Neural Machine Translation) (Kadiu, 2019:83-84). Interactive and adaptive MT will also be mentioned as the newest addition to the MT system family.

MT first occurred in form of rule-based systems which are "based on linguistic information about source and target languages basically retrieved from (bilingual) dictionaries and grammars covering the main semantic, morphological, and syntactic regularities of each language respectively" (Okpor, 2014:160). Some of the notable RBMT systems are Lucy LT, Systran and Apertium. Even though the process of translation is automated, human intervention is crucial for it, since it is linguists and language experts who need to establish rules and norms and implement them in order for this system to produce good-quality output. That being said, one can conclude, that MT requires an extensive knowledge of linguistics and much time to implement those rules. As the main downsides to this type of MT, Okpor (2014) names the lack of good dictionaries, the fact that some linguistic rules still need to be established and the problematic translation of ambiguous and idiomatic phrases because of their metaphorical meaning which does not abide the established rules. (Okpor, 2014:161) However, a big advantage of this system is the syntactic and partially semantic analysis of both source and target languages, which makes RBMT appropriate for translating language pairs where parallel corpora still do not exist (Okpor, 2014:164).

As the "more advanced" type of MT there is SMT or statistical machine translation. Systems that represent this type of MT are, for example, Moses and GIZA++. The idea behind SMT is that "every sentence in one language is a possible translation of any sentence in the other and the most appropriate is the translation that is assigned the highest probability by the system" (Okpor, 2014:163). In other words, each word or phrase from the source language is transferred to the target language as the input the system chooses as most probable in the assigned context. Some of the main problems when analysing SMT output and the process in general are sentence

alignment, statistical anomalies, data dilution, idioms and word order (Macketanz et al., 2017:32). What is grouped together as one sentence in the source language can be divided into various sentences in the target text and vice versa, which makes sentence alignment difficult. Statistical anomalies describe the occurrence of most frequent proper names being replaced with false inputs because some other word or phrase occurs more frequently in correlation to that specific word. Data dilution is a major problem when using SMT for specific brand-related translations if the trained system does not include brand-specific terminology. Idioms often lose their idiomatic meaning and the word order of the output often does not correspond to the word order specific for the target language in question.

Hybrid systems developed as a consequence of flawed single techniques, namely RBMT and SMT. They combine the best features of the two systems to enhance the quality and performance (Okpor, 2014:164).

> "In some cases, translations are performed in the first stage using a rule-based approach followed by adjusting or correcting the output using statistical information. In the other way, rules are used to pre-process the input data as well as post-process the statistical output of a statistical-based translation system. This technique is better than the previous and has more power, flexibility, and control in translation." (Okpor, 2014:164)

The most recent and the most promising type of MT is neural machine translation or NMT, which has nothing to do directly with the human brain even though the name suggests that. The name stems from a metaphor used to explain how NMT actually works. The computational principles used by NMT are the so-called *neural networks*, which are actually artificial units personified as neurons in our brain since their activation depends on the stimuli received from other units and connections that pass these stimuli on (Forcada, 2017:292). Some of the examples of NTM systems are the worldwide used Google Translate, DeepL, Sockeye, OpenNMT, Neural Monkey, etc. NMT is often compared to SMT since it is also trained on huge corpora of parallel pairs of source and target text sentences, but it uses different computational approach. NMT consists of three networks: an encoder network, a decoder network, and an attention network.

> "The encoder transforms a source sentence into a list of vectors, one vector per input symbol. Given this list of vectors, the decoder produces one symbol at a time, until the special end-of-sentence symbol (EOS) is produced. The encoder and decoder are connected through an attention module which allows the decoder to focus on different regions of the source sentence during the course of decoding." (Wu et al., 2016:3)

Both NMT and SMT choose the target language sentence based on likelihood, but the process of choosing the appropriate equivalent differs. In NMT there is no segmentation (as opposed to SMT where sentences are divided into chunks or phrases which are then each translated separately). The decoding mechanism operates based on analysing the source sentence and the words preceding in the target sentence (Forcada, 2017:299-300). Since NMT is similar to SMT, similar errors are likely to occur in the unedited output: "…inconsistencies in numerical expressions and URLs, mistranslation of proper nouns (particularly compound proper nouns such as United Nations or Bank of England), terminological inconsistencies, misplacing of formatting tags, etc. …" (Forcada, 2017:301). Some of these errors, especially the ones concerning proper nouns could cause serious change in the meaning and convey a false message (Forcada, 2017:302).

The newest addition to the TM system family are adaptive and interactive NMT. Adaptive refers to the system's ability to adapt to the translator's style, that is, to change in response to confirmed segments. It also means adapting chosen segments to the context of the whole document instead of taking into consideration only the preceding or succeeding words or the sentence-level context. An example of an adaptive MT system is ModernMT, which is compatible with various CAT tools such as SDL Trados, MateCAT or MemoQ (https://www.modernmt.com/translators/). Another Example is Lilt, which in addition to being adaptive is also interactive, which enables suggestions to integrate into the translation based on the analysis of the source segments (https://support.lilt.com/hc/en-us/articles/360058203753-What-is-Interactive-Adaptive-MT-).

2. Features of MT

When asked about downsides of human translation (HT), one may say that it is slow and expensive and there is also a lack of professional translators compared to the amount of texts that need to be translated. But even so, it is still more appreciated than MT, which is negatively perceived in general. MT is inevitably getting compared to HT, which is always perceived as superior, mainly because of the better quality and the style of the unedited output. However, results of the World MT Championship from 2020 show that certain language pairs (German-English, Chinese-English and English-Inuktitut) mirror the quality of HT and can be put in the same quality class as the human-produced translation with smaller alterations (https://aclanthology.org/2020.wmt-1.1.pdf, 2020:36). The following two sections will list positive and negative sides of machine translation based on a realistic comparison to HT. The last section will additionally analyse some of the findings of different MT quality estimation research, which compare different systems and the quality of the output they produce.

2.1. Disadvantages of MT

In her comparative analysis of HT and MT, Kadiu (2019) decided to point out that MT is not that different from HT. She analyses Meschonnic's poetics of translation, which claims that HT involves decision-making, is creative, reflexive and comprehensive, while MT denotes non-reflexive use of language, relies on automation and is actually a repetition of "pre-existing codes" (Kadiu, 2019:72-75). According to Meschonnic, MT "has contributed to proliferation of non-reflective uses of language in translation" (Kadiu, 2019:72), because, in his opinion, the process of translation requires decision-making, which again reflects the "translator's perception of language and translation". Hence, if there is no living agent in machine translation, there is no decision making. For him language is always a creative activity, and MT denotes a mechanic use of language. Therefore, MT distorts the beauty and creativity of the language because it is based on the "repetition of pre-existing codes", while human translation "necessarily involves a comprehensive theory of language" (Kadiu, 2019:75).

One of the biggest advantages of HT, and therefore a disadvantage of MT, is the ability to preclude grammatical errors and use suitable lexical and semantical constructions which ensure loyalty to the source text and easy comprehension (Kadiu, 2019:82), which also means less post-editing in the long run.

In Meschonnic's opinion the ability "to choose beyond existing propositions", i.e. decision making, which not only means avoiding mistakes but also coming up with alternative solutions

(Kadiu, 2019:83), applies only to HT, since MT is based on fixed rules and terminological correspondences. However, this has proven false. MT data base actually comprises of segments translated by human agents and added to these huge pools from which the system chooses the right equivalent. Another argument against this statement is the fact that RBMT is based on language norms written by experts and that SMT chooses its TL equivalents from the basis of segments translated by human translators. The author even suggests that the human translator is omnipresent since he also provides the source text, selects the target language and post-edits the machine output (Kadiu, 2019:84).

MT is portrayed as inferior to HT since computers do not have the knowledge of the real world and usually focus only on one segment at the time. However, there are improvements to this aspect which will be discussed in the next section. The system's inability to recognise and take into account the context of the source text represents a major interpretation problem since one and the same idea can be interpreted or understood in several different ways. (Kadiu, 2019:81)

> "Problems in comprehension arise because of the translation's contextual inadequacy to the source text, because of the system's incapacity to offer an English translation that matches the meaning and/or the construction of the source text, because of its inability to take into account the context in which the utterance takes place."(Kadiu, 2019:81)

Machine translation is dependent on pre-existing terminology and is mostly unable to deal with lexical ambiguities or syntactic complexity and idiomatic singularity (Kadiu, 2019:79-80). Another big challenge for MT are new terms since they are not included in the system's data base (Doherty, 2016:953), but human translators must also put in extra effort to find an appropriate equivalent when encountered with an unknown term.

Both MT and HT require post-editing before putting them to use, but they are being revised for different set of errors made during the process of translation. (Hutchins and Somers, 1992:3). Human intervention is mandatory in order to ensure quality, accuracy and fluency of the output.

> „If not handled properly, machine translation could even pose a risk to a brand, causing, for example, a cultural faux pas. The volatile quality of MT suggestions means that sometimes the opposite meaning of the original is communicated, idiomatic expressions are not properly expressed, content becomes unintentionally offensive, and words can be missed." (Memsource, 2020:17).

The final, and probably the biggest downsides of MT are "[…] issues of quality, legality, responsibility, and remuneration…" (Doherty, 2016:963) arise when translating specific text types such as legal or confidential documents where the clause of confidentiality must be signed and originals are forbidden from being uploaded to such platforms.

2.2. Advantages of MT

Although it may seem that there are too many disadvantages of MT so that it is not worth bothering post-editing such an output, there are just as many advantages of it, if not more.

Repetition is often observed as a negative side of MT, but repetition is to be found in HT too. Automation, codification and repetition are also a part of HT, which is observable in the parts of the text or phrases that have fixed translations usually found in technical or specialized text types (Kadiu, 2019:86). Human translators rely on their memory and scan it just like the machine scans its data base while searching for an equivalent (Kadiu, 2019:86).

Decision making is certainly included in the process of MT when considering the principles different MT systems operate on. It is also human language experts who train these MT systems and feed them segments previously translated by human translators (Doherty, 2016:953).

MT is generally perceived as not being capable of considering context in the same way the human translator does, but there are improvements in the way NMT operates compared to RBMT or SMT principles. As mentioned in the section dealing with types of MT, NMT systems have at dispense the attention module which does take context into account to a certain extent, and the most recent ones, adaptive NMT systems take the context of the whole text into consideration before choosing equivalents for the final output.

One of the biggest advantages of MT is its capacity for development. It may be true that sometimes these systems have trouble translating new terminology, but they are constantly fed new information, which makes them reliable and more efficient. When stumbled upon a new term or a term that has no equivalent in target language, human translator needs to come up with a new one, which is then also transferred to MT – surveys show that in a few months' time MT systems contain suggestions they didn't have earlier thanks to the new segments uploaded to the data base (Kadiu, 2019:86-87).

Post-editing is not always mandatory for MT outputs. There are certain situations where raw MT output can be used. For instance, such output can be used as a draft for translators, to save time and give the rough idea of the translation. (Hutchins and Somers, 1992:3). There are also

situations when the client only wants a brief overview of the content – if the text type is appropriate and the raw output is understandable and does not alter the meaning of the original, saving time by using MT system is a good choice. Some surveys even show that the extent of PE required for MT output is lower than post-editing outputs translated using CAT tools, even by 60% (Doherty, 2016:960). According to Memsource (2020) analyses, machine translation is the base of almost 35% of content translated by human translators. Such cooperation could be illustrated as shaking hands with a robot – the machine part provides the translation quickly, saves time and money, while the human part ensures its quality. (Memsource, 2020:5).

As already mentioned, the undeniable advantages of MT are the savings of money and time, boosted with enhanced productivity and efficiency. In 2020 Memsource published a detailed article analysing these two aspects of MT and suggested some strategies for incorporating MT as a translation strategy for language service providers. MT is considered as a translating strategy more often because of its improved quality due to constant investment and research in the field of neural machine translation and deep learning and its application continues to grow each year (Memsource, 2020:4). "The MT market has been valued between USD 130 million to USD 400 million, and is estimated to exceed USD 1.5 billion by 2024." (Memsource, 2020:3) They conducted research including all translation companies that have translated 10 000 or more words and have used MT as a translation strategy in at least 1% of their work between 2018 and 2019 to analyse the percentage of MT application. The results show a very favourable percentage for MT – 82.16% of these organisations using it to save money and time and increase their efficiency (Memsource, 2020:4). These savings depend on the language pair and the type of text, but they are evident (Memsource, 2020:7). Studies conducted on MT post-editing have shown an increase of 40% productivity per hour and suggest that even if the output is of lower quality, the productivity rate is still higher than when translating from scratch (Memsource, 2020: 10).

2.3. Quality Estimation

It is a widely accepted opinion that machine translation "tends to produce poor-quality results, despite the wide variety of tools available" (Kadiu, 2019:80). But not all platforms and systems are of the same efficiency. Non-revised translations are always observed to be of low quality. However, high quality output does not rule out PE since there is always room for progress and improvement. What does matter for MT, according to Slocum, is that the output is of "sufficient quality for the intended use" (1985:110) and that with the process of PE it is still cost-efficient. The most important criteria that define if a translation is of good quality or not, but also sets the bar for the amount of PE required, are "fidelity, accuracy, intelligibility, appropriate style

and register" (Hutchins and Somers, 1992:2). An experiment including language professionals conducted by Popović et al. in 2014 lists following categories as the most frequent interventions during the process of PE: correcting word form, correcting word order, adding missing word, deleting extra word and correcting lexical choice – wrong lexical choice, and wrong word order being on the top of the list of errors (Popović et al, 2014:548).

Surveys have shown that RBMT exceeds SMT in those aspects that can be "grammatically determined", i.e. that are based on different language norms: function words, verb tense, mood and aspect, composition, non-verbal agreement, subordination and verb valency (Macketanz et al, 2017:41-42). An experiment conducted by Thurmair in 2005 shows that RBMT exhibits better quality results, i.e. produces more grammatically correct sentences, which do not need post-editing, while SMT produces mostly understandable quality sentences which demand some alteration to be acceptable (2005:93). However, there is still space for improvement of SMT. What is also interesting is that both produce approximately the same amount of incorrect translations.

Comparative analyses of SMT and NMT outputs (language pair English-German) have shown that NMT produces better quality output, i.e. that there are 50% less errors concerning word order, 19% less morphological errors and 17% less lexical errors. The findings also report on better fluency in NMT systems, which, however, does not prove adequacy of the output for the purpose (Forcada, 2017:303). Forcada in his article mentions studies conducted by Bojar et al. in 2016 and Toral and Sanchez Cartagena in 2017 which have shown first indications that the neural system is capable of learning several aspects of the language that are coded in the rules of the RBMT in a better way than the phrase-based SMT systems. Research on the amount of PE have shown that the PE time does not differ much depending on the type of MT, but the "technical effort, i.e. the number of keystrokes or minimum number of edits goes in favour of NMT" (Forcada, 2017:304).

When talking about automatic evaluation methods for determining the quality of the MT output, there are different types of methods based on the type of measure computed. The principle they are based on is comparing the MT output with the reference segment translated by a human translator and computing the similarity between them. The more similar the output and the reference segment are, the better (or more human-like) the quality of the MT output (https://blog.taus.net/automated-mt-evaluation-metrics). Giménez and Màrquez (2010) name four categories of evaluation methods: those measuring edit distance (WER, PER, TER), those computing lexical precision (BLEU, NIST), those based on lexical recall (ROUGE, CDER), and those based on balance between precision and recall (GTM, METEOR, BLANC, SIA, MAXSIM).

The use of automatic evaluation methods has allowed "[…] for fast, inexpensive, and objective numerical measurements of translation quality on demand" (Giménez; Màrquez, 2010:210). The most commonly used methods are BLEU, NIST, METEOR and TER. All of the methods use n.grams as the measuring units, which are defined as "[…] contiguous sequence[s] of n items from a given sample of text or speech" (https://blog.taus.net/automated-mt-evaluation-metrics).

The BLEU score is the most popular method due to its high correlation with human translation. However, its reliability has been questioned since it is word based. The BLEU score is reported on a scale from 0-1 or alternatively 0-100. "BLEU's n-gram matching requires exact word matches, meaning that if different vocabulary or phrases are used in reference translation, the score will be lower." (https://blog.taus.net/automated-mt-evaluation-metrics) NIST is an improved version of BLEU, differentiating in the n-gram precision calculation. While the BLEU score treats all n-grams as equal, NIST gives more weight to n-grams that are less common. METEOR takes recall into consideration along with precision. It allows for multiple reference translations and deals with the problem of flexibility in word matching. TER is a character based method for measuring the number of edits performed on a MT output to provide human like quality product.

Several automatic evaluation metrics-algorithms have been developed to improve the estimation of the time and effort required for the PE.

> "MT error identification […] is a first step towards a fully automatic MT post-editing system (APE). The error identification module is used to identify incorrect segments and pass them to the second step (error correction). The final goal of our APE system is to automatically correct the most common repetitive errors in raw MT output, allowing the human post-editors to focus on the more essential changes […]" (de Jesus Martins and de Medeiros Caseli, 2015:21)

The research showed that application of the filters improved the quality between 69 and 100% and has reduced the errors. What makes quality estimation different from classical MT evaluation is the absence of reference translations. QE collects data on the estimated quality of the product from the information about the input and the output text and the translation process itself. (Gonzalez-Rubio et al., 2013:282)

> "From the MT evaluation perspective, results show that, given a model trained on data from any MT system, language-pair and text domain, it is possible to obtain quality estimates for any number of new sentences, since reference translations are not

> necessary, and these estimates correlate significantly better with human evaluation than reference-based metrics commonly used for MT evaluation." (Specia et al., 2010:48)

Training data is still difficult for some language pairs since there are not enough available data for smaller languages (Eetemadi et al., 2015:190). Solution to this is human translators providing training data by means of translating monolingual sentences to improve quality of the output.

3. Post-Editing – Definition and Criteria

According to ISO 17100:2015, post-editing means to "edit and correct machine translation output (ISO, 2015)" (Hu and Cadwell, 2016:347). PE has established itself as a separate service in the language industry, but its full potential has still not been embraced. In a survey among project managers of different LSPs, some said that they would never include PE as part of their service, while the others did not negate the possibility, but still did not show much enthusiasm. The main reason for such reactions is the reluctance of translators towards PE because it is considered boring, demeaning and it is less paid than translation from scratch (Sakamoto, 2019:206).

It is a fact that clients are more and more interested in PE, since it means getting higher productivity at a lower rate, which starts to pose a problem for LSPs who fear that experienced translators will not accept PE as their future job and might even change profession. The annual survey of the European language industry, conducted by EUATC, ELIA, FIT Europe and GALA and published in June 2020, compares the state in the industry before and after the occurrence of the coronavirus. The analysis of the responses has shown that independent language professionals still prefer CAT tools and they do not usually use MT as translation strategy, but the responses of language service companies show that 78% of them plan on starting implementing MT or increasing the usage of it for translation tasks, along with post-editing. High percentage of them (66%) also announces investing in development of MT usage in their companies (https://ec.europa.eu/info/sites/default/files/2020_language_industry_survey_report.pdf, 44-51). This shows an increase of 54% compared to the study on the attitudes towards MT and PE conducted by Sakamoto back in 2019 which showed that only 24% of the participating LSPs offer PEMT (post-edited machine translation) as a service, while a large number of LSPs not offering it even in one way or another mentions it as demeaning in the description of their services online (Sakamoto, 2019:207).

Nitzke et al. (2019) mention in their article that, as any other business, PEMT must necessarily include risk management as part of the process. This means deciding which risks are allowed, and what must be avoided. "This initial analysis should consider the negative consequences of failures in the translation, such as impaired communication, loss of reputation, property damage, lawsuits or other legal consequences, injuries, which could even amount to danger to life and limb, etc." (Nitzke et al., 2019:242)

Nitzke et al. (2019) name five criteria that help deciding if a certain text is suitable for MT and PE or not. Rule number one deals with the text type: "[…] when the texts seem suitable for

the use of Translation Memory software, they can also be used for MT" (Nitzke et al., 2019:243). Creative texts such as literary or marketing texts usually do not go hand in hand with MT because of their style – they require creative solutions, which would require too much PE effort. It is easier to translate them from scratch, while different technical texts produce solid or high quality MT output with lower PE effort to make them suitable for the purpose. MT product quality affects the PE effort, so the time and the cost of PE of lower quality output must be considered before embarking on it. Deadlines and "the lifetime" of a translation are also important. If the deadline is short, not many translators are available, and the final product will be replaced by another translation in a certain amount of time (e.g. website content), MT could be the right choice. Data security is directly connected to risk management, since not many people should have access to confidential data. In such case, an in-house trained MT system would be acceptable, but open and free platforms such as Google Translate are forbidden. The last criterion is controlled-language use since it "increase[s] readability, translatability, and the reusability of texts by consistent, clear, and target-oriented writing" (Nitzke et al., 2019:245).

According to Sakamoto (2019), there are three stages of PE process. Firstly, the usability of the MT output must be assessed. The second step is the editing of the MT product. At last the quality of edited segments must be tested (Sakamoto, 2019:247).

3.1. Types of PE

PE is usually divided into light PE and full PE, but there are no universal guidelines for the process. Light PE usually means editing or correcting segments up to a level of being understandable and appropriate for the use, while full PE usually means "human-like" translation (Hu and Cadwell, 2016:348). Hu and Cadwell (2016) give an overview of different guidelines sets for both light and full PE published by different LSPs. Further on, the tables of guidelines for both light and full PE will be introduced and analysed based on their suggestions and differences in the approach.

Table 1: Comparison of guidelines for light PE (Hu and Cadwell, 2016:349)

LIGHT POST-EDITING	TAUS (2016) (FIANAGAN & CHRISTENSEN, 2014)	O'BRIEN (2010)	MESA-LAO (2013)	DENSMER (2014)
Accuracy	TT communicates the same meaning as ST	Important	Important	Factually accurate
Terminology		No need to research	No need to spend too much time researching if incorrect	Be consistent
Grammar	May not be perfect	Not a big concern	No need to correct unless the information has not been fully delivered	Correct only the most obvious errors
Semantics	Correct			Correct
Spelling	Apply basic rules	Apply basic rules		
Syntax	Might be unusual	Can be ignored	Do not change	
Style	No need		No need	
Restructure	No need if the sentence is correct		No need if can be understood	Rewrite confusing sentences
Culture	Edit if necessary	Edit if necessary		
Information	Fully delivered			
Others	Use as much raw MT output as possible	Textual standards are not important; very high throughput expectation; low quality expectations	No need to change a word if correct	Fix machine-induced mistakes; delete unnecessary or extra machine-generated translation alternatives

Table 1 is a comparative overview of light PE guidelines. It is evident that accuracy and semantic correctness are extremely important, while grammar, syntax and style do not play much of a role. There are some discrepancies on account of terminology, with some models arguing for consistency, albeit without the need to research terminological accuracy.

Table 2: Comparison of guidelines for full PE (Hu and Cadwell, 2016:350)

FULL POST-EDITING	TAUS (2016)	O'BRIEN (2010)	FLANAGAN & CHRISTENSEN (2014)	MESA-LAO (2013)	DENSMER (2014)
Accuracy	TT communicates same meaning as ST	Important	Important		Absolutely accurate
Terminology	Key terminology is correct	Key terminology is correct	Key terminology is correct	Apply the term as used in the term database for any incorrect terminology	Consistent and appropriate
Grammar	Correct	Accurate	Correct	Correct	Correct
Semantics	Correct		Correct	Correct	Correct
Punctuation	Correct	Apply basic rules	Apply basic rules		Correct
Spelling	Apply basic rules	Apply basic rules	Apply basic rules		Correct
Syntax	Normal		Correct		Make modifications in accordance with practices for the TL
Style	Fine	Ignore stylistic and textual problems		Not important	Consistent, appropriate and fluent
Restructure			No need if the language is appropriate	No need if the sentence is semantically correct	Rewrite confusing sentences
Culture	Edit if necessary	Edit if necessary	Edit if necessary		Adapt all cultural references
Information	Fully delivered	Fully delivered	Fully delivered		
Formatting	Correct	All tags are present and in the correct positions	Ensure the same ST tags are present and in the correct positions;		Correct (including tagging)
Others	Basic rules apply to hyphenation; human translation quality	Apply basic rules to hyphenation; high throughput expectation; medium quality expectations	Use as much raw MT output as possible; ensure the untranslated terms belong to the client's list of 'Do not translate' terms	No need to change a word if it is correct; accept the repetitive MT output	Perfect faithfulness to the source text; fix machine-induced mistakes; delete unnecessary or extra machine-generated translation alternatives; cross-reference translations against other resources; human translation quality

Table 2 shows a comparative overview of full PE guidelines. The process of full PE entails punctuation and formatting of the text along with the other criteria included in the light PE. The table shows discrepancies concerning style, which then directly affects the quality of the product after PE – should the quality of the final product match the quality of a human translation or it is allowed to be of a slightly lower quality. It is also interesting that Mesa-Lao's guidelines leave most criteria open to discussion and do not set specific rules, while Densmer's guidelines have very strict criteria which raise the question of efficiency – If the quality needs to be the same for PEMT and HT, can PE still be cost-efficient?

3.2. Competences of Post-Editors

Competences required from professional post-editors are mostly the same as for professional translators: “translation competence, linguistic and textual competence, competence in research/information mining, cultural competence, technical competence, and domain competence” (Yamada, 2019:87). Nitzke et al. (2014) provide a figure of all competences a professional post-editor must “have at his or her dispense” in order to be suitable for the job required.

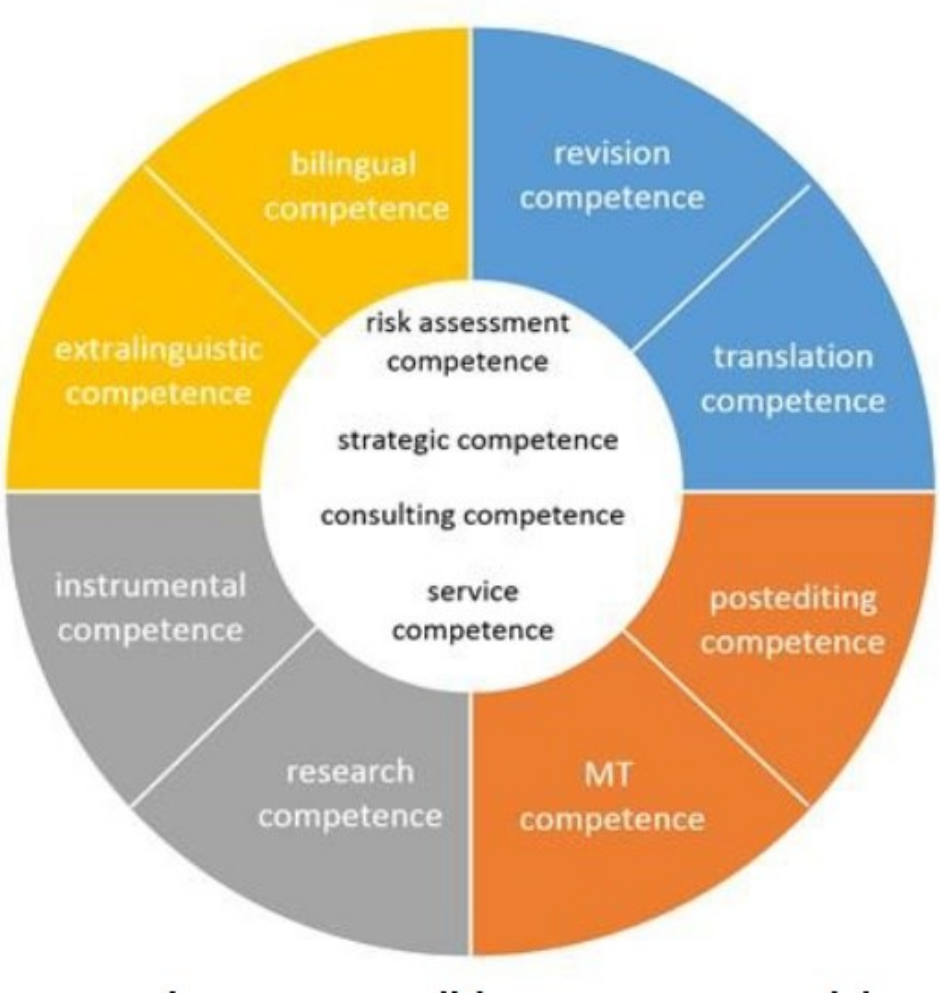

Figure 1. Competences of post-editors (Nitzke et al., 2019:250)

The core competences entail the risk assessment competence (the ability to assess risks and be able to choose the right translation strategy according to this), strategic competence (choosing the right level of PE), consulting competence (ability to communicate possible risks and strategies to the project manager and the client as well as justifying those and the price) and service competence (familiarity with the market, stepping up to the requirements, ability to set the time and price right, etc.) (Nitzke et al., 2019:247).

In addition to these competences the post-editor must:

- have proficient knowledge of both source and target language (bilingual competence),
- have the general knowledge and domain knowledge (extralinguistic competence),
- be familiar with CAT tools that MT system is a part of and be able to use simple text processing functions (instrumental competence),
- have efficient research strategies in order to quickly find reliable information not available in the MT system (research competence),
- acquire the ability to read the text written by others and distinguish between necessary and optional changes (revision competence),
- have knowledge on the process of translation itself, i.e. be familiar with domain-specific, stylistic, cultural and other differences (translation competence),
- know how MT systems work, how to assess the quality of the materials and even be able to improve training data (machine translation competence),
- be able to spot all types of mistakes a certain MT system could produce (post-editing competence). (Nitzke et al., 2019:249-50)

The attitude of project managers towards introducing PE as a service mentioned previously shows that professional translators do have necessary skills for the job, but they are lacking training and experience which would enable them to successfully cope with PE.

Memsource report on MT suggest training in PE as a good way of making translators more comfortable with MT. Such training had a positive feedback, by means of post-editors claiming that at first, it was hard to correct only what is necessary (making no subjective changes), but over a longer period of time, MT combined with post-editing strategies has shown higher efficiency (Memsource, 2020:13-14).

3.3. Efficiency of PE

There are numerous studies confirming the efficiency of PE. It has been proven that PE saves time, therefore also money, increases productivity and lowers the cognitive effort of post-editors compared to cognitive effort when translating from scratch.

As already mentioned, PE increases productivity, but studies have shown certain mismatches in some cases. Memsource favours PEMT as translation strategy because it doubles the level of productivity compared to HT (Memsource, 2020:3) in general, while the “productivity triples for short sentences and quadruples for long sentences, compared to translating from

scratch.“ (Memsource, 2020:10). Even though the productivity in general is higher than for the translation from scratch, the actual improvement depends on the language pair, MT product quality, different types of texts and it also varies between translators (Groves and Schmidtke, 430).

Surveys among translation studies students show that PEMT speeds up the process of non-professionals, too (Screen, 2019:135). There is a significant decrease in the amount of edits, but the same competence is required as for the translation from scratch (Yamada, 2019:87). Surveys show that there is a difference in time when post-editing domain-specific texts and general language texts. Domain-specific texts show better results concerning time than HT, while general language texts are not always faster post-edited than translated (Jia et al., 2019:62).

There are also various studies that suggest that cognitive effort is significantly lower for PE than translation from scratch. Jia et al. (2019) give Krings's (2001: 179) definition of cognitive effort as the "[…] type and extent of those cognitive processes that must be activated to remedy a given deficiency in a machine translation" (Jia et al., 2019:62). PE effort is directly connected with MT product quality, since it is the quality of the raw MT output that determines the level of PE required for the text to be fit for the purpose (Carl and Toledo Baez, 2019:110). "Records of gaze data reveal that the reading time of, and thus presumably also the allocation of cognitive resources to the source text and target text is very different in post-editing compared to from-scratch translation (Mesa-Lao 2014; Carl et al. 2015; Daems et al. 2017; da Silva et al. 2017)." (Jia et al., 2019:62) In other words, during PE, the focus is on the target text, while during translation from scratch the focus is on the source text (Jia et al., 2019:62). Another factor of the cognitive effort is the density and duration of pauses during the translation or PE. Longer duration and density suggest higher cognitive effort (Jia et al., 2019:63). Studies have shown that PE involves fewer pauses per word for both domain-specific and general language texts compared to the from-scratch translation. It is also observable that the density of pauses during PE varies, with general language texts demanding more pauses (Jia et al., 2019:69-70). This concludes that due to all these factors, PE poses less of a challenge than translating from scratch.

There are even some studies done on the end-users, since it is them after all who use the final product. MT outputs with different levels of post-editing as well as a translation-from-scratch were presented to the participants, who were asked to choose the best version of the translation. The analysis of the results shows that a high level of PE is not always the winning choice. Participants in some cases chose the text with lower level of PE. A study of the language pair English-Welsh has shown that there is no difference in the reading experience, readability and

comprehensibility of translated vs. post-edited text (Screen, 2019:147). In another research comparing reactions of end users to HT and PEMT, when asked to choose the best translation, participants chose the from-scratch translation, although there were no significant differences in accuracy and fluency. When asked about the reason for such a choice, the assessors named style as an important criterion.

Taking all of the elements into consideration, it is important to think thoroughly through, if the time, money and effort saved are the reasons enough to choose PEMT over HT, bearing in mind that the style obviously does matter to those who, in the end, will use the translation in question – end users.

4. Research on the Quality of MT Output - Purpose

As an addition to this master thesis, a short study on the efficiency of the MT and PE was conducted. All research papers used as literature in this thesis contained different studies and surveys conducted on different language pairs – English was combined with Chinese, Japanese, Spanish, French, some of the Nordic languages, German – and what all these languages have in common is a large number of native speakers, but also speakers in general, since they are popular international languages and people often take on learning them as their foreign language. These studies have shown that the quality of the MT output varies due to the language combination, the available data for that language pair, but also the MT system used and text type. Therefore, it was decided to take the language pair English-Croatian and use NMT as the most "human-like" system and analyse the data.

This research was inspired by the research conducted by Nataša Pavlović back in 2016. She decided to analyse the differences between HT and MT based on the type of errors found in both versions of translation. Participants were beginner translators and the MT system used for the purpose was Google Translate (which at the time still operated as SMT, while today it is an NMT system). She observed that the most common type of errors in HT are lexicological errors and in MT morphological or syntactical errors. She also concluded that, even considering that MT systems may not be free, it is still more efficient to post-edit an MT output than translate from scratch since HT requires PE too, which prolongs the process even more. Her study has shown that post-editing saves between 50-80% of time usually invested in conventional translation. (Pavlović, 2016: 5).

The presumption is that the MT output will be of a lower quality due to the fact that Croatian is considered to be a small language and therefore the corpus for MT training for the language pair English-Croatian may not be sufficient for good quality MT output for each text type. There are two purposes of the research: to compare the types of errors found in MT and HT versions with Pavlović's conclusion on the typical error types and to test the premise of the BLEU method, which suggests that the higher the score, the less post-editing is needed to improve the MT output quality in order to match that of the conventional translation. Therefore, this part of the thesis will compare the quality of the raw MT output and conventional translation, as well as the time of the process to compare the effort needed to post-edit both versions to give them the same level of quality and make them suitable for publishing.

4.1. Methods

For the purpose of this research four different text types were chosen (excerpts from a legal text, a newspaper article, a package leaflet of a medicine and a tourist board text, i.e. commercial text). The chosen texts are not culture-specific, the text types are familiar to the wide public and no translation into Croatian is available online. All text excerpts were taken from the official webpages and range from ca. 150 to ca. 250 words. The layout of the text is intact, if it was written in bullet points (like the package leaflet and the excerpt of the law) it was entered into the MT system in its original form. The system chosen for the process is Google Translate, which operates as an NMT system since 2016 and therefore should provide better quality output than before. Each translated segment was then compared to the original one (segments listed in a table) and the types of errors are briefly discussed for each text type. There are 7 steps of this research.

1) Translating the chosen excerpts using Google Translate and comparing the segments of the MT output to the original. The original texts can be found at the end of this thesis, in the Appendix.
2) Analysis of qualitative errors found in the raw MT output for each text.
3) Translation from scratch (HT) of each text and recording the time required. A beginner translator was asked to translate the original excerpts from scratch and measure the time needed for the process.
4) Post-editing of HT and MT and recording the time required. Another beginner translator post-edited the MT output and measured the time needed for the process. The aim of the PE process was to level the quality of the HT and MT version and to ensure that the final MT product would be appropriate for publishing.
5) Calculating the BLEU score for MT compared to the HT provided by the colleague translator using the BLEU calculator (available at https://www.letsmt.eu/Bleu.aspx)
6) Comparing the total time required for PEHT and PEMT
7) Drawing conclusion based on the analysis of differences in time required and the BLEU score.

4.2. Analysis of the raw MT output

This part of the thesis deals with the MT outputs generated by Google Translate and types of mistakes found. The following four chapters will each analyse the specific text and categorise the errors found in the text.

4.2.1. Commercial Text

Commercial texts are usually not suitable for MT because of their specific style, which leans more towards literary texts than technical texts. They often require creative solutions and therefore is HT a better option. Especially for that reason such a text was chosen to test the quality of the MT output. The text was taken from the official website of the city of Wolverhampton, England. The raw MT output follows:

Original text (English)	**MT output (Croatian)**
Wolverhampton...	*Wolverhampton ...*
A welcoming city centre to Enjoy	*Ugodno središte grada za uživanje,*
Wolverhampton is a vibrant City, perfectly located in the heart of the West Midlands.	*Wolverhampton je živahni grad, savršeno smješten u srcu West Midlands.*
Known for its rich cultural diversity and is a thriving centre for arts and entertainment.	*Poznat po bogatoj kulturnoj raznolikosti i uspješan je centar umjetnosti i zabave.*
In the Grand Theatre, it has the only traditional theatre in the Black Country.	*U Velikom kazalištu ima jedino tradicionalno kazalište u Crnoj zemlji.*
An art-house cinema and a gallery that boasts one of the finest collections of Pop Art in Europe and how could we forget to mention the home for the rising star of the English Premier League, Wolverhampton Wanderers FC.	*Art-house kino i galerija koja se može pohvaliti jednom od najboljih zbirki pop-arta u Europi i kako bismo mogli zaboraviti spomenuti dom za zvijezdu u usponu engleske Premier lige, Wolverhampton Wanderers FC.*
So whether you want to enjoy the arts, invest in some well-earned retail therapy, perhaps a special meal, a drink or just a great day and night out - this website provides all the information you need.	*Dakle, želite li uživati u umjetnosti, uložiti u neku dobro zarađenu maloprodajnu terapiju, možda u poseban obrok, piće ili samo u izvrstan dan i noć - ova web stranica nudi sve potrebne informacije.*
You'll find a wealth of friendly independent retailers, high street brands and markets in addition to two impressive purpose built shopping centres.	*Uz dva impresivna namjenska trgovačka centra pronaći ćete bogatstvo prijateljskih neovisnih trgovaca, robnih marki i tržnica.*
The city centre offering is an attractive one to visitors, residents and businesses alike.	*Ponuda centra grada privlačna je posjetiteljima, stanovnicima i poslovnim subjektima.*

Wolverhampton is a safe, clean and welcoming city in which to shop, visit and live.	*Wolverhampton je siguran, čist i gostoljubiv grad u kojem možete kupovati, posjećivati i živjeti.*

Compared to the original layout, the text is unchanged, all the sections of the original are preserved in the output. There are also no spelling or grammar errors and the punctuation is fine. The accuracy of the text is decent, the translation still communicates the same meaning as the original and presents the town in the same light.

There are no changes in the meaning, just a few clumsy translations which affect the style of the translation. There are several geographic or cultural terms, which could be unknown to an average Croatian reader due to geographic distance (the West Midlands, the Black Land, the Grand Theatre, Wolverhampton Wanderers FC). Some of them are just transferred from ST to TT without an explanation or apposition which would make clearer what this term really denotes, and others are translated word-for-word, which does not bring the meaning to the reader any closer either. There were also several wrong semantic choices (*dobro zarađenu* from *well-earned*, *noć* from *night out*, *maloprodajnu terapiju* from *retail therapy*), which do not alter the meaning of the original and do not affect readability and comprehensibility, but there certainly are better options.

The syntax of the output is fine, the system even made an inversion in the first sentence of the last section to make it more natural in the Croatian language. The syntax and grammar are somewhat compromised, not because of the system, but because of the lower quality of the original, where several sentences are verbless and sound rather clumsy, which then mirrors in the target text too.

4.2.2. Newspaper Article

For the newspaper article an excerpt from the Guardian's webpage was taken, an article reporting on the drop of retail sales in Great Britain in July. Here is the unedited translation followed by the analysis:

Original text (English)	**MT output (Croatian)**
Retail sales in Great Britain suffered a sharp unexpected fall in July after a mini-boom during football's European Championship a month earlier, according to official figures.	*Maloprodaja u Velikoj Britaniji pretrpjela je nagli neočekivani pad u srpnju nakon mini buma tijekom Europskog nogometnog*

	prvenstva mjesec dana ranije, prema službenim podacima.
The Office for National Statistics said the volume of retail sales fell by 2.5% between June and July as spending declined across much of the high street and food sales fell after the end of the tournament.	*Ured za nacionalnu statistiku priopćio je kako je obujam maloprodaje pao za 2,5% u razdoblju od lipnja do srpnja jer je potrošnja opala na većem dijelu glavnih ulica, a prodaja hrane pala je nakon završetka turnira.*
City economists had forecast a modest 0.4% rise on the month.	*Gradski ekonomisti predviđali su skroman rast od 0,4% u mjesecu.*
The drop in sales also came as consumers raised their spending in pubs, cafes, and restaurants at the expense of shopping for food and drink in supermarkets after the easing of coronavirus restrictions across the country.	*Do pada prodaje došlo je i kada su potrošači povećali potrošnju u pubovima, kafićima i restoranima na račun kupovine hrane i pića u supermarketima nakon ublažavanja ograničenja koronavirusa u cijeloj zemlji.*
The latest snapshot showed sales fell across almost every category, with a 4.4% drop in non-food stores, driven by weaker sales in clothes shops, secondhand stores and spending on computers and telecoms.	*Najnoviji snimak pokazao je da je prodaja pala u gotovo svim kategorijama, s padom od 4,4% u trgovinama neprehrambenim proizvodima, uzrokovanim slabijom prodajom u trgovinama odjeće, rabljenim trgovinama i potrošnjom na računala i telekome.*
In a potential sign that supply shortages caused by Covid-19 and Brexit are having an impact on retail sales, the ONS pointed to Bank of England evidence that showed there had been delays to shipments of electrical goods in recent months.	*U potencijalnom znaku da nestašice opskrbe uzrokovane Covid-19 i Brexitom utječu na maloprodaju, ONS je ukazao na dokaze Banke Engleske koji su pokazali da je bilo kašnjenja u isporuci električne robe posljednjih mjeseci.*
Heavy rainfall in early July also led to a fall in spending on fuel at petrol stations by 2.9% on the month amid a decline in traffic on Britain's roads.	*Obilne oborine početkom srpnja također su dovele do smanjenja potrošnje na gorivo na benzinskim postajama za 2,9% u mjesecu usred pada prometa na britanskim cestama.*

The layout of the text has remained unchanged and there are no errors of punctuation or spelling. Grammar is generally fine, there was only one instance of wrong case usage (*uzrokovane*

Covid-19, should have been *uzrokovane Covidom-19*), and one instance of a syntactic error in form of a prepositional phrase placed at the end of the sentence, which does not function in Croatian (*prema službenim podacima*).

The accuracy of the text is somewhat compromised because of several mistranslations. *Across much of the high street* was translated literally as *na većem dijelu glavnih ulica* (but it refers to fashion as *in high street fashion*), *easing of coronavirus restrictions* was translated as *ublažavanja ograničenja koronavirusa* (as if the virus was given more freedom). These also affect the semantics of the text along with some word-for-word translations (*secondhand stores – rabljenim trgovinama*, *in a potential sign – u potencijalnom znaku, supply shortages – nestašice opskrbe*). There is also *ONS* (Office for National Statistics) which remained untranslated, and there are some discrepancies concerning words that are officially foreign words but are used in everyday language: *mini bum* which is phoneticised and later in the text there is *pubovima* which does not correspond to Croatian.

4.2.3. Technical Text

For the domain of technical texts, a package leaflet of a medicine was chosen as it is hard to find a device or a tool available on our market but without a user manual translated into Croatian. The package leaflet for Adderall, a prescription medicine often misused by young people in America, was translated. Here is the translated excerpt:

Original text (English)	**MT output (Croatian)**
WHAT IS ADDERALL®?	*ŠTO JE ADDERALL®?*
ADDERALL® IS A CENTRAL NERVOUS SYSTEM STIMULANT PRESCRIPTION MEDICINE. IT IS USED FOR THE	*ADDERALL® JE CENTRALNI ŽIVČANI SUSTAV, STIMULANTNA MEDICINA OPISA. KORISTI SE ZA*
TREATMENT OF ATTENTION-DEFICIT HYPERACTIVITY DISORDER (ADHD).	*LIJEČENJE POREMEĆAJA HIPERAKTIVNOSTI POZDRAVLJANJA (ADHD).*
ADDERALL® MAY HELP	*ADDERALL® MOŽE POMOĆI*
INCREASE ATTENTION AND DECREASE IMPULSIVENESS AND HYPERACTIVITY IN PATIENTS WITH ADHD.	*POVEĆAJTE PAŽNJU I SMANJITE IMPULSNOST I HIPERAKTIVNOST U BOLESNIKA S ADHD –om.*

ADDERALL® SHOULD BE USED AS A PART OF A TOTAL TREATMENT PROGRAM FOR ADHD THAT MAY INCLUDE	*ADDERALL® TREBA KORISTITI KAO DIO UKUPNOG PROGRAMA LIJEČENJA ZA ADHD koji može uključivati*
COUNSELING OR OTHER THERAPIES.	*SAVJETOVANJE ILI DRUGE TERAPIJE.*
ADDERALL® IS ALSO USED IN THE TREATMENT OF A SLEEP DISORDER CALLED NARCOLEPSY.	*ADDERALL® SE KORISTI I U LIJEČENJU POREMEĆAJA SPAVANJA KOJI SE ZOVE NARKOLEPSIJA.*
ADDERALL® IS A FEDERALLY CONTROLLED SUBSTANCE (CII) BECAUSE IT CAN BE ABUSED	*ADDERALL® JE FEDERALNO UPRAVLJANA SNAGA (CII) JER SE MOŽE ZLOSTAVATI*
OR LEAD TO DEPENDENCE. KEEP ADDERALL® IN A SAFE PLACE TO PREVENT MISUSE AND	*ILI VODITE DO Ovisnosti. ČUVAJTE ADDERALL® NA SIGURNOM MJESTU DA SPRIJEČITE ZLOČINE I*
ABUSE. SELLING OF GIVING AWAY ADDERALL® MAY HARM OTHERS AND IS AGAINST THE LAW.	*ZLOSTAVLJANJE. PRODAJA DALJANJA ADDERALL® -a MOŽE ŠTETITI DRUGIMA I PROTIV JE ZAKON.*
Tell your doctor if you or your child have (or have a family history of) ever abused or been	*Obavijestite svog liječnika ako ste vi ili vaše dijete bili (ili imate obiteljsku povijest) zlostavljani ili ste bili zlostavljani*
dependent on alcohol, prescription medicines or street drugs.	*ovisno o alkoholu, lijekovima na recept ili uličnim drogama.*
WHO SHOULD NOT TAKE ADDERALL®?	*TKO NE SMIJE UZIMATI ADDERALL®?*
ADDERALL® SHOULD NOT BE TAKEN IF YOU OR YOUR CHILD:	*ADDERALL® se NE SMIJE UZIMATI AKO VI ILI VAŠE DIJETE:*
• have heart disease or hardening of the arteries	*• imate srčane bolesti ili otvrdnuće arterija*
• have moderate to severe high blood pressure	*• imate umjeren do jak visok krvni tlak*
• have hyperthyroidism	*• imate hipertireozu*
• have an eye problem called glaucoma	*• imate problem s očima koji se naziva glaukom*
• are very anxious, tense, or agitated	*• su jako zabrinuti, napeti ili uznemireni*
• have a history of drug abuse	*• imate povijest zlouporabe droga*

• are taking or have taken within the past 14 days an anti-depression medicine called a monoamine oxidase inhibitor or MAOI.	*• uzimate ili ste uzimali u posljednjih 14 dana lijek protiv depresije koji se naziva monoaminooksidaza inhibitor ili MAOI.*
• are sensitive to, allergic to, or had a reaction to other stimulant medicines	*• su osjetljivi na, alergični na ili su imali reakciju na druge stimulativne lijekove*
ADDERALL® IS NOT RECOMMENDED FOR USE IN CHILDREN LESS THAN 3 YEARS OLD.	*ADDERALL® se NE PREPORUČUJE ZA KORIŠTENJE U DJECE MANJE OD 3 GODINE.*

The system preserved the layout of the text – small/upper case and symptoms written out in bullet point form. There are no spelling or punctuation errors, but accuracy is a big issue in this translation, due to the layout of the original. The text was weirdly spaced and many sentences are divided into several lines, which poses a problem for the system since it was decided not to change the layout. The system could not recognize which segments belong together and translated them as separate phrases which caused change of the meaning and the use of unedited output in this case would be extremely dangerous.

The output definitely does not convey the same meaning as the original because of the numerous mistranslations present throughout the text. There are completely wrong translations (*central nervous system stimulant prescription medicine – centralni živčani sustav, preskriptivna medicina opisa, attention deficit hyperactivity disorder – poremećaja hiperaktivnosti pozdravljanja, impulsiveness – impulsnost, federally controlled substance – federativno upravljanje snaga, misuse and abuse – zločine I zlostavljanje, less than 3 years old – manje od 3 godine*), there is an non-existent word (*selling of giving away – prodaja daljanja*), and there are wrong word forms due to the inability of the system to pair segments together to provide the right translation. Instead of providing infinitive forms of the verbs, the system provided imperative forms (*povećajte pažnju i smanjite impulsnost*, while it all depends on the verb phrase *may help* from the previous line). These wrong word choices affect semantics and syntax of the text immensely, there is person and number inconsistency, and the whole text must be post-edited before use.

4.2.4. Legal Text

It was hard to find a legal text on Eurlex available in English but not in Croatian so the British laws were inspected in hope to find something published after Brexit and which does not

affect any other country. The website legislation.gov.uk entails all legal acts applied to the territory of the United Kingdom, those originating from the European Union and those applying to just one country of the UK only. Here is the translation of the excerpt from Wildlife and Natural Environment (Scotland) Act 2011, regulating i.e. prohibiting keeping invasive animals or plants:

Original text (English)	**MT output (Croatian)**
Prohibition on keeping etc. of invasive animals or plants	*Zabrana držanja i dr. Invazivnih životinja ili biljaka*
(1)Subject to the provisions of this Part, any person who keeps, has in the person's possession, or has under the person's control—	*(1) Podložno odredbama ovog dijela, svaka osoba koja drži, ima u posjedu osobe ili ima pod njenom kontrolom—*
(a)any invasive animal of a type which the Scottish Ministers, by order, specify; or	*(a) bilo koju invazivnu životinju tipa koju su škotski ministri, po nalogu, naveli; ili*
(b)any invasive plant of a type so specified, is guilty of an offence.	*(b) svaka invazivna biljka tako specificirane vrste je kriva za djelo.*
(2)An order under subsection (1) may make different provision for different cases and, in particular, for—	*(2) Naredba iz pododjeljka (1) može sadržavati različite odredbe za različite slučajeve, a posebno za -*
(a)different types of invasive animal or invasive plant;	*(a) različite vrste invazivnih životinja ili invazivne biljke;*
(b)different circumstances or purposes;	*(b) različite okolnosti ili svrhe;*
(c)different persons;	*(c) različite osobe;*
(d)different times of the year; and	*(d) različita doba godine; i*
(e)different areas or places.	*(e) različita područja ili mjesta.*
(3)Subject to subsection (4), it is a defence to a charge of committing an offence under subsection (1) to show that the accused took all reasonable steps and exercised all due diligence to avoid committing the offence.	*(3) Podložno pododjeljku (4), to je obrana optužbe za počinjenje kaznenog djela iz pododjeljka (1) kako bi se pokazalo da je optuženi poduzeo sve razumne korake i uložio svu dužnu revnost kako bi izbjegao počinjenje djela.*
(4)Where the defence provided by subsection (3) involves an allegation that the commission of the offence was due to the act or omission of another person, the person charged must	*(4) Ako obrana iz pododjeljka (3) uključuje navod da je počinjenje kaznenog djela posljedica djela ili propusta druge osobe, optužena osoba ne smije se, bez dopuštenja*

not, without leave of the court, be entitled to rely on the defence unless, within a period ending 7 days before the hearing, the person has served on the prosecutor a notice giving such information or assisting in the identification of the other person as was then in the person's possession.	*suda, osloniti na obrane, osim ako je u roku od 7 dana prije ročišta osoba uručila tužitelju obavijest koja daje takve podatke ili pomaže u identifikaciji druge osobe koja je tada bila u posjedu te osobe.*

The layout of the text remained intact and there is no need for formatting, there are also no spelling or punctuation errors. There are various grammatical errors, concerning case incongruence between bullet points (segment 1a written in dative – *bilo koju invazivnu životinju* and 1b then in accusative *svaka invazivna biljka*) and person incongruence (*životinju tipa koju su* and it should have been *koji su* since the reference word is *tipa*). Accuracy is somewhat compromised because of the complexity of the text which often results in system's poor choice of the right equivalent.

Although the simple legal terminology (e.g. *subject to, subsection, provisions, allegation*) is translated correctly, the more complex phrases, i.e. multi-word phrases are mistranslated (*guilty of offence – kriva za djelo, a defence to a charge – obrana optužbe, has in the person's possession – ima u posjedu osobe*) which also alters the meaning of the original and does not send out the same message. The syntax of the output is generally fine, but the style does need to be corrected to make it human like and appropriate for publishing.

4.3. The BLEU Scores

This chapter provides tables of compared raw MT and HT versions as well as the BLEU score for each segment and the score for the whole text excerpt. The scores are compared with the time needed for the whole process and listed in tables below.

Table 3. Representation of the BLEU scores for each segment in the commercial text

Raw MT output	**Translation from scratch**	**BLEU Score**
Wolverhampton ...	Wolverhampton...	100.00
Ugodno središte grada za uživanje.	Ugodan centar grada u kojem možete uživati.	8.46

Wolverhampton je živahni grad, savršeno smješten u srcu West Midlands. Poznat po bogatoj kulturnoj raznolikosti i uspješan je centar umjetnosti i zabave. U Velikom kazalištu ima jedino tradicionalno kazalište u Crnoj zemlji. Art-house kino i galerija koja se može pohvaliti jednom od najboljih zbirki pop-arta u Europi i kako bismo mogli zaboraviti spomenuti dom za zvijezdu u usponu engleske Premier lige, Wolverhampton Wanderers FC.	Wolverhampton je živahan grad koji se nalazi u srcu grofovije West Midlands. Poznat je po svojoj bogatoj kulturološkoj raznolikosti te je rastuće središte umjetnosti i zabave. Jedino tradicionalno kazalište u regiji Black Country nalazi se unutar kazališta Grand Theatre. U Wolverhamptonu nalazi se umjetničko kino i galerija koju krasi jedna od najboljih Pop Art kolekcija u Europi, a kako bismo mogli zaboraviti spomenuti da je ovo i dom zvijezde u usponu engleske Premier lige, nogometnog kluba Wolverhampton Wanderers FC.	26.25
Dakle, želite li uživati u umjetnosti, uložiti u neku dobro zarađenu maloprodajnu terapiju, možda u poseban obrok, piće ili samo u izvrstan dan i noć - ova web stranica nudi sve potrebne informacije.	Neovisno o tome želite li uživati u umjetnosti, upustiti se u zasluženu maloprodajnu terapiju, otići na posebnu večeru ili piće ili jednostavno želite imati odličan dan i noćni izlazak – ova web-stranica pruža sve informacije koje su Vam potrebne.	27.92
Uz dva impresivna namjenska trgovačka centra pronaći ćete bogatstvo prijateljskih neovisnih trgovaca, robnih marki i tržnica. Ponuda centra grada privlačna je posjetiteljima, stanovnicima i poslovnim subjektima. Wolverhampton je siguran, čist i gostoljubiv grad u kojem možete kupovati, posjećivati i živjeti.	Pronaći ćete mnoštvo ljubaznih nezavisnih trgovaca na malo, poznate high street brendove i tržnice te, osim toga, dva dojmljiva trgovačka centra. Centar grada ima atraktivnu ponudu za sve posjetitelje, građane i poduzeća. Wolverhamtpon je siguran, čist i ugodan grad kojeg možete	18.25

	posjetiti, gdje možete otići u kupovinu i živjeti u njemu.	
		The average BLEU score: **23.65**

Table 4. Representation of the BLEU scores for each segment in the newspaper excerpt

Raw MT output	**Translation from scratch**	**BLEU Score**
Maloprodaja u Velikoj Britaniji pretrpjela je nagli neočekivani pad u srpnju nakon mini buma tijekom Europskog nogometnog prvenstva mjesec dana ranije, prema službenim podacima.	Prema službenim podacima, maloprodaja u Velikoj Britaniji pretrpjele su oštar i neočekivan pad u srpnju, mjesec dana nakon kratkog ekonomskog rasta tijekom Europskog prvenstva u nogometu.	24.68
Ured za nacionalnu statistiku priopćio je kako je obujam maloprodaje pao za 2,5% u razdoblju od lipnja do srpnja jer je potrošnja opala na većem dijelu glavnih ulica, a prodaja hrane pala je nakon završetka turnira. Gradski ekonomisti predviđali su skroman rast od 0,4% u mjesecu.	Državni zavod za statistiku Ujedinjenog Kraljevstva tvrdi da je volumen maloprodaje pao za 2,5% između lipnja i srpnja, uzimajući u obzir da je potrošnja opala u većini trgovina odjećom i prehrambenim proizvodima opala nakon završetka prvenstva. Lokalni ekonomisti bili su predvidjeli skroman rast od 0,4% u tom mjesecu.	22.77
Do pada prodaje došlo je i kada su potrošači povećali potrošnju u pubovima, kafićima i restoranima na račun kupovine hrane i pića u supermarketima nakon ublažavanja ograničenja koronavirusa u cijeloj zemlji.	Smanjenje opsega prodaje također se dogodilo jer su potrošači povećali potrošnju u gostionicama, kafićima i restoranima nauštrb kupovine hrane i pića u supermarketima nakon popuštanja mjera protiv koronavirusa diljem države.	47.76
Najnoviji snimak pokazao je da je prodaja pala u gotovo svim kategorijama, s padom od 4,4% u	Najnoviji snimak pokazao je smanjenje opsega prodaje u gotovo svakoj kategoriji, uključujući pad od 4,4% u	33.14

trgovinama neprehrambenim proizvodima, uzrokovanim slabijom prodajom u trgovinama odjeće, rabljenim trgovinama i potrošnjom na računala i telekome.	neprehrambenim trgovinama, koji je pokrenut slabijom prodajom u trgovinama odjećom, trgovinama rabljenom odjećom te slabijom potrošnjom na računala i teleoperatere.	
U potencijalnom znaku da nestašice opskrbe uzrokovane Covid-19 i Brexitom utječu na maloprodaju, ONS je ukazao na dokaze Banke Engleske koji su pokazali da je bilo kašnjenja u isporuci električne robe posljednjih mjeseci.	Kao potencijalni znak da nestašica zaliha uzrokovana Covidom i Brexitom ima utjecaj na maloprodaju, Državni zavod za statistiku Ujedinjenog Kraljevstva ističe navode Engleske Banke, koji pokazuju da je došlo do zakašnjenja isporuke električnih dobara u posljednjih nekoliko mjeseci.	7.22
Obilne oborine početkom srpnja također su dovele do smanjenja potrošnje na gorivo na benzinskim postajama za 2,9% u mjesecu usred pada prometa na britanskim cestama.	Obilne kiše početkom srpnja također su dovele do pada u potrošnji goriva od 2,9% na mjesečnoj bazi uslijed smanjenja prometa na britanskim cestama.	41.08
		The average BLEU score: **29.84**

Table 5. Representation of the BLEU scores for each segment in the package leaflet

Raw MT output	**Translation from scratch**	**BLEU Score**
ŠTO JE ADDERALL®?	ŠTO JE ADDERALL®?	53.73
ADDERALL® JE CENTRALNI ŽIVČANI SUSTAV, STIMULANTNA MEDICINA OPISA. KORISTI SE ZA	ADDERALL® JE LIJEK NA RECEPT U OBLIKU STIMULANSA ZA SREDIŠNJI ŽIVČANI SUSTAV. KORISTI SE ZA	30.46
LIJEČENJE POREMEĆAJA HIPERAKTIVNOSTI POZDRAVLJANJA (ADHD). ADDERALL® MOŽE POMOĆI	LIJEČENJE POREMEĆAJA HIPERAKTIVNOSTI I DEFICITA PAŽNJE (ADHD). ADDERALL® MOŽE POMOĆI	56.88

POVEĆAJTE PAŽNJU I SMANJITE IMPULSNOST I HIPERAKTIVNOST U BOLESNIKA S ADHD -om.	PRI POVEĆANJU POZORNOSTI I SMANJENJU IMPULZIVNOSTI I HIPERAKTIVNOSTI KOD PACIJENATA S ADHD-OM.	24.62
ADDERALL® TREBA KORISTITI KAO DIO UKUPNOG PROGRAMA LIJEČENJA ZA ADHD koji može uključivati	ADDERALL® TREBA KORISTITI KAO DIO UKUPNOG PROGRAMA LIJEČENJA ADHD-A KOJI MOŽE UKLJUČIVATI	71.41
SAVJETOVANJE ILI DRUGE TERAPIJE.	SAVJETOVANJE ILI DRUGE TERAPIJE.	100.00
ADDERALL® SE KORISTI I U LIJEČENJU POREMEĆAJA SPAVANJA KOJI SE ZOVE NARKOLEPSIJA.	ADDERALL® SE TAKOĐER KORISTI U LIJEČENJU POREMEĆAJA SPAVANJA TJ. NARKOLEPSIJE.	43.91
ADDERALL® JE FEDERALNO UPRAVLJANA SNAGA (CII) JER SE MOŽE ZLOSTAVATI	ADDERALL® JE FEDERALNO KONTROLIRANA SUPSTANCA (CII) JER SE MOŽE ZLOUPORABITI	53.28
ILI VODITE DO Ovisnosti. ČUVAJTE ADDERALL® NA SIGURNOM MJESTU DA SPRIJEČITE ZLOČINE I	ILI DOVESTI DO OVISNOSTI. ADDERALL® DRŽITE NA SIGURNOM MJESTU KAKO BI SE IZBJEGLA ZLOUPOTREBA.	14.40
ZLOSTAVLJANJE. PRODAJA DALJANJA ADDERALL® -a MOŽE ŠTETITI DRUGIMA I PROTIV JE ZAKON.	PRODAJA ILI RASPAČAVANJE ADDERALLA® MOŽE UGROZITI DRUGE TE JE PROTUZAKONITO.	9.31
Obavijestite svog liječnika ako ste vi ili vaše dijete bili (ili imate obiteljsku povijest) zlostavljani ili ste bili zlostavljani, ovisno o alkoholu, lijekovima na recept ili uličnim drogama.	Obratite se liječniku ukoliko ste Vi ili Vaša djeca zloupotrebljavali (ili imate anamnezu) ili bili ovisni o alkoholu, lijekovima na recept ili uličnim drogama.	43.29

TKO NE SMIJE UZIMATI ADDERALL®?	KADA NE UZIMATI ADDERALL®?	15.62
ADDERALL® se NE SMIJE UZIMATI AKO VI ILI VAŠE DIJETE:	ADDERALL® NEMOJTE UZIMATI UKOLIKO VI ILI VAŠA DJECA:	20.45
• imate srčane bolesti ili otvrdnuće arterija	- imate srčane bolesti ili aterosklerozu	46.17
• imate umjeren do jak visok krvni tlak	- imate umjereno povišen ili jako povišen krvni tlak	7.03
• imate hipertireozu	- imate hipertireozu	63.89
• imate problem s očima koji se naziva glaukom	- imate glaukom	4.46
• su jako zabrinuti, napeti ili uznemireni	- imate problema s uznemirenošću ili napetošću	3.80
• imate povijest zlouporabe droga	- imate povijest zlouporabe droge	39.76
• uzimate ili ste uzimali u posljednjih 14 dana lijek protiv depresije koji se naziva monoaminooksidaza inhibitor ili MAOI.	- uzimate ili ste uzimali u posljednjih 14 dana antidepresiv inhibitor monoaminooksidaze ili MAOI.	42.46
• su osjetljivi na, alergični na ili su imali reakciju na druge stimulativne lijekove	- ako ste osjetljivi, alergični ili ste imali reakciju na druge medicinske stimulanse	30.28
ADDERALL® se NE PREPORUČUJE ZA KORIŠTENJE U DJECE MANJE OD 3 GODINE.	ADDERALL® SE NE PREPORUČUJE DJECI MLAĐOJ OD 3 GODINE.	35.41
		The average BLEU score: **37.31**

Table 6. Representation of the BLEU scores for each segment in the law excerpt

Raw MT output	**Translation from scratch**	**BLEU score**
Zabrana držanja i dr. Invazivnih životinja ili biljaka	Zabrana držanja itd. invazivnih životinja ili biljaka	63.40

(1) Podložno odredbama ovog dijela, svaka osoba koja drži, ima u posjedu osobe ili ima pod njenom kontrolom -	(1)U skladu s odredbama ovog Dijela, osoba koja drži, posjeduje ili upravlja-	32.11
(a) bilo koju invazivnu životinju tipa koju su škotski ministri, po nalogu, naveli; ili	(a)invazivnom životinjom vrste koju škotski ministri, po nalogu, navode; ili	38.09
(b) svaka invazivna biljka tako specificirane vrste je kriva za djelo.	(b)invazivnom biljkom navedene vrste, kriva je za kazneno djelo.	15.40
(2) Naredba iz pododjeljka (1) može sadržavati različite odredbe za različite slučajeve, a posebno za -	(2)Naredba iz pododjeljka (1) može propisati različite odredbe u različitim slučajevima i, poglavito, za-	49.93
(a) različite vrste invazivnih životinja ili invazivne biljke;	(a)različite vrste invazivnih životinja ili invazivnih biljaka;	77.44
(b) različite okolnosti ili svrhe;	(b)različite okolnosti i svrhe;	70.71
(c) različite osobe;	(c)različite osobe;	100.00
(d) različita doba godine; i	(d)različita doba godine; i	100.00
(e) različita područja ili mjesta.	(e)različita područja i mjesta.	76.92
(3) Podložno pododjeljku (4), to je obrana optužbe za počinjenje kaznenog djela iz pododjeljka (1) kako bi se pokazalo da je optuženi poduzeo sve razumne korake i uložio svu dužnu revnost kako bi izbjegao počinjenje djela.	(3)U skladu s pododjeljkom (4), obrana od optužbe za počinjeno kazneno djelo prema pododjeljku (1) važeća je ukoliko je optuženik poduzeo sve razumne korake te savjesno izbjegao počinjenje kaznenog djela.	31.54
(4) Ako obrana iz pododjeljka (3) uključuje navod da je počinjenje kaznenog djela posljedica djela ili propusta druge osobe, optužena osoba ne smije se, bez dopuštenja suda, osloniti na obrane, osim ako je u roku od 7 dana prije ročišta osoba uručila tužitelju obavijest koja daje takve podatke ili	(4)Ukoliko obrana predviđena pododjeljkom (3) uključuje navod da je počinjenje kaznenog djela uzrokovano radnjom ili propustom druge osobe, optuženik ne smije bez sudskog odobrenja pozivati se na obranu, osim ako u razdoblju koje završava 7 dana prije saslušanja, optuženik ne	27.82

pomaže u identifikaciji druge osobe koja je tada bila u posjedu te osobe.	obavijesti tužitelja o stvarnom stanju ili pomogne pri identifikaciji druge osobe.	
		The average BLEU score: **42.14**

The results show that the two excerpts which seemed the most demanding based on the type and the amount of errors, namely package leaflet and law excerpt, actually have the highest similarity to the conventional translation and therefore should the whole process be less demanding than for the two other text types.

When talking about the time required for the process, the results show that it is faster to post-edit the MT output than translate from scratch. Table 7 and Table 8 show the time measured for the process of translation from scratch and for the process of post-editing of MT outputs. Since it only takes a few seconds for an MT system to provide translation, the table concerning PEMT shows only the duration of post-editing. However, translation from scratch also requires post-editing to avoid possible errors that occurred during the process of translation so the table concerning HT has a column with the time required for revision and another column with the total amount of time invested in the process. Duration in all tables is expressed in minutes and seconds.

Table 7. Duration of PEMT

Time measured for post-editing MT output	
Commercial text	9:25
Newspaper article	11:26
Package leaflet	11:37
Law excerpt	10:50

Table 8. Duration of HT and PE

Time measured for translation from scratch			
	Translation	Post-editing	Total time
Commercial text	37:30	5:29	42:59
Newspaper article	34:44	7:33	42:16
Package leaflet	30:28	8:28	38:56
Law excerpt	52:09	9:25	61:34

Table 9 shows the amount of time saved by post-editing MT outputs instead of translating these excerpts from scratch. The results confirm Pavlović's statement that post-editing saves between 50-80% of time. However, the results do not confirm the premise of the BLEU method that the higher similarity score means less work to improve quality. Even though the commercial text had the lowest score (only 23.65), it also required the least amount of time for the process of post-editing, while the law excerpt which had the highest score (42.14) took almost as much time as the other two text types with lower scores. In this case, the duration is expressed in seconds to simplify the process of calculating the duration and time saved.

Table 9 Time saved in PEHT vs. PEMT

Difference in time needed for the whole process			
	PEHT	PEMT	Time saved
Commercial text	2579 seconds	565 seconds	2017 seconds (78% less time)
Newspaper article	2536 seconds	686 seconds	1850 seconds (73% less time)
Package leaflet	2336 seconds	697 seconds	1639 seconds (70% less time)
Law excerpt	3694 seconds	650 seconds	3044 seconds (82% less time)

4.4. Final Comparison

The purpose of the research was to compare differences in errors occurring in HT and MT as well as the time required for the whole process. The analysis has confirmed Pavlović's thesis that lexical errors are more likely for conventional translation, while morphological or syntactical errors are more common for MT. In HT not many interventions were necessary due to grammar or syntax, a few of them concerned formatting or punctuation, but most of them concerned lexical choices. Errors in MT outputs were more complex and will be more closely elaborated.

When talking about the impact of a text type on the quality of the MT output, it is observable that it does not affect quality of the MT output much. The commercial text, which, according to other studies, should have been of the poorest quality is indeed of pretty good quality, while the technical text and the legal text, which are more suitable for MT because of the repetitions and fixed translations, were of pretty low quality and even changed the meaning of the original. The reason for such quality results may be due to the fact that the commercial text did not contain many creative sections. There were no complex constructions, which is helpful since research has shown that shorter and simpler sentences are of better quality than longer and complex ones (the reason why the machine translations of technical and legal texts are of lower quality). The layout of the text could possibly be a great factor for the quality since the system cannot

recognise which segments belong together and translate them as a whole. The quality would have been better with different spacing because the system could recognize the whole verb phrase and it would not be translated separately just because it is separated into two lines.

Spelling, formatting and punctuation were not an issue for the system, it even "intelligent" changed the decimal point in English to the decimal comma in Croatian when dealing with percentages in the newspaper excerpt.

Accuracy varied from very good for the commercial text to very bad for the Adderall description, the same is true of semantic choices. In the less complex texts (commercial and newspaper excerpts), semantic choices were decent, they did not denote the exact same thing as the original but they were similar, while in the more texts (legal and technical texts) they bordered with extremely bad and the meaning was at times completely different from the original, which also affects readability and comprehensibility of the translation.

There are not many inconsistencies concerning syntax, sentences were pretty decently arranged, but grammar posed a bit of a problem in some cases as there were some incongruences in case/person/number.

The style of the MT product was decent for the commercial text, but was lacking for other text types. The MT outputs are informative enough, but they can be used only for internal purposes, otherwise the style needs polishing to make them publishable and more human-like.

The time required for the process goes in favour of MT. The results confirmed Pavlović's study on the efficiency of MT and proved that post-editing of MT output can save up to 80% of time usually invested in conventional translation. However, there are some discrepancies in accordance of the BLEU score and the time required for post-editing of MT outputs. Even though the BLEU score varies from 23 to 42, the time measured does not change much, it varies from 9 to 12 minutes. The method suggests that the higher the score, the less post-editing the output requires to match the human version, but the time measured in this research does not confirm it. Although the time does not change much, the amount of post-editing done might have confirmed the BLEU method. Further research could focus on the number of edits in the process instead of the time needed and analyse the results to see if the higher BLEU score really denotes less PE effort and an even higher efficiency of this process.

Conclusion

Back in 1985, Slocum presumed in his article that MT will grow continuously and it will become widely accepted because it shortens the time needed for translation, lowers the cost of translation and "reduces the burden on human translators" (Slocum, 1985:115) and he was right. Machine translation and post editing are claiming their place in the language industry and are becoming more and more prominent as a language service.

MT has gone a long way from the beginning as RBMT, operating based on a set of rules for a specific language and SMT dependent on reference translations to choose the best equivalent, all the way to hybrid systems and NMT, which has a potential of providing almost human-like quality. It can be described as boring, as a non-reflexive use of language because of repetitions and fixed translations, it cannot take context into account and does not understand metaphorical meaning, it produces different types of errors, and PE is obligatory before putting such a text to use, but MT and PE do have many advantages as well. They can take the burden of translating repetitive content with fixed translations produced by human translators. MT has an incredible capacity for development, which means better quality and less errors in the long run. With the right system training methods, and most importantly, with the right combination of MT quality, language combination and text type, it does save time and money and reduces the cognitive effort while enhancing productivity and efficiency.

The key is being able to distinguish when PEMT is more efficient than HT, taking into consideration the possible risks and knowing the terms of the service required, such as the deadline and the pricing. In case of choosing PEMT, it is important not to "over-do", i.e. to distinguish between necessary changes and those that are subjective and under consideration for change just because it would sound better. Such changes may improve the style but they put in danger the efficiency of the whole process and HT may show as the better and cheaper version.

The research has also shown that MT has a good potential for smaller languages, too. Training a system and feeding it a larger amount of translated data could surely produce a better quality output than Google Translate since it "collects" its data from all over the Internet, which also means including amateur translations which can be of significantly lower quality than the professional ones.

The question that remains unanswered is the future of the language industry. "The age of MT is here, and it's here to stay." (Memsource, 2020:18). Translators are aware of this, but most of them still frown upon MT and PE for two reasons. Some of them diminish its potential due to

the fact that human intervention is still obligatory and the others are concerned because of the ever growing demand for this approach. Introducing PE courses to translation studies programmes and trainings for professional translators could be the answer to it. They would make students and translators more familiar with the way MT functions and more comfortable with revising and correcting its product to make it acceptable for the end users. With the present state of the things, changes in the translator profession seem inevitable.

Works cited

Adaptive Neural Machine Translation. Available at https://translated.com/adaptive-neural-machine-translation (Visited on 6 September 2021).

Barrault, Joïc; Blesialka, Magdalena; Bojar, Ondřej; Costa-Jussà, Marta R.; Federmann, Christian; Graham, Yvette; Grundkiewicz, Roman; Haddow, Barry; Huck, Matthias; Joanis, Eric; Kocmi, Tom; Koehn, Philipp; Lo, Chi-kiu; Ljubešić, Nikola; Monz, Christof; Morishita, Makoto; Nagata, Masaaki; Nakazawa, Toshiaki; Pal, Santanu; Post, Matt; Zampieri, Marcos (2020). Findings of the 2020 Conference on Machine Translation (WMT20). *Proceedings of the 5th Conference on Machine Translation (WMT):*1-55. Available at: https://aclanthology.org/2020.wmt-1.1.pdf (Visited on 6 September 2021).

Carl, Michael; Toledo Baez, M. Christina (2019). Machine Translation Errors and the Translation Process: a Study across Different Languages. *The Journal of Specialised Translation* 31: 107-132. Available at

https://www.researchgate.net/publication/335929192_Machine_translation_errors_and_the_translation_process_A_study_across_different_languages (Visited on 10 July 2021).

De Jesus Martins, Debora Beatriz; de Medeiros Caseli, Helena (2015). *Automatic Machine Translation Error Identification.* Machine Translation. Volume 29, Number 1 (2015), 1-24. Springer Science+Business Media Dordrecht. Available at

https://www.jstor.org/stable/44113783 (Visited on 2 November 2020).

Doherty, Stephen (2016). The Impact of Translation Technologies on the Process and Product of Translation. *International Journal of Communication* 10: 947-969. Available at https://www.researchgate.net/publication/284725157_The_impact_of_translation_technologies_on_the_process_and_product_of_translation (Visited on 2 November 2020).

Eetemadi, Sauleh; Lewis, William; Toutanova, Kristina; Radha, Hayder (2015). *Survey of Data-Selection Methods in Statistical Machine Translation.* Machine Translation. Volume 29, Number 3/4 (2015), 189-223. Springer Science+Business Media Dordrecht. Available at https://www.jstor.org/stable/44113790 (Visited on 2 November 2020).

Forcada, Mikel L. (2017). Making Sense of Neural Machine Translation. *Translation Spaces* 6:2 (2017) 291–309, DOI. Available at https://www.dlsi.ua.es/~mlf/docum/forcada17j2.pdf (Visited on 10 July 2021).

Gonzalez-Rubio, Jesus; Navarro-Cerdán, Jose Ramón; Casacuberta, Francisco (2013). *Dimensionality Reduction Methods for Machine Translation Quality Estimation.* Machine Translation. Volume 27, Number 3/4 (2013). 281 – 301 Springer Science+Business Media Dordrecht. Available at

https://www.jstor.org/stable/42628819 (Visited on 2 November 2020).

Groves, Declan; Schmidtke, Dags (2009). *Identification and Analysis of Post-Editing Patterns for MT.* MT Summit XII: Proceedings of the Twelfth Machine Translation Summit, 429–436. Ottawa, Canada. Available at

https://www.researchgate.net/publication/228789664_Identification_and_Analysis_of_Post-Editing_Patterns_for_MT (Visited on 5 June 2021).

Hu, Ke; Cadwell, Patrick (2016). *Comparative Study of Post-editing Guidelines.* Baltic J. Modern Computing, Volume 4, Number 2 (2016), 346-353. ADAPT Centre, Dublin City University. Available at

https://www.researchgate.net/publication/303564681_A_Comparative_Study_of_Post-editing_Guidelines (Visited on 5 June 2021).

Hutchins, W. John; Sommers, Harold L. (1992). *An Introduction to Machine Translation.* London: Academic Press Limited. Available at

https://ug1lib.org/book/1270107/bea9d7?id=1270107&secret=bea9d7&dsource=recommend (Visited on 20 March 2021).

Jia, Yanfang; Carl, Michael; Wang, Xiangling; (2019). How Does the Post-Editing of Neural Machine Translation Compare with From-Scratch Translation? A Product and Process Study. *The Journal of Specialised Translation* 31: 60-86. Available at https://www.jostrans.org/issue31/art_jia.pdf (Visited on 10 July 2021).

Kadiu, Silvia (2019) "Human vs. Machine Translation: Henri Meschonnic's poetics of Translating." *Reflexive Translations Studies,* edited by Timothy Mathews, UCL Press, 2019, 71-94. Available at https://www.jstor.org/stable/j.ctv6q5315.9 (Visited on 3 April 2021).

Macketanz, Vivien; Avramidis, Eleftherios; Burchardt, Aljoscha; Helcl, Jindrich; Srivastava, Ankit (2017). *Machine Translation: Phrase-Based, Rule-Based and Neural Approaches*

with Linguistic Evaluation. Cybernetics and Information Technologies. Volume 17, Number 2 (2017), 28-43. Sofia. Available at

https://www.researchgate.net/publication/318184162_Machine_Translation_Phrase-Based_Rule-Based_and_Neural_Approaches_with_Linguistic_Evaluation/link/595c170b0f7e9bf415b4acb8/download (Visited on 10 July 2021).

Nitzke, Jean; Hansen-Schirra, Silvia; Canfota, Carmen (2019). Risk Management and Post-Editing Competence. *The Journal of Specialised Translation* 31: 239-259. Available at

https://www.jostrans.org/issue31/art_nitzke.pdf (Visited on 10 July 2021).

Okpor, M.D. (2014). *Machine Translation Approaches: Issues and Challenges*. IJCSI International Journal of Computer Science Issues, Volume 11, Number 2 (2014), 159-165. Available at

https://www.academia.edu/20654928/Machine_Translation_Approaches_Issues_and_Challenges (Visited on 3 April 2021).

Panić, Milica (2020). Automated MT Evaluation Metrics. Available at https://blog.taus.net/automated-mt-evaluation-metrics (Visited on 6 September 2021).

Pavlović, Nataša (2016). "Strojno i konvencionalno prevođenje s engleskoga na hrvatski: usporedba pogrešaka." *Jezik kao predmet proučavanja i jezik kao predmet poučavanja*, edited by Diana Stolac and Anastazija Vlastelić, HDPL, 2017, 279-295. Zagreb. Available at

https://www.researchgate.net/publication/329184636_Strojno_i_konvencionalno_prevodenje_s_engleskoga_na_hrvatski_usporedba_pogresaka (Visited on 6 September 2021)

Popović, Maja; Avramidis, Eleftherios; Burchardt, Aljoscha; Hunsicker, Sabine; Schmeier, Sven; Tscherwinka, Cindy; Vilar, David; Uszkoreit, Hans (2014). *Involving Language Professional in the Evaluation of Machine Translation.* Language Resources and Evaluation. Volume 48, Number 4 (2014), 541-559. Springer Science+Business Media Dordrecht. Available at https://www.jstor.org/stable/24710064 (Visited on 2 November 2020).

Sakamoto, Akiko (2019). Why Do Many Translators Resist Post-Editing? A Sociological Analysis Using Bourdieu's Concepts. *The Journal of Specialised Translation* 31: 201-216. Available at https://www.jostrans.org/issue31/art_sakamoto.pdf (Visited on 10 July 2021).

Screen, Benjamin (2019). What Effect Does Post-Editing Have on the Translation Product from an End-User's Perspective? *The Journal of Specialised Translation* 31: 134-157. Available at https://www.jostrans.org/issue31/art_screen.pdf (Visited on 10 July 2021).

Slocum, Jonathan (1985). *Machine Translation*. Computers and the Humanities. Volume 19, Number 2 (1985), 109-116. Paradigm Press Inc. Available at https://www.jstor.org/stable/30204397 (Visited on 2 November 2020).

Specia, Lucia; Raj, Dhwaj; Turchi, Marco. (2010). *Machine Translation Evaluation versus Quality Estimation.* Machine Translation. Volume 24, Number 1 (2010), 39-50. Springer Science+Business Media B.V. Available at https://www.jstor.org/stable/40926411 (Visited on 2 November 2020).

Tan, Zhixing; Wang, Shuo; Yang, Zonghan; Chen, Gang; Huang, Xuancheng; Sun, Maosong; Liu, Yang (2020) Neural machine translation: A Review of Methods, Resources and Tools. *AI Open 1*: 5-21. Available at

https://reader.elsevier.com/reader/sd/pii/S2666651020300024?token=3FCB4C89042961C74A93E8BD78D9D4409C48C2F83C2CCC786DEDE5EE643B57A8E27C802F995B59485158DABA1EE08523&originRegion=eu-west-1&originCreation=20210912130041 (Visited on 6 September 2020).

Thurmair, Gregor (2005). *Hybrid Architectures for Machine Translation Systems.* Language Resources and Evaluation. Volume 39, Number 1 (2005), 91-108. Springer Science+Business Media Dordrecht. Available at https://www.jstor.org/stable/30200544 (Visited on 2 November 2020).

What is Interactive, Adaptive MT? Available at https://support.lilt.com/hc/en-us/articles/360058203753-What-is-Interactive-Adaptive-MT- (Visited on 6 September 2021).

Wu, Yonghui, Schuster, Mike; Chen, Zhifeng; Le, Quoc V.; Norouzi, Mohammad (2016). *Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation.* Available at https://arxiv.org/pdf/1609.08144.pdf (Visited on 10 July 2021).

Yamada, Masaru (2019). The impact of Google Neural Machine Translation on Post-editing by Student Translators. *The Journal of Specialized Translation* 31: 87-106. Available at https://www.researchgate.net/publication/330831614_The_impact_of_Google_Neural_Machine_Translation_on_Post-editing_by_student_translators (Visited on 10 July 2021).

European Language Industry Survey 2020 Before and After Covid-19 (June 2020). Available at https://ec.europa.eu/info/sites/default/files/2020_language_industry_survey_report.pdf (Visited on 6 September 2021).

How to Unlock the Potential of Machine Translation. Your Guide to Implementing MT Post-Editing (2020). Memsource. Available at

https://www.memsource.com/uploads/2020/11/13/memsource-unlock-the-potential-of-mt-with-post-editing-guide.pdf (Visited on 20 March 2021).

Appendix: Excerpts from the original texts

1. Commercial text (downloaded from https://enjoywolverhampton.com/ on 20 August 2021)

Wolverhampton...

A welcoming city centre to Enjoy.

Wolverhampton is a vibrant City, perfectly located in the heart of the West Midlands. Known for its rich cultural diversity and is a thriving centre for arts and entertainment. In the Grand Theatre, it has the only traditional theatre in the Black Country. An art-house cinema and a gallery that boasts one of the finest collections of Pop Art in Europe and how could we forget to mention the home for the rising star of the English Premier League, Wolverhampton Wanderers FC.

So whether you want to enjoy the arts, invest in some well-earned retail therapy, perhaps a special meal, a drink or just a great day and night out - this website provides all the information you need.

You'll find a wealth of friendly independent retailers, high street brands and markets in addition to two impressive purpose built shopping centres. The city centre offering is an attractive one to visitors, residents and businesses alike. Wolverhampton is a safe, clean and welcoming city in which to shop, visit and live.

2. Newspaper article (downloaded from https://www.theguardian.com/business/2021/aug/20/retail-sales-in-great-britain-fall-as-people-turn-to-dining-out on 20 August 2021)

Retail sales in Great Britain suffered a sharp unexpected fall in July after a mini-boom during football's European Championship a month earlier, according to official figures.

The Office for National Statistics said the volume of retail sales fell by 2.5% between June and July as spending declined across much of the high street and food sales fell after the end of the tournament. City economists had forecast a modest 0.4% rise on the month.

The drop in sales also came as consumers raised their spending in pubs, cafes, and restaurants at the expense of shopping for food and drink in supermarkets after the easing of coronavirus restrictions across the country.

The latest snapshot showed sales fell across almost every category, with a 4.4% drop in non-food stores, driven by weaker sales in clothes shops, secondhand stores and spending on computers and telecoms.

In a potential sign that supply shortages caused by Covid-19 and Brexit are having an impact on retail sales, the ONS pointed to Bank of England evidence that showed there had been delays to shipments of electrical goods in recent months.

Heavy rainfall in early July also led to a fall in spending on fuel at petrol stations by 2.9% on the month amid a decline in traffic on Britain's roads.

3. Technical text (downloaded from https://www.accessdata.fda.gov/drugsatfda_docs/label/2007/011522s040lbl.pdf on 20 August 2021)

WHAT IS ADDERALL®?

ADDERALL® IS A CENTRAL NERVOUS SYSTEM STIMULANT PRESCRIPTION MEDICINE. IT IS USED FOR THE

TREATMENT OF ATTENTION-DEFICIT HYPERACTIVITY DISORDER (ADHD). ADDERALL® MAY HELP

INCREASE ATTENTION AND DECREASE IMPULSIVENESS AND HYPERACTIVITY IN PATIENTS WITH ADHD.

ADDERALL® SHOULD BE USED AS A PART OF A TOTAL TREATMENT PROGRAM FOR ADHD THAT MAY INCLUDE

COUNSELING OR OTHER THERAPIES.

ADDERALL® IS ALSO USED IN THE TREATMENT OF A SLEEP DISORDER CALLED NARCOLEPSY.

ADDERALL® IS A FEDERALLY CONTROLLED SUBSTANCE (CII) BECAUSE IT CAN BE ABUSED

OR LEAD TO DEPENDENCE. KEEP ADDERALL® IN A SAFE PLACE TO PREVENT MISUSE AND

ABUSE. SELLING OF GIVING AWAY ADDERALL® MAY HARM OTHERS AND IS AGAINST THE

LAW.

Tell your doctor if you or your child have (or have a family history of) ever abused or been

dependent on alcohol, prescription medicines or street drugs.

WHO SHOULD NOT TAKE ADDERALL®?

ADDERALL® SHOULD NOT BE TAKEN IF YOU OR YOUR CHILD:

• have heart disease or hardening of the arteries

• have moderate to severe high blood pressure

• have hyperthyroidism

• have an eye problem called glaucoma

• are very anxious, tense, or agitated

• have a history of drug abuse

• are taking or have taken within the past 14 days an anti-depression medicine called a monoamine oxidase

inhibitor or MAOI.

• are sensitive to, allergic to, or had a reaction to other stimulant medicines

ADDERALL® IS NOT RECOMMENDED FOR USE IN CHILDREN LESS THAN 3 YEARS OLD.

4. Legal text (downloaded from https://www.legislation.gov.uk/asp/2011/6/section/14 on 20 August 2021)

Prohibition on keeping etc. of invasive animals or plants

(1)Subject to the provisions of this Part, any person who keeps, has in the person's possession, or has under the person's control—

(a)any invasive animal of a type which the Scottish Ministers, by order, specify; or

(b)any invasive plant of a type so specified,is guilty of an offence.

(2)An order under subsection (1) may make different provision for different cases and, in particular, for—

(a)different types of invasive animal or invasive plant;

(b)different circumstances or purposes;

(c)different persons;

(d)different times of the year; and

(e)different areas or places.

(3)Subject to subsection (4), it is a defence to a charge of committing an offence under subsection (1) to show that the accused took all reasonable steps and exercised all due diligence to avoid committing the offence.

(4)Where the defence provided by subsection (3) involves an allegation that the commission of the offence was due to the act or omission of another person, the person charged must not, without leave of the court, be entitled to rely on the defence unless, within a period ending 7 days before the hearing, the person has served on the prosecutor a notice giving such information or assisting in the identification of the other person as was then in the person's possession.