



# Contemporary Issues of Machine Translation in the Field of Law – A Contrastive Analysis<sup>1</sup>

**Kordić L<sup>1\*</sup> and Jokić A<sup>2</sup>**

<sup>1</sup>Faculty of Law Osijek, Josip Juraj Strossmayer University of Osijek, Croatia

<sup>2</sup>Faculty of Humanities and Social Sciences, Josip Juraj Strossmayer University of Osijek, Croatia

**\*Corresponding author:** Ljubica Kordić, Associate Professor, Faculty of Law Osijek, Josip Juraj Strossmayer University of Osijek, S. Radića 13, 31 000 Osijek, Croatia, Tel: +385 91 126 6287; Fax: +385 224 530; Email: kljubica@pravos.hr

**Research article**

**Volume 5 Issue 1**

**Received Date:** December 20, 2021

**Published Date:** January 26, 2022

**DOI:** 10.23880/phij-16000222

## Abstract

We live in the age of constant changes and tremendous development of information technologies in all spheres of human activity. The field of translation is not exempted from these changes and developments. In this paper we shall explore the achievements of computer technologies in the field of machine translation. The research is conducted on the machine translations of four legal text in the field of EU legislation from German into Croatian language. The method used is a contrastive analysis of translations performed by two online translators: Google Translate and Microsoft Bing Translator.

In the introductory part, specific features of the translation in the field of law are presented. In the theoretical introduction to the research, approaches to machine translations are discussed with specific reference to the editing and post-editing phases of translation procedure. Based on a number of features that represent challenges in translating legal texts, our hypothesis is that in translating legal texts these specific features have to be considered and that an intervention of a skilled human translator mastering both the linguistic and extra-linguistic knowledge is necessary, at least in a post-editing phase. In the empirical part, four different texts from the field of EU legislations are translated by using Google Translate and Microsoft Bing Translator. The translations are then contrastively analysed against the error analysis criteria developed by Costa and Popović Popović. In the final part of the paper, qualitative and quantitative analysis of findings are conducted, types of errors occurring in both machine translations discussed, and conclusions drawn.

**Keywords:** Machine Translation; Legal Texts; Errors; Google Translate; Microsoft Bing Translator

## Introduction: Specific Features of the Translation in the Field of Law

Translation of legal texts represents “...an act of communication in the mechanism of law” that leads “to legal

effects and may induce peace or prompt war” [1-3].

In the 1970s, the concept of the equivalence of terms in the source language and the target language was put aside as secondary, while the function of the translation in

<sup>1</sup>This paper is a product of work that has been fully supported by the Jean Monnet Module Language and EU Law Excellence LEULEX, project no. 620231-EPP-1-2020-1-HR-EPPJMO-MODULE



the target language and the target culture moved into the focus of modern translation. This approach found favour in translators dealing with legal translations who have been facing problems when translating the texts stemming from different legal systems and different cultures. Within the new functionalist approach, the translation was observed in the socio-cultural context of the target culture, and extra-linguistic factors were for the first time in history concerned as decisive factors influencing the translation process [4]. Accordingly, German authors Katarina Reiss and Hans Vermeer defined translation as a specific type of cultural transfer (“Sondersorte des kulturellen Transfers”). In that context, they understood the translation in the field of law as a transfer from one legal system and one culture into another (pp: 13) [5]. The differences between legal systems and cultures are reflected in linguistic phenomena such as polysemy of terms, collocations, metaphorical expressions, idioms, and other culture-bound terms that are difficult to translate into another language (and culture) without the translator’s extra-linguistic knowledge. Terminology differences as a result of different legal systems and culture-bound differences represent the major causes of problems that translators in the field of law are faced with.

This approach, also known as the functionalist approach or the “skopos-theory” (skopos being a Greek word for purpose), placed the purpose of translation, the target language, and the final user of the translation into the focus of translation. The complexity of legal translation has been further emphasized by the awareness of the existence of numerous types of legal texts entailing different legal effects. Christiane Nord distinguishes between two main translation types in the field of law regarding the function of the translated text: a documentary translation that serves as a document in communication between the source and the target culture, and instrumental translation, serving for communicative purposes in the target language [6]. These challenges represent the main ground for our hypothesis that the correct and reliable translation in the field of law is not possible by using exclusively machine translators, without any interference of human translators who possess extra-linguistic knowledge about the differences between the legal systems and cultures of the source language and the target language.

All the above mentioned factors demand specific competencies in translators dealing with legal translation. According to Kelly, translators’ competencies comprise the following sub-competencies: communicative and textual sub-competence, cultural and intercultural sub-competence, professional and instrumental, psychophysiological, strategic, subject area sub-competence, as well as interpersonal sub-competence (pp: 162) [7]. Similarly, Šarčević sees the

following competencies as necessary in legal translation: knowledge of legal terminology, in-depth understanding of logical principles, logical reasoning, the ability of problem-solving, the ability of text analysis and the knowledge of the target legal system, and the source legal system (pp: 13–14) [3]. In the contemporary situation of the implementation of computer technologies in the translation process, both of them would certainly add the skill of using computer technologies and databases as a necessary competence of the translators.

In the following analysis of the machine translations of four texts by Google Translate and Microsoft Bing Translator we shall assess the quality of machine translation of those texts from German into Croatian. By comparing the number of errors in parallel translations and their contrastive analysis against the objectively established error typology we shall explore which machine system is better and to which extent they can replace a human translator. For that purpose, an official translation available in the Eurlex database will be used as the reference text.

### Human Translation and Machine Translation

The tremendous development of information technologies that we have witnessed in recent years has changed the field of translation as well. The achievements of computer technologies in the field of machine translation have made translation tools more available and more efficient in terms of time and money. On the other hand, in modern multinational and multilingual political communities, the use of multilingual machine translation controlled by expert translators has become an indispensable factor in the functioning of the community. This particularly refers to the translation procedure in the European Commission. The translation practice in the Directorate-General for Translation of the European Commission includes the simultaneous drafting of legal texts in the three working languages of the EU (French, English, and German) followed by a translation into other official languages. This is the reason why the procedure is called co-drafting rather than translating legal texts (pp: 271) [3]. Because of the linguistic features of legal texts that we discussed in the introductory part, legal translators are faced with serious challenges in the EU translation procedures. The procedure is even more complicated by the principles that have to be respected in the translation process in the EU Commission. The most important principles are the principle of legal uniformity and the principle of legal certainty of all language versions. The reason for that is that the legal provisions of the EU law should be interpreted in the same way across the Union to harmonize different legal concepts between different national

legal systems across the Union. The Multilingual Lexicon of European Law has been created to meet that requirement, as well as EurLex - a terminological database of the European Commission. Legal certainty and legal uniformity as basic target principles of translation activities in the European Commission are joined by the principle of standardization of terminological, phraseological, and graphical aspects of translation into all official languages of the EU. The graphical equalization of all translations implies the same graphical design, the same structure and sequence of articles and sections, the same number of sentences and full stops in the translations into all 24 languages. This approach, which is often criticized by linguists, is known as 'the full stop rule' (pp: 152) [8]. The rules of standardization and equalization of legal texts can be understood as reasons for introducing neologisms and international terms into EU law that have been recommended in the Joint Practical Guide published by the European Commission for its DGT [9,10]. The translation rules of the DGT also include syntactic simplification of translations to enable an easier manipulation of legal texts in the case of amendments to legal texts in the EU legislation procedure [11]. Of course, massive translation machinery of the European Commission is using the benefits of the IT translation tools and CAT tools to control the multilingual translation process. To be able to follow rapid changes in the functioning of the EU law and to update legal terminology within the EU, the EURAMIS translation memory has been created, which is regularly updated on a daily basis.

Although machine translation has been implemented a lot both for business and private purposes, there are numerous objections relating to errors occurring in machine translations. We shall analyze different machine translations of the same source texts by using a contrastive analysis to evaluate the quality of the chosen machine translators. For that purpose, a reliable set of error categories are applied as criteria against which the errors are evaluated. In recent years, numerous sets of error categories have been developed by different authors for two reasons: firstly, to evaluate the quality of specific machine translators available online, and secondly, to enable automatic evaluation, i.e. developing programs that produce a score based on the similarity between the reference translation and the translation output by using error categorization as assessment criteria. Popović [2] explains that the score is usually produced either in the form of a percentage of matched n-grams between the reference and the output or as edit distance between them. Since automatic evaluation is much faster, cheaper, and more consistent than human evaluation, a number of automatic evaluation metrics have been investigated based on word n-gram precision, the chrF system, etc [12].

To identify and classify errors in a translated text, several error classifications have been developed. Their

purpose was to provide a better foundation for decisions on the quality of specific machine translation on one hand and for development, purchase, or use of computer programs to be applied in the post-editing phase of machine translation on the other. The main goal in the error analysis is to obtain an error profile for a translation output of a specific machine translator and to establish a distribution of errors over the defined error classes for a specific machine translator. In our research, the taxonomy of errors developed by Costa [1] will be combined with the typology developed by Popović [2]. That new error categorization will be used as a criterion of the quality assessment of two machine translation systems: Google Translate and Microsoft Bing.

### Research: Goals, Corpus, and Methodology

The questions that should be answered by the results of our contrastive analysis are the following: What are the particular advantages and weaknesses of the specific translation system? Which errors indicate the most striking problems of a specific translation system?

As a corpus of the research we shall use four randomly chosen texts excerpted from the EU agreement *Übereinkommen zwischen den Regierungen der Staaten der Benelux-Wirtschaftsunion, der Bundesrepublik Deutschland und der Französischen Republik betreffend den schrittweisen Abbau der Kontrollen an den gemeinsamen Grenzen*. The source text is available in the EurLex databases.<sup>1</sup>

For the sake of scientific reliability and objectivity, the errors occurring in translations by two online translators will be analyzed against the list of errors that we have created based on the error taxonomy developed by Costa [1] that we have complemented with the elements of the error taxonomy by Popović [2]. After the categorization of detected errors, the errors of the two machine translators will be contrasted and conclusions drawn concerning the reliability and quality of each machine translator. As a reference text, we shall use the official Eurlex translation of the respective text from German into the Croatian language.

Before the contrastive analysis, error categories should be defined. This is important because the errors should reflect all strengths and weaknesses of the respective machine translation system. In recent years, several authors have developed various error typologies. They range from rather simple ones [13,14] that included five basic error categories (lexical errors, morphological errors, orthographic errors,

<sup>1</sup> <https://eur-lex.europa.eu/legal-content/HRDE/TXT/?from=HR&uri=CELEX%3A22017A0117%2801%29&qid=1628789600949>, accessed 22.10.2021

syntactic errors, and semantic errors) to more complex ones that have been developed by Federico, et al. [15], Kirchhoff, et al. [16], etc. The most precise error typology seems to be that of Costa [1], which could be applied to legal texts as it includes culture-bound errors such as *Semantic Confusion*, *Wrong choice*, *Collocational errors*, *Idioms* [12]. We have complemented Costa's typology with the elements of the error taxonomy by Popovic [1] that, in our opinion, contains

the necessary criteria to assess the quality of legal texts, such as terminology, mistranslation, and a more detailed elaboration of morphological errors such as wrong tense, aspect, parts of speech, etc.

Costa's error typology contains the following error categories distributed on two levels:

Level 1	Level 2
Orthography	Punctuation, Capitalization, Spelling
Lexis	Omission, Addition, Untranslated
Grammar	Word Class, Verbs, Agreement, Contraction
	Misordering
Semantic	Confusion of senses, Wrong choice, Collocational errors, Idioms
Discourse	Style, Variety, Should not be translated

By combining different error taxonomies, Popović has developed her own general error taxonomy that she finds applicable on technical (LSP) texts, as it includes a

terminology error as a relevant error variable [12].

General error taxonomy by Popović [12]:

Level 1	Level 2	Level 3
Lexis	mistranslation	terminology
	addition	
	omission	
	untranslated	
	should not be translated	
Morphology	inflection	tense, number, person, case, number, gender
	Derivation	POS, verb aspect
Syntax	word order	range
	phrase order	range
	collocations	
	disambiguation	
Orthography	capitalisation, punctuation, spelling	
	Too many errors	

She has developed the above taxonomy for developing programs that should produce a score based on the similarity between the reference translation and the translation output in which her error categorization serves as assessment criteria. Her taxonomy includes broad classes and enables various possibilities for expansion [2], so for this research conducted on the four examples of short legal texts we

have combined a rather simple taxonomy by Costa with the specific elements of Popović's typology that are relevant for legal texts. It includes the elements of terminology and more detailed descriptions of morphological errors referring to tense, number, person, case, gender, POS (parts of speech), and verb aspect:

Level 1	Level 2
Orthography	Punctuation, Capitalization, Spelling
Lexis	Omission, Addition, Untranslated, mistranslation, terminology
Grammar	Word Class, Verbs, Agreement/Congruence, Contraction
Morphology	Tense, person, case, number, gender
Syntax	Misordering (word order, phrase order)
Semantic	Confusion of senses, Wrong choice, Collocational errors, Idioms
Discourse	Style, Variety, should not be translated

WE hypothesize that, due to specific features of legal texts and all the complex competencies that legal translators should master, a good and reliable translation of legal texts is not possible without human intervention, i.e. human error correction in the post-editing phase of the translation process. Before conducting the research, we cannot predict which online translation system would offer translations of a better quality nor which of them will have a better translation performance.

### Contrastive Analysis of Errors: Google Translate v. Microsoft Bing Translator

The analysis is conducted on four examples excerpted from the EU agreement in German language - Übereinkommen zwischen den Regierungen der Staaten der Benelux-Wirtschaftsunion, der Bundesrepublik Deutschland und der Französischen Republik betreffend den schrittweisen Abbau der Kontrollen an den gemeinsamen Grenzen.<sup>2</sup> Here we shall contrastively analyze the translation outputs by Google Translate and Microsoft Bing Translator and discuss the errors from the qualitative and the quantitative point of view.

#### Original Text

##### Artikel 6

*Die Vertragsparteien ergreifen - unbeschadet weitergehender Regelungen - die notwendigen Maßnahmen, um den Verkehr der Angehörigen der Mitgliedstaaten der Europäischen Gemeinschaften zu erleichtern, die in Gemeinden an den gemeinsamen Grenzen leben, um ihnen zu gestatten, die Grenzen außerhalb der zugelassenen Grenzübergangsstellen und außerhalb der Öffnungszeiten zu überschreiten. Für den begünstigten Personenkreis gelten diese Vorteile nur, wenn die mitgeführten Waren innerhalb der Freigrenzen liegen und die geltenden Devisenbestimmungen beachtet werden.*

#### Google Translate:

##### Članak 6

<sup>2</sup> <https://eur-lex.europa.eu/legal-content/HRDE/TXT/?from=HR&uri=CELEX%3A22017A0117%2801%29&qid=1628789600949> accessed 22.10.2021

*Ne dovodeći u pitanje daljnje propise, ugovorne strane poduzet će potrebne mjere kako bi olakšale kretanje građanima država članica Europskih zajednica koji žive u općinama na zajedničkim granicama, kako bi im se omogućio ulazak u granice izvan odobrene granice prijelaza i izvan radnog vremena premašuju.*

*Ove prednosti vrijede samo za korisničku skupinu ako je roba koja se prevozi unutar granica izuzeća i ako se poštuju važeći devizni propisi.*

#### Microsoft Bing

##### Članak 6

*Ugovorne stranke, ne dovodeći u pitanje daljnje odredbe, poduzimaju potrebne mjere kako bi olakšale kretanje državljanima država članica Europskih zajednica koji žive u komunama na zajedničkim granicama kako bi im se omogućilo prelazak granica izvan odobrenih graničnih prijelaza i izvan radnog vremena.*

*Za skupinu osoba korisnica te se prednosti primjenjuju samo ako je roba koja se prevozi unutar ograničenja izuzeća i ako se poštuju primjenjivi devizni propisi.*

#### The Reference Text (Human translation EurLex)

##### Članak 6

*Ne dovodeći u pitanje primjenu povoljnijih dogovora između stranaka, one poduzimaju potrebne mjere za olakšavanje kretanja državljanima država članica Europskih zajednica koji borave na lokalnim upravnim područjima duž zajedničkih granica, kako bi im se na taj način omogućio prelazak tih granica na mjestima izvan ovlaštenih graničnih prijelaza i izvan radnog vremena kontrolnih točaka.*

*Dotične osobe mogu iskoristiti ove prednosti pod uvjetom da prevoze jedino onu robu koja je dopuštena u okviru dogovora o nultoj stopi carine i da poštuju devizne propise.*

The first thing that we can notice in Google Translate is the omission of the point after the article number, which is a tradition in writing laws in the Croatian legal system – according to the Croatian rules of legal writing, articles are denoted by ordinal numbers (orthography error). The Bing



translator did not make this omission.

Both machine translators use a different equivalent for the term *regulations*, but both terms are acceptable from the legal point of view. On the other hand, the Croatian word *dogovor* used in human translation is not the corresponding equivalent to the legal term *Regelung* and can be classified as a wrong choice word (terminology error), as it covers a narrower semantic field than the words *propisi* and *odredbe* that were used by the two machine translators. The term *dogovor* used by a human translator is a rather everyday speech term, while a more appropriate legal correspondent would be *sporazum*.

Google Translate used a wrong tense for the verb *ergreifen* – the Future Tense – while according to Croatian legal tradition the normative obligations are expressed by the Present Tense (morphology error - tense). The Bing Translator and the human translator followed that rule. Legal term *komuna* chosen by the Bing Translator is considered a mistranslation (lexis - wrong choice), as it is an archaic term that is not in use anymore, while Google Translate made a good choice in this case. On the other hand, a human translator used a semantically wider term *lokalna upravna područja* but a corresponding Croatian legal phrase in colloquial use is *jedinice lokalne uprave* rather than *lokalna upravna područja* (terminology error; wrong phrase). Google Translator used a wrong translation for the collocation *ulazak u granice* for *die Grenzen überschreiten* (semantic error, wrong collocation) while Bing's translation was correct. Similarly, the term *Grenzübergangsstellen* was wrongly translated by Google as *granice prijelaza*. This Croatian collocation holds a completely different meaning than in the source text (mistranslation, semantic error,

wrong collocation). On the other hand, Bing used a correct equivalent *granični prijelaz*. Google has also added the unnecessary verb *premašuju* (lexical error, addition). Google has also mistranslated the collocation *begünstigten Personenkreis* as *korisnička skupina*, which is a collocation from computer or mobile phone technologies (semantic error, wrong collocation). Bing uses a more appropriate poly-lexical term *skupina osoba korisnica*, which can be seen as semantically not quite precise, while the official human translator completely avoided using a precise translation and used a general term *dotične osobe* without any specific reference to persons enjoying special benefits (omission error, avoidance). Similar case is with the phrase *innerhalb der Freigrenzen liegen*. Neither the choice by Google *unutar granica izuzeća* nor the variation by Bing *unutar ograničenja izuzeća* do not cover the semantic field of the phrase and lead to unclear translation (semantic error, terminology error, confusion of senses). Human translation offers the best solution by using a lexical description of that phenomenon in the EU internal frontier regime in the form of poly-lexical phrase *u okviru dogovora o nultoj stopi carine*. In the case of legal collocation *die geltenden Devisenbestimmungen*, Google has offered a better translation *važeći devizni propisi*, while Bing instead of the adjective *važeći* offers *primjenjivi*, which covers a narrower semantic field and refers to legal rules that can be applied instead of valid rules in the sense of the source text (semantic error, wrong choice). Interestingly, the human translator has omitted this term from its translation (omission). It has also wrongly translated the collocation *zugelassene Grenzübergangsstellen* as *ovlašteni granični prijelazi* in which wrong choice was made: *ovlašteni* instead of *odobreni granični prijelazi* (*ovlašteni granični prijelazi* = *empowered* border crossings instead of *permitted*) which was the choice by Bing and Google Translate.

Error typology	Google Translate	Microsoft Bing
Orthography	Omission	
Lexis	Addition	wrong choice, mistranslation (2)
Grammar		
Morphology	wrong tense	
Syntax		
Semantic	wrong collocation (2) wrong phrase (terminology)	wrong phrase (terminology)
Discourse		

Although the errors in the human translation were not included in contrastive analysis because it was used as a reference text, we have seen that it also contains some errors. Most of them refer to terminology choice and omissions or additions that are used to clarify the meaning of the sentences. The nature of lexical and semantic errors

indicates that the translator did not master Croatian legal terminology sufficiently. This could be explained by the fact that the document was translated at an early stage of the EU integrations when the whole DGT machinery and Croatian terminological databases were not as developed as today.

**Example 2: Original Text****Artikel 7**

Die Vertragsparteien bemühen sich, so bald wie möglich ihre Sichtvermerkspolitik anzunähern, um mögliche negative Folgen bei der Erleichterung der Kontrollen an den gemeinsamen Grenzen auf dem Gebiet der Einreise und der inneren Sicherheit zu vermeiden. Sie ergreifen - möglichst bis zum 1. Januar 1986 - die notwendigen Schritte bei der Anwendung ihrer Verfahren zur Sichtvermerkerteilung und der Einreiseerlaubnis, um so den Schutz der Gesamtheit der Hoheitsgebiete der fünf Vertragsparteien vor unerlaubter Einreise und vor Handlungen, die die innere Sicherheit beeinträchtigen können, sicherzustellen.

**Google Translate**

## Članak 7

Ugovorne strane nastojat će približiti svoju viznu politiku što je prije moguće kako bi se izbjegle moguće negativne posljedice u olakšavanju kontrola na zajedničkim granicama u području ulaska i unutarnje sigurnosti. Oni će poduzeti potrebne korake u primjeni svojih procedura za izdavanje viza i dozvola za ulazak - ako je moguće do 1. siječnja 1986. - kako bi zaštitili cjelokupno suvereno područje pet ugovornih strana od neovlaštenog ulaska i od radnji koje mogu narušiti unutarnje sigurnost, osigurati.

**Bing Microsoft**

## Članak 7

Ugovorne stranke nastoje što prije približiti svoju viznu politiku kako bi izbjegle moguće negativne posljedice u olakšavanju provjera na zajedničkim granicama u području ulaska i unutarnje sigurnosti. Oni poduzimaju potrebne korake, ako je moguće do 1. siječnja 1986., u primjeni svojih postupaka izdavanja viza i dozvola za ulazak kako bi osigurali zaštitu svih područja pet ugovornih stranaka od nezakonitog ulaska i od radnji čiji učinak na unutarnju sigurnost može utjecati na unutarnju sigurnost.

**Human Translation**

## Članak 7

Stranke nastoje, što je prije moguće, uskladiti svoje vizne politike, kako bi se izbjegle nepovoljne posljedice na području useljavanja i sigurnosti koje mogu nastati zbog smanjenih kontrola na zajedničkim granicama. Po mogućnosti do 1. siječnja 1986. one poduzimaju potrebne korake za primjenu vlastitih postupaka za izdavanje viza i dozvolu ulaska na njihova državna područja, uzimajući u obzir potrebu da se osigura zaštita cjelokupnog državnog područja pet država od ilegalnog useljavanja i aktivnosti koje bi mogle ugroziti sigurnost.

Error typology	Google Translate	Microsoft Bing
Orthography	Omission	
Lexis	Addition (wrong collocation) Wrong choice (literal trans.)	Addition Omission Wrong choice (literal trans.)
Grammar	Noun-pronoun congruence Adjective-noun congruence	Noun-pronoun congruence
Morphology	wrong tense (2)	
Syntax		
Semantic		Terminology (simplification)
Discourse	Style (loanword)	Style (redundancy)

The contrastive analysis of the two machine translators has shown that more errors occurred in the Google translate translations. Again, the Croatian legal tradition that articles are ordinal numbers has been ignored (orthography error, omission). Accordingly, instead of the Present Tense that is used to express normativity in Croatian laws, Google Translate has used the Future Tense in two sentences (morphology error/ terminology error – wrong tense). This error is not merely a morphology problem but primarily the problem of the lack of knowledge on structures used in the target legal system and culture. Both translators have made a wrong choice of words when they translated the word *Erleichterung* (der Kontrolle) as *olakšavanje*, which

should be translated in the way the human translator has interpreted it as a part of collocation: *smanjenje kontrola* (lexical error – wrong choice/ semantic error). Both machine translators have wrongly added unnecessary words (lexical error - addition): Google Translate has translated correctly the collocation containing the function verb *sicherstellen* - *den Schutz sicherstellen* (to secure protection) by using a one-word term *zaštititi* (protect). Additionally it added at the end of the sentence the unnecessary function verb *osigurati* – *secure*, while Bing Translator has used the phrase *osigurati zaštitu unutarnje sigurnosti*, in which case the complete poly-lexical expression (*na unutarnju sigurnost*) was repeated (addition). A surprising error that occurred in

both translations was the incongruence of the plural noun in the first sentence *ugovorne strane/ugovorne stranke* and the pronoun replacing it in the next sentence, which was used in a wrong gender in the Croatian language (*oni* instead of *one*). Both translators have made similar stylistic errors as well. Although IATE system as well as the Croatian Institute for Language and Linguistics suggest avoiding loanwords as much as possible, Google Translate used the term *procedure* instead of *postupci*, while Bing's connecting the verb *utjecati* with the noun *učinak* that was derived from the same verb made the whole sentence redundant and unnatural (style errors, redundancy). The contrastive analysis of translations by two machine translators has also revealed that in one case Bing used a general language term *sva područja* instead of a more precise and generally acceptable legal terms *suvereno područje* (Google Translate) or *državno područje* (human translation) respectively. Additionally, Google made another grammar error of wrong adjective-noun congruence: *narušiti unutarnje sigurnost*.

In this example, human translators applied legal terms commonly used in the EU legislation. In several cases, they also made a better choice of words than that of the machine translators (e.g. *uskladiti* used by human translator vs. *približiti* that was used by both machine translators; *smanjenje kontrole granica* instead of *olakšanje kontrole granica* – a solution offered by machine translator, etc.). Additional words and phrases that often occur in the human translation are used to make the target text more clear and precise, which can be perceived as a specific style of the respective translator.

### Example 3

#### Artikel 9

*Die Vertragsparteien verstärken die Zusammenarbeit zwischen ihren Zoll- und Polizeibehörden insbesondere im Kampf gegen Kriminalität, vor allem gegen den illegalen Handel mit Betäubungsmitteln und Waffen, gegen die unerlaubte Einreise und den unerlaubten Aufenthalt von Personen, gegen Steuer- und Zollhinterziehung sowie gegen Schmuggel. Zu diesem Zweck bemühen sich die Vertragsparteien im Rahmen ihres jeweiligen innerstaatlichen Rechts, den Austausch von Informationen zu verstärken, die für die anderen Vertragsparteien insbesondere im Kampf gegen die Kriminalität von Interesse sein könnten.*

*Die Vertragsparteien verstärken im Rahmen ihrer bestehenden nationalen Gesetze die gegenseitige Unterstützung im Hinblick auf illegale Kapitalbewegungen.*

#### Google Translate

##### Artikel 9

*Ugovorne strane jačaju suradnju između svojih carinskih i policijskih tijela, posebno u borbi protiv kriminala, posebno protiv ilegalne trgovine opojnim drogama i oružjem, protiv*

*neovlaštenog ulaska i boravka osoba, protiv utaje poreza i carine te protiv krijumčarenja. U tu svrhu, ugovorne stranke nastojat će, u okviru svojih odgovarajućih domaćih zakona, povećati razmjenu informacija koje bi mogle biti od interesa za druge ugovorne stranke, posebno u borbi protiv kriminala.*

*Ugovorne strane će, u okviru svojih postojećih nacionalnih zakona, jačati međusobnu pomoć u pogledu nezakonitog kretanja kapitala.*

#### Bing Microsoft

##### Članak 9

*Stranke jačaju suradnju između svojih carinskih i policijskih tijela, posebno u borbi protiv kriminala, posebno protiv nezakonite trgovine opojnim drogama i oružjem, protiv nezakonitog ulaska i boravka osoba, protiv utaje poreza i carine te protiv krijumčarenja. U tu svrhu stranke će nastojati, u okviru svojih nacionalnih zakona, ojačati razmjenu informacija koje bi mogle biti od interesa za druge stranke, posebno u borbi protiv kriminala.*

*Stranke će, u okviru svojih postojećih nacionalnih zakona, ojačati uzajamnu pomoć u pogledu nezakonitog kretanja kapitala.*

#### Human Translation

##### Članak 9

*Stranke pojačavaju suradnju između njihovih carinskih i policijskih tijela, osobito u suzbijanju kriminala, posebno protiv nedopuštene trgovine drogom i oružjem, neovlaštenog ulaska i boravka osoba, carinskih i poreznih prijevара i krijumčarenja. U tu svrhu i u skladu sa svojim nacionalnim zakonodavstvima stranke se obvezuju poboljšati razmjenu informacija i intenzivirati tu razmjenu, ako se radi o informacijama koje bi drugim strankama mogle biti korisne kod suzbijanja kriminaliteta.*

*U okviru njihovih nacionalnih zakonodavstava stranke intenziviraju uzajamnu pomoć u vezi s nedopuštenim kretanjem kapitala.*

In the contrastive analysis of these texts translated by the two machine translators, the differences that we can observe are a matter of variations in the translation and a different choice of words rather than errors (*pojačati razmjenu* vs. *ojačati*, *povećati* vs. *poboljšati*...). The Google Translate again does not use a point after the number of the article neglecting the fact that it is a traditional rule in the Croatian lawgiving (orthography error – punctuation). Additionally, Google Translate has used an everyday language word *domaći zakon* instead of the usual legal term *nacionalni zakoni/ nacionalno zakonodavstvo* (wrong term choice; terminology error). In the second case of using the term in the same text, the translation was the proper



one (*nacionalnih zakona*). In the human translation, we can see some stylistic variations differing from those in machine translations. Some collocations used by the human translator can be seen as more colloquial legal terms in law than those used by machine translators (*suzbijanje kriminaliteta* instead of *borba protiv kriminala*, *nacionalnih zakonodavstava* instead of *postojećih nacionalnih zakona*) but both are acceptable. The human translator is consistent in using the Present Tense in the Croatian text instead of Future Tense. The use of Future Tense in only one sentence

by both machine translators cannot be observed as an error as it relates to future attempts of the parties. What we can resent against the human translator is that it prefers using loanwords (*intenzivirati razmjenu*, *intenziviraju uzajamnu pomoć*). The only error that can be observed in this example of translation is a wrong usage of the possessive pronoun *njihov* by the human translator (*njihovih carinskih propisa*, *njihovih nacionalnih zakonodavstava*) instead of the Croatian reflexive pronoun *svoj* that was correctly used by both machine translators.

Error typology	Google Translate	Microsoft Bing
Orthography	omission	
Lexis	wrong term choice	
Grammar		
Morphology		
Syntax		
Semantic		
Discourse		

#### Example 4

##### Artikel 13

*Die Vertragsparteien bemühen sich, bis zum 1. Januar 1986 das zwischen ihnen im grenzüberschreitenden Straßengüterverkehr geltende Genehmigungssystem mit dem Ziel der Vereinfachung, der Erleichterung und der Möglichkeit der Umstellung von Fahrtgenehmigungen auf Zeitgenehmigungen mit einer Sichtkontrolle beim Grenzübertritt zu verbessern. Die Modalitäten der Umwandlung von Einzelfahrtgenehmigungen in Zeitgenehmigungen werden bilateral vereinbart, wobei der Bedarf des Straßengüterverkehrs der beteiligten Länder berücksichtigt wird.*

##### Google Translate

###### Članak 13

*Ugovorne strane nastojat će do 1. siječnja 1986. poboljšati sustav licenciranja koji se primjenjuje među njima za prekogranično cestovno prijevoz s ciljem pojednostavljenja, olakšavanja i omogućavanja pretvaranja putnih dozvola u privremene sa vizualnim pregledom pri prelasku granice. Načini pretvaranja pojedinačnih putnih dozvola u vremenske dozvole dogovaraju se bilateralno, uzimajući u obzir potrebe cestovnog teretnog prijevoza u uključenim zemljama.*

##### Bing Microsoft

###### Članak 13

*Ugovorne stranke će do 1. siječnja 1986. nastojati poboljšati sustav licenciranja koji se između njih primjenjuje u međunarodnom cestovnom prijevozu robe s ciljem pojednostavljenja, olakšavanja i omogućavanja prelaska s odobrenja putovanja na privremena odobrenja uz vizualni*

*pregled pri prelasku granice. Modaliteti pretvaranja pojedinačnih putnih dozvola u vremenske dozvole dogovaraju se bilateralno, uzimajući u obzir potrebe cestovnog teretnog prijevoza zemalja sudionica.*

##### Human Translation

###### Članak 13

*Stranke se obvezuju da do 1. siječnja 1986. usklade sustave koje će primjenjivati na međusobno izdavanje dozvola za komercijalni cestovni prijevoz u prekograničnome prometu, kako bi pojednostavile, olakšale i, po mogućnosti, zamijenile dozvole za putovanja dozvolama za određeno razdoblje uz vizualnu kontrolu vozila prilikom prelaska zajedničkih granica. Postupci zamjene dozvola za putovanja dozvolama za određeno razdoblje se dogovaraju na bilateralnoj osnovi uzimajući u obzir zahtjeve za cestovni prijevoz u dotičnim različitim državama.*

Again, Google Translate does not obey the Croatian rule for writing articles as ordinal numbers (Orthography error – punctuation). In the first sentence, it uses the phrase *sustav licenciranja*, which can be used in different contexts and belongs to the business sphere and the patent law rather than to international law in the context of a cross-border commercial transport (semantic error – wrong collocation). Probably leaning on its existing databases, Microsoft Bing makes the same error as Google and translates the phrase in the same way (semantic error – wrong collocation). The human translator has used a descriptive approach here by choosing a dependent clause to clarify the meaning of the source information expressed by a compound noun

*Genehmigungssystem*: sustave koje će primjenjivati na međusobno izdavanje dozvola (...). In the same sentence, Google has used the prepositional phrase *za prekogranično cestovno prijevoz* which is grammatically wrong because of the incongruence of the noun and the adjectives modifying it (grammar error; adjective-noun congruence). In the same example, Google has used a wrong preposition *za*, because the preposition *u* is a better choice with reference to the verb *primjenjuju* that is postmodified by the respective prepositional phrase (lexical error – wrong preposition). Bing has offered here a good solution *u međunarodnom cestovnom prijevozu robe*, which semantically suits better to the verb *primjenjivati*. Human translator has used the same preposition in this phrase as Google Translate, but it is a good choice here as it is used as a postmodifier of the word *dozvola*, which requires the use of the preposition *za* in the Croatian language (*dozvola za komercijalni cestovni prijevoz u prekograničnome prometu*). In the quoted phrase we recognize a human touch again, as the whole legal rule is clarified by using additional adjective *komercijalni* instead of *prekogranični*. The same term *prekogranični* is used again in a commonly used EU Law collocation *prekogranični promet*. We can conditionally concern the choice of the word *pretvaranje* instead of *zamjena* (it was used twice in the text) as a word choice error in Google's translation (*pretvaranja putnih dozvola u privremene...*), because it is a literal translation of the word (lexis; word choice error). Bing had a better solution here and used the term *prelazak* (transfer from permanent permission to a temporary one), but then in the next sentence it used the same word *prelazak* as an equivalent for *border crossing*, which can lead to misunderstandings (word choice error). On the other hand, a human translator has applied legal reasoning while

translating this text. His/her choice of the word *zamjena* indicates that he/she has understood the changes in the cross-border transport rules, which was not the case with the machine translators. Both machine translators have offered a good solution by using the collocation *privremena dozvola* (Bing has chosen the term *odobrenje* instead of *dozvola*) while the human translator has chosen a less appropriate phrase *dozvole za određeno razdoblje*. That the human translator tends to use collocations commonly used in the EU law even when it is not necessary can be illustrated by the example *zajedničke granice* instead of *granice*. A rather serious grammar error can be observed in the translation by Google Translate in which it used the wrong preposition *u* (in) that could change the meaning of the whole legal rule described by the text: *prijevoz u uključenim zemljama* (road traffic *in those countries* instead of *between those countries*) (grammar error – wrong preposition). Bing has translated this sequence correctly. It has also translated correctly the collocation *beteiligte Staaten*, for which Google's solution was a less appropriate expression *uključene zemlje*. Only in one case, both machine translators used a less appropriate collocation *vremenske dozvole* (= *time permission*) rather than a more common collocation *privremene dozvole* (*temporary permission*) which was correctly used by the human translator (discourse error- style: wrong collocation). It should be noted that in this respect, in our opinion, a human translator made a mistake that does not occur in two machine translations. The phrase *der Bedarf des Straßengüterverkehrs der beteiligten Länder* was translated by using the legal term *zahtjev* for *Bedarf* but in the context of the whole text, it was unnecessary to use the legal term instead of the corresponding general language term *potreba* (*need*) that was used in both machine translations.

Error typology	Google Translate	Microsoft Bing
Orthography	Omission	
Lexis	Wrong word choice (2)	Wrong word choice
Grammar	Noun-adjective congruence Prep. phrase instead of Genitive	
Morphology		
Syntax		
Semantic	Wrong collocation Wrong collocation	Wrong collocation (2)
Discourse		

## Discussion

As our analysis has revealed, several types of errors occurred in the machine translations. Although some errors occurred in the human translation as well, they mostly refer to additions and using legal terms and collocations suggested

by the IATE in the cases when it was not necessary.

The contrastive analysis of errors in two machine translators reveals various types of errors at all levels. Not all of them influence the final information for the target receiver in the same way, so it is difficult to qualify one or

the other machine translator as better or more reliable. To be able to draw that kind of conclusion, a deeper and more complex analysis of every text and its informational value for the respective legal document it was excerpted from should be conducted, which is beyond the scope of this research.

If we observe the findings from the quantitative point of view, there are more errors in Google Translate (23 errors in four legal texts) than in Microsoft Bing (12). In terms of the language quality of the translation, Bing had a better score, as Google Translate made some serious errors on the level of grammar and morphology. The same translator had more difficulties in choosing proper legal terms and collocations, while in one case Bing's mixing up terms in the same translation could lead to misunderstandings, though the wider context of the whole text has enabled proper understanding of the information. On the other hand, wrong usage of prepositional phrases by Google Translate could have led to the wrong interpretation of the whole article, and as such, to legal consequences. Our contrastive analysis has also revealed that errors most often occurred on the level of lexis (both machine translators made five errors of wrong choice each) and on the semantic level. Here, the cause of the problem was a translator's insufficient knowledge of legal collocations and phrases. As for the gravity of errors, Google Translate made a number of errors in the field of grammar and morphology (eight out of altogether 23 errors). These levels of linguistic analysis along with the semantic level, especially errors referring to legal collocations and phrases, should be reviewed and improved by experts who develop translation programs for Google.

Before the conclusion, we need to stress our awareness of the limitations of this study. Our findings have revealed only a tiny piece of numerous layers of the complex issue of machine translation. That is why our results should be seen as indicative rather than conclusive.

## Concluding Remarks

Translation the field of law, especially translating legal rules of the EU that have to be implemented in all its member states is a demanding job, not only within the EU Commission dealing with official translations in 24 languages but also for private citizens who want to be informed about those rules for their private or professional purposes. That is why the quality of machine translations accessible online is very important and the errors occurring in them can strongly influence the lives of their users in different ways. The contrastive analysis of errors in two machine translators has shown that specific types of errors appear more often than some other types. It has also revealed that in translating legal texts the translator Microsoft Bing is more reliable than Google Translate.

Of course, we have to keep in mind the limitations of this study. Firstly, it was conducted on the translations of four short texts from German into the Croatian language. Maybe quite different results and different types of errors could have been revealed if other language pairs had been involved or if larger sequences of legal texts had been explored. There are some other variables that could have influenced the results of the study, such as the specific languages involved, specific branches of law, whether the source text stems from one legal system and culture and the target text from quite different ones, etc. In spite of that, we believe that our contrastive analysis that was conducted against clearly defined criteria and error taxonomy has highlighted specific problems in machine translation that should be solved in the close future. It is obvious that machine translation should be improved, especially that of Google Translate which, according to the results of our small study, made twice as many errors as Microsoft Bing Translator [17,18].

In that respect, the work of scientists in the field of information technology and their cooperation with linguists in developing computer systems for correcting errors is of utmost importance. For that purpose, further investigation of error typology for different language pairs and for different combinations of languages should be conducted in the future. By developing specific rule-based types of errors, they could create computer programs for automatic error correction in translation outcomes. Having in mind the complexity of languages for specific purposes like the language of the law, it is not realistic to expect a generally applicable tool for all types of texts. This implies that defining an appropriate error taxonomy is a very challenging task. Based on the results of this study, we must be aware that human annotators can distinguish a larger number of errors than the existing automatic tools. However, in the context of multilingual communities like the European Union, that cannot be a satisfying solution for the future. In this respect, we agree with Popovic [2] that the future work of expert teams should be focussed on developing large error taxonomies that include appropriate sub-sets for specific language pairs and, as we have shown by this research, for specific professional fields, one of them being the language of the law.

## References

1. Costa A, Ling W, Luis T, Correia R, Coheur L (2015) A linguistically motivated taxonomy for Machine Translation error analysis. *Machine Translation* 29(2): 127-161.
2. Popović M (2018) Error Classification and Analysis for Machine Translation Quality Assessment. In: Moorkens J, Castilho S, et al. (Eds.), *Translation Quality Assessment*, Springer, pp: 129-158.

3. Šarčević S (1997) New Approach to Legal Translation. *Kluwer Law International* 13: 308.
4. Baker M, Saldanha G (2009) *Routledge Encyclopedia of Translation Studies*. Routledge, pp: 900.
5. Reiss K, Vermeer H (1984) *Grundlegung einer allgemeinen Translationstheorie*. Max Niemeyer Verlag, pp: 8-245.
6. Nord C (1997) *Translating as a Purposeful Activity. Functionalist Approaches Explained*. St Jerome, Manchester.
7. Kelly D (2005) *A Handbook for Translator Trainers*. St Jerome, Manchester, pp: 186.
8. Trosborg A (1997) *Translating Hybrid Political Texts*. In: Trosborg A (Ed.), *Text Typology and Translation*. John Benjamins, pp: 145-158.
9. Kordić LJ (2018) Specific Issues and Challenges in Translating EU Law Texts. *Athens Journal of Humanities & Arts* 7(3): 235-254.
10. Kordić, LJ, Barna Z (2019) Lingvistička obilježja njemačkog jezika u kontekstu prava Europske unije na primjeru Amsterdamskog ugovora. *Pravni vjesnik* 35(3-4): 223-241.
11. Yankova D (2008) On Some Aspects of Prescriptive Legal Texts in Continental, Common Law and Supranational Jurisdictions. In: Sočanac L, Goddard C, et al. (Eds.), *Curriculum, Multilingualism and the Law*. Nakladni zavod Globus, pp: 483- 495.
12. Popović M (2015) chrF: character n-gram F-score for automatic MT evaluation. *Proceedings of the Tenth Workshop on Statistical Machine Translation (WMT)*, Lisbon, pp: 392-395.
13. Vilar D, Xu J, D'Haro LF, Ney H (2006) Error Analysis of Statistical Machine Translation Output. *Proceedings of the 5<sup>th</sup> International Conference on Language Resources and Evaluation (LREC 2006)*, European Language Resources Association (ELRA), pp: 697-702.
14. Farrus M, Costa Jussa MR, Marino JB, Fonollosa JAR (2010) Linguistic-based Evaluation Criteria to Identify Statistical Machine Translation Errors. *Proceedings of the 14<sup>th</sup> Annual Conference of the European Association for Machine Translation (EAMT 2010)*, St Raphael, France, pp: 167-173.
15. Federico M, Negri M, Bentivogli L, Turchi M (2014) Assessing the Impact of Translation Errors on Machine Translation Quality with Mixed-effects Models. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP 2014)*, Doha, Qatar, pp: 1643-1653.
16. Kirchhoff K, Capurro D, Turner A (2012) Evaluating user preferences in machine translation using conjoint analysis. *Proceedings of the 16<sup>th</sup> Conference of European Association for Machine Translation (EAMT-12)*, Trento, Italy, pp: 119-126.
17. Fishel M, Bojar O, Zeman D, Berka J (2011) Automatic Translation Error Analysis. Pilsen, Czech Republic, pp: 72-79.
18. Papineni K, Roukos S, Ward T, Zhu WJ (2002) BLEU: a method for automatic evaluation of machine translation. *Proceedings of the 40th Annual Meeting of Error classification and analysis for machine translation quality assessment 29 the Association for Computational Linguistics (ACL)*, Philadelphia, pp: 311-318.

