

A Comparative Analysis of the Use of ‘Thereof’ in an English Non-translated Text and the English Machine- and Human-translated Versions of the Hungarian Criminal Code¹

Tímea Kovács

Károli Gáspár University of the Reformed Church in Hungary

kovacs.timea@kre.hu

Abstract

Owing to the recent rise of neural language translation, a paradigm shift has been witnessed regarding the role of translators and reviewers. As neural machine translation is increasingly more capable of modelling how natural languages work, the traditional tasks of translators are being gradually replaced by new challenges. More emphasis is placed on pre- and post-editing (revision) skills and competences, presumably enabling the production of higher quality and near human-made translations.

In my paper, I attempt to demonstrate through the qualitative and quantitative comparison of machine-translated legal texts (acts) to human-translated ones the relevant challenges and dynamic contrasts arising in the process of translating. Through the qualitative and quantitative analysis of the original Hungarian (source language) Criminal Code and its English (target language) machine and human translations, I aim to highlight the peculiar challenges emerging in the process of translation. I also aim to demonstrate what patterns can be observed in translations produced by human and non-human translators.

Keywords: neural machine translation, human-made translation, compound pronouns, thereof, structural differences

Introduction

Due to its specific nature, legal language and legal translation represent a distinct branch within the field of translation studies. The explosion in machine translation in the last few years is posing new challenges to translation studies and to translators in practice. If a constantly increasing number of texts and their translations is becoming easily accessible in electronic form, what new challenges and expectations does this pose for translators? If access is becoming easier to a steadily increasing number of texts and their translations in electronic form, what new challenges and expectations does this pose for translators? Parallel to this, is the concept of a quality standard for translators also evolving? How can we define the quality standard for both machine and human translation? How can we define the quality standard for both machine and human translation?

Legal Translation

One of the most important characteristics of legal translation is to preserve the source language content as accurately as possible and to make it interpretable as clearly as possible, while making its meaning as clear as possible in the target language. Therefore, in the conflict often arising in the process of translation between faithfulness to source language content and linguistic and cultural appropriateness of the target language text, the former is as important, if not more important, than the latter factor. In this process, translators have to take into account

¹ This paper is the revised version of a Hungarian paper titled *Névmási szerkezetek kvantitatív és kvalitatív elemzése forrásnyelvi és fordított büntetőjogi szövegekben* (Quantitative and qualitative analysis of compound pronouns in source language and translated criminal law texts), Kovács (2021).

a number of linguistic, cultural, pragmatic and legal specificities affecting the translatability and interpretability of target and source language texts. The genre specificities of English-Hungarian and Hungarian-English legal translation (Balogh, 2020) and relevant strategies adopted in legal translation (Kovács, 2020b; Kovács, 2020c) are discussed by several authors. Nevertheless, in this paper, the author will focus on one particular aspect of legal translation, i.e., the use of complex pronoun structures in English source language texts and in the machine and human translated texts of Hungarian source language texts. Since the referential function of language elements used in the legal register is crucial, it is worth examining the use and translation of pronoun structures in legal texts, as they have a significant impact on the interpretability and the actual meaning of translated texts.

The present study attempts to compare and analyse the frequency and use of complex pronoun structures in English source-language normative (prescriptive) texts and their (raw) machine and human translations. The study seeks to answer the question of which of the analysed texts correspond more to the language use trends observed in the original English text used as a reference text. In other words, is it the human- or the machine- translated language text that “gravitates” (Halverson 2003, 2010a, 2017) more towards the original, English-language text?

Machine translation, post-editing, corpus linguistics

Machine translation is evolving dynamically thanks to the proliferation of applications based on neural network models (neural-machine translation). Since neural machine translation takes far less time to translate the same source language text into the target language, and the quality of its output is improving day by day, its use is becoming increasingly widespread among translation service providers. The unprecedented pace at which neural machine translation is improving and modifying the translation setting has also led to the emergence of a new translation process, i.e., post-editing. The definition of post-editing according to the ISO 17100:2015 standard is “the editing and correction of a machine translation”. However, the definition of the quality standards and reference texts to be considered in the process of post-editing is still rather imprecise. Although there are quantifiable indicators (metrics), their mere application is not sufficient to accurately assess the quality of a given target language text. Commonly used approaches are BLEU (Papineni et al., 2002) and (H)TER (Snover et al., 2006). In fact, no generally accepted or standard principles exist in post-editing (DePalma, 2013; TAUS, 2016). Practical and theoretical experts involved in translation have developed several possible quality levels for defining post-editing. These distinguish between partial and full post-editing (on the basis of Hu & Cadwell, 2016). However, this two-fold division regarding the depth of post-translation editing aimed at linguistic, textual, and cultural aspects does not seem to be sufficient for a precise qualitative definition of post-editing.

Thanks to the unprecedented progress in neural machine translation, the quality of translated texts is increasing. Nevertheless, the quality of translations hinges on a myriad of factors, such as the type and register of the text, the target and source languages involved, and the machine translation software used (Google, Deep L, etc.). According to some theorists, neural machine- translated texts can even reach a level of near human translation quality (Lample et al., 2018). The question is whether they can ever exceed it.

Parallel to the rapid expansion and development of machine translation, translation competence is increasingly being transformed into post-editing competence (Pym, 2013). However, a precise and standardised definition of what is expected of post-editing is still to be defined. One possible way of carrying out post-editing, either 'partial' or 'full', could be by means of reference texts that can be considered as expected standards in translation. But the question arises whether a text written in the source language and in the relevant register can

function as an actual reference or standard text for a translated text. In an attempt to offer some insight into this question, the use of certain linguistic elements in original and translated target language will be analysed in this paper.

Corpus linguistics is playing an increasingly important role in research, thanks to the proliferation of computer applications in translation methodology since the 1990s. As corpus linguistic tools allow quantitative research to be carried out on a vast range of electronically accessible texts, they are justifiably popular among researchers. At the same time, several theorists point to the dangers inherent in the use of corpus linguistic tools. Solely quantitative and item-based research can make the interpretation of results one-sided (Heltai, 2014). For this reason, the corpus linguistic (Sketch Engine) research used to collect and analyse the data presented in this study is complemented by the qualitative analysis and interpretation of the results.

Aim and methodology of the research

The aim of the present study is therefore to quantitatively investigate and qualitatively analyse particular linguistic elements collected from source language (non-translated English) texts and the machine- and human-translated English texts of source language Hungarian texts, using corpus linguistic tools. In particular, the analysis will seek possible answers to the question of what characteristics can be observed in the use of a specific linguistic element in the source (non-translated) and (human- or machine-translated) target language. In several previous studies, the author of this paper has investigated the interaction of source and target language texts in English-Hungarian and Hungarian-English legal translations in both human (Kovács, 2020b) and machine (Kovács, 2020a, 2020c) translations. As the use of one specific compound pronoun, i.e., ‘thereof’, in English-Hungarian legal translation often poses challenges both for translators and post-editors, its use will be subject to quantitative and qualitative analysis in the present paper.

A total of 4 monolingual ad-hoc corpora has been used for the analysis. The English source language (non-translated) corpus is the *Penal Code of California* (hereinafter: PCC_EN). The Hungarian source (non-translated) language corpus is the *2012. évi C. Törvény a Magyar Büntetőjogról* (Act C of 2012 on the Hungarian Criminal Code) (hereinafter: CC_HU). The human English translation of the above Hungarian text subject to analysis is available on the Internet at the National Legislation Library (hereinafter: CC_HU_EN_HT), and the Deep L Pro (neural network based) machine translation of the text is available online (hereinafter: CC_HU_EN_MT). The Criminal Penal Codes selected for analysis are normative, prescriptive texts according to the classification of Šarčević (1997) and other experts (Kjaer, 2007; Zödi, 2017).

Four corpora have been created using the above four texts. The general characteristics of the four corpora are summarised in the first table.

Table 1: *General characteristics of the corpora*

	PCC_EN	CC_HU_EN_HT	CC_HU_EN_MT	CC_HU
number of words	391257	66326	73872	53130
number of sentences	9336	1430	1481	1531
lexical density	41,9	46,4	49,9	34,7

As can be inferred from data in Table 1, the Penal Code of California (391,257 words) is much more extensive than the Hungarian Criminal Code (53,130 words). As the present

analysis does not cover the similarities and differences in the structure and content of the two criminal codes, this difference is not addressed in the paper. However, it could be the subject of future investigation. The present study focuses exclusively on the original English (non-translated) use of compound pronouns and the use in English (human- and machine-made) translated versions of the Hungarian source text.

By comparing the Hungarian source language text and its two (human- and machine-made) translations, it can be seen that both translated texts (human translation: 66,326 words, machine translation: 73,872 words) are longer than the original Hungarian text (53,130 words). Of the three texts, the machine-made Hungarian translation is the most extensive. Taking a look at the lexical density of the texts, i.e., the average number of words used in a sentence, it is evident that the Hungarian Criminal Code is the least dense (34.7) and, interestingly, the two translations exceed the source language text in lexical density. On the other hand, the two (human- and machine-translated) English target language texts are closer in terms of lexical density to the density observed in the Penal Code of California.

The use of compound pronoun structures in the corpora studied

In this section, the results of the analysis of the use of compound pronouns will be discussed. According to Pavlíčková (2008), the elements beginning with there- and here- are legal deictic linguistic devices that refer to a given speech situation or to a situation, element or relation outside of it. Indeed, they are used instead of the indicative pronouns ‘this’ and ‘that’, associated only with the prepositions required in legal language use. The most frequently used prepositions attached to there and here are -after, -by, -upon, -in, -on, and -to. In these compound pronoun structures, which are most commonly used in legal language, the first element is a locative, ‘there’ or ‘here’, to which a preposition is attached functioning as a post-positioned element (ending). The main difference between structures starting with ‘there’ and ‘here’ is that the element replacing ‘that’ refers to something beyond the context of the text, while this refers to something within it.

As claimed by another interpretation, the prefix ‘there’ refers to another document or element beyond the given text, while the elements starting with ‘here’ refer to the given document (Bázlik et al., 2010). According to Pavlíčková (2008), these linguistic devices are adverbs, which can be used as both conjunctions and adjectives. They denote the relation between the participants in a legal communication situation and the relation to the text and the texts referred to within the text. They may refer to one or more parties, times, places or even parts of them. They can also refer to certain things, such as statements, sentences or ideas. Since their function is to substitute and refer to specific linguistic elements, the author of this paper also prefers to refer to them as compound pronoun structures (Balogh, 2020).

The structures formed by attaching prepositions to the words ‘here’ and ‘there’ are specific examples of legal language use. Therefore, examining their use in the English non-translated source-language legal text and their English (human and machine) translations of the Hungarian source-language legal text could highlight relevant tendencies. Table 2 illustrates the most frequently occurring compound pronoun structures and their frequencies in the three English language corpora: PCC_EN and its human CC_HU_EN_HT and machine CC_HU_EN_MT translations. The Hungarian source language text is excluded from the comparison.

Table 2: *The use of compound pronoun structures in the three English language corpora*

	PCC_EN	CC_HU_EN_HT	CC_HU_EN_MT
1 thereof	296	37	20

2 therein	66	29	23
3 thereafter	39	0	0
4 thereby	39	19	24
5 thereto	34	0	6
6 hereby	27	0	0
7 herein	22	0	0
8 therefrom	20	6	1
9 therewith	13	3	2
10 thereupon	10	0	0
12 hereafter	8	0	0
13 hereinafter	6	12	17
14 thereunto	2	0	0
15 hereinabove	1	0	0
16 hereof	1	7	0
17 hereto	1	0	0
18 heretofore	1	0	0
19therethrough	1	0	0
20 theretofore	1	0	0

The table above shows the frequency of compound pronoun structures. In two texts, ‘thereof’ is the most frequently used structure. It is most frequently found in the English (non-translated) source text (296 instances), followed by the human translation (37 instances) and then the machine translation (20 instances) of the Hungarian source language text. It is interesting to note that ‘hereto’ occurs only once in the English source language reference text, compared to seven times in the human translation, and none in the machine-translated text. Taking a look at the overall number of different compound pronoun structures in each text, it is shown that the English (non-translated) source-language text manifests the greatest variety with 19 elements in total. However, in terms of variety, the two translated texts are identical, though the use of the elements differs. In both texts, ‘thereof’, ‘therein’, ‘thereby’, ‘therefrom’, ‘therewith’, ‘hereinafter’ are used, but ‘hereof’ appears only in the human translation and ‘thereto’ only in the machine translation.

Following this, the position of the phrase ‘thereof’ in respective sentences has been analysed, using the N-grams application of *Sketchengine*. The results of this analysis are summarised in Table 3 below. Table 3 includes the ten most common two-to-six-word combinations in which ‘thereof’ is used.

Table 3: *N-grams in the Penal Code of California (PCC_EN) and their frequency*

PCC_EN	Frequency
part thereof	36
or any part thereof	27
any part thereof	27
conviction thereof	26
upon conviction thereof	25

and upon conviction thereof	22
thereof shall	22
thereof to	19
upon conviction thereof shall	17
conviction thereof shall	17

It is clear from the table that ‘thereof’ appears most often in conjunction with such nouns as ‘part’ and ‘conviction’ and their respective modifications, as well as with the auxiliary ‘shall’.

In Table 4 below, word combinations including the highest numbers of elements, i.e., six, are summarised, together with their frequency.

Table 4: *Six-element N-grams in the Penal Code of California (PCC_EN)*

and upon conviction thereof shall be	13
conviction thereof shall be punished by	11
upon conviction thereof shall be punished	11
thereof shall be punished by a	8

It can be seen that the six-element N-grams including ‘thereof’ contain ‘conviction’ and ‘shall’ and their supplements.

In Table 5 below, N-grams occurring in the human translation of the Hungarian Criminal Code (CC_HU_EN_HT) are listed.

Table 5: *N-grams and their frequency in the human English translation of the Hungarian Criminal Code (CC_HU_EN_HT)*

CC_HU_EN_HT	Frequency
1 facilitating thereof	6
2 or facilitating thereof	6
3 part thereof	5
4 gaining knowledge thereof	4
5 after gaining knowledge thereof	4
6 promptly after gaining knowledge thereof	4
7 required therefor or facilitating thereof	4
8 offense promptly after gaining knowledge thereof	4
9 therefor or facilitating thereof	4
10 knowledge thereof	4

The table above shows that the corpus under study displays lower variance in word combinations. The two most frequent combinations (‘facilitating thereof’, and ‘or facilitating thereof’) occur six times, although the tenth most frequent combination (“knowledge thereof”) also occurs four times. Interestingly, ‘thereof’ appears with or without the combinations ‘facilitating’ and ‘gaining knowledge’ and their supplements. Table 6 contains the six-element word combinations in the human translation of the Hungarian Criminal Code.

Table 6: *Six-element N-grams in the human translation of the Hungarian Criminal Code (CC_HU_EN_HT)*

offense promptly after gaining knowledge thereof	4
impacts thereof increasing or decreasing the	2
and the impacts thereof increasing or	2
the impacts thereof increasing or decreasing	2
financial year and the impacts thereof	2
authorities before they become aware thereof	2
thereof increasing or decreasing the profit	2
programs required therefor or facilitating thereof	2
year and the impacts thereof increasing	2

The table above shows that the most frequently used six-element combinations include ‘impacts’, ‘becoming aware’ and ‘increasing’.

Following this, an analysis of N-grams in the machine translation of the Hungarian Criminal Code has been implemented. The results of the analysis are summarised in Table 7.

Table 7: *N-grams and their frequency in the English machine translation of the Hungarian Criminal Code (CC_HU_EN_MT)*

CC_HU_EN_MT	Frequency
part thereof	7
thereof shall	5
part thereof shall be	5
part thereof shall	5
thereof shall be	5
remaining part thereof shall	4
respect thereof	4
in respect thereof	4
the remaining part thereof	4
the remaining part thereof shall	4

Based on the data in the table above, it can be concluded that ‘part’ and ‘shall’ are the most frequent elements in word combinations with ‘thereof’. Therefore, the syntactic embedding of the lexical items in the machine-translated text displays more similarity to the

original non-translated English text (the Penal Code of California). In addition, the word ‘respect’ also appears. The six-element structures are summarised in Table 8 below.

Table 8: *Six-element N-grams in the machine translation of the Hungarian Criminal Code (CC_HU_EN_MT)*

the remaining part thereof shall be	4
part thereof shall be converted into	4
or the remaining part thereof shall	4
thereof shall be converted into imprisonment	4
remaining part thereof shall be converted	3
service or the remaining part thereof	3
a third person in respect thereof	3
unlawful advantage or a promise thereof	3

Table 8 shows that the linguistic units ‘part’ and ‘shall’ appear the most frequently in word combinations with ‘thereof’, together with ‘respect’ and ‘promise’.

In summary, the above table shows that of the two analysed texts, the frequency with which ‘thereof’ is used converges more in the human English translations with the original (non-translated) English Penal Code than in the machine translation. On the other hand, examining the syntactic embeddedness of the words and their frequency in the most frequent word combinations, it is clear that the machine translation converges more with the original (non-translated) English text.

Qualitative analysis of complex pronoun structures used in translations

The quantitative analysis was followed by a qualitative analysis of the human and machine translations of the Hungarian Criminal Code. The purpose of the analysis was to compare the use of ‘thereof’ in extracts of the two translations. Based on the data from the above analysis, ‘thereof’ appeared 37 times in the human translation and 20 times in the machine translation. What these elements have in common is that ‘thereof’ appeared in word combinations with ‘part’.

Example 1: *The use of ‘thereof’ in the human and machine translation extracts of the Hungarian Criminal Code*

CC_HU	CC_HU_EN_HT	CC_HU_EN_MT
Ha a fiatalok a munkakötelezettségének önként nem tesz eleget, a közérdekű munkát vagy annak hátralévő részét szabadságvesztésre kell átváltoztatni.	If a juvenile offender fails to voluntarily comply with his work obligation, the community service work, or the remaining part thereof, shall be substituted to imprisonment.	If the juvenile does not voluntarily fulfil his/her work obligation, the community service or the remaining part thereof shall be converted into imprisonment.

It is clear that in the Hungarian text, ‘remaining part thereof’ appears as the translation of **annak hátralévő részét** (“the remaining part thereof”) of the Hungarian text in both

translations. However, in only four cases was there an overlap in the use of ‘thereof’ in the human- and machine-translated texts. Examining the fragments of the Hungarian source text which contain translation of ‘thereof’, it can be seen that in most cases it appears as the translation of the conjugated forms of the demonstrative pronouns ‘ez’ (‘this’) or ‘az’ (‘that’) (‘ezek’, ‘azok’, ‘ennek’, ‘annak’, ‘ezt’, ‘azt’, etc.) (20 elements in total). This is illustrated in Example 2:

Example 2: *The use of ‘thereof’ in the human and machine translations of Hungarian conjugated demonstrative pronouns (‘annak’)*

CC_HU	CC_HU_EN_HT	CC_HU_EN_MT
[...] a csoportot olyan életfeltételek közé kényszeríti, amelyek azt vagy annak egyes tagjait pusztulással fenyegetik,	[...] constrains the group into living conditions threatening the demise of the group on the whole or certain members thereof ,	[...] forces the group into living conditions that threaten to destroy it or some of its members ,

‘Thereof’ also appears in the English human translation of the target language as the translation of the possessive conjugated form of the noun (‘melléklet’, ‘Annex’) in the Hungarian source text (in six cases). This is illustrated in Example 3:

Example 3: *The use of ‘thereof’ in the human and machine translations of Hungarian possessive conjugations of nouns (‘melléklet’)*

CC_HU	CC_HU_EN_HT	CC_HU_EN_MT
egyek, a halálbüntetés, a kínzás vagy más kegyetlen, embertelen vagy megalázó bánásmód vagy büntetés során alkalmazható áruk kereskedelméről szóló, 2019. január 16-i (EU) 2019/125 európai parlamenti és tanácsi rendelet - II. mellékletében meghatározott áruk tekintetében - 33. cikkének,	Article 17 of Council Regulation (EC) No. 1236/2005 of 27 June 2005 concerning trade in certain goods which could be used for capital punishment, torture or other cruel, inhuman or degrading treatment or punishment, as regards the goods specified in Annex II thereof ;	17. Article 33 of Regulation (EU) 2019/125 of the European Parliament and of the Council of 16 January 2019 concerning trade in certain goods which could be used for capital punishment, torture or other cruel, inhuman or degrading treatment or punishment, as regards the goods defined in Annex II ,

Interestingly, ‘thereof’ also appears in the human translation as the counterpart of a Hungarian source language subordinate clause (four instances in total), see Example 4:

Example 4: *The use of ‘thereof’ in the human and machine translations of Hungarian clauses*

CC_HU	CC_HU_EN_HT	CC_HU_EN_MT
[...] nem tett meg a hatáskörében álló minden szükséges és indokolt intézkedést annak érdekében, hogy megakadályozza elkövetését, vagy haladéktalanul nem tett feljelentést azt követően, hogy a bűncselekmény elkövetéséről tudomást szerzett.	[...] the public executive did not take the measures deemed necessary and justified within his power to prevent the criminal offense, or did not report the criminal offense promptly after gaining knowledge thereof.	[...] failed to take all necessary and reasonable measures within his or her power to prevent the commission of the offence or to report the offence promptly after becoming aware that it had been committed.

It can be observed that in the human-made translations ‘thereof’ mostly appears as the English counterpart of Hungarian conjugated demonstrative pronouns, e.g. ‘ez’, ‘az’ (Example 2 and 3) or as reference to a whole clause (Example 4). However, another tendency can also be

noted. In machine-translated texts, the literal translation of a given Hungarian reference word and the marking of possessive relations by using possessive adjectives (its, their) (Example 2), and that of a whole clause (Example 4) are more prevalent than the use of ‘thereof’.

It is interesting to look at extracts of translations where the term ‘thereof’ appears in the machine-made (13 times) but not in the human-made translation. In the machine-translated version, ‘thereof’ appears most often as part of a word combination together with ‘part’ (3 times in total). Such an example is highlighted in Example 5 below.

Example 5: *The use of ‘thereof’ in the human and machine translations of Hungarian clauses*

CC_HU	CC_HU_EN_HT	CC_HU_EN_MT
Ha az elítélt a pénzbüntetést nem fizeti meg, illetve részletfizetés engedélyezése esetén egyhavi részlet megfizetését elmulasztja, a pénzbüntetést vagy annak meg nem fizetett részét szabadságvesztésre kell átváltoztatni.	If the defendant did not pay the fine, or if failed (sic. orig.) to pay a monthly installment where payment by installment had been authorized, the fine or the unpaid portion shall be substituted by the appropriate term of imprisonment.	If the convicted person has ceased to serve his sentence, the degree of execution of the sentence or the remaining part thereof shall be imprisonment.

In addition, ‘thereof’ is also used in combinations with ‘possession’, ‘promise’, ‘respect’, and ‘members’. An example of this is shown in Example 6 below.

Example 6: *The use of ‘thereof’ in the machine translation of the Hungarian Criminal Code*

CC_HU	CC_HU_EN_HT	CC_HU_EN_MT
Aki gazdálkodó szervezet részére vagy érdekében végzett tevékenységével kapcsolatban jogtalan előnyt kér, avagy a jogtalan előnyt vagy ennek ígéretét elfogadja,	Any person who requests or receives an unlawful advantage in connection with his activities performed for or on behalf of an economic operator, for himself or for a third party, or accepts a promise of such an advantage,	Any person who, in connection with the activities of an economic entity for or on behalf of the entity, requests an unlawful advantage or accepts an unlawful advantage or a promise thereof,

Summary

The above data clearly shows that the frequency of ‘thereof’ in the human translation of the Hungarian source language text (*Act C of 2012 on the Hungarian Criminal Code*) corresponds more closely to the frequency pattern observed in the original (non-translated) English text, the *Penal Code of California*. At the same time, the syntactic embeddedness of ‘thereof’ in two-to-six-element word combinations in the machine-translated text corresponds better to the original (non-translated) English text. When comparing the human-made and machine-translated English texts to the source-language Hungarian text, it is apparent that in the machine translation, ‘thereof’ appears exclusively as a translation of word combinations formed by the conjugated forms of demonstrative pronouns and nouns. In the human translation, there is greater variation: in addition to word combinations formed by the conjugated forms of demonstrative pronouns and nouns, it also appears as a reference to conjugated nouns and even to whole phrases. There are even examples of ‘thereof’ appearing as redundant additions to the English translation of the Hungarian source language text.

Conclusion

In terms of the frequency of the use of ‘thereof’, the human translation text corresponds more closely to the frequency patterns observed in the original English source language text. Nevertheless, its syntactic positioning in word combinations and sentences in the machine-translated English text shows a more striking similarity to the original (non-translated) English text. This may confirm the gravitational effect observed in translated text by Halverson (2003, 2010, 2017), whereby lexical or grammatical items specific to a given register are more likely to be chosen by translators, and thus their use may be over-represented in translations, or even redundant. In the human-made translation, the human translator tends to use ‘thereof’ as a characteristic element of legal language use, although it is not used in the same way as in the original English language text. The use of ‘thereof’ in the machine-translated text follows more closely its syntactic positioning in the original English text than in the human-made translation. Nevertheless, in the human translation the frequency of using ‘thereof’, is more in line with the frequency patterns observed in the original English text. Further data and analysis are needed to confirm this claim. However, when defining a quality standard for post-editing legal texts, it is crucial to take into account that the actual use of some elements characteristic of the legal register may be over-represented or even redundant in human-made translations. Therefore, the machine-translated versions of such source language texts may correspond better to actual (non-translated) source language use.

Literature:

- Balogh, D. (2020). *Műfajtudatosság a jogi szakfordításban és szakfordítóképzésben*. ELTE Doktori Iskola, unpublished doctoral dissertation. <https://doi.org/10.15476/ELTE.2020.035>
- Bázlik, M. & Ambrus, P. & Beclawski, M. (2010). *The Grammatical Structure of Legal English*. Translegis.
- DePalma, D. (2013). *Post-editing in practice*. <http://www.tcworld.info/e-magazine/translation-and-localization/article/post-editing-inpractice/>
- Halverson, S. (2003). The cognitive basis of translation universals, *Target*, 15(2), 197–241. <https://doi.org/10.1075/target.15.2.02hal>
- Halverson, S. (2010). Cognitive Translation Studies and the merging of empirical paradigms: The case of ‘literal translation’, *Translation Spaces* 4(2), 310–340, <https://doi.org/10.1075/ts.4.2.07hal>
- Halverson, S. (2017). Gravitational pull in translation. Testing a revised model. In De Sutter, G. & Lefer, M. A. & Delaere, I. (Eds.), *Empirical Translation Studies. New Methodological and Theoretical Traditions* (pp. 9–47). De Gruyter Mouton. <https://doi.org/10.1515/9783110459586>
- Heltai, P. (2014). Pszicholingvisztikai és neurolingvisztikai eljárások a fordítástudományban, *Fordítástudomány*, 16(2), 5–23.
- Hu, K. & Cadwell, P. (2016). A Comparative Study of Post-editing Guidelines, *Baltic J. Modern Computing*, 4(2), 346–353. <https://aclanthology.org/W16-3420.pdf>
- ISO, ISO 17100:2015: Translation services – Requirements for translation services, http://www.iso.org/iso/catalogue_detail.htm?csnumber=59149
- Kjaer, A. L. (2007). Legal Translation in the European Union. A Research Field in Need of a New Approach. In Kredens, K., Goźdz-Roszkowski, S. (Eds.) *Language and the Law: International Outlooks* (pp. 69–95). Frankfurt am Main: Peter Lang.

- Kovács, T. (2020a). Humans, machines, and texts: The implications of the rise of neural machine translation for the educators of future translators. In Besznyák, R., Fischer, M. & Szabó, Cs. (Eds.) *Fit- For-Market Translator and Interpreter Training in a Digital Age* (pp. 135–151). Vernon Press.
- Kovács, T. (2020b). Interferenciajelenségek összehasonlító vizsgálata nem fordított magyar és angolról magyarra fordított jogi szaknyelvi szövegekben. In Szabóné Papp, J. (Ed.). *Alkalmazott Nyelvészeti Közlemények, Interdiszciplináris tanulmányok, A Miskolci Egyetem Közleményei. Publications of the University of Miskolc*, 13(2), 150–165. Miskolci Egyetemi Kiadó.
- Kovács, T. (2020c). Jogi szaknyelvi és fordítási feladatok. In Dróth, J. (Ed.). *Fordítási és szaknyelvi gyakorlatok távoktatásban.* (pp. 122–179). KRE–L'Harmattan, Károli Könyvek.
- Kovács, T. (2021). Névmási szerkezetek kvantitatív és kvalitatív elemzése forrásnyelvi és fordított büntetőjogi szövegekben, *Modern Nyelvoktatás*, XXVII. évfolyam, 3 – 4. szám, p. 58–72.
- Lample, G., Ott, M., Conneau, A., Denoyer, L., Ranzato, M.A. (2018). *Phrase-based and Neural Unsupervised Machine Translation.* Cornell University, [online]. <https://aclweb.org/anthology/D18-1549>
- Pavličková, E. (2008). Ways of addressing in legal discourse. *Discourse and Interaction*, 1(1), 87–98. <https://journals.muni.cz/discourse-and-interaction/article/view/6901/8267>
- Papineni, K., Roukos, S., Ward, T., and Zhu W.J. (2002). BLEU: A Method for Automatic Evaluation of Machine Translation. *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), Philadelphia, July 2002*, pp. 311–318.
- Pym, A. (2013). Translation Skill-Sets in a Machine-Translation Age. *Meta: Translators' Journal*, 58(3), 487–503.
- TAUS. (2016). TAUS Post-Editing Guidelines, <https://www.taus.net/think-tank/articles/postedit-articles/taus-post-editing-guidelines>
- Snover, M. & Dorr, B. & Schwartz, R. & Micciulla, L. & Makhoul, J. (2006). A Study of Translation Edit Rate with Targeted Human Annotation, file:///Users/tita/Downloads/A_study_of_translation_edit_rate_with_targeted_hum.pdf
- Zódi, Zs. (2017). Jogi szövegtípusok. *Magyar Jogi Nyelv.* (2), 20–30. <https://joginyelv.hu/jogi-szovegtipusok/>

Sources used:

Penal Code of California

<https://leginfo.ca.gov/faces/codesTOCSelected.xhtml?tocCode=PEN>, last accessed on 31 August 2022

The human translation of Act C of 2012 on the Hungarian Criminal Code, available: https://njt.hu/translation/J2012T0100P_20210708_FIN.pdf, last accessed on 31 August 2022

The machine (Deep L Pro) translation of Act C of 2012 on the Criminal Code, last accessed on 31 August 2022