

Tímea Kovács

Károli Gáspár University of the Reformed Church
Faculty of Humanities and Social Sciences
Institute of English Studies / Department of English Linguistics

Human and machine translation: a comparative analysis of neural machine- and human-translated EN-HU and HU-EN legal texts

<https://doi.org/10.48040/PL.2022.1.5>

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 (Castilho et al., 2019). Consequently, more emphasis is placed on pre- and post-editing (revision) skills and competences (Pym, 2013; Robert et al., 2017), enabling the production of higher quality and near human-made translations. Therefore, the efficiency of pre- and post-editing largely depends on how aware translators are of the mechanisms and limitations of neural machine translation tools adopted in given language pairs (Lample et al., 2018). This paper aims to demonstrate through the comparison of the neural machine and human-translated English and Hungarian translations of Hungary's Fundamental Law and the U.S. Constitution, respectively, the different challenges arising in the course of translation and posed by post-editors, especially from the perspective of comprehensibility and well-formedness.

Key words: *neural machine translation, human-made translation, low-resource language pair, comprehensibility, meaning, well-formedness*

Introduction

In this paper, I intend to compare texts translated by humans and through neural machine-based (Google) translation application. In particular, target language texts produced by means of neural machine translation will be compared to those produced by means of human translation and will be evaluated from the perspective of their comprehensibility and well-formedness. Through examples, the relevant challenges and dynamic contrasts arising in the process of translating in the specific legal language domain will also be highlighted.

Legal language use and translation

There has been a myriad of scholars doing research on the peculiarities of English legal language use. Šarčević (1997) and other scholars (Kjaer, 2007; Cao, 2012; Zódi, 2017) examined legal language use as to how legal texts can be classified as descriptive, perspective, or hybrid texts based on the theory of speech acts (Austin, 1962; Searle, 1979). Other scholars have focused on the grammatical and structural aspects of English legal language use (Pavličková, 2008; Bázlik et al., 2010). In addition to the pragmatic, grammatical, and structural features of legal language, legal translation has also received considerable attention. A significant development can also be observed in the field of presenting new research methods in legal translation and the possible applications of corpus linguistic tools (Biel – Engberg, 2013; Biel 2014, 2019; Khaydarova, 2019). There has been significant research into the use of Hungarian legal language (B. Kovács, 1999; Dobos, 2014; Minya-Vinnai, 2018; Stíluskönyv, 2014; Tóth – Kurtán, 2017), as well. Nevertheless, there are relatively fewer academic publications on the translation of English-Hungarian and Hungarian-English legal texts (Balogh, 2020; Kovács, 2018, 2020). The recent developments in (neural) machine translation and artificial intelligence pose new challenges for translators, so it is increasingly more relevant to carry out more comparative analysis about the recurring patterns in human and neural machine translation.

Research methodology

The basic hypothesis of this paper is that neural machine translation (NMT) produces a near-human quality translation (Lample et al., 2018). Nevertheless, different recurring patterns can be observed in human (HT) and NMT translated texts.

In order to compare the different patterns in NMT and HT legal texts, six texts were compared by applying corpus-linguistic devices (Sketchengine). In the analysis, monolingual and translated corpuses were examined. In the Hungarian–English direction, *Magyarország Alaptörvénye*¹ (The Fundamental Law of Hungary, hereinafter referred to as FUND_HU) served as the Hungarian source-language (SL) text which was also used as a reference corpus for the English–Hungarian human (CONS_HT_HU) and neural machine (CONS_NMT_HU) translated texts. In the English–Hungarian direction, the U.S. Constitution² was selected as the English SL text (hereinafter referred to as CONS_EN) and used as the English monolingual reference corpus. Its human³ (hereinafter referred to as FUND_HT_EN) and neural machine English translated texts were subject to quantitative and qualitative analysis. First, the six texts were subject to qualitative analysis by means of Sketchengine. Table 1 below contains the basic statistical data of the six texts.

Table 1. Corpus statistics of the analysed texts

	FUND_H U	FUND_HT_ EN	FUND_NMT_ EN	CONS_E N	CONS_HT_ HU	CONS_NMT_ HU
words	11, 596	17, 814	16, 532	4, 376	3, 445	3, 282
sentences	565	585	606	119	77	123
words/sentences	20.5	30.4	27.2	36.7	44.7	26.7

As can be inferred from the table above, with regard to lexical density, that is the number of words per sentences, the human Hungarian translation of the U.S. Constitution (CONS_HT_HU) contains the highest number of words per sentence while the SL Hungarian Fundamental Law of Hungary (FUND_HU) is the least dense lexically.

In the dimension of source- and target-language texts in the Hungarian–English direction, it can be seen that the NMT text (FUND_NMT_EN) contains more sentences, more words, and more words per sentences than its Hungarian source-language counterpart (FUND_HU). Interestingly though, the human-based translation (FUND_HT_EN) contains more words and fewer sentences than the SL text. Therefore, among these three texts, the human-translated text has the highest lexical density. A potential explanation to this might be that human translators tend to prefer analytical versus synthetic word formation strategies. In the English–Hungarian direction, the HT translation contains impressively fewer sentences and words than the SL text, while its NMT counterpart contains fewer words than the SL text and more sentences than the SL text. Therefore, in terms of lexical density, the NMT text has a lower while the HT has a higher density than the SL text. Therefore, based on the above results, the NMT texts tend to be lexically less dense than their HT counterparts. This might be due to the fact the NMT translations tend to converge more to a simplified language use. Nevertheless, the tendency of NMT to produce lexically less dense sentences, while in human translation the use of denser sentences can be observed needs more data and research.

¹ <https://net.jogtar.hu/jogszabaly?docid=a1100425.atv>, last accessed: 15 January 2021

² [archives.gov/founding-docs/constitution-transcript](https://www.archives.gov/founding-docs/constitution-transcript), last accessed: 15 January 2021

³ https://njt.hu/translated/doc/TheFundamentalLawofHungary_20191213_FIN.pdf, last accessed: 15 January 2021

Next, the six texts were analysed and compared in terms of the specific word types found in them. First, the most frequently occurring nouns in the Hungarian HT and NMT translations of the U.S. Constitution and the SL text were examined.

Table 2: Nouns in CONS_EN, CONS_HT_HU, and CONS_NMT_HU

CONS_EN	CONS_HT_HU	CONS_NMT_HU
States	a	az
United	az	a
state	állam	állam
President	áll	áll
Congress	kongresszus	amely
section	amely	törvény
year	törvény	minden
House	elnök	kongresszus
case	személy	elnök
Senate	minden	szakasz

It can be seen in the table above that for some reason, the corpus-linguistic analysis application, Sketchengine classified some articles and relative pronouns as nouns. It could be due to the fact that the application is based on the English classification of word classes. Nevertheless, when it comes to comparing the list of the most frequently used nouns, almost the same terms can be found. It is interesting though that *törvény* ('law') appears among the most frequently used nouns in both NMT and HT texts, while in the source-language English text, it does not. This could be due to the fact that not only 'law', but also other English words (for example 'act', 'regulations') are regularly translated as *törvény* ('law'). Therefore, the frequency of using this term is higher in the Hungarian translation.

In Table 3, the most frequently used nouns are compared in the source-language Hungarian Fundamental Law of Hungary and its human and neural-machine made English translations.

Table 3: Nouns in FUND_HU, FUND_HT_EN, and FUND_NMT_EN

FUND_HU	FUND_HT_EN	FUND_NMT_EN
a	National	law
az	Assembly	government
törvény	Act	President
elnök	government	National
országgyűlés	President	Assembly
cikk	law	right
jog	right	state
Magyarország	state	Republic
kormány	Republic	Hungary
alaptörvény	Hungary	member

The table above suggests that there are no significant differences in the list of the ten most frequently used nouns in NMT and HT texts.

Next, the list of the ten most frequently used verbs was examined. Table 4 contains the comparative list of the ten most frequently used verbs in the SL U.S. Constitution and its Hungarian HT and NMT translations.

Table 4: Verbs in CONS_EN, CONS_HT_HU, and CONS_NMT_HU

CONS_EN	CONS_HT_HU	CONS_NMT_HU
be	van	van
have	egyesül	egyesül
make	kell	kell
provide	választ	rendelkezik
hold	tesz	lesz
enter	fogad	tölt
appoint	szabályoz	ad
elect	rendelkezik	tesz
choose	gyakorol	előír
grant	megillet	választ

The table above shows that the three most frequently used verbs in both the NMT and HT texts are the same. There are some terms, for example, *tesz* ('to do' or 'make'), *választ* ('to elect'), *rendelkezik* ('to impose', 'to order') which can be found in the list, though their ranking differs. It should be noted, though, that the Hungarian NMT text contains the term *lesz* ('will be'), which is not likely to occur in Hungarian legal texts.

In Table 5, the list of the 10 most frequently used verbs can be seen in the SL Fundamental Law and its English HT and NMT translations.

Table 5: Verbs in FUND_HU, FUND_HT_EN, and FUND_MT_EN

FUND_HU	FUND_HT_EN	FUND_NMT_EN
van	be	be
határoz	have	have
gyakorol	lay	specify
kell	provide	take
választ	adopt	elect
hoz	elect	determine
nemz	exercise	establish
tart	establish	exercise
alkot	decide	decide
dönt	take	declare

Data in the table above suggests that almost the same verbs are used in the HT and NMT translations. In the HT text, there are three verbs, namely *lay*, *provide*, and *adopt* while in the NMT three, that is, *specify*, *determine*, *declare*, respectively, which do not occur among the ten most frequently used verbs in the other text.

Last, the distribution of adverbs was examined. Table 6 contains the comparative list of the ten most frequently used adverbs in the SL U.S. Constitution and its two Hungarian HT and NMT translations.

Table 6: Adverbs in CONS_EN, CONS_HT_HU, and CONS_MT_HU

CONS_EN	CONS_HT_HU	CONS_NMT_HU
not	nem	nem
thereof	kivéve	sem
as	ahogy	de
then	csak	kivéve
herein	sem	ahogy
so	amikor	amikor
faithfully	jelen	fel
prior	de	akkor
respectively	fel	mielőtt
together	arra	amint

It can be inferred from the table above that there are some differences in the list of the adverbs. Three terms can only be found in the HT text, *csak*, ('only'), *jelen* ('present'), and *arra* (which is a demonstrative pronoun, 'there'), while *amint* ('as soon as') is present only in the NMT text.

Table 7 contains data about the use of adverbs in the HT and NMT English translations of the Fundamental Law of Hungary.

Table 7: Adverbs in FUND_HU, FUND_HT_EN, and FUND_MT_EN

FUND_HU	FUND_HT_EN	FUND_NMT_EN
nem	not	not
ahogy	as	as
is	only	only
csak	well	well
amint	also	also
de	no	no
haladéktalan	longer	immediately
akkor	freely	annually
már	simultaneously	independently
amíg	abroad	later
legfeljebb	autonomously	thereof

It can be observed that the six most frequently used adverbs in the two translations are the same. The last four are different, but any of these terms is likely to be used in legal texts. It should be highlighted though that in the source-language English text (the U.S. Constitution), which could be used as a source-language reference text (see Table 6), there are two adverbs, that is, *thereof* and *herein*, which are among the ten most frequently used adverbs. As these terms are frequently used in legal texts to refer to something that has already been mentioned, they can be expected to occur in the translated texts as well. Nevertheless, only the NMT text contains a similar term, *thereof*, as the 11th most frequently used adverb. In the human-translated text, it cannot be found among the twenty most frequently used adverbs.

Relying on the findings presented in the tables (Tables 2–7) above, it can be established that regarding the use of words in specific word classes, the category of adverbs, both in the EN-HU and HU-EN translation language directions shows the most differences. This could be due to the fact that as a result of structural differences in the English and Hungarian, in translated texts, the use of adverbs is more inconsistent than that of other word classes. This finding is based on a relatively small set of data, so more research is needed to reinforce it.

Findings of the qualitative analysis

Next, the six texts were subject to qualitative analysis. The appropriateness of the translations was evaluated from the perspective of faithfulness (F) of the translated (TL) utterance to the underlying meaning of the source language text (ST) and the well-formedness (WF) of the target text (TT). Some examples are listed as follows:

Table 8: An example of the HT and NMT translation of FUND_HU

FUND_HU	FUND_HT_EN	FUND_NMT_EN
Az ilyen törekvésekkel szemben törvényes úton mindenki jogosult és köteles fellépni.	Everyone shall have the right and obligation to resist such attempts in a lawful way.	Everyone is entitled and obliged to act on* a lawful manner against such efforts.

The example in Table 8 demonstrates that the NMT translation is comprehensible. In terms of its faithfulness to the source-language text, it is appropriate. There is one inappropriate preposition, *on* instead of ‘in’, but the overall meaning is the same as in the target-language text. As for the use of verbs, however, there is a significant difference. The human-translated text uses *shall* as a way of expressing obligation, while in the NMT text, third person indicative form, *is*, is used.

Table 9 shows how a special term, *sarkalatos*, referring to acts that can come into force only if two-thirds of the members of the Hungarian Parliament approve it, has been translated in the HT and NMT English texts.

Table 9: The comparison of the HT and NMT translations of FUND_HU

FUND_HU	FUND_HT_EN	FUND_NMT_EN
Sarkalatos törvény a magyar állampolgárság keletkezésének vagy megszerzésének más eseteit is meghatározhatja.	A cardinal Act may specify other instances of the origin or acquisition of Hungarian citizenship.	Corollary law may also define other cases of the formation or acquisition of Hungarian citizenship.

In the HT text, it has been translated as *cardinal law* and has been used consistently in the entire text. In the NMT text, though, it has been translated as *corollary*. There are some other terms whose usage differs in the two texts, for example, *specify* and *define*, and *origin* and *formation*, but those terms can be used interchangeably depending on the context.

In the table below, more examples are provided to illustrate the translation of *sarkalatos*.

Table 10: The translation of *sarkalatos* (‘cardinal’) in HT and NMT texts

FUND_HU	FUND_HT_EN	FUND_NMT_EN
Az állampolgárságra vonatkozó részletes szabályokat sarkalatos törvény határozza meg.	The detailed rules for citizenship shall be laid down in a cardinal Act.	Detailed rules on nationality are laid down in a law of the past .
A címer és a zászló használatának részletes szabályait, valamint az állami kitüntetések sarkalatos törvény határozza meg.	The detailed rules for the use of the coat of arms and the flag, as well as state decorations, shall be laid down in a cardinal Act.	The detailed rules for using the coat of arms and the flag, as well as the state awards, are determined by the [...]law.

As can be seen in Table 10, in the HT text, the English counterpart of *sarkalatos*, *cardinal* is consistently used. Nevertheless, in the NMT text, it is translated as a *law of the past* or in the second sentence, it is simply left out.

Next, the differences in the HT and NMT Hungarian translations of the U.S. Constitution are examined through examples.

Table 11: An example of the HT and NMT translations of CONS_EN

CONS_EN	CONS_HT_HU	CONS_NMT_HU
Article. I.	I. Cikk	Cikk. ÉN .
Section. 1.	1. §	Szakasz. ÉN .
All legislative Powers herein granted shall be vested in a Congress of the United States, which shall consist of a Senate and House of Representatives.	Minden itt meghatározott törvényhozó hatalom ezennel az Egyesült Államok Kongresszusát illeti, amely Szenátusból és Képviselőházból áll.	Az itt megadott minden jogalkotói hatáskört az Egyesült Államok Kongresszusa kapja, amely Szenátust és Képviselőházat foglal magában.

As is illustrated in Table 11, *Article I*. (in the original text, a full stop is used) has been appropriately translated into Hungarian in the HT text as *I. cikk* (*‘Article I’*), while in the NMT text, it is used as *Cikk. Én.* (*‘Article. I’*, that is, the first-person singular form of the personal pronoun). Other than that, the Hungarian sentence is comprehensible, though there are some post-positional suffixes which are used inappropriately. It is caused by the fact that in the NMT text, no difference is made between definite and indefinite conjugation.

In Table 12, another example demonstrates the relevant differences in the HT and NMT texts.

Table 12: An example of the HT and NMT translations of CONS_EN

CONS_EN	CONS_HT_HU	CONS_NMT_HU
The Senate shall have the sole Power to try all Impeachments .	Kizárólag a Szenátus jogosult a közjogi felelősségre vonás esetén a tárgyalás lefolytatására.	A szenátus kizárólagos hatáskörrel rendelkezik az összes végrehajtás kipróbálására .

The NMT text above is hardly comprehensible in Hungarian. The Hungarian sentence is a transliteration of the English source-language text: it complies with the English linear SVO word order. As a result, there is no focus in the NMT Hungarian sentence. Furthermore, the special legal term, ‘impeachment’ is translated in its literal sense, as *végrehajtás*, ‘execution’. Also, ‘try’ is translated as the noun form of the literal sense of ‘to attempt’, *kipróbálására*, and not in the sense of ‘to prosecute’ or ‘to hold a trial’.

Conclusion

In this paper, NMT translation output was compared to that of human-made translation by means of conducting a quantitative and qualitative text-based micro-analysis on two legal texts, the Fundamental Law of Hungary and the U.S. Constitution. On the basis of the qualitative analysis, it can be stated that both in the HT and NMT English and Hungarian translated texts, the use of adverbs is more inconsistent than that of other word classes (namely, nouns and verbs). It is interesting, though, that with regard to the use of referential adverbs, ‘thereof’, the NMT text is more congruent with the source-language reference text than its human made counterpart. Findings of the qualitative text suggest that there is a significant difference in the quality of the English-Hungarian and Hungarian-English machine translated texts. Both the human and machine Hungarian-English translated texts are comprehensible and fairly appropriate in terms of source-language faithfulness, and target-language well-formedness.

Nevertheless, inconsistencies in the use of special terminology (Haque – Hasanuzzaman – Way, 2020) can be observed in the Hungarian-English NMT texts though: *corollary* or *arctic* are used instead of ‘cardinal’ (law) and *define* instead of ‘specify’. Furthermore, some ellipsis can also be found (see Table 10) in the Hungarian-English NMT texts. Overall, as a result of comparing human and machine English-Hungarian texts, a significant difference in quality can be established. The English–Hungarian NMT text is hardly comprehensible. In the English-Hungarian NMT texts, sentences tend to follow the English word order, are unsusceptible to definite and indefinite conjugations, and contain fuzzy lexical matches. Overall, the Hungarian text produced by neural-machine application underperforms the human translated text both in terms of source language appropriateness and target language well-formedness.

References

- Austin, J. L. (1962): *How to do things with words*. Clarendon Press: Oxford.
- B. Kovács, M. (1999): A funkciógés szerkezetek a jogi szaknyelvben. *Magyar Nyelvőr* 123. évf. 388–94.
- Balogh, D. (2020): *Műfajtudatosság a jogi szakfordításban és szakfordítóképzésben*. ELTE Doktori Iskola, kiadatlan doktori disszertáció. <https://doi.org/10.15476/ELTE.2020.035>
- Bázlik, M. – Ambrus, P. – Beclawski, M. (2010): *The Grammatical Structure of Legal English*. Translegis.
- Biel, Ł. (2014): The textual fit of translated EU law: a corpus-based study of deontic modality. *The Translator*, 20(3), 332–355. DOI: <https://doi.org/10.1080/13556509.2014.909675>
- Biel, Ł. – Engberg, J. – Ruano, R. M. – Sosoni, V. (eds.) (2019): *Research Methods in Legal Translation and Interpreting*. Routledge: London and New York. DOI: <https://doi.org/10.4324/9781351031226>
- Biel, L. – Engberg, J. (2013): Research models and methods in legal translation. *Linguistica Antverpiensia*, New Series–Themes in Translation Studies, (12). DOI: <https://doi.org/10.52034/lanstts.v0i12.316>
- Cao, D. (2012). *Legal translation. The encyclopedia of applied linguistics*. DOI: <https://doi.org/10.1002/9781405198431.wbeal0679>
- Castilho, S. – Gaspari, F. – Moorkens, J. et al. (2019): Editors’ foreword to the special issue on human factors in neural machine translation. *Machine Translation*. 33. 1–7. DOI: <https://doi.org/10.1007/s10590-019-09231-y> [Last accessed: 15 January 2021].
- Dobos, Cs. (2014): Lexikai átváltási műveletek a nyelven belüli fordítás- ban. *Alkalmazott Nyelvészeti Közlemények*, IX. évf., 2. szám. 63–80
- Haque, R. – Hasanuzzaman, M. – Way, A. (2020): Analysing terminology translation errors in statistical and neural machine translation. *Machine Translation* 34. 149–195. DOI:
- Khaydarova, U. (2019): Specific peculiarities of translation in legal document <https://doi.org/10.1007/s10590-020-09251-zs>. *Journal of Legal Studies and Research*, 5(5), 157–165.
- 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*. Frankfurt am Main: Peter Lang.
- Kovács, T. (2018): Átváltási műveletek (megoldások) az angol-magyar jogi szaknyelvi fordításban. *Modern nyelvoktatás. Fordítástudomány különszám*, 25(2–3). 40–53.
- Kovács, T. (2020): 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.
- Lample, G. – Ott, M. – Conneau, A – Denoyer, L. – Ranzato, M.A. (2018): *Phrase-based and Neural Unsupervised Machine Translation*. Cornell University. DOI: <https://doi.org/10.18653/v1/D18-1549> [Last accessed 15 January 2021].
- Minya, K. – Vinnai, E. (2018): Hogyan írjunk érthetően? Kilendülés a jogi szaknyelv komfortzónájából. *Magyar jogi nyelv*, 1. szám. 13–18.
- Papineni, K. – Roukos, S. – Ward, T. – 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. 311–318. DOI: <https://doi.org/10.3115/1073083.1073135>
- 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>
- Pym, A. (2013): Translation Skill-Sets in a Machine-Translation Age. *Meta: Translators' Journal*. 58/3. 487–503. DOI: <https://doi.org/10.7202/1025047ar>
- Robert I.S. – Terryn A.R. – Ureel, J.J. – Remael, A. (2017): Conceptualising translation revision competence: A pilot study on the ‘tools and research’ subcompetence. *The Journal of Specialised Translation*. 28. 293–316.
- Šarčević, S. (1997): *New Approach to Legal Translation*. Hague: Kluwer Law International.
- Searle, J. (1979): *Expression and Meaning: Studies in the Theory of Speech Acts*. Cambridge: University Press. DOI: <https://doi.org/10.1017/CBO9780511609213>
- „Stiluskönyv”: A joggyakorlat-elemző csoport összefoglaló véleménye a határozatszerkesztés tárgykörében 2014. évben végzett elemzés megállapításairól (2014). Kúria, pp. 1–19. https://kuria-birosag.hu/sites/default/files/joggyak/osszefoglalo_1.pdf (Letöltve: 2019. 04. 30.)
- Tóth, J. – Kurtán, Zs. (2017): A jogszabályok szövege – avagy a jogalkotók szövege. *Magyar Jogi Nyelv*, 1. szám. 34–36.

Sources

Magyarország Alaptörvénye: <https://net.jogtar.hu/jogszabaly?docid=a1100425.atv>.

[Last accessed: 31 October 2021].

The Fundamental Law of Hungary:

https://njt.hu/translated/doc/TheFundamentalLawofHungary_20191213_FIN.pdf.

[Last accessed: 31 October 2021].

U.S. Constitution: [archives.gov/founding-docs/constitution-transcript](https://www.archives.gov/founding-docs/constitution-transcript).

[Last accessed: 31 October 2021].

The human translation of the U.S. Constitution: Bódy, P. and Urbán, A. (szerk.): *Szöveggyűjtemény az Amerikai Egyesült Államok történetéhez*, Budapest–Pécs, Dialóg Campus, 2001, 479–495.