



10.2478/topling-2021-0006

Sentence-structure errors of machine translation into Slovak

Katarína Welnitzová*

Constantine the Philosopher University in Nitra, Slovakia

Daša Munková

Constantine the Philosopher University in Nitra, Slovakia

Abstract

The study identifies, classifies and analyses errors in machine translation (MT) outputs of journalistic texts from English into Slovak, using error analysis. The research results presented in the study are pioneering, since the issue of machine translation – with its strong interdisciplinary character and novelty – has not yet been studied in the Slovak academic environment. The evaluation of the errors is based on a framework for classification of MT errors devised by Vaňko, which was arranged for the Slovak language. The study discusses and explains the issues of sentence structure, including predicativeness, syntactic-semantic correlativeness, and a modal and communication sentence framework. We discovered that the majority of the errors are related to the categories of agreement, word order and nominal morpho-syntax. This fact clearly correlates with features of journalistic texts, in which nominal structures and nouns in all realizations are used to a great extent. Moreover, there are some serious differences between the languages which limit and affect the quality of translation.

Key words

machine translation, error typology, publicistic style, synthetic language, analytic language

1. Introduction

In recent decades, we have gradually become conscious of the fact that machines and devices are taking over some human activities, mainly in the fields of industry and medicine. The present digital tools are successively substituting tasks that were primarily determined by human thinking and intelligence. Due to globalization, a parallel situation can be seen in translation – there has been a significant increase in translations, and more sophisticated processes and methods are being adopted to fulfil the market's demands.

House (2015, p.12) construes translation as a cognitive process that relies on the individual's social, interlingual and intercultural experience. The concept of equivalence in translation is considered to be crucial for both translation and its further evaluation. Nowadays, the term "equivalence" seems to be quite controversial, as some theorists are in favour of "equivalence" (e.g. Jakobson, Nida, Catford, and House), and others do not consider it essential to translation (Reiss, Vermeer, and Snell-Hornby). Varied conceptions of translation quality and evaluation methods inspired the establishment of different conceptions and attitudes to translation. In our research, we follow House's conception (2015, p. 14) in which translation is considered to be the substitution of an original text in a source language by a text written in a target language.

* Address for correspondence: Katarína Welnitzová, Department of Translation Studies, Faculty of Arts, Constantine the Philosopher University in Nitra, Štefánikova 67, 949 01 Nitra, Slovakia. E-mail: kwelnitzova@gmail.com

From this aspect, translation is not merely a linguistic act but also a communicative act. It represents the relation between particular language signs of source and target texts (micro-text level) and between source and target texts as a whole (macro-text level) (Müglová, 2009, pp. 226-227).

According to Valdeón (2018, p. 558), media translation (involving journalistic texts) is no longer a marginalized area of translation studies research. Although the main theoretical approaches have traditionally focused on literary translation, many of their tenets have been successfully applied to media translation and have contributed to consolidating it.

In translation studies, there are two other challenging factors: language and identity. Conway (2015, pp. 521-535) posits that news translation has been analysed from three perspectives: political (referring to what news texts travel, why and how); linguistic (referring to language choices) and cultural (referring to the role of journalists and their relationship with intended audiences).

Translation represents an effective intercultural and interlingual communication tool. To get a “raw” translation of news written in “minor” languages, some news agencies, e.g. CNN or BBC have been using systems of machine translation. The research of machine translation (MT) in “major” languages (e.g. English) is well examined and described; on the contrary, MT evaluation for minor languages (e.g. Slovak) is lacking.

To fill the gap and contribute to the research and development in the field of machine translation, this paper highlights and evaluates the difficulties and major errors in machine translation from English into Slovak in publicistic texts.

The aim of the given study is to identify and analyse the most frequent and significant errors in MT in publicistic texts which are frequently translated by machine. We study the issue in MT outputs from English into Slovak and explain it in terms of typological differences between the languages and the characteristics of the publicistic style. In our study, we detect and evaluate errors in MT outputs in the category of sentence structure, including predicativeness, syntactic-semantic correlativeness, and the modal and communication sentence framework of Vaňko’s framework for error identification.

The paper is organized as follows: section 2 describes a theoretical overview of typological characteristics of Slovak and English languages; it briefly characterizes English and Slovak in the context of morpho-syntactic features. It also defines the publicistic style in English and Slovak contexts, highlighting common features and differences between definitions and characteristics of both languages. Section 3 describes the methodology, in which we were interested in the difference in the errors incidence among the categories of predicativeness, modal and communication sentence framework, syntactic-semantic correlativeness, compound/complex sentences and lexical semantics. Section 4 presents the results and section 5 discusses the results/findings in relation to the sentence structure and their effect on error rate. Section 6 summarizes errors found in MT outputs in the category of sentence structure, including predicativeness, syntactic-semantic correlativeness, and the modal and communication sentence framework of Vaňko’s framework for error identification.

2. Typological characteristics of Slovak and English languages

Slovak is characterized mainly as a synthetic language, whereas English is characterized as an analytic language (Vaňko 2015, p. 24; Dolník 2013, p. 87). According to Vaňko (2015, p. 24), grammatical meaning in analytic languages is expressed analytically – i.e. by specific verbs – one carrying lexical meaning and the other auxiliary or grammatical meaning. For example, the English structure ‘he did not write’ has its Slovak equivalent ‘*nepísal*’ and the English structure ‘they will not go’ the Slovak form ‘*nepôjdu*’. A low degree of inflection of main verbs in English correlates with a relatively high number of multiple-word verbs.

The issue of analytic languages can be found in suffixes which indicate grammatical and semantic function of subject and object. Analytic languages do not formally express the difference between nominative and accusative case; word-order in such languages is firmly fixed. In English, there is only one way of expressing the meaning ‘Peter loves Eve’, whereas in Slovak there are two possibilities: ‘*Peter ľúbi Evu.*’ (word-order S+V+O) and ‘*Evu ľúbi Peter.*’ (word-order O+V+S). Suffixes of the proper nouns, e.g. *-0, -u* (*Peter-0, Ev-u*) reliably determine the subject (*Peter-0*) and the object (*Ev-u*) of the sentence. Dolník (2013, p. 89) adds that due to the low degree of verb inflection in English, names and personal pronouns used in the position of subjects are obligatory: ‘I go’ (English), ‘*chodím*’ (Slovak).

According to Ondruš and Sabol (1984, p.186), Slovak has a synthetic morphology. It is characterized by numerous forms and morphemes, derivational affixes changing word bases, and modification of words expressing different grammatical categories (e.g. gender, number, case), typically by one formal feature. For example, the morpheme ‘-u’ in the form ‘žen-u’ (woman in accusative) reflects feminine gender, singular number and accusative case. For synthetic languages, common synonymy and homonymy of case affixes is typical.

Vaňko (2015, p. 27) explains that grammatical meaning in Slovak is expressed by inflection, i.e. by suffixes (e.g. ‘knih-a’ (a book), ‘knih-y’ (of a book/books)). Suffix can distinguish grammatical meaning of a given word, e.g. the form ‘ruk-e’ (to a hand or about a hand) as dative or locative case. Further, Slovak is characterized by numerous verb patterns expressing person, number and tense by identical suffixes, e.g. the suffix -m: ‘číta-m’ (I am reading), ‘robí-m’ (I am making).

2.1 Characteristic features of publicistic style in Slovak and English contexts

According to Mistrík (1989, p. 460), the most significant aim of publicistic style in Slovak is to inform the reader, and then to provide them with a varied, concise and accurate text. The publicistic style is characterized by numerous nouns, numbers, abbreviations, names and marks, but conversely, by a low number of verbs. He states that within the verb categories, the category of aspect (the ratio of perfective and progressive aspect is 1:2), person (the most frequently used person is third-person singular) and tense (mostly past tense) are important. Syntactic and grammatical constructions are not as fixed. The publicistic style does not use connectors much; verbs and notional words are arranged next to each other; facts and terms are emphasized. Texts are condensed and concise. Findra (2013, p. 267) states that the publicistic style offers much new information, a wide range of topics and has a low index of repetition.

Biber and Conrad (2009, p. 109) define “newspaper style” (also called “newspaper writing”) in English as a style characterized by a written register, with emphasis on information. Reader and author do not communicate directly, albeit the most crucial task of newspaper style is to inform. The aim of the style is to report and describe events which have happened, rather than to interpret them (unlike e.g. the role of an essay). Based on large-scale corpus analyses, the authors characterize particular features of newspaper writing. In terms of nominal features, nouns in all realizations (nouns, nouns as pre-modifiers, post-modifiers, nouns in noun phrases) are more common. Nominalization, prepositional phrases after nouns, and attributive adjectives are represented in newspaper style as well. In terms of verbal features, newspaper style uses present simple and past tense to narrate sequences of events. Conversely, the use of modals is minimal (the most frequent is ‘will’ and ‘would’), and passive forms represent 15% of all finite verbs.

Adverbs of time and place are represented in the newspaper style as well. The style uses standard syntax constructions, mainly simple sentences and clauses, and questions are less common. Due to its written register, there are many noun phrases and well-structured syntactic constructions. It prefers the active voice to the passive, as the focus is on the agent of an event and the event itself. Naturally, some discrepancies in genres and subgroups of the newspaper style might occur, which are – due to their functions – different. For example, news and editorials substantially vary in the use of modals and conditionals. As informing about events has a mainly narrating character, constructions similar to hypotheses are not usually used.

Additionally, attributes of the Slovak “publicistic style” according to Mistrík (1989) and Findra (2013) and English “newspaper writing” according to Biber and Conrad (2009) are very similar. We can assume that the features and specifications of the style in both languages are almost identical. In our study, we will refer to it as the publicistic style and publicistic texts.

3. Methodology

The main goal of the research was to identify, classify and analyse the occurrence and causation of the most frequent errors in the category of sentence structure in MT outputs.

According to Schwarzl (2001, p. 321), error analysis is more objective and reliable than translation quality evaluation, since for evaluators error identification is relatively objective and consistent (compared to the category of “correctness”). The understanding of an “error” is identical in both human

and machine translation. An “error” is defined as a deviation from a norm in a particular communication situation. It is caused by misunderstanding (by human or machine) or low linguistic competence.

For our research purposes, we excerpted journalistic texts – 59 newspaper articles from the British online newspaper “The Guardian”. We analysed 3,376 segments (54,442 words). Data were pre-processed by tokenization and segmentation, and then translated by Google Translate (GT, a free service translating between English and over 100 other languages). MT outputs were subsequently imported with their original texts into a virtual environment OSTPERE – a system for translation, post-editing and evaluation of machine translation (Munková et al., 2016; Benko and Munková, 2016). Here, the texts were post-edited by professional translators and later analysed.

Identification and classification of errors was carried out by two Slovak language experts who evaluated and considered MT errors in terms of Vaňko’s categorical framework for errors analysis from a foreign language into Slovak (Vaňko, 2017). Generally applied error typology frameworks (i.e. Dynamic Quality Framework or Multidimensional Quality Metrics) do not consistently map and specify errors of MT in inflectional languages like Slovak; therefore, we used the adjusted error typology for Slovak devised by the Slovak linguist Juraj Vaňko (2017). His framework allows the evaluation of morpho-syntactic and syntactic-semantic relations – the areas with significant errors – in more detail.

Consequently, after identification and classification of MT errors according to Vaňko’s framework (2017, p. 100), we calculated the error occurrence in the examined publicistic texts.

The framework (Table 1) maps five main issues: 1. Predicativity, 2. Modal and communication sentence framework, 3. Syntactic-semantic correlativity, 4. Compound/complex sentences, 5. Lexical semantics.

Table 1. Vaňko’s framework (2017)

-
- I. Predicativity
 - 1. Predicative categories
 - i. tense
 - ii. mood
 - 2. Agreement categories
 - i. agreement in person
 - ii. agreement in number
 - iii. agreement in gender
 - iv. others
 - II. Modal and communication sentence framework
 - 1. Modality
 - i. necessity
 - ii. obligation
 - iii. possibility
 - iv. intention
 - v. epistemic modality
 - vi. evaluativity
 - vii. emotionality
 - viii. others
 - ix. negation
 - 2. Communication and function
 - i. assertiveness
 - ii. interrogativeness
 - iii. directiveness
 - iv. optativeness (wish)

- III. Syntactic-semantic correlativeness
 - 1. Nominal morpho-syntax
 - i. agreement within a noun phrase
 - ii. 'of' genitive
 - 2. Pronominal morpho-syntax
 - 3. Numeral morpho-syntax
 - 4. Verbal morpho-syntax
 - i. non-prepositional phrases
 - ii. prepositional phrases
 - 5. Word-order
 - 6. Others
 - IV. Compound/complex sentences
 - 1. number of sentences
 - 2. semantic relations between sentences
 - 3. connectiveness of sentences (elision of connectors)
 - 4. time shifts
 - 5. others
 - V. Lexical semantics
 - 1. inadequate meaning
 - 2. polysemy
 - 3. semantic compatibility
 - 4. stylistic compatibility
 - 5. homonymy
 - 6. terms
 - 7. derivation
 - 8. omission
 - 9. other phenomena
 - 10. phraseologisms
 - 11. proper names
-

Due to the scope of the study, we focus on the issue of sentence syntax – including predicativeness, modal and communication sentence framework and syntactic-semantic correlativeness.

4. Results

We tested a statistically significant difference in the occurrence of errors in the examined categories. We stated the following null hypothesis:

H0: There is no statistically significant difference in the occurrence of errors among the examined categories (14 categories).

Table 2. Frequency analysis – Cochran's Q test and Kendall's W test

	Sum	Percent 0's	Percent 1's
Predicative categories	196	89.11	10.89
Agreement categories	684	61.98	38.02
Modality	37	97.94	2.06
Communication functions	4	99.78	0.22
Nominal morpho-syntax	502	72.10	27.90
Pronominal morpho-syntax	211	88.27	11.73
Numeral morpho-syntax	74	95.89	4.11
Verbal morpho-syntax	278	84.55	15.45
Word-order	604	66.43	33.57

	Sum	Percent 0's	Percent 1's
Other morpho-syntactic phenomena	391	78.27	21.73
Syntactic-semantic correlativeness others	296	83.55	16.45
Compound/complex sentences	568	68.43	31.57
Lexical semantics	1469	18.34	81.66
Lexical semantics (phraseologisms and proper names)	314	82.55	17.45

Cochran Q Test: N = 1799, Q = 6126,202, df = 13,
p < 0,0001; Kendall Coeff. of Concordance = 0.26195

The results show (Table 2) that the highest number of errors in MT outputs occurred in the category of lexical semantics (1469), then agreement categories (684), word-order (604), compound/complex sentences (568), and nominal morpho-syntax (502).

Based on the results of the Cochran Q test (Table 2), we reject the null hypothesis with 99.9% reliability, i.e. there is a statistically significant difference in the incidence of errors among the examined categories. The value of the Kendall coefficient of concordance (Table 2) is 0.26195; 0 meaning a trivial degree of concordance, and 1 a total concordance in the occurrence of errors in the individual examined categories, which confirms the results of the Cochran Q test.

5. Discussion

Following the scope of the study, we will only discuss the error rate and issue of agreement categories, word-order and nominal morpho-syntax.

Errors in the category of *agreement* correlate with the fact that the third person and past tense are used in publicistic texts very frequently. Particularly in Slovak, errors in the category *agreement in person* are highly connected with the problem of verbs in the third-person singular. Nouns are linked with verbs, and the gender of the noun determines the form of the verb (its inflectional morpheme), e.g. ‘*on bol*’ (he was), ‘*ona bol-a*’ (she was), ‘*ono bol-o*’ (it was). The situation in Slovak is more complicated since it uses two natural and three grammatical genders (Vaňko, 2015, p. 32). The nouns e.g. ‘*muž*’ (a man) and ‘*hrniec*’ (a pot) are of masculine gender; ‘*žena*’ (a woman) and ‘*trieda*’ (a classroom) of feminine gender; ‘*dievča*’ (a girl) and ‘*srdce*’ (a heart) of neuter gender. In some cases, a noun’s inflectional morpheme might indicate a noun’s gender, but in many cases the situation is more complicated and unclear. For example, the morpheme ‘*-a*’ is used with the masculine gender ‘*ten hrdin-a*’ (hero), the feminine gender ‘*tá žen-a*’ (woman) and even with the neuter gender ‘*to dievč-a*’, (girl). As explained, the errors in the category of agreement are closely related to the typological differences between English and Slovak.

The errors in the category *word order* and *nominal morpho-syntax* are also highly connected with the features of the publicistic style and the differences between the languages. The publicistic style is characterized by nouns, numbers, abbreviations, names and marks, and a low number of verbs. In terms of nominal features, nouns in all positions (nouns, nouns as pre-modifiers, post-modifiers, nouns in noun phrases) are more common. Nominalization, prepositional phrases after nouns and attributive adjectives are also present in the publicistic style.

As a synthetic language, Slovak uses numerous inflectional morphemes, affixes, and inflectional word classes (nouns, adjective, pronouns, numerals, verbs). It is characterized by significant gender differentiation of morphemes and by elements expressing a complex of grammatical meanings (gender, number, case; e.g. the morpheme ‘*-ov*’ is used with masculine gender, plural number and genitive case).

The errors of word order are connected with the different use of nouns in the position of pre-modifiers, post-modifiers, and noun phrases in English and Slovak. The translation situation (both by human and machine) worsens with multi-word sentence elements and complicated sentence structures (e.g. a group of south Pacific islands, Europe’s centuries-long history, local agricultural business leaders, a positive contribution to developing an effective post-Brexit immigration policy).

5.1 MT errors and their effect on readers' comprehension

To understand the errors of machine translation and their effect on readers' comprehension, we will highlight and explain some frequent and significant MT errors in publicistic texts from English into Slovak. Each example consists of an original sentence in English (marked ST – source text), its machine translation output in Slovak (MT – machine translation output created by Google Translate) and post-editing of MT output (PEMT – minimal intervention of post-editor to make MT output comprehensible).

5.1.1 Predicativeness

Considering the category of *predicativeness*, the quality of machine translation is highly dependent on two basic prerequisites: identification of subject and verb in the original text and their translation into the target text (MT). Correct identification and further translation of subject and verb – essential elements of sentence – has a significant impact on grammar correctness and comprehension of the text.

Predicative categories

Since English is characterized by a fixed word-order (S-V-O) and the subject is always explicitly expressed – using a proper name, noun or personal pronoun – issues related to identification and translation from English into Slovak were quite rare. We assume that more errors would occur in the direction Slovak – English, as Slovak is defined as a language with less fixed word order and the subject can be lexically unexpressed. According to Vaňko (2015, s. 23-26), difficulties can be noticed in the identification of suffixes which are the only indicators of the grammatical and semantic function of subject and object. The problems with unexpressed subject can be outlined e.g. in the sentence '*Beží.*'. The subject is not explicitly expressed; and actually there are three possible meanings: '*On beží.*' (He is running.); '*Ona beží.*' (She is running.); or '*Ono beží.*' (It is running.) In the past tense, all persons take the same inflectional morpheme '-í'. The identification of person in such sentences depends on the context or references.

More issues and errors occurred in the translation of verbs. This correlates with the fact that English has a relatively high number of tenses and verb forms, and many of them do not have counterparts in Slovak, e.g. present perfect, future perfect, past perfect, continuous aspect (in all tenses), multi-word verbs (e.g. will have been writing) and by the fact that in English, some auxiliaries can occur as main verbs (e.g. have).

The errors in predicative categories are related to the abovementioned issues – the number of tenses in English and the lack of equivalent tenses and verb forms in Slovak. The differences between the languages affect the translation of verb forms from the source language (original) into the target language (translation). For example, present perfect tense in English can be translated into past tense or present tense in Slovak, following the meaning of the sentence. Inadequate translations cause logical discrepancies and shifts in meaning.

The issue can be demonstrated in the translation of the verb form 'has dominated' (1). GT's suggestion '*dominovala*' (it dominated) refers to an event which happened in the past, the form '*dominuje*' (dominates / is dominating) to an action which happens (or is happening) in the present. In this context, the verb form '*dominuje*' is a more appropriate equivalent as the present perfect tense suggests something that happened in the past and affects the situation that exists now (also a connection between what happened in the past and happens at the present time, Hewings, 1999, p. 8). For a better understanding of MT output in Slovak, we recommend adding an adverb '*doteraz*' (till now). Otherwise, the form '*dominovala*' indicates that 'domination is over'.

- (1) ST: But two of the most significant results were in the north, where Labour **has dominated** local government for many decades.
 MT: *Ale dva z najvýznamnejších výsledkov boli na severe, kde práca **dominovala** miestnym vládam po mnoho desaťročí.*
 PEMT: *Ale dva z najvýznamnejších výsledkov boli na severe, kde Labouristická strana **dominuje** miestnym vládam už mnoho desaťročí.*

As mentioned above, Slovak does not use perfect tenses and their translation into the language is an arduous task even for a human translator. Obviously, GT cannot identify their whole forms properly,

thus translate them reliably. As a result, consequences and errors in MT outputs are evident and corrections are inevitable.

Obviously, the situation with the translation of verb forms seems to be more complicated with multi-word verbs (e.g. had been patrolling, has been detained, or was behaving). They are usually translated by GT literally or some parts of the verb forms are omitted. Generally, their translations are inaccurate, vague, or even incomprehensible.

The category of *mood* reflects whether the action marked by the verb is real, unreal, whether it expresses condition, or requirement (see Vaňko, 2015, p. 90). Due to the style of the examined texts – for which indicative mood, and marginally conditional mood are typical – we did not anticipate a high number of errors. Our assumptions were confirmed.

‘Would’ is basically used when expressing willingness, likelihood or certainty, or in reported speech (will – would). Most MT errors were related to the latter. For example (2), the structure *she would brief* as a part of the reported sentence “(it) said she would brief” was translated as ‘*kancelária by informovala*’ (it would inform) which indicates conditional in Slovak, not reporting. The correct meaning of the structure in this context is different as it refers to an action in the future ‘*povedala, že bude informovať*’.

- (2) ST: Her office **said she would brief** EU leaders over her US visit.
 MT: *Jej kancelária povedala, že by informovala* vedúcich predstaviteľov EÚ cez jej návštevu USA.
 PEMT: *Jej kancelária povedala, že bude informovať* vedúcich predstaviteľov EÚ počas jej návštevy v USA.

The errors related to the incorrect use of ‘would’ in reporting belong to the category of *compound/complex sentences* which is not in the scope of the paper.

Agreement categories

Most of the mistakes in the category of *agreement* overlapped and they occurred in different combinations, e.g. agreement in person and number, or agreement in person, number and gender. During the process of evaluation, the mistakes were precisely recorded in the framework mapping each category individually. The error in the category *agreement in person* can be demonstrated by the structure ‘*výnos klesáš*’ instead of ‘*výnos klesá*’ (3). The inflectional morpheme ‘-š’, ‘*kles-á-š*’ is typical for the second-person singular of the present tense; the inflectional morpheme ‘-á’, ‘*kles-á*’ is used with the third-person singular (with all genders) of the present tense. The example demonstrates that GT did not match the noun (*revenue – it*) with the person of the verb (‘*klesáš*’ – second-person singular) correctly.

- (3) ST: Revenue **has been falling** too, with \$4.8bn pulled in last quarter, compared to an all time high of \$8.8bn in the second quarter of 2013 (again, annualised averages).
 MT: *Výnos klesáš príliš, s 4,8 mld \$ vytiahol v poslednom štvrtroku v porovnaní s všetkých čias vysoká vo výške \$ 8,8 miliárd v druhom štvrtroku 2013 (opäť anualizované priemery).*
 PEMT: *Výnos takisto klesá, s 4,8 miliardami amerických dolárov v poslednom štvrtroku v porovnaní s najvyššou sumou 8,8 miliárd amerických dolárov v druhom štvrtroku 2013 (opäť ročné priemery).*

The error in *agreement in number* can be observed in MT’s suggestion ‘*my sme nezávislá*’ as an equivalent for the English structure ‘we are independent’ (4). A flecional morpheme – á in the form ‘*nezávisl-á*’ indicates the feminine gender singular ‘*ja som nezávisl-á*’ (I am independent), ‘*ty si nezávisl-á*’ (you are independent), ‘*ona je nezávisl-á*’ (she is independent); always referring to a female.

The form ‘*nezávisl-í*’ is connected to the masculine gender inanimate in plural ‘*my/vy/oni sú nezávisl-í*’ (we/you/they are independent) or for mixed groups of males and females (‘*muži sú nezávisl-í*’ (men are independent), ‘*ženy sú nezávisl-é*’ (women are independent), ‘*muži a ženy sú nezávisl-í*’ (men and women are independent). The correct form is ‘*my môžeme byť úplne nezávislí*’ (not ‘*my môžeme byť úplne nezávislá*’).

- (4) ST: Only together can we be fully **independent**.
 MT: *Iba spoločne môžeme byť úplne **nezávislá**.*
 PEMT: *Iba spoločne môžeme byť úplne **nezávislí**.*

In Slovak, errors in the *category of agreement in person* are highly connected with the problem of verbs in the third-person singular in the past tense. For GT, a distribution of endings was quite challenging due to the identification of gender (5). In Slovak, the inflectional morpheme of the verb depends on the person's gender: 'on hlasoval-' (he voted), 'ona hlasoval-a' (she voted), 'ono hlasoval-o' (it voted). The proper name 'Británia' (Britain) is considered feminine in Slovak, and the verb form 'hlasoval' refers to the masculine gender. The structure 'Britain' voted was translated as 'Británia hlasoval-' instead of 'Británia hlasoval-a'.

- (5) ST: "**Britain voted** for change, especially on free movement, but there has been very little debate about what kind of reforms or immigration control that should now mean or how we get the best deal for the country," Cooper said.
 MT: *"**Británia hlasoval** pre zmenu, a to najmä v oblasti voľného pohybu, ale tam bola veľmi malá debata o tom, akú reformy alebo imigračné kontroly, ktorá by teraz mala znamenať alebo ako sa dostať to najlepšie riešenie pre túto krajinu," povedal Cooper.*
 PEMT: *„**Británia hlasovala** za zmenu, a to najmä v oblasti voľného pohybu, ale bola len veľmi malá debata o tom, aké reformy alebo imigračnú kontrolu teraz myslia, alebo ako dostať to najlepšie riešenie pre túto krajinu,“ povedal Cooper.*

The group covers the errors related to the category of *predicativeness* which had not been anticipated and thus implemented into the framework at the time it was being designed. They are the following: *incorrect word class transfer, incorrect verb form, and missing verb*.

Within the category of *incorrect word class transfer*, we noticed several errors in MT outputs. They occurred mainly in the incorrect use of nouns and verbs (a noun used as a verb and vice versa), rarely between other word classes. For example (6), the structure 'the committee's chair', in Slovak consisting of two nouns 'predseda výboru', was translated as 'predsedá výboru' (verb + noun; she is chairing). The structure 'predsedá', representing the present tense, is illogically followed by the verb form 'said', referring to the past. In this context, MT's suggestion makes the meaning of the sentence strange and confusing. Moreover, the noun 'predseda' in this context should be substituted by the noun 'predsedníčka' as the noun 'predseda' refers to male, and the noun 'predsedníčka' to the female. In PEMT, we keep the word 'predseda' to see the issue of incorrect word class transfer clearly.

- (6) ST: Yvette Cooper, **the committee's chair**, said she wanted to encourage people to talk frankly about immigration.
 MT: *Yvette Cooper, **predsedá výboru**, povedal, že chce, aby sa ľudia otvorene hovorili o imigrácii.*
 PEMT: *Yvette Cooperová, **predseda výboru** povedala, že chce, aby ľudia otvorene hovorili o imigrácii.*

Concerning the *category of incorrect word class transfer*, an interesting error occurred in the translation of the surname of the former British PM Theresa May (7). In contexts when her surname occurred without her first name, GT could not ascertain reference to the person and translated the surname *May* as a modal verb 'môcť' (*may*), or as the name of the month 'May' ('máj'), using lower-case letter 'm' (correct for Slovak). This issue can be further analysed in terms of lexical semantics.

- (7) ST: **May** will miss the afternoon session of the Valletta summit, at which the remaining 27 EU leaders will resume discussion of how Brexit can be handled, and preparations for next month's 60th anniversary of the Treaty of Rome, which set up the EU's precursor, the European Economic Community.

- MT: *Môže chýbať popoludňajšie zasadnutie summitu Valletta, pri ktorom sa zostávajúce 27 vedúci predstavitelia EÚ pokračovať diskusia o tom, ako možno Brexit manipulovať a prípravy na 60. výročie budúci mesiac Rímskej zmluvy, ktorá zostavou predchodcu EÚ, Európskeho hospodárskeho spoločenstva.*
- PEMT: *Mayová bude chýbať na popoludňajšom zasadnutí summitu vo Vallette, pri ktorom zostávajúci 27 vedúci predstavitelia EÚ budú pokračovať v diskusii o tom, ako možno zvládnuť Brexit a budúco-mesačné prípravy na 60. výročie Rímskej zmluvy, ktorá je predchodcom EÚ, Európskeho hospodárskeho spoločenstva.*

Errors associated with *incorrect verb forms* were quite frequent, compared to other errors of predicativeness defined in Vaňko's framework. Basically, they were cases of the inadequate transformation of a verb form – given by the category of person, number, gender – from the source language into the target language. Instead of the correct verb form, GT suggested an infinitive of the given verb (8). Although such mistakes were numerous, they did not have a significant impact on the comprehension of the machine translation output as the form and meaning of the sentence is relatively predictable from its infinitive.

- (8) ST: The extraordinary radiation readings **highlight** the scale of the task confronting thousands of workers, as pressure builds on Tepco to begin decommissioning the plant – a process that is expected to take about four decades.
- MT: *Mimoriadne meranie radiácie **upozornit'** na rozsah tejto úlohy konfrontácii tisíce robotníkov, ako vzrastá tlak na TEPCO začať vyradenie elektrárne - proces, ktorý sa očakáva trvať asi štyri desiatky rokov.*
- PEMT: *Mimoriadne merania radiácie **upozorňujú** na rozsah tejto úlohy, konfrontujúcej tisíce robotníkov, ako vzrastá tlak na spoločnosť TEPCO, aby začala vyradovať elektrárne - proces, ktorý by mal trvať asi štyri desiatky rokov.*

The issue of *missing verbs* in MT output (comparing the verb in the source text) was also considered an error of the category *other mistakes*. Generally, *missing verbs* are not typical errors of the GT which are based on a statistical approach (they appear to be a significant problem of neural machine translation), and they occurred rarely. They occurred when the verb form from the source text was not translated into the target text (the verb was missing), or when the verb form from the target text was not translated into the target text as a whole (some parts were missing).

As can be observed in the following example (9); the form '*there's*' was translated as '*tu*' (there) without the verb '*is*' which is a part of the existential '*there*'. The missing verb in MT output represents a serious mistake in the comprehension of the text.

- (9) ST: But **there's** bad news for Apple too.
- MT: *Ale **tu** pre Apple zlé správy príliš.*
- PEMT: *Ale **sú tu** pre Apple príliš zlé správy.*

5.1.2 Modal and communication sentence framework

Modality expresses the existence of modality from the point of view of reality and it reflects an author's attitude to his utterance. Basically, each statement has its communication task; it can be a statement, negation, wish, or requirement (Vaňko, 2017, p. 87). In the examined publicistic texts, the category of modality does not play a key role as the texts are primarily unmarked (this can be proved by the number of errors which is 41).

Within Vaňko's category *modal and communication sentence framework*, the greatest number of errors were detected in the category of *negation*. In the examined languages, negation is used differently. English uses simple negation – both not-negation and no-negation; and it is more than twice as common in conversation as in written registers (Biber, Conrad, and Leech, 2002, p. 239). On the contrary, Slovak uses simple, double or even multiple negation – '*žiadne výbušniny neboli nájdené*' (no explosives had been found) represents a double negation; '*nikto nikam nepôjde*' (nobody would go anywhere) represents a triple negation. In Slovak, sentence negation is expressed by the negative form of the verb (with a prefix '*ne-*' and highlighted with the expressions e. g. '*nikto*' (nobody), '*nik*' (nobody), '*nič*'

(nothing), ‘*nijaký*’ (no), ‘*nikde*’ (nowhere), or with the words ‘*žiadne*’ (no), ‘*ani jeden*’ (not a single one) (Pauliny, Ružička, and Štolc, 1968, p. 357).

MT’s suggestions *žiadne výbušniny boli nájdené* and *tam bolo žiadna hrozba* are for a Slovak reader quite incomprehensible and illogical (10).

- (10) ST: The man was wounded, including in the stomach, but he was still alive. Cadot said **no** explosives **had been found** in the man’s bag and there **was** “**no** threat”.
- MT: *Ten človek bol zranený, vrátane žalúdka, ale bol stále nažive. Cadot uviedol žiadne výbušniny boli nájdené v mužovej sáčku a tam bolo "žiadna hrozba".*
- PEMT: *Ten človek bol zranený, vrátane žalúdka, ale bol stále nažive. Cadot uviedol, že žiadne výbušniny neboli predtým nájdené v mužovej taške a nebola tam „žiadna hrozba“.*

5.1.3 Syntactic-semantic correlativeness

In this category, we attempted to evaluate the syntactic-semantic relationships and correlativeness of lexical words in the sentences. Due to the typological characteristics of Slovak (inflectional with a high number of morphemes; different ways of expressing relationships between words and multiple-word structures) we assumed that the errors would be numerous and significant. The errors can be divided into several groups:

Nominal morpho-syntax

The category of *nominal morpho-syntax* involves errors of nouns (in the position of superordinate sentence elements) with adjectives, syntagms of nouns which are in different cases (agreement within a noun phrase, ‘of’ genitive), or structures of two or more nouns when at least one word has the form of an adjective.

In the following example (11) we can notice the problem of GT with the translation of the noun phrase ‘South Pacific islands’. The translation ‘*Južné tichomorských ostrovov*’ is grammatically incorrect as the word ‘islands’ was not identified as the subject of the sentence (according to the inflectional morpheme ‘-ov’ it was probably considered the object). The inflectional morphemes of the attributes do not agree in this case; ‘*južné*’ represents the nominative, ‘*tichomorských*’ the genitive and ‘*ostrovov*’ the genitive case.

Numerous errors in the *category of nominal morpho-syntax* are related to the correct identification of sentence elements as discussed above (incorrect identification of subject and object and consequent incorrect use of inflectional morphemes in Slovak); problems with identification of gender in Slovak (‘another soldier’ translated by GT as ‘*d’alšie vojak*’ instead of ‘*d’alší vojak*’ (the adjective) ‘*d’alši-e*’ carries the flectional morpheme of neuter gender and the noun ‘*vojak*’ is of masculine gender); different ways of structuring words and phrases in English and Slovak (‘*Paris police prefect*’ translated as ‘*parížska polícia prefekt*’ instead of ‘*prefekt parížskej polície*’). Another problem can be seen in the identification of word classes in English (‘*Paris anti-terrorist police*’ translated as ‘*Paríž anti-terroristická polícia*’ instead of ‘*parížska antiteroristická polícia*’; ‘*Paríž*’ is a proper name, ‘*parížska*’ is an adjective). The problem worsens in multiple-noun phrases in which there is a partial or no agreement between the words (e.g. ‘a huge former royal palace’ translated as ‘*obrovské bývalý kráľovský palác*’ instead of ‘*obrovský bývalý kráľovský palác*’).

- (11) ST: **South Pacific islands** ban western junk food and go organic
- MT: *Južné tichomorských ostrovov zákaz západnej nezdravé jedlo a ist’ organické*
- PEMT: *Južné tichomorské ostrovy zakazujú západnej nezdravé jedlo a smerujú k organickému*

Pronominal morpho-syntax

Problems in pronominal morpho-syntax can be noticed in the use of the determiner ‘its’, which is typical for the third-person singular. There are numerous problems with its translation into Slovak due to the problems with genders as discussed above. Slovak expresses relation to masculine gender – to both inanimate and animate nouns – by the pronoun ‘*jeho*’: ‘*muž – jeho kabát*’ (man – his coat), ‘*hrniec –*

jeho farba (pot – its colour), to feminine by the pronoun ‘jej’: ‘žena – jej vlasy’ (woman – her hair), ‘trieda – jej priestory’ (classroom – its space), to neuter by the pronoun ‘jeho’: ‘dievča – jeho postava’ (girl – her figure), ‘srdce – jeho funkcia’ (heart – its function).

In the source text (12), the determiner ‘its’ refers to the structure ‘Torba province (it)’. In Slovak, the noun ‘*provincia*’ (province) is of feminine gender, so the possessive pronoun ‘*jej*’ needs to be used with it. Actually, GT suggested the pronoun ‘*jeho*’ (reference to masculine) which is incorrect.

Unlike in English, the use of the possessive pronoun in Slovak is less frequent as it is sometimes considered redundant and unnatural: ‘*Torba... si kladie za cieľ využiť jej produktívnu poľnohospodársku pôdu*’. Instead, the use of the possessive pronoun ‘*svoj*’ is more appropriate. According to Oravec, Bajzík, and Fudík (1984, p. 120), the use of ‘*svoj*’ (even ‘*svoja*’, ‘*svoje*’) depends on the sentence structure and as it can only be used with the subject of the sentence ‘*Zjedol svoj obed.*’ (He ate his lunch. and not ‘*Zjedol jeho obed.*’).

- (12) ST: Torba province, part of Vanuatu, aims to impose restrictions on the import of western foodstuffs and instead take advantage of **its** productive agricultural land and rich natural resources.
- MT: *Provincia Torba, časť Vanuatu, si kladie za cieľ zaviesť obmedzenia na dovoz západných potravín a namiesto toho využiť jeho produktívnej poľnohospodárskej pôdy a bohatými prírodnými zdrojmi.*
- PEMT: *Provincia Torba, časť Vanuatu, si kladie za cieľ zaviesť obmedzenia na dovoz západných potravín a namiesto toho využiť svojú produktívnu poľnohospodársku pôdu a bohaté prírodné zdroje.*

Numeral morpho-syntax

Some negligible errors occurred even in the category of numeral morpho-syntax. GT successfully avoided a literal translation of the structure ‘six years ago’ (‘šest’ rokov dozadu’) which would have sounded quite unnatural in Slovak (12). Actually, it was not able to find the correct inflectional morpheme for the numeral ‘šest’ (six) in the structure ‘*pred šest’ rokov*’ (correct: ‘*pred šiestimi rokmi*’).

- (12) ST: Radiation levels inside a damaged reactor at the Fukushima Daiichi nuclear power station are at their highest since the plant suffered a triple meltdown almost **six years ago**.
- MT: *Úroveň radiácie vnútri poškodeného reaktora v jadrovej elektrárni Fukushima Daiichi sú na najvyššej úrovni, pretože závod utrpela trojitou zrútením pred takmer šest’ rokov.*
- PEMT: *Úroveň radiácie vnútri poškodeného reaktora v jadrovej elektrárni Fukushima Daiichi je na najvyššej úrovni, pretože elektrárňu zažila trojitú roztavenie pred takmer šiestimi rokmi.*

Verbal morpho-syntax

The category of verbal morpho-syntax maps appropriate use of verb structures in Slovak. A missing preposition, redundant preposition or incorrect preposition in Slovak verb structure was assessed as an error. Even though the issue seems trivial, in many cases it caused incomprehensibility of the MT output. The structure ‘British PM will also urge leaders’ was translated by GT as ‘*Britský premiér bude tiež naliehať lídrov*’. The verb ‘*naliehať*’ (to urge) is used with the preposition ‘*na*’ (‘*naliehať na*’) so the correct form would be ‘*Britský premiér bude tiež naliehať na lídrov*’. The first part of the sentence is quite vague, even to-phrase does not make any sense; the MT output is incomprehensible (13).

- (13) ST: British PM will also **urge leaders** at Malta summit to spend more on defence after Donald Trump’s Nato scepticism.
- MT: *Britský premiér bude tiež naliehať lídrov na summite Malta minúť viac na obranu po tom, čo Donald Trump v NATO skeptou.*
- PEMT: *Britský premiér bude tiež naliehať na lídrov na summite v Malte, aby minuli viac na obranu po skepticizme Donalda Trumpa nad NATO.*

Word-order

The analysis has shown that errors in the category of word-order are evidently related to the category of morpho-syntax as the highest number of errors occurred in the translation of noun phrases. The structure of the phrases is different in both languages. For example (14), the phrase ‘head of the local tourism council’ was translated as ‘*hlava miestnej cestovného ruchu rady*’ which represent a literal translation, incomprehensible in Slovak. To make it clear, we need to use a different word-order: ‘*hlava miestnej rady cestovného ruchu*’.

- (14) ST: Father Luc Dini, a community leader and **head of the local tourism council**, said a ban on foreign food imports would improve the health and wellbeing of islanders.
 MT: *Otec Luc Dini, vodcu spoločenstvo a hlava miestnej cestovného ruchu rady, povedal zákaz dovozu zahraničných potravín by zlepšilo zdravie a pohodu ostrovanov.*
 PEMT: *Otec Luc Dini, vodca spoločenstva a hlava miestnej rady cestovného ruchu vydal zákaz dovozu zahraničných potravín, aby sa zlepšilo zdravie a pohoda ostrovanov.*

In MT outputs, GT basically copies the word order of the source sentence. Although Slovak is specified as a language with less fixed word order, in some contexts the word order needs to be kept. These are situations of underlying important facts or words, or the use of adverbs. There are also some GT’s suggestions in which translations sound unnatural (15).

- (15) ST: He has had spasms before.
 MT: *Mal kŕče predtým.*
 PEMT: *Predtým mal kŕče.*

Others

The category *Others* involves errors which are related to the issue of *syntactic-semantic correlativeness* and are not specified in the previous items. These are e.g. errors of punctuation in simple sentences. The most frequent errors were those with writing commas and quotation marks which are used differently in both languages. For example (16), Slovak does not separate personal names in sentences with commas as outlined in the example below. Also, it uses a different method of quotation (17): in direct speech (or references to direct speech) we enclose what is said within a pair of single or double quotation marks – in Slovak the marks are like „...“ or ‚...‘; English uses ‘...’ or “...”).

- (16) ST: Ukip’s leader, Paul Nuttall, said:
 MT: *Vedúci Ukip, Paul Nuttall, uviedol:*
 PEMT: *Líder strany Ukip Paul Nuttall uviedol:*
- (17) ST: “Retention of Brits is very low.”
 MT: *“Zachovanie Britov je veľmi nízka.”*
 PEMT: *„Udržanie si Britov nie je ľahké.“*

6. Conclusion

Machine translation represents a new method of translation which requires a new approach. MT’s users need to understand its advantages and disadvantages, and adequate and inadequate suggestions, and then, to find a compromise between quality and quantity. Machine translators (e.g. Google Translate) are able to translate large volumes of text in a short time, but their quality (although still developing) is not comparable with human translation. The difference between human and machine translation is in the fact that human translation can maintain coherence and cohesion, as a human translator understands the source text and can convey the meaning into the target text. On the contrary, machine translation cannot “understand” the text: the source text is divided into smaller segments (one segment representing one sentence, or one phrase) and they are matched with the most likely translation equivalents (based on language and translation models).

We can state that in our research the most frequent issues occurred in agreement categories (684), in the word-order category (604), and in the nominal morpho-syntax category (502). The results highly

correlate with the characteristics of the publicistic style: the greatest number of errors detected are in the translation of nouns (and all forms containing nouns) which is the most frequent word class of the publicistic style.

The category of predicativeness saw errors in identification of verbs (or whole verb forms) when translating them into Slovak. Moreover, English has a wider range of tenses and forms than Slovak, and sometimes MT translations did not convey the same meaning as expressed in English. It caused various vague and incomprehensible suggestions which needed correction (e.g. 'he had shouted' translated literally as '*on mal kričal*'). Incorrect identification and translation of the verb form in many cases caused errors in subject – verb agreement.

The journalistic style is characterized by the frequent use of past tense and the third-person singular, and many errors in the category of *predicativeness* are related to this. Although the inflectional morphemes are identical for all persons in the past tense in English (e.g. he worked, she worked, or it worked), the situation in Slovak differs: it uses different inflectional morphemes according to the person, e.g. '*on hovoril*' (he said), '*ona hovorila*', (she said), or '*ono hovorilo*' (it said).

As the number of genders in Slovak is higher than in English and their use is more complicated, the errors were numerous.

Errors in the category of agreement are also connected with the fact that nouns in all realizations (nouns, nouns as pre-modifiers, post-modifiers, nouns in noun phrases) are more common for the publicistic style. Differences between English and Slovak cause many discrepancies and errors in the category of agreement and consequently in the category of word order. For example, the phrase 'Paris knife attack' cannot be translated as '*Paris útok nožom*' (literal translation of the structure), neither '*parížsky nožový útok*' (literal translation using adjectives as pre-modifiers) but as '*parížsky útok nožom*' (pre-modifier, noun, post-modifier).

The category *Modal and communication sentence framework* does not evince serious errors. Most of them are related to the category of negation. In both languages, negation is used in a different way: English uses simple negation, Slovak simple or multiple negation according to the intention or the speaker. Differences in MT outputs can lead to misunderstanding of the text or its vagueness.

The category of *syntactic-semantic correlativeness* revealed the most errors in word order and nominal morpho-syntax. As nouns are the typical word class of the journalistic style and there are different specific features in both languages, the errors are logical. The basic issue of Slovak is in its gender – there are two natural and three grammatical genders. Incorrect identification of a noun's gender can affect the predicative categories (in the case of a subject) or the agreement categories (in the case of an object).

Finally, we can state that the differences between languages and the characteristics of style have a significant effect on the quality of machine translation. The given study represents a pilot study based on previous research (Welnitzová, Munková, and Wrede, 2020) and it opens new perspectives for further research. The study can be used for a deeper analysis: evaluation of GT errors in the categories of compound/complex sentences and lexical semantics, or for different types of texts, e.g. administrative.

Acknowledgment

This work was supported by the Slovak Research and Development Agency under the contract no. APVV-18-0473 and Scientific Grant Agency of the Ministry of Education of the Slovak Republic (ME SR) and of Slovak Academy of Sciences (SAS) under the contract no. VEGA - 1/0792/21.

References

- Biber, D. and Conrad, S., 2009. *Register, genre, and style*. Cambridge: Cambridge University Press.
- Biber, D., Conrad, S. and Leech, G. 2002. *Student grammar of spoken and written English*. Harlow: Longman.
- Conway, K., 2015. What is the role of culture in News Translation? A materialist approach. *Perspectives. Studies in Translatology*, vol. 23, no. 4, pp. 521-535.
- Dolník, J., 2013. *Všeobecná jazykoveda. Opis a vysvetľovanie jazyka*. Bratislava: VEDA.
- Findra, J., 2013. *Štylistika súčasnej slovenčiny*. Martin: Osveta.
- Hewings, M., 1999. *Advanced grammar in use*. Cambridge: Cambridge University Press.
- House, J., 2015. *Translation quality assessment: Past and present*, London and NY: Routledge.

- Mistrík, J., 1989. *Štylistika*. Bratislava: Slovenské pedagogické nakladateľstvo.
- Müglová, D., 2009. *Komunikácia, tlmočenie, preklad alebo Prečo spadla Babylonská veža*. Nitra: ENIGMA.
- Munková, D. et al. 2017. „*Mýliť sa je ľudské (aj strojové)*“. *Analýza chýb strojového prekladu do slovenčiny*. Nitra: Univerzita Konštantína Filozofa.
- Ondruš, Š. and Sabol, J., 1984. *Úvod do štúdia jazykov*. Bratislava: Slovenské pedagogické nakladateľstvo.
- Oravec, J., Bajzíkova, E. and Furdík, J., 1984. *Súčasný slovenský spisovný jazyk*. Bratislava: Slovenské pedagogické nakladateľstvo.
- Pauliny, E., Ružička, J. and Štolc, J., 1968. *Slovenská gramatika*. 5th ed. Bratislava: Slovenské pedagogické nakladateľstvo.
- Schwarzl, A., 2001. *The (Im)Possibilities of machine translation*. Oxford: Peter Lang Publishing.
- Valdeón, R., A., 2019. Translation and culture in mainstream media and journalism. In S.-A. Harding and O. Carbonell Cortés, eds. *The Routledge handbook of translation and culture*. London: Routledge, pp. 558-573.
- Vaňko, J. and Auxová, D., 2015. *Morfológia slovenského jazyka*. Nitra: Univerzita Konštantína Filozofa.
- Welnitzová, K., Munková, D. and Wrede, O., 2020. Chybovosť v predikatívnosti a kvalita strojového prekladu. *Jazyk a kultúra*, vol. 11, no. 41-42, pp. 160-172.