

Univerzita Karlova

Filozofická fakulta

Ústav anglického jazyka a didaktiky

Bakalářská práce

Anna Kenderová

**The subject in Czech and in English as seen through Google
Translate**

Český a anglický podmět pohledem překladače Google

Praha, 2019

Vedoucí práce: doc. PhDr. Markéta Malá, Ph.D.

Ráda bych moc poděkovala vedoucí své práce, doc. PhDr. Markétě Malé, Ph.D., za trpělivost, cenné rady a ochotu se mnou podstoupit takové dobrodružství.

Ráda bych také poděkovala Bc. Petru Kasalickému za konzultace a dohlédnutí na faktickou přesnost pasáží týkajících se strojového překladu a neuronových sítí.

Prohlášení:

Prohlašuji, že jsem bakalářskou/diplomovou/rigorózní/dizertační práci vypracoval/a samostatně, že jsem řádně citoval/a všechny použité prameny a literaturu a že práce nebyla využita v rámci jiného vysokoškolského studia či k získání jiného nebo stejného titulu.

I declare that the following BA thesis is my own work for which I used only the secondary literature that is listed in the resources. This thesis was not used as a part of any other university study, nor was it used to gain a different university degree.

V Praze, dne 24.5.2019

.....

Anna Kenderová

Abstract

This BA thesis aims to describe the differences between the Czech and English subject through the prism of Google Translate and the divergent strategies used by it to translate the subject from Czech to English. GT output is then compared to strategies used by a human translator.

The theoretical part describes the history and methods of machine translation, strategies generally employed by human translators and finally, compares the form and function of the Czech and English subject. Special emphasis is put on the translation of Czech subjectless sentences and on the way the information structure influences the word-order in the languages.

In the analytical part, 100 cases of non-corresponding translation of the subject by GT are excerpted and judged on the accuracy of the solution. They are then compared to the strategies used by the human translator.

It was found that the most frequent non-corresponding translation strategy was in fact condensation (32%), which was not included in the initial hypothesis. In addition to condensation, construing a different clause element as the English object was the second most common (26% of excerpts, the majority of which was formed by construing Czech object as English subject.) In regard to human translation strategies, GT used the same strategy as a human translator in 37% of the cases. In 30% of the cases, human translator preserved the syntactic structure of the source text instead. The human translation used a completely different strategy in 32% of the cases. The main difference found between human translations and GT's translations was higher level of literal translation in GT and more creative solutions in human translation.

Key words: subject, subject constancy, machine translation, Google translate, syntax, translation strategies

Abstrakt

Tato bakalářská práce si klade za cíl popsat rozdíly mezi českým a anglickým podmětem pohledem Překladače Google, a popsat také nekorespondentní strategie, které překladač Google užívá.

Data k analýze byla čerpána z dramatu R. U. R. Karla Čapka a tří románů - Europeany Patrika Ouředníka, Miráklu Josefa Škvoreckého a Hrdého Budžese Ireny Douskové. Tyto romány byly

postupně překládány Překladačem Google, než bylo vyňato 100 příkladů nekorespondenčního řešení podmětu. Tyto příklady pak byly porovnány s jejich anglickým překladem.

České typy podmětu, o kterých se předpokládalo, že budou způsobovat nekorespondenční řešení, zahrnovaly postverbální podmět, nevyjádřený podmět a věty jednočlenné (bezpodmětné).

V analytické části se potvrdila hlavní role těchto typů podmětu. Strategie využívané Překladačem Google oproti očekávání zahrnovaly kondenzaci, jinak odpovídaly odhadům – přetvoření jiného českého větného členu na anglický subjekt (nejčastěji předmětu, méně často příslovečného určení a přísudku) a existenciální konstrukce. Překladač často chyboval vynecháním podmětu, využitím neodpovídajícího pronominálního podmětu v případě českého nevyjádřeného podmětu a slučováním klauzí souvětí dohromady.

Co se týče porovnání s lidskými překladateli, ve 34% případů využíval překladač Google stejné strategie jako lidský překladatel. Lidský překladatel ponechal větné členy ve stejné struktuře jako ve zdrojovém textu ve 30% případů. Rozdílná metoda byla použita v 36% případů, ale nebyly nalezeny žádné významné podskupiny, spíše individuální řešení.

Klíčová slova: podmět, stálost podmětu, strojový překlad, Překladač Google, syntax, strategie překladu

Table of Contents

List of tables and figures	7
1. Introduction	8
2. Theoretical background	9
2.1. Machine Translation (MT).....	9
2.1.1. Grammar theories and their relationship to NLP and MT	10
2.1.2. Statistical MT (SMT)	11
2.1.3. Neural MT (NMT)	12
2.1.3.1. What is NMT? What are neural networks (NNs)?	12
2.1.3.2. NMT and linguistics.....	13
2.1.3.3. Weaknesses of NMT	13
2.1.4. Evaluation of MT	14
2.1.4.1. Common Error-Types introduced by Bojar in 2011	14
2.2. Human Translation (HT).....	15
2.2.1. Criteria for adequate translation	15
2.2.2. Translation Strategies, levels of equivalence.....	15
2.3. The subject.....	17
2.3.1. The subject in different syntactic theories	17
2.3.2. Common characteristics of the subject	18
2.3.3. The English subject.....	19
2.3.3.1. Form of the English subject.....	19
2.3.3.2. Absence of subject in English.....	20
2.3.4. Czech subject	21
2.3.5. General differences between English and Czech.....	21
2.3.5.1. Czech subjectless sentences and their English equivalents.....	22
2.3.5.2. Verbo-nominal Czech subjectless sentences	24
2.3.5.3. Subject constancy in the languages	25
2.3.5.4. Subject syntactic and FSP constancy across languages	25
3. Material and Methodology	27
3.1. Material.....	27
3.2. Method	27
4. Analysis	30
4.1. Non-corresponding Czech subject types	30
4.2. GT's strategies.....	31
4.2.1. Condensation	32

4.2.2.	Czech Object and other clause elements construed as the English subject.....	34
4.2.3.	Existential construction.....	34
4.2.4.	Pronominal subject	35
4.2.5.	Merging.....	36
4.2.6.	Missing subject.....	37
4.3.	GT Strategies compared to the human translator.....	38
4.3.1.	Difference overall.....	38
4.3.2.	Cases solved with the same strategy as GT	40
4.3.3.	Cases with structure corresponding to ST	41
4.3.4.	Cases solved by a divergent strategy.....	42
4.3.5.	Strategies not found in GT	44
4.3.5.1.	Ellipsis.....	44
4.3.5.2.	Cleft construction.....	45
4.3.5.3.	Inversion.....	45
4.3.5.4.	Communicative translation.....	46
5.	Conclusion.....	46
6.	Resumé	48
7.	Sources and literature:.....	50
8.	Appendix	52

List of tables and figures

Table 1 Non-corresponding Czech subject types and their adequacy rate	30
Table 2: GT's strategies and their adequacy rate	32
Table 3: The overview of HT strategies	39
Tabulka 4: Overview of the cases corresponding with ST subject	41
Table 5: Overview of cases solved by a divergent strategy	43

List of abbreviations

Adv.....	Adverbial
ALPAC	Automatic Language Processing Advisory Committee
Cs, Co.....	Subject complement, object complement
GT	Google Translate
HT	Human translation
MT.....	Machine translation
NLP	Natural language processing
NMT.....	Neural Machine Translation
NN.....	Neural Networks
O.....	Object
S	Subject
SL.....	Source language
SMT	Statistical Machine Translation
ST.....	Source text
TL.....	Target language
V.....	Verb
WMT.....	Workshop on Statistical Machine Translation

1. Introduction

This BA thesis explores the differences between the subject in Czech and English through the perspective of machine translation (MT).

The main goal of this approach is to view the differences between the Czech and English subject through the lens of state-of-the-art artificial intelligence by seeing the strategies it employs to translate it and compare them to strategies used by human translators.

The non-correspondences arising in MT output are signs of divergence between the two languages and are further analysed to uncover the underlying tendencies both in the languages studied and in the functioning of MT.

The software used for the analysis (Google Translate, GT henceforth) was chosen for its easy accessibility, world-wide reach and the fact that it is on the edge of technological advancement. Although there is other software that is comparable and, according to 2017 results of annual Conference of Machine Translation (WMT2017), even superior in terms of quality (developed e.g. at the University of Edinburgh or at the Prague Institute of Formal and Applied Linguistics), none of it is so wide-spread and relevant to the average computer user.

The theoretical part of the thesis will first supply background information on the manner in which computers process and translate text: the development and the widely used mechanisms of machine translation with their common weaknesses and the methods used to judge their accuracy. The human-approach sub-chapter will focus on the general differences between Czech and English from the typological perspective and what they mean for the form and function of the languages' subjects. In accordance with the expected results of the practical part of the thesis, the principal phenomena discussed will be the influence of functional sentence perspective on word order in Czech and English and the existence of the subjectless sentence in Czech.

The analytical part of the thesis will be performed on four Czech-language texts: Karel Čapek's drama *R.U.R.* and three novels: Patrik Ouředník's *Europeana*, Josef Škovercý's *Mirákl*, and *Hrdý Budžes* by Irena Dousková. Each text will be translated into English by GT and 25 non-correspondences in the translation of the subject will be excerpted, totalling 100 non-correspondences for analysis. Strategies used by GT to deal with Czech subjects conducive to non-correspondence will be described and analysed. The thesis will then attempt to uncover the reasons behind the divergences from the point of linguistics drawing mainly on the findings of Dušková (2002, 2012, 2015). GT's strategies will then be compared to those used by human translators.

2. Theoretical background

2.1. Machine Translation (MT)

Machine translation was one of the first non-numerical tasks for the newly-invented computers. The approaches and motives were numerous, from the simply utilitarian ones provoked by the Cold War to more ambitious academic ones that aspired to uncover the principles of natural language and even the functioning of the human mind.

According to Hutchins, the first efforts were centred mainly on Russian and English and were dictionary-based (Hutchins, 2000: 4). However, it was soon realised that word-for-word translation is not enough, even more so when non-Indo-European languages are considered, and by early 1960's, most of the research groups had re-focused on syntax instead. The ultimate goal of most of them was fully automatic high-quality translation, which proved to be a much more complicated achievement than was initially expected. For illustration, the baby steps of MT were made in the time when the computers took up entire rooms and the only way to interact with them were punch-cards – not only did the program need be stored on some, but the input and the dictionaries had to be converted into binary and punched onto cards as well, all taking up enormous amounts of time, money and space.

As the research was funded by the government, the pressure to display visible results was great. In 1963, Automatic Language Processing Advisory Committee (ALPAC) was established by the United States government to evaluate the progress in the field. Majority of its members was still basing their judgement on the unattainable expectation of fully automatic high-quality translation. In 1966, ALPAC recommended discontinuation of funding, which was widely considered “biased and short-sighted” (Hutchins, 2014: 2). As the news of its decision spread, state financing was mostly halted not only in the US, but also in the USSR and Europe. MT was, for the time being, considered a failure (Hutchins, 2000: 4).

The efforts were renewed in 1980's and 90's, no longer motivated by the Cold War, but by an increasingly interconnected global market and administrative demands of multilingual and multinational bodies like the Commission of the European Communities. This was accompanied by sobering of expectations held for MT. The previous fairy-tale vision of completely autonomous computer translator able to deal even with *belles-lettres* was replaced by a more reality-grounded image: MT as a “merely” useful aid, the output of which none the less must be carefully edited by a human translator.

The most wide-spread systems at the time were based on an indirect translation using interlingua: intermediary symbols containing semantic, morphological and syntactic information

supplemented by non-linguistic “knowledge bases” – micro-glossaries etc. Such approaches are called rule-based: they rely on the assumption that the rules for the production of language can be systematically described and new, fully grammatical sentences can be automatically generated on their basis. However, this assumption proved to be true only to a certain degree. The number of exceptions is too great for this method of describing language to be effective and the results fluent.

The end of 1980’s brought a new direction to MT research, represented by two significant events. The results of the first purely statistical translation system, IBM’s grammatically “naïve” Candide, were published, and Japanese research groups started to use corpora of translation examples in the development of their software. Although the rule-based approaches continued to be investigated both separately and in combination with the statistical models, the general direction of the research turned toward statistics and machine-learning (Hutchins, 2014: 4).

Due to the increasing availability of computers, the development of MT gradually ceased being a mostly academic affair and spread to the commercial sector. The first online translation service Babel Fish was launched in 1997 (Hutchins, 2014: 4). Google Translate, currently the most significant software for the majority of computer users, was introduced in 2006 and, for the first ten years of its existence, used the statistical MT model with English as its default language, which means that even in the case of translation between two languages other than English, the source text was first translated to English and only then to the target language. This approach had many obvious shortcomings that needed to be addressed.

Luckily, in 2012, there was a sudden breakthrough in machine learning – an old idea called “neural networks” was re-introduced in the field of image recognition with astonishing results. This spurred new advances in all aspects of artificial intelligence development in the following years, including natural language processing (NLP henceforth), a field which covers such disciplines as speech recognition and production, opinion analysis or, most relevantly for this thesis, machine translation. In 2016, Google introduced a new version of Google Translate based on neural networks for several world languages. (Turovsky, 2016) This was accompanied by a sudden advancement in the quality of the translations, which was caused, among other factors to be detailed in following subchapters, by English no longer being GT’s “default language” – all the translations are now made directly from one language to another.

2.1.1. Grammar theories and their relationship to NLP and MT

In the development of NLP and NLP-motivated thinking about language, syntax occupied a special role, and understandably so. As soon as it was realized that mere word-for-word

translation would not suffice, it was the next logical direction of thought to take. Especially influential in this field were the theories of Noam Chomsky and Robert Montague, who both tried to approach natural language as a formal language. (Čermák, 2011: 130) The concept of Chomsky's generative grammar seemed precisely like what NLP needed – a set of rules according to which any possible grammatical sentence of the given language could be produced – the ultimate description of a language.

However, the generative approach is often criticised by structural linguists as mistakenly and “nonsensically” reducing grammar merely to syntax (Čermák, 2014: 74). It was thought that through parsing the sentence according to its syntactic structure, the semantic structure of the text could be revealed (Isabelle and Foster: 763), but this turned out not to be the case. As Čermák (2011: 130) states, “[Grammar] itself is determined by the formal rules of valency which lie in the lexicon,” and further warns that “considering the production and creation of a text, the semantic rules have always the priority over the formal rules” (Ibid.: 167).

2.1.2. Statistical MT (SMT)

As was stated in the section detailing the history of MT, SMT was the first translation approach that did not have to rely on linguistics as such and relied instead on the rules of probability - in its purest form, the system did not contain any traditional grammar rules. In this regard, current state-of-the-art NMT systems discussed in the following subchapter are the same.

There are two steps to SMT: the input is first processed by the translation model that has been trained on parallel human-aligned corpora to find the most probable equivalent to each translation unit individually, then a linguistic model is applied to gauge the probability of the output sentence occurring in the target language. The linguistic model is generally easier to train well as it needs only target-language data, which is easier to obtain. The larger the dataset is, the better the accuracy. This also true for NMT.

The most prominent problem with this method is the aforementioned translation unit. How is it defined? Where do its boundaries lie? It has been defined a single word ((Brown et al., 1990) cited in Bojar, 2017: 9), which carried many obvious problems. It didn't reflect the fact that human language is generated and processed in phrases and lexical bundles, and did not account for the fact that e.g. reference or morphological dependencies can often span across the whole sentence or even the whole text, which resulted in a large number of errors in morphology. Also, the language model was linear - only able to consider one word

at a time with regard only to the words preceding it (Bojar, 2017 [lecture]). The mechanism simply presumed too much independence of individual words.

There were attempts to remedy this, namely the later development of phrase-based systems which tried to parse the input not into individual words, but instead into phrases that were hoped to correspond to some semantic units. This was often successful. SMT was also frequently augmented by a set of rules, e.g. for morphology (Bojar, 2017: 11). However, although the quality of translations improved, the true revolution came around 2010 with the implementation of neural networks.

2.1.3. Neural MT (NMT)

2.1.3.1. What is NMT? What are neural networks (NNs)?

Despite the word “neural” in the name, the nature of the translation mechanism remains statistical - it still searches for the most suitable translation of the source sentence based on probability. However, there are two main advancements: it is able to consider each word with regard to every other word in the sentence, which leads to better disambiguation of meaning in the case of polysemic words. Also, it is able to generalize and infer the probable behaviour of words in contexts other than those present in the training data, which is useful for relatively rare words all forms of which were not necessarily represented even in large training data (Bojar, 2017 [lecture]).

A neural network (NN) is a deep machine learning architecture based on representation (Goldberg, 2017: 2). To put it more simply, NN is a programme designed to train itself to recognize patterns on a large number of examples with no rules given beforehand. Currently, NNs are experiencing a boom and are widely used in fields such as speech processing or optical character recognition. Despite the fact that the notion of NNs has been around for a long time, the NN boom is a fairly recent phenomenon. The reason for their sudden rise in popularity in 2010's is twofold: the rapidly growing volume of data that now comprises the Internet can be, after careful selection, used in training the models, and the technological advances that make the calculations possible.

The input first goes through the encoding layer which converts the words into vectors of ones and zeros and whole sentences into matrices so that they can be subject to

mathematical operations.¹ The matrices map the way in which the words co-occur. This has been explored in a revolutionary way by Tomáš Mikolov in his *Word2vec* tool, which can create, from any corpus, a projection of words and their linguistic and, perhaps even more remarkably, semantic relations to each other in multidimensional space. This provides a solution to the major problem of older SMT, which saw each word as a discrete unit and was completely unaware that e.g. the words *tree* and *trees* had something in common (Wilks, 2006: 417). NNs are not only able to spot this type of obvious morphological connection, but they are also able to identify relationships not readily recoverable from the forms of the words only, as between e.g. *king* and *queen*, and even articulate them in an equation - “‘*king*’ - ‘man’ + ‘woman’ = ‘*queen*’” (Mikolov, 2013: 746).

2.1.3.2. NMT and linguistics

On the basis of the previously stated facts, one could ask whether traditional linguistics, which had been indispensable in the earlier approaches, is in any way still relevant to the future development of NLP and MT. Is it to become a strictly mathematical domain?

Currently, certain linguistic awareness is needed in selecting and aligning the training data, which is a vital part of the process. It is also useful in the evaluation of the MT output and naming and categorizing the inevitable surfacing problems. Additionally, remnants of the rule-based approach, which had the most to do with traditional linguistics, are also still present in some systems in some form.

However, it seems that its significance for the design of the actual translation mechanism will only lessen.

2.1.3.3. Weaknesses of NMT

When comparing the results of SMT and NMT, it may seem that NMT brought an enormous rise in quality of the translations. It is true that NMT mostly solves the problem of fluency. Indeed, sometimes the results appear so natural that it can lead one to believe that the problem of MT is solved for good. However, the actual adequacy of translations is improved only to a smaller extent. (Bentivogli et al., 2016a; Bojar et al., 2016a; Castilho et al., 2017b,a in Bojar, 2017: 44) The output is

¹ This is the “representation” mentioned in the beginning of this section.

“more often seemingly perfect but including a semantic flaw” (Bojar, 2017: 44), which is harder to spot.

In the present, the problem that NMT encounters the most often, due to it being trained on a limited amount of data, are rare words. They can be proper nouns, simply less frequent words or creative coinages. To deal with them, GT utilizes the method of “wordpieces” (Wu et al, 2016: 7). The default approach is to process a word as a whole outlined by spaces on each side. However, if the word is not recognized, every possible way of dividing it into smaller parts is analysed and the most probable outcome is used. The parts making no sense at all are left as they were. This can lead to some bizarre outcomes.

2.1.4. Evaluation of MT

2.1.4.1. Common Error-Types introduced by Bojar in 2011

The errors outlined in Bojar’s paper, based on preceding typology by Vilar et al., are the result of a translation in the direction from English to Czech. As such, they cannot be applied directly onto errors made in Czech to English translation because of the differences of the target languages.

Bojar distinguishes five classes of errors: “Missing words”, “Word order”, “Incorrect words”, “Unknown words” and “Punctuation” errors. This typology still operates with SMT, so some types will not be detailed further as they mostly do not apply to NNs, others are not a part of this thesis’ subject matter, like “Punctuation” errors and “Word order” (Bojar, 2011: 70).

2.1.4.1.1. Missing Words

This type of error is expected to occur more in NNT systems than in previous SMT systems as the biggest weakness of SMT is in this respect its biggest asset: if it is set to consider each word independently, it has a lower chance of missing it. Bojar distinguishes between missing content words and missing auxiliary words. NNT is prone especially to the former – simply disregarding a word because it does not fit into the trained phrases. The latter is an area NNT is very good at – to make the sentence seem as fluent as possible.

2.1.4.1.2. Incorrect Words

According to Vilar et al. (2006), this is the widest category of errors in which four subtypes can be distinguished: “Extra word”, “Bad word form”, “Not translated” and “Sense” (Vilar et al., 2006: 699), or what Bojar (2011) calls “Bad word sense.”

Out of these, “Bad word form” and “Extra word” and “Not Translated” seem unlikely to be very common in NNT and will not be detailed further, as NNT output is characterized by the tendency to appear fluent at the first glance. In fact, NNTs may be overly active in trying to translate each word or at least make it seem like an organic part of the target language. This is an excellent trait and a great improvement from SMT in some cases, as when translating a compound undocumented in the training data, but an inappropriate approach to be taken with proper names or numbers.

Bad Word Sense

“Bad word sense” describes either a wrong choice between several meanings of a homograph or the choice of an unsuitable synonym. This is still an area in which NNTs are expected to fail from time to time as they do not try to interpret the sentence but rather to blindly “approximate” the output to the form of target language without analysing it. The question is whether anything else can ever be expected from computers.

2.2. Human Translation (HT)

2.2.1. Criteria for adequate translation

According to Grygová (Knittlová et al., 2010: 14-15), the three criteria for a quality translation are to a) sound completely natural and organic in the target language, b) be as close to the original as possible both in meaning (denotation) and in tone (connotation) and c) preserving the dynamic of the ST – the translation should provoke the same reaction in the recipient of TT as in ST.

2.2.2. Translation Strategies, levels of equivalence

The methods and strategies of translation by humans have been described and categorized in many different ways. On a general level, there are four main types of translation: literal

translation (selects the equivalents of each item without any regard to context), interlinear translation (which is essentially an extreme version of literal translation – it does not even necessarily follow the grammar rules of the target language and is used mainly in descriptive linguistics), free translation (only approximates the meaning of source text and often disregards stylistics or register and leaves information out) and communicative translation (pragmatically oriented, used e.g. when translating set phrases and conventional formulas) (Knittlová et al., 2010: 17)

Regarding the strategies that can be used, Vinay and Darbelnet's categorization, on which many others are based, distinguishes direct and oblique methods of translation. Use of the direct method, i. e. literal translation of each unit, is often not possible as languages express different concepts in different manner. In absence of a direct equivalent, to deal with the "gap" between the languages, the translator needs to employ an oblique strategy.

In *Překlad a překládání* (2010), Knittlová et al. summarize Vinay and Darbelnet's oblique strategies in regard to their usage in Czech linguistic tradition by e. g. Poldauf or Bareš. The relatively straightforward ones include direct borrowings (e.g. *déjà vu*), calquing (e.g. *potflower - hrnková květina* (Knittlová et al., 2010: 19)), substitution (e.g. of a noun by a corresponding pronoun or a word by its synonym, which may or may not result in a semantic shift) and transposition ("necessary changes in grammar due to the different language systems" (Knittlová et al., 2010: *ibid.*), i. e. "replacing one word class with another without changing the meaning of the message" (Vinay and Darbelnet, 1995: 36)). The more complex ones include modulation (a change in the viewpoint, e.g. *elbow of the pipe – koleno potrubí* (Knittlová et al., *ibid.*) or *It is not difficult to show... - Il est facile de démontrer...*²(Vinay and Darbelnet, *ibid.*)), equivalence ("usage of means stylistically and structurally deviating from the original", as e.g. *my sweet girl – děvenka*, or translations of idioms and sayings (Knittlová et al., 2010: *ibid.*)) and adaptation (substitution of an entire situation with a different, analogous one, e. g. when translating a pun or a situation not occurring within the culture of the target language (Knittlová et al., 2010: *ibid.*))

Based on the previous subchapter, an assumption can be made that the more complex and specific a strategy is, the sparser in the training data and thus less likely to be adopted by Google Translate. While modulation may still include set phrases which can be supposed to occur somewhere in the training data, equivalence is not expected to be encountered

² Translation: It is easy to show. This has also been described by ... as vector substitutin – looking at the same situation from the opposite angle (A in front of B, B behind A).

much and adaptation seems totally impossible for GT as it is based on extra-lingual knowledge.

Interesting with regard to MT output may also be Komissarov's five levels of equivalence also described by Knittlová et al. (2010) – categorization comparing the correspondence of source text (ST) and target text (TT). The first level is the pragmatic level on which there is no visible correspondence at all, only in the communicative intentions (*How do you do* – *Dobrý den*), then comes the situational level on which the situation is the same, but different aspects of it are highlighted (*He answered the phone* – *Zvedl sluchátko*). The third level of equivalence is the level of announcement³ which may diverge in syntax and hierarchization of parts of the text, but the units are the same. The fourth level is the level of utterance⁴ on which the syntactic structure corresponds, but some slight changes are still made, as e.g. passivisation. The fifth is correspondence on the level of the sign - literal translation (Knittlová, 2010: 26). Out of these, GT's output is expected to exhibit overall higher level of correspondence than the HT output.

2.3. The subject

2.3.1. The subject in different syntactic theories

There are two approaches to parsing a sentence. It can be parsed by the form, in which case the sentence is divided into phrases determined by the word class of their head. As was already stated, the formal viewpoint was preferred by the early NLP researchers, inspired by Chomsky's generative grammar. However, the continental linguistic tradition has always been closer to the functional viewpoint. Beyond the level of the formal structure lies the distinction of constituents by their function into clause elements. (Petr Sgall's creation of Functional-generative Description (FGD) was an attempt to bring the best of both worlds into the NLP field. Projects like *Prague Dependency Treebank* continue this legacy.)

Connected to these two views are the two perspectives on the hierarchy of the constituents. Phrase-structure grammars, shaped by fixed-order-language native speakers, view the noun phrase and the verb phrase as central and in a special, predicative relationship in which both of the constituents are equal. This view is also taken in primary and secondary education, and understandably so – “while the number of predicates determines the number of the clauses in the complex sentence, the subject determines the grammatical (or possibly

³ „Rovina oznámení“ in Knittlová et al., 2010

⁴ „Rovina výpovědi“ in Knittlová et. a., 2010

also semantic) categories of the predicate.” (Panevová et al., 2014: 19) ⁵ This makes them both good candidates for the status of the elementary constituents. (ibid.) In contrast, the continental functional tradition puts more emphasis on the valencies of each clause element and the dependencies which always stem from the predicate. The subject is then seen primarily as an argument of the verb (Čermák, 2011: 184). According to the Czech linguist Havránek, the bipartite structure of the sentence is merely a superstition (Čermák, 2014: 74). However, it is acknowledged that the position of the subject “stands the highest in the hierarchy of valency positions.” (Daneš, 1987: 27)

2.3.2. Common characteristics of the subject

In both languages, the subject most frequently takes the form of a noun phrase or a nominal clause. It determines the number and the person of the predicate. It also determines grammatical categories of other clause elements, e.g. direct and indirect objects in the form of reflexive pronouns or the number of the subject complement. (Quirk et al., 1985: 725)

Its semantic role can be generally characterized as that of the agent, i.e. "the animate being instigating or causing the happening denoted by the verb" (Quirk et al. 1985: 741, cf. also the Czech term 'původce děje' (Daneš et al. 1987: 45)).” Apart from the prototypical agentive role, it can also take the role of an external causer (typically an inanimate cause of an event, as in *The avalanche destroyed several houses*), an instrument (the entity which is used by the agent to perform the action, as in *A car knocked them down – Srazilo je auto*), an affected (where the action happens to the subject rather than being initiated by it, as in *Jack fell down* or *The pencil was lying on the table*), an experiencer (as in *He suffered a heavy blow – Utrpěl těžkou ránu*), or a recipient (*I was given no choice – Nebylo mi dáno na vybranou*). The subject can also be locative (*The garden swarms with bees – Zahrada se rojí včelami*), temporal (*Léta jsou horká, zimy studené – Summers are hot and winters cold*) or eventive (*The match is tomorrow – Zápas je zítra*). (Quirk et al., 1985: 743-749; Dušková et al. 2012: 13.14; Daneš et al., 1987: 44-50)

In regard to information structure and functional sentence perspective (FSP henceforth), the subject takes typically the role of the theme both in Czech and in English (Dušková et al. 2012: 13.15.1) – it expresses the information “regarded by the speaker as given” (“Quirk et al., 1985: 726).

⁵ translation from Czech, „Podle počtu predikátů se určuje počet vět v souvětí, na druhé straně subjekt ovlivňuje svými mluvnickými (popř. významovými) kategoriemi podobu predikátu...“

2.3.3. The English subject

In English, the subject is regarded as the most important clause element after the predicate. It is “obligatory in finite clauses except in imperative clauses, where it is normally absent but implied.” (Quirk et al., 1985: 724-725).

2.3.3.1. Form of the English subject

The most common form of the subject is the noun phrase (headed by a morphological noun or a syntactic noun which is primarily a different part of speech, as e.g. ...*the old may persist alongside the new* – ...*může staré přetrvávat vedle nového*) The head noun can also be substituted by a personal, demonstrative, interrogative or a relative pronoun, or any pronoun ending with *-one*, *-body*, *-thing*. (Dušková, 2012: 13.13.1)

It can also be realised by an adverb phrase, as in *Now is the right time*.

Non-finite verb forms can also form the subject. This includes the infinitive, as in “To reach an agreement proved impossible – Dosáhnout dohody se ukázalo nemožným”, and the gerund, e. g. Confiding in him was a mistake – Svěřit se mu byla chyba. (Dušková, 2012: 15.21)

Subordinate clauses are other units that can function as the subject, e. g. in That such cases do occur has been proved beyond doubt. – Že se takové případy vyskytují, bylo dokázáno mimo vší pochybnost. (Dušková, 2012: 16.21.11)

Apart from the anaphoric function typical for pronouns in general (e. g. *Be careful with the typewriter, it's quite new.* – *Buďte s psacím strojem opatrná, je úplně nový*), *it* also has other, special functions. It can be used deictically, pointing to the reality outside of the language, e. g. *Is it your suitcase?* – *Je to váš kufr?* (Dušková et. al., 2012: 13.13.2) It can also serve as a purely formal prop in the cases where the agent is unclear, typically in a place where Czech would use the subjectless sentence, as in e. g. *It had turned chilly* – *Ochladilo se*. The fifth kind of *it*-subject is the anticipatory *it*. It can either be used when an infinitive or gerund would be inconvenient in the preverbal position because of its length or complexity, as in *It was better not to invite too many important people at the same time.* – *Bylo lépe nezvat příliš mnoho důležitých lidí zároveň* (Dušková et. al., 2012: 13.13.2), or in a cleft sentence, where the actual subject that the *it* refers to is the rheme (the new information being said about the theme of the sentence (Daneš et al., 1987: 550, cf. *focus* in Quirk et al.,

1985: 1361)) and thus needs to be highlighted, e.g. *It was Dr. Long who had rung the bell. – Tím, kdo zazvonil, byl doktor Long* (Quirk et al., 1985: 349).

Another formal device to make possible the placing of the actual subject in the postverbal position is the existential construction *there is/there are*. It is used either to express either the general existence of a thing or a concept, as in *Contrary to popular belief, there are many more than five senses. - Na rozdíl od rozšířeného názoru je (existuje) mnohem více než pět smyslů* (Mackay, 60 in Dušková et al., 2012: 12.21.4), or a location-bound existence of a thing, as in *There is a girl in the waiting room. There* is understood as the formal subject and what is formally the subject complement is considered “subject proper” or “notional subject” (Dušková et al., 2012: 13.13).

This can be also achieved by inversion – placing of the subject in postverbal position. While this is standard in Czech, it is highly stylistically marked for English (although standard in questions). In spoken English, it is common only in such expressions as *Here comes Peter! – Tady přichází Petr!* (Dušková, 2012: 14.37) In literature, it is used as a stylistic means for continuous storytelling.

2.3.3.2. Absence of subject in English

There are two possible cases in which the subject may be absent from a sentence. It can be a result of an ellipsis or in an imperative sentence.

Ellipsis is the process during which a part of the grammatical sentence is dropped (elided), but the specific words omitted are readily recoverable (Dušková 1999: 176). It standardly occurs during coordination of the sentence elements, as in *David loves Joan and wants to marry her* (Quirk et al., 1985: 938). It is also frequently found in colloquial English. In the majority of cases, the subject is then elided together with the finite part of the verb, producing utterances and lexicalised phrases, such as *Sorry* instead of *I am sorry* or *Glad to hear it* instead of *I am glad to hear it*. Instances where only the subject is left out are much rarer and include examples as *Thank you* or *Serves him right*. (Dušková, 1999:176)

Quirk et al. consider elliptical sentence “grammatically ‘defective’” as “some normally obligatory element of a grammatical sentence is missing.” (Quirk et al. 1985: 885)

A similar thing can be said about the imperative sentences in the second person: while they do not explicitly include the subject, they cannot be considered “subjectless” (Dušková et al., 2012:), but merely as having an “unexpressed subject, comparable to Czech sentences with a pronominal subject that is left out.” Although the subject of imperative sentences cannot be distinguished by any ending, the uniqueness of the form of the absent subject in the imperative sentence makes it immediately identifiable. In some cases, the identity of the subject is made even clearer, as e.g. with a reflexive verb or an appended tag question. (Dušková, 1999: 176)

2.3.4. Czech subject

While the Czech subject is commonly realised by a noun or a pronoun in the nominative case or a clause or an infinitive, there are other forms it can take.

When it is realised by the genitive case instead, the syntactic relationship does not change, but the genitive carries additional information, mostly of quantitative character. (Daneš et al., 1987: 43) There are three distinguishable types: “záporový” genitive - *Neminulo dne, aby nepřišla nějaká další nepříjemná zpráva*, partitive genitive - *Tam bylo hraček!* and numerative genitive - *Z vozu vystoupilo pět ozbrojených mužů.*

It can be also realised by a prepositional case (*Náhle se rozsvítilo na sta světel.*) (Daneš, 1987: 43)

2.3.5. General differences between English and Czech

The differences stem from the distinct historical development of the languages. While Czech remains an inflectional language, which means that it uses affixes to express the dependencies between the words, English evolved into an analytical, isolating language, which has reduced flexion in favour of the fixed word order S-V-O.

This means that, in Czech, the pronominal subject can often be elided since the verb suffixes are in most cases sufficient to ensure the coherence of the text. Czech also uses subjectless sentences for expressing the absence of an agent, which is not possible in English. Its largely absent flexion and reliance on fixed word order means that the “slot” for subject generally must be filled for the sentence to be intelligible. (This is elaborated on in the following subchapter.)

Correspondingly, a lexical item in the role of the Czech subject implies agency much more strongly than in English. Action verbs are customarily used exclusively with agentive subjects. However, this is not true for English, as can be seen e. g. in the sentence *The diver burst his eardrums – Potapěči praskly bubínky*, where the diver, who is the subject in the English sentence and the indirect object in the Czech, is not the agent, but a passive patient. As English moved away from inflectional to analytical structure with fixed word order and the need for the presence of a subject grew, the subject started to take on other semantic roles more frequently, as e. g. the locative subject – *The path is swarming with ants – Na cestě se to hemží mravenci*. or any participant involved in the action as in *The meat cuts well – Maso se dobře krájí*. (Dušková et al. 2012: 13.14) This is useful either when there is no specific agent or where the agent is rhematic and therefore needs to take the role of a different clause element for which it is possible to occur later in the sentence.

Although the subject is mostly thematic in both Czech and English, the actual percentage of thematic subject occurrence is higher in English due to its fixed word-order. In Czech, “the subject fairly often assumes the function of the rheme” (Dušková, 2015: 15) - the part of the sentence with the highest degree of communicative dynamism, which means that it carries the new information the sentence is supposed to convey. Rheme typically stands at the end. The subject assumed this role and position in 22,4% of the observed Czech subjects in Dušková, 2015: 15 (also Mathesius: 1947, 278 and Dušková et al., 2012: 13.15.1.).

The hypothesis is that these differences in the structure of the text will be among the reasons for GT’s mistakes, as they are not that typical and NMT’s weakness is data sparseness.

The above differences between the English and Czech subjects suggest that the areas which are likely to cause problems in translation from Czech to English involve the English counterparts of the Czech subjectless sentences, null subjects and non-thematic (post-verbal) subjects.

2.3.5.1. Czech subjectless sentences and their English equivalents

This specific topic is thoroughly investigated by Dušková in the 1999 paper “English Counterparts of Czech Verbal Subjectless Sentences.” It is structured on the basis of the Czech subjectless sentence system. First, they are categorized with regard to the form of the sentence into subjectless verbal sentences (which include non-agentive verbal subjectless sentences and verbal subjectless sentences with the reflexive form of the verb) and subjectless verbo-nominal sentences. Each category is then divided into semantic groups.

2.3.5.1.1. Verbal subjectless sentences

Non-agentive verbal subjectless sentences

These sentences can be divided into two semantic groups: those expressing weather conditions, such as *Stmívá se*. – *It is getting dark* or *Prší*. – *It is raining*, or those expressing physical states, like *Svrbí mě* – *I itch*.

Sentences related to weather are generally translated in one of the three ways - with empty *it* as the subject (*Ochladilo se* – *It has grown cold*), with the existential construction (*Zablesklo se* - *There was a flash of lightning*), or with a noun related to the Czech predicate functioning as the subject (*Hustě sněžilo* – *The snow was falling pretty thick*). Out of these three, English generally prefers the use of the existential(-locative) construction with a copula– for instance, instead of *It snowed heavily last night*, a more natural-sounding sentence would be *There was heavy snow last night*. (Curme, 1935: 68 in Dušková, 1999: 178)

Sentences semantically expressing physical or psychic states are slightly different – in Czech, they feature the experiencer or the affected part as an object in the accusative, dative or locative case, thus diminishing their agency and leaving the originator of the state unknown. In these cases, English loses this fine distinction and the experiencer is typically construed as the subject, e.g. *Svrbí mě* translates as *I itch*. If they contain the location as well, it is possible to use the same method with added local adjunct, or the adjunct can become the subject of the English sentence and the experiencer will be expressed by the possessive form, as in *Teče mi z nosu* and *My nose is running* (ibid., 181).

Formally ungeneralizable are translations of verbal subjectless constructions with prepositional objects. On this topic, Dušková notes

“There are verbs like *jde o*, *jedná se o*, *běží o* [...]. In English, they correspond to different subject-predicate constructions, all of which serve to indicate the rhematic character of the semantic element operating as the complement of the Czech verb, e.g. ‘*tady jde o spravedlnost*’ [translates as] ‘*this is a matter of justice*’, ‘*běželo o něco jiného*’ [as] ‘*the point was something else*’ (Dušková, 1999: 183).

Reflexive verbal Czech sentences

Reflexive subjectless verbal sentences are, unlike non-reflexive subjectless sentences, not non-agentive. Instead, they imply the general human agent (*O Persově dívčí škole se v té době nemluvilo dobře – The Perse School for Girls was not well spoken of, at that time*). They are one of the two ways to express the passive voice in Czech and can be translated as such into English, “provided that that the verb has a complement that can operate as subject” (ibid., 185). If there is not any, as in the case of “šlo se rovnou k nejbližšímu konci zeleninové zahrady,” English has two possibilities, namely construing a general human agent as the subject (e.g. *You went right to the furthest end of the kitchen garden*) or relying on the existential construction (e.g. *Dosud se nebojovalo – There had been, as yet, no fighting* (ibid.)).

Less frequently, reflexive subjectless verbal sentences can express a disposition, in which case the subject can take the form of deverbal noun (*Nyní se žije lépe – Life is easier now*), the Czech object (*Těžko se s ním vychází – He is difficult to get on with*)

2.3.5.2. Verbo-nominal Czech subjectless sentences

This type of sentences consists of a copula (mostly *be*) and a nominal part (ibid., 186). It can again describe both physical states (*Je mi nevolno – I feel unwell*) and psychic states (*Bylo jí teskno – She felt lonely*) where the role of the subject is taken on by the experiencer. States of wider and local natural environment often use the anticipatory *it* (*Bylo pozdě a dost tma – It was late and rather dark*) or “the locality of the state in question” (*Je jasno – The sky is clear* (ibid., 187)) as their subject. Verbo-nominal subjectless sentences referring to time or space also frequently use the anticipatory *it* (*Je čtvrt na pět – It is quarter past four*) or, less often, a temporal or locative subject (*Bylo po škole – The school was over.*)

Lastly, there is a semantic group of sentences expressing modality, the complements of which are formed in present-day-Czech mainly by modal adjectives (*Je nutno odejít – It is necessary to leave*). English equivalents of these sentences either use the anticipatory *it* exemplified above, the passive voice with the appropriate modal verb (*Nyní lze položit následující otázku – The following question can now be asked*), the existential construction (*Nebylo třeba svolávat valné shromáždění – There was no need to summon a general gathering.*) Relative clauses can be condensed into the attributive infinitive (*První problém, který je třeba řešit, je tento – The first*

problem to be solved is this.) Dušková then points out the phrase “je třeba” and remarks, that this can be supplemented by the verb *require* (*K tomu bylo třeba značné odvahy – This required considerable courage.* (ibid., 190))

2.3.5.3. Subject constancy in the languages

Since the theme is usually constant for at least several sentences, we can deduce that in English, the identity of the subject has a tendency to be quite constant as well. (Mathesius 1947, 278) On the other hand, since the Czech subject (especially postverbal) frequently takes the role of the rheme and rheme changes from sentence to sentence, the entity construed as the subject is not as constant throughout the Czech text. This may potentially be a cause for confusion for an NMT system.

2.3.5.4. Subject syntactic and FSP constancy across languages

This topic was studied by Dušková in the 2003 article “Constancy of syntactic function across languages”. Preservation of syntactic function of a lexical item in translation was found to be quite high (95,2% in the Czech-English direction). In the non-corresponding cases, Czech subject was found to be construed as the direct object in 28% of the cases. This degree of occurrence is still somewhat common, so NMT is not expected to err much in these cases either. 19% of Czech subjects was translated without any explicit English counterpart. 16% of Czech subjects was reflected in the possessive determiners. In 10%, the subject corresponds to the indirect object.

Given Czech’s more frequent use of rhematic subjects and limited means of English to achieve this, the results are more diverse when analysing the correspondence of Czech rhematic subject in particular. The majority (80%) of rhematic subjects still remain constant, which means that the same lexical entity is construed as the subject despite not following the typical FSP layout of theme-transition-rheme (*a tu ho náhle zachvátí jakási lítost – a wave of sudden pity came over him*). Dušková interprets it as “the principle of end focus [being] to a large extent overruled by the grammatical principle.” (Dušková, 2015: 24) However, the rheme still remains constant most of the time due to its context independence and the semantics of the word (*a wave of sudden pity* from previous example is not contextually bound and therefore it would be apparent for the recipient that it is a new piece of information – the rheme).

However, sometimes the translation is not as successful in preserving the FSP roles, which means a slight shift in meaning. (*tehdy panovala přísná morálka – morals were pretty strict in those days, ibid.*)

3. Material and Methodology

3.1. Material

The material for the analytical part of this thesis was taken from four original Czech texts: one drama – Karel Čapek’s *R.U.R.* (referred to as text A in the appendix) – and three novels – Patrik Ouředník’s *Europeana* (text B), Josef Škvorecký’s *Mirákl* (text C) and Irena Dousková’s *Hrdý Budžes* (text D). Their English translations by David Wyllie, Gerald Turner, Paul Wilson and Melvyn Clarke respectively were also used for comparison of GT’s techniques.

3.2. Method

As GT allows 5 000 characters per translation, the texts were inserted into GT in passages of approximately this size in such a way as not to divide paragraphs. This was done until 25 cases of non-correspondence of the subject had been collected from each text.

Non-correspondence was defined as any instance in which the entity construed as the subject in Czech was not construed as the subject in English, including the cases where the role was filled by some other entity, when the English subject was completely absent in GT’s output and where there was actually no Czech subject to translate (the case of Czech subjectless sentence.)

Cases with null subject in Czech translated by the addition of an adequate pronoun were not included in the analysis, since they have a Czech subject corresponding to the English subject, only it is not expressed in Czech, but indicated by the verbal inflection and the context (ex. (1))

- (1) Jmenoval se Hrdý Budžes, [on] byl velice statečnej a vytrval, i když [on] měl všelijaký potíže. (D 76)⁶

His name was Proud Budesha, he was very brave and persevering, even though he had all kinds of trouble.

Cases in which the whole clause was missing from the translation were not included either as there is not much to describe in terms of correspondence between the two languages (ex. (2)).

- (2) A jiní zase říkali, že dvacáté století začalo, když se zjistilo, že lidé pocházejí z opic...

And others said the twentieth century began when [] people came from monkeys...

⁶ This example, however, has been included on different grounds – the non-correspondence in the first clause. (D76) refers to the label of the example in the Appendix.

During the excerpt, cases were encountered in which there actually was correspondence between the entities construed as the subject, which, however, resulted in an inadequate translation (ex. (3)). These cases, although notable, also were not counted, since they are not non-correspondent.

(3) Zkrátka – slečno Gloryová, nenudí vás to?

In short - Miss Glory, wasn't it bored?⁷

Then, the corresponding English translations by the human translator were collected.

The data collected from these texts during the analysis were the grammatical forms of the Czech subject, GT's subject and HT subject and the adequacy of GT's solution.

The line between adequate and inadequate solutions is thin as the topic of this thesis is only the subject. The general aim was to focus on the subject as much as possible, but it is impossible to judge it completely separately from the rest of the sentence. Solutions involved in a clause resulting as inadequate were judged as inadequate, even though the subject itself could be considered adequate had it been followed by a different translation of the rest of the clause. This is exemplified by ex. (4), where a completely valid strategy of transforming the Czech object into the English subject is used. However, for the overall clause to keep its meaning, passivisation of the predicate would be required.

(4) Nechtěnou aktuálnost nerušila ani dosti středověká písmena, jež z pekelně černého pozadí ohnivě hlásala: ...(C55)

Unwanted topicality did not disturb even the rather medieval letters, which from the hellish black background were praising fire: ...

On the other hand, the translation of the subject in ex. (5) was judged adequate even though the following clause (ex. (6)) shows that it leads to an inadequate solution for the overall sentence.

(5) O první světové válce se říkalo,...

The First World War was said... (B29)

(6) ...že v ní lidé padali jako semena, ...

to have fallen as seeds in it, ...

HT resolves this by a cleft construction in place of ex. (5) and preserving of the Czech syntactic structure in ex. (6) - *It was said of the First World War that people in it fell like seeds*. However,

⁷ In this case, the inadequacy of GT's translation was caused by the preference of English, instead of using *bore* as a verb, to use it in a copular construction, *be bored*. (This again points to the tendency of English to lean towards nominality.) While *be bored* was the more probable option of translating *nudit* according to the training data, the necessary transposition of transforming the Czech object into English subject was not carried out.

if the subject of the following clause was identical to the one in the first, the translation in ex. (3) would be functional. That is why it was judged as adequate.

Inadequacy was judged both on the semantic level and the level of grammar, however, the semantic level was considered more important than fluency and ease of phrasing. This means that excerpts similar to ex. (7), although somewhat awkward, were deemed adequate.

(7) Nešlo mu o nic víc než podat důkaz, ...

He was no more than giving evidence that

4. Analysis

This chapter of the thesis will be structured on the strategies used by GT to render Czech subjects conducive to non-correspondence. First, there will be an overview of the Czech subject types causing non-correspondence in Google Translate (see 4.1) and of the strategies used by GT in general (4.2) Each strategy will then be described and further divided into emerging subtypes. The adequacy of its results will be analysed and possible causes of non-adequacy described.

Methods of GT will then be compared to the methods of human translators. (4.3) A short section will be dedicated to the specifics of each text (4.4)

4.1. Non-corresponding Czech subject types

null subject		37
	adequate	20
	inadequate	17
postverbal subject		30
	adequate	17
	inadequate	13
subjectless sentence		22
	adequate	18
	inadequate	4
regular subject		8
	adequate	4
	inadequate	4
relative pronoun		3
	adequate	1
	inadequate	2
total		100

Table 1 Non-corresponding Czech subject types and their adequacy rate

As hypothesised in the theoretical part of the thesis, the major subject types found to lead to non-corresponding translation are the null subject (37% of the non-correspondences, ex. (8)), postverbal subject (30%, ex. (9)) and subjectless sentence (22%, ex. (10)). Surprisingly, a whole 8% of sentences with non-correspondent subject belong to the category of “regular”, or preverbal subject (ex. (11)). The smallest group are relative pronouns, which form 3% of the excerpted non-correspondences (ex. (12)).

The overall success rate of GT's non-correspondent strategies was 60%. Regarding the difficulty each type of Czech subject poses, the relatively most complex proved to be relative pronouns (GT's solution was judged inadequate in 67% of the cases, ex. (12)). However, as there were only three of them in total, it is necessary to treat this percentage cautiously.

The null subject, postverbal subject and regular subject all have similar inadequacy rates, 47%, 43% and 44% respectively.

The least complicated were, surprisingly, subjectless sentences, in which GT erred in only 18% of the cases. This suggests that they have relatively regular equivalents.

(8) A chudí lidé se těšili, že [oni] pojedou vlakem,...

And the poor people were looking forward to take the train,...

(9) Jinak se mi včera stala taková nemilá věc

Otherwise, yesterday I had such an unfortunate thing.

(10) Třídou zašumělo.

There was a noise in the classroom.

(11) Ale nemilá věc se mi líbí, to se prostě tak říká.

But I like the unpleasant thing, it just says so.

(12) ..., ne symbolickým, ale mým vlastním, zle postiženým gonorrhoeou, jež nemohla pocházet od nikoho jiného než od profesorky ruštiny Ireny Znenáhlíkové.

..., not symbolic but my own, badly afflicted gonorrhoea, who could not come from anyone other than Irena Znenáhlíková, a Russian professor.

4.2. GT's strategies

condensation	32
adequate	25
inadequate	7
Cz O → S	23
adequate	15
inadequate	8
pronominal subject	14
adequate	5
inadequate	9
existential construction	11

	adequate	11
merging		10
	adequate	2
	inadequate	8
missing subject		6
	inadequate	6
Cz Adv → S		3
	adequate	1
	inadequate	2
CZ verb → S		1
	adequate	1

Table 2: GT's strategies and their adequacy rate

The most frequent strategy which GT chooses to deal with the previously described types of subject is condensation (32%), which has been overlooked in the initial hypothesis, but its use is not surprising, given the tendency of English to lean towards nominalisation. The second most frequently used strategy is construing the Czech object as the English subject (24%). Pronominal subject is used in 14% of the cases and the existential construction in 11%. Other comparatively minor strategies include merging (10%), leaving out the subject (6%), and extracting the subject from other sentence elements (2% from adverbial and 1% from the predicate). Surprisingly, cleft construction was not used at all.

4.2.1. Condensation

The strategy of condensation was fairly successful, it was judged adequate in 78% of the cases. GT's solutions have been classified as condensation when there was an "implied predication, which makes it possible to expand the condensed structure into a clause" (Jelínek, 2017). It can be realized by any non-finite verb form or by an action noun (Dušková, 2012: 15)

GT most frequently uses the infinitive (16 cases), then the present participle (9) and to a lesser extent the gerund (3), past participle (3), and exceptions are prepositional phrases and action nouns (1 occurrence each).

It was used the most with Czech null subject (52% of its occurrences, see ex. (8)), which is connected to the typology of the languages – both are standard, neutral solutions in their respective language. While the Czech subject is standardly not explicitly stated in each

sentence as it would be superfluous, because the reference to it is contained in the finite verb suffixes, it is common for English to use non-finite verb forms with the subject expressed in superordinate structures.

All the inadequacies occurred because the subject from the clause which has been condensed by GT was misidentified as identical to a clause element of the clause which is modified by the condensate. This could be a subject, as in example (13) in which the *gonococci* are not the subject *explaining the resilience*, it is rather the fact, that the gonococci had travelled all over the globe - the notional subject of the Czech relative clause is the whole preceding clause. (This has been categorized by Daneš et al. as “nepravá vedlejší věta vztažná navazovací” (Daneš et. al., 1987: 538)).

(13) Moje gonokoky obcestovaly tedy celou zeměkoulí, což snad vysvětlovalo jejich jedinečnou odolnost, málem nezničitelnost.

So my gonococci traveled all over the globe, perhaps explaining their unique resilience, almost indestructible.

Other examples include the misidentification of the subject of the condensed clause as identical to the object of the Czech clause. In ex. (14) this inadequacy is due to the clause-final position of the Czech subject.

(14) Jejich zkřivené tváře a ručičky vztažené k šikmému světci měly patrně naznačit extatickou zbožnost: v rámci celkové kompozice daly se však také vyložit jako přirozená reakce lidí, jimž padá na hlavu kolos rhodský.

Their distorted faces and hands relative to the oblique saint were likely to indicate ecstatic piety: but within the overall composition they could also be interpreted as the natural reaction of the people falling upon the head of the Rhodes.

However, when the subject is included in the absolute construction (ex. (15)), the results of the condensation are adequate.

(15) Představte si ho, že sedí nad zkumavkou a myslí na to, jak z ní vyroste celý strom života, jak z ní budou vycházet všechna zvířata, počínajíc nějakým vírníkem a končíc – končíc samotným člověkem.

Imagine him sitting over a tube and thinking of how the whole tree of life will grow out of it, all the animals coming out of it, starting with some rotifers and ending - ending with the man himself.

4.2.2. Czech Object and other clause elements construed as the English subject

Construing other Czech clause elements as the English subject was used most frequently the with Czech postverbal subject (41%). This can be explained by the tendency of the postverbal subject to perform the rhematic role. The strategy of transposition of the sentence elements sometimes resulted in preservation of the FSP roles, as in ex. (16), where the English subject corresponds to the Czech object *vás*, making it possible to retain the clause-final position of the noun phrase *historie vynálezu – the history of the invention*.

(16) Zajímala by vás historie vynálezu?

Are you interested in the history of the invention?

As was stated in the theoretical part, another device to preserve FSP roles is the anticipatory *it* used in cleft constructions. Curiously, this strategy was not used at all by GT in the 100 excerpts.⁸

The sentences judged as inadequate also often displayed lack of passivisation where it was needed for the translation to preserve the meaning (ex. 17). Without passivisation, it appears more likely that the subject was simply misidentified as an object due to the postposition.

(17) Sošku samu poznamenal -- rukou pana Řeřichy -- obrodný proces, a znamení dosud odolávala času i jeho staronovým osvícením.

The statue itself - Mr. Řericha's - revived the process, and the signs still resisted the time and its old-age enlightenment.

4.2.3. Existential construction

The existential construction proved an exceptionally reliable strategy – the resulting translations were judged adequate in 100% of the cases. This suggests that it is a highly

⁸ Although there had been one instance of GT using anticipatory *it*, it had to be excluded from the analysed examples, as it was a translation of a quote originally in English.

Žádný člověk není symbol... proto je tak snadné *zabijet skutečné lidi* jménem té nebo oné zatracené ideologie;

No man is a symbol... so *it* is so easy to kill real people on behalf of that or that damned ideology.

It was used with a postverbal subject and nominal predicate. “Be” functioned as the copula in both languages. Out of the included examples, none of these display this pattern.

frequent and quite straightforward translation strategy. However, comparison with the HT reveals that it is not as frequent when translating fiction. This will be elaborated on in 4.3.

It was used exclusively with postverbal subjects (73%) (example 18) and subjectless sentences (27%) (example 19), making it the most specifically used one from the more frequent strategies.

(18) V kontrastu k této výzdobě stěn je na zemi nádherný turecký koberec.

In contrast to this wall decoration, there is a beautiful Turkish carpet on the ground

(19) Třídou zašumělo.

There was a noise in the classroom.

In this example of usage with the subjectless sentence, the rheme, which lies in the Czech predicate *zašumělo*, is preserved in the noun *noise* construed as the notional subject.

4.2.4. Pronominal subject

Pronominal subject was, on the whole, one of the less successful strategies (36% adequacy rating). The majority was applied to sentences with Czech null subject (50% of the pronominal subject cases, ex. (20)) where it involved adding a pronoun, but also to the subjectless sentences (29% of pronominal subject cases), postverbal subject (14%, ex. (22)) and relative pronoun (7%, see ex. (12)).

(20) [Ona] má takový krátký černý vlasy a hrozně hezky kreslí.

He has such a short black hair and draws very nicely.

With the null subject cases, the inadequacy is frequently caused by adding the wrong personal pronoun. Choosing the correct one would require chance or understanding the reference between entities spanning much larger distances across text, which so far seems impossible for MT.

(21) Zato bylo otevřenými dveřmi vidět daleko do krajiny, ... (C58)

On the other hand, through the open door, he could see far into the countryside, ...

(22) (Vtom spustí tovární píšťaly a sirény.)

(Then he starts the factory pipes and sirens.)

Example (22) would again require awareness of context for the adequate solution to be reached, as there are two possible interpretations of the sentence as such – the verb suffix -í can both refer to third person singular and plural. GT chose the wrong interpretation – it presumed that there was a null subject in the Czech sentence and that *tovární píšťaly a sirény* were the object instead of subject in postposition).

On the whole, it can be said that as pronominal subjects always refer to another entity which may not be present in the sentence or the clause, it is very easy to choose the inadequate translation without context.

4.2.5. Merging

Merging could be, strictly speaking, also classified as a subcategory of condensation, as, at the end of the process, two clauses are merged into one. However, it does not use methods generally associated with condensation to do so. Instead, it appears to be a result of the tendency of the GT algorithm to simply disregard parts of ST for which it cannot find a probable solution fitting into a fluent sentence in the target language. The resulting merging then does not try to mirror the examples of actual condensation in the training data, but instead ignores a problem and hopes for the best. It is mostly unsuccessful (83%), with only 2 examples judged as adequate, if awkwardly phrased.

(23) Člověkem z jiné látky, než jsme my. Slečno Gloryová, to byl ohromný okamžik.

Man from a substance *other than us*. Miss Glory, that was a great moment.

(24) ...to prej je nejlepší skladatel hudby ze všech, *co jsou na celým světě*, něco jako Bedřich Smetana, ale ještě mnohem lepší.

...he is the best music composer *of all the world*, something like Frederick Smetana, but much better.

Both of the Czech examples are postmodifying clauses.

However, the majority of the mergings were different. Some of them resulted in grammatical sentences with shifted or changed meaning, such as the following example,

(25) Co myslíte, *jaký dělník* je prakticky nejlepší?

What do you think is practically the best?

in which the subject *jaký dělník* is completely left out to make the first clause of the sentence the subject of the second clause. While the result is fluent, the meaning is very generalized. Although the specifics of the question may still be deduced from the context, this example has still been qualified as inadequate translation.

In ex. (26) of merging it also appears that the algorithm assumed that the identity of the subject remained constant throughout the sentence and did not change between clauses. This, however, is less frequent in Czech than in English.

(26) Prosím, má i typické chmýří blondýnek. Jen oči jsou drobátko – Ale zato
vlasý! Obráťte se, Sullo!
Please have a typical blonde fluff. Just the eyes are a crumb - But the hair!
Turn around, Sullo!

This is a tendency observed more often in the failed cases of condensation and the following strategy used by GT, the missing subject.

4.2.6. Missing subject

Missing subject is not very frequent as it only comprises 6% of the non-correspondences. It is similar to merging in that the same mechanism is employed by the GT algorithm to drop a problematic part, but the resulting clause then ends up incomplete and without any subject.

Four of these cases (67% of missing subjects, see (27)) actually result in awkward, but grammatical sentences in which the absence of a subject can be interpreted as a functional ellipsis used with multiple predication. They again point to the tendency of English to retain the same subject throughout the sentence or longer stretches of text. However, none of them were considered adequate as they failed to convey the same meaning as the ST (some to a lesser degree as ex. (28), some failed totally, as ex. (29)).

(27) Pak se to vypálí v peci a je socha.
Then it burns in the oven and *is a statue.*

While the resulting shift in meaning in ex. (24) does not pose a problem from the pragmatic point of view as *to* from the first clause and *socha* from the second clause refer to the same material entity in Czech, it is too great for fiction. There being two distinct subjects for the same thing, one before the process and the other after, highlights its transformative nature. The English equivalent loses this semantic detail.

Example (25) is completely inadequate from every point of view, as the resulting sentence does not make any sense. The null subject from the second clause referring to the object of the previous clause of ST is again interpreted instead as the subject being preserved from the first sentence.

(28) Zaviklal jsem kolíkem, a napůl se vylomil.

I screamed through the peg, and broke out halfway.

Example (16) can be hypothesized to result from the misidentification of the informal verb ending *-aj* in the third person as the imperative of second person singular. In that case, the resulting sentence is also awkward yet grammatical, but the meaning changes totally.

(29) *Taky si tam povídaj zajímavý věci a já myslím, že já budu sochař, až budu velká.*

Also, talk interesting things there and I think I'll be a sculptor when I'm big.

4.3. GT Strategies compared to the human translator

4.3.1. Difference overall

While the human translator is limited by the same grammatical means and thus uses similar strategies of translation on purely grammatical level, the execution itself frequently differs. (Although there were some grammatical types of strategies not found in GT's output, they form only 11% of the examples, see 4.3.2) The main characteristic of HT compared to GT is creativity.

Creative solutions belong either to Komissarov's category or situational equivalence, in which the situation described is the same, but it is looked at from another perspective (ex. (30)). These strategies do not work only with the words found in the ST, but with the picture of the situation the words create in the reader's mind. In other words, they rely on an actual understanding of the text and world knowledge, which is still out of reach of artificial intelligence.

(30)

CZ: ... jak z ní vyroste celý strom života, jak z ní budou vycházet všechna zvířata, počínajíc nějakým vířníkem a končíc – končíc samotným člověkem.

GT: and thinking of how the whole tree of life will grow out of it, all the animals coming out of it, starting with some rotifers and ending - ending with the man himself.

HT: Thinking about how it could grow out into an entire tree of life made of all the animals starting with a tiny coil of life and ending with . . . ending with man himself.

34% of excerpts of HT selected the same strategy as GT. Surprisingly, 30% of the subjects conducive to non-correspondent solutions in GT remained constant between ST and HT. This may suggest that GT overuses transformational strategies. However, preserving the ST sentence structure may not be a straightforward task, as explained in 4.3.2.

Same strategy as GT		34
	condensation	17
	Cz O --> S	8
	existential construction	3
	pronominal subject	4
	merging	1
	Cz Adv --> S	2
Correspondence with ST		30
	postverbal subject	13
	null subject	13
	regular subject	3
	relative pronoun	1
Different method		36
	ellipsis	7
	Cz O --> S	7
	pronominal subject	6
	Cz Adv --> S	3
	condensation	4
	existential construction	2
	cleft construction	2
	communicative translation	1
	inversion	1
	Cz Cs --> S	1
	CZ V --> S	1

Table 3: The overview of HT strategies

4.3.2. Cases solved with the same strategy as GT

Out of the cases where the strategy of HT and GT was the same (37% of all cases), 84% of GT's translations were considered accurate.

Even when the general type of the strategy coincided, there were often divergences in the “substrategy” used. However, this by itself did not necessarily lead to inadequacy (ex. (27), (28))

(31)

- CZ: Mohu si dovolit představit se tomuhle malému, šeříky obrostlému městečku tím, že hned první den navštívím venerologickou ordinaci?
- GT: Can I afford to imagine this small, lilac-grown town by visiting the vener's office the very first day?
- HT: Could I afford to begin my stay in this small, lilac-bowered town with a visit to the venereological ward of the local hospital on my very first day?

Some differences in the substrategies are more pronounced, as seen in ex. (27) in the previous section and some are relatively minor, as in ex. (28). Both GT and HT use the general strategy of condensation and both the solutions are considered adequate. GT condenses the adverbial clause of manner from ST by a gerund while HT uses the morphologically related noun *visit*.

The cases in which GT follows the general strategy of HT but the result is still judged inadequate (6 cases) include 3 cases of the pronominal subject inadequately selected by GT (see ex. (21)) and incorrect transformation of one sentence element to another (2 cases with the Czech object construed as the GT's subject, see ex. (17) and 1 with the Czech adverbial of place construed as GT's subject (ex. 32)

(32)

- CZ: Předkožkou, podobnou napučené pneumatice, proklouzávaly šedivé kapičky, popsané ve všech příručkách. (C65)
- GT: The puffy tire-like foreskin, slipped through the gray droplets described in all manuals.
- HT: and my foreskin resembled an over-inflated inner tube oozing grey drops—exactly as described in the clinical textbooks.

This case also lacks the necessary modulation of the rest of the sentence to be considered adequate. However, the modulation necessary is far beyond the skills of GT as it requires the actual understanding of the situation and cannot be simply achieved by switching around the lexis found in ST.

4.3.3. Cases with structure corresponding to ST

Correspondence with ST			
Czech subject type			
	HT strategy	Count	Percentage from subject type
postverbal subject		13	40%
	merging	4	
	existential construction	4	
	missing subject	1	
	Cz O --> S	1	
	pronominal subject	1	
	condensation	1	
	Cz Adv --> S	1	
null subject		13	35%
	Cz O --> S	5	
	condensation	4	
	missing subject	2	
	pronominal subject	1	
	merging	1	
regular subject		3	38%
	merging	2	
	condensation	1	
relative pronoun		1	34%
	condensation	1	

Tabulka 4: Overview of the cases corresponding with ST subject

30% of the excerpts solved in a non-corresponding manner by GT were actually solved by HT in a way corresponding to the source text. The two biggest groups with congruent HT counterparts were the postverbal subject (13% of all the excerpts and 40% of all the postverbal subject instances) and the null subject (also 13%, 35% of null subjects). Not

very surprisingly, the structure of Czech sentences featuring a regular subject was also frequently copied (38%), as well as that of sentences with relative pronouns (1% of the excerpts which is 34% of relative pronouns).

In these cases of the postverbal subject, the position of the subject in HT was not kept postverbal, but the identity of the subject remained constant. As the postverbal position of the subject is standard and neutral in Czech and the preverbal one in English, this would not be counted as a non-correspondence by the criteria of this thesis.

(33)

CZ: Taky tam chodily jiný děti, ...

GT: There were other children, ...

HT: Other kids went there too, ...

In this case, the preverbal position of the subject in HT does not change the its FSP role of the rheme, instead, the rhematicity of the subject is indicated by it not being a part of the context (see Dušková et al., 2012: 14.31). This is a more subtle strategy than the existential construction used by GT.

The cases of null subject correspondence are the adequately assigned pronouns expressing the null subject (ex. (34))

(34)

CZ: Lebek a koster se [já] bojím ze všeho nejvíc.

GT: Skulls and skeletons are most afraid of everything.

HT: I'm afraid of skulls and bones more than anything else.

4.3.4. Cases solved by a divergent strategy

Divergent strategies		33
GT strategy		
	HT strategy	
pronominal subject		8
	ellipsis	4
	inversion	1
	condensation	1
	Cz Cs --> S	1

	anticipatory it	1
	condensation	8
	Cz O --> S	4
	ellipsis	2
	existential construction	1
	Cz Adv --> S	1
	Cz O --> S	7
	pronominal subject	3
	anticipatory it	1
	existential construction	1
	communicative translation	1
	condensation	1
	existential construction	4
	pronominal subject	2
	Cz Adv --> S	1
	CZ V --> S	1
	missing subject	3
	condensation	1
	ellipsis	1
	Cz O --> S	1
	merging	2
	pronominal subject	1
	Cz Adv --> S	1
	CZ verb --> En S	1
	condensation	1

Table 5: Overview of cases solved by a divergent strategy

As is apparent from the table, cases in which the translation strategy differs both from the ST structure and GT's solution do not contain many large groups. Rather, there appear to be a lot of individual solutions.

Nevertheless, some tendencies can be identified – what is solved by GT by condensation is often solved in HT by transformation of ST object into subject (4 cases). This suggests that the strategy of condensation may be overused by GT.

Other minor group includes GT's pronominal subjects solved by ellipsis in HT. (Also 4 cases). However, this is hardly standard as 2 of the examples are stage directions where ellipsis is acceptable as opposed to normal language use and 1 case of interrupted dialogue.

The adequacy rate of GT's solutions divergent from HT was slightly lower than the overall adequacy rate but still surprisingly high (54%, while its overall adequacy is 60%). This is a proof that there is rarely only one right solution.

4.3.5. Strategies not found in GT

Purely grammatical categories of HT strategies not found in GT's output are a small group including ellipsis (which is essentially GT's category of missing subject, only successful, see ex. (27), 7%), cleft construction (ex. (28), 2%), inversion (ex. (), 1%) and communicative translation (ex. (), 1%). As is apparent, all of these are quite rare.

4.3.5.1. Ellipsis

Ellipsis was used in three types of cases. The less standard one includes stage directions occurring only in text A (2 % of the total excerpts) and situations when a character interrupts another in mid-sentence (1 %). The more frequent type occurs when there is coordination between two clauses (4%, ex. (35)).

(35)

CZ: Otočil jsem se zády k oltáři, abych se rozhlédl po zbývajících oknech.

GT: I turned my back to the altar to look around the remaining windows

HT: I turned my back to the altar and looked around at the remaining windows.

Only one case of HT ellipsis coincides with GT's use of missing subject (ex. (36)). GT's result was judged as inadequate because of the loss of the emphasis on transformation (see 4.2.6). HT is a solution on the level of situational equivalence – the same situation is looked at from different perspective and the emphasis on the transformative nature of the process is preserved by choice of a different predicate – *make them into statues*.

(36)

CZ: Pak se to vypálí v peci a je socha.

GT: Then it burns in the oven and is a statue.

HT: I fire things in the kiln and make them into statues

4.3.5.2. Cleft construction

The decision to use cleft construction also requires a comprehensive understanding of the situation described in the text, which GT is incapable of. In both the cases, ex. (37), the cleft construction is used to keep the FSP structure of the text together with the principle of end-weight – the longest and most complex phrase gravitates towards the end to make the sentence more readily understandable.

(37)

CZ: Bude skutečně v jejich silách pomlčet o tom, že nový profesor jejich al
mater se do Kostelce přijel léčit na kapavku?

GT: Will he really be silent about the fact that the new professor of their
alma mater came to Kostelec to heal for gonorrhoea?

HT: Was it realistic to hope they would keep quiet about a new teacher
who came all the way to Hronov seeking a cure for the dap?

4.3.5.3. Inversion

(38)

CZ: Trochu mě to vyvedlo z míry, ale pohled na úplně všemi mastmi maza
liščí tvář mě okamžitě poučil, že jde o provokaci.

GT: It made me a little bit annoyed, but the look of all the ointment-
smeared fox face taught me immediately that it was a provocation.

HT: ..., and at first I didn't know what to think. But one more look at that
sly, foxy face told me that here was another test of my endurance

This one specific case of inversion involves adverbial fronting (*here*). According to Dušková et al., it is often used in written language as stylistic means of continuous storytelling (Dušková et al., 2012: 14.37) In this case, both HT and GT solutions were accepted as adequate.

4.3.5.4. Communicative translation

(39)

CZ: Těší mne. Je vám asi hrozně smutno tak daleko od světa, vidíte?

GT: I'm pleased. You may feel so sad from the world, don't you?.

HT: Pleased to meet you. It must be very hard for you out here, cut off from the rest of the world.

Communicative translation involves the translation of a set phrase by another set phrase. While GT's solution is closer to the actual words of ST, HT offers a phrase which is actually used in these circumstances. However, GT's solution was judged as adequate as well, as the ease of language is not the main concern.

5. Conclusion

The aim of this thesis was to identify and describe differing traits of Czech and English subject influencing the strategy of translation employed by Google Translate and then to compare GT's strategies to strategies employed by human translators.

It was found that, in accordance to the prior hypothesis, the traits of Czech subject conducive to use of divergent strategy by GT included the null subject, the subjectless sentence and the postverbal subject. In addition to these, other types sometimes leading to non-correspondence were discovered: regular (preverbal) subject and relative pronoun, but those were comparatively minor.

In regard to the strategies used, the most frequent one was condensation (37%), which was not initially included in the hypothesis. Other strategies (construing a different clause element as the English subject, the existential construction and use of pronominal subject) were expected, as were the less suitable strategies (they could also be described as types of errors): missing subject (missing word being a frequent error type for NMT) and merging. The overall adequacy of the strategies was 60%, with the most successful being the existential construction (100% adequacy rate), condensation (78% adequacy) and construing the Czech object as English subject (68% adequacy). The least successful (100% inaccuracy) was the missing subject.

When comparing the strategies used by human translators, some additional strategies were found – ellipsis, cleft construction, inversion and communicative translation, although together they formed only 11% of the cases. GT used the same strategy as a human translator in 37% of the cases. In 30% of the cases, human translator preserved the syntactic structure of the source text instead. A completely different strategy was used in 32% of the cases, however, this included mainly individual solutions without any large subgroups, although one minor group could still be observed (using the object from the Czech clause as the subject of the English clause instead of condensation).

6. Resumé

Tato práce se zabývala rozdíly mezi českým a anglickým podmětem pozorovanými na strategiích překladu užitých Google Translate.

V teoretické části je popsána historie strojového překladu od jeho pravidly řízených počátků v padesátých letech, kdy se věřilo, že je možné jazyk popsat pouze jako soubor gramatických pravidel na jejichž základě je možné vygenerovat všechny možné věty v cílovém jazyce, po jeho statistické pokračování, které vyhodnocuje pravděpodobnost překladu každé překladové jednotky na základě trénovacích dat. Dřívější jednodušší technologie klasického statistického překladu, která brala v úvahu pouze předcházející překladové jednotky, byla nahrazena technologií statistického neuronového překladu, který je schopný brát v úvahu všechny jednotky překládané věty najednou. Neuronové překlady díky tomu působí plynuleji, ale jejich sémantická přesnost nedosáhla takového zlepšení, v jaké se doufalo. Slabou stránkou neuronového překladu je hlavně vynechávání částí věty, pro které nenalezl statisticky pravděpodobné řešení.

Dále je v teoretické části popsány Vinayovy a Darbenetovy strategie lidského překladu a Komissarovy roviny ekvivalence. Ekvivalence buď mohou být na rovině pragmatické, ve které se slova užitá ve zdrojovém jazyce vůbec neshodují se slovy cílového jazyka, ale pragmatický účinek je stejný, na rovině situace, ve které se popsaná situace shoduje, jen je nazírána z jiného úhlu, nebo na dalších třech rovinách z nichž finální je rovina znaku – doslovný překlad. Předpokládalo se, že první dvě roviny budou Překladačem Google využívány méně a celková úroveň korespondence bude častěji vyšší – bude dosahovat úrovně znaku.

Dále jsou v teoretické rovině popsány typologické rozdíly mezi češtinou, flektivním jazykem, který využívá sufixů k vyjádření reference a má volný slovosled, a angličtinou, která patří k analytickým jazykům s pevným slovosledem. Tyto rozdíly mají také za výsledek jiné formy a role podmětu – v češtině standardně dochází k výskytu nevyjádřeného podmětu, který je obsažen ve verbálních sufixech, angličtina podmět může postrádat pouze v rozkazovacím způsobu a při koordinaci. Anglický podmět je standardně vždy na začátku věty (inverze je bezpříznaková pouze v tázacích větách) a tudíž častěji zaujímá tematickou roli, zatímco český podmět se může bezpříznakově vyskytovat i v postverbální pozici. V případech, kdy není znám činitel, má čeština možnost využít jednočlenné věty, ve které žádný podmět není, zatímco angličtina musí v těchto případech využít existenciální konstrukce, „prázdného“ *it.*“

Data k analýze byla čerpána z dramatu *R. U. R.* Karla Čapka a tří románů - *European* Patrika Ouředníka, *Mirácku* Josefa Škvoreckého a *Hrdého Budžese* Ireny Douskové. Tyto romány byly postupně překládány Překladačem Google, než bylo vyňato 100 příkladů nekorespondenčního řešení podmětu. Tyto příklady pak byly porovnány s jejich anglickým překladem.

České typy podmětu, o kterých se předpokládalo, že budou způsobovat nekorespondenční řešení, zahrnovaly postverbální podmět, nevyjádřený podmět a věty jednočlenné (bezpodmětné).

V analytické části se potvrdila hlavní role těchto typů podmětu. Strategie využívané Překladačem Google oproti očekávání zahrnovaly kondenzaci, jinak odpovídaly odhadům – přetvoření jiného českého větného členu na anglický subjekt (nejčastěji předmětu, méně častěji příslovecného určení a přísudku) a existenciální konstrukce. Překladač často chyboval vynecháním podmětu, využitím neodpovídajícího pronominálního podmětu v případě českého nevyjádřeného podmětu a slučováním klauzí souvětí dohromady.

Co se týče porovnání s lidskými překladači, ve 34% případů využíval překladač Google stejné strategie jako lidský překladač. Lidský překladač ponechal větné členy ve stejné struktuře jako ve zdrojovém textu ve 30% případů. Rozdílná metoda byla použita v 36% případů, ale nebyly nalezeny žádné významné podskupiny, spíše individuální řešení.

7. Sources and literature:

Source texts:

- A. Čapek, Karel. R. U. R.: rossum's universal robots: kolektivní drama o vstupní komedii a třech dějstvích. Praha: Fr. Borový, 1940.
- B. Ouředník, Patrik (2012) *Europeana: stručné dějiny dvacátého věku*. Praha: Volvox Globator
- C. Škvorecký, Josef (1997) *Mirákl: politická detektivka*. Praha: Ivo Železný.
- D. Dousková, Irena (2002) *Hrdý Budžes*. Brno: Petrov.

Translations:

- Čapek, Karel., Translated by Wyllie, David (2006) *R. U. R.: Rossum's universal robots: A play in introductory scene and three acts*. Adelaide: eBooks@Adelaide. [online]. URL: <https://ebooks.adelaide.edu.au/c/capek/karel/rur/> [cit. 2019-05-08]
- Dousková, Irena. Translated by Clarke, Melvin (2016) *B. Proudew*. Brno: Pálava Publishing.
- Ouředník, Patrik. Translated by Turner, Gerald (2005) *Europeana: A Brief History of the Twentieth Century*. Champaign: /Dalkey Archive Press.
- Škvorecký, Josef. Translated by Wilson, Paul (1992) *The Miracle Game*. New York: W.W. Norton

Literature:

- Bojar, Ondřej (2011) „Analyzing Error Types in English-Czech Machine Translation.“ *The Prague Bulletin of Mathematical Linguistics* 95, April 2011. Pp 63–76.
- Bojar, Ondřej (2017) “Úvod do neuronového strojového překladu nejen pro korpusové lingvisty.” 2017-11-21, Filozofická fakulta, Ústav českého národního korpusu. Lecture.
- Bojar, Ondřej (2017) *English-to-Czech MT: Large Data and Beyond*. Habilitation thesis. Charles University, Faculty of Mathematics and Physics. [online]. URL: <http://ufallab.ms.mff.cuni.cz/~bojar/hab/bojar-habilitation-text.pdf>. [cit 2019-02-22]
- Čermák, František (2011) *Jazyk a jazykověda: přehled a slovníky*. Vyd. 4., v Karolinu 2., doplněné. Praha: Karolinum. Print.
- Čermák, František (2014) *Jazyk a slovník: vybrané lingvistické studie*. Praha: Karolinum. Print.
- Cho, Kyunghyun et al. (2014) “On the Properties of Neural Machine Translation: Encoder–Decoder Approaches“ [online]. [cit. 2018-01-15].
- Daneš, František, Miroslav Grepl, and Zdeněk Hlavsa (1987) *Mluvnice češtiny - 3, Skladba*. Praha: Academia. Print.
- Dušková, Libuše (1975) “English Counterparts of Czech Verbal Subjectless sentences.“ *Philologica Pragensia* 18: 199-221. Reprinted in Dušková, Libuše (1999) *Studies in the English Language*. Part 2. Prague: Karolinum - Charles University Press. Chapter 26.

- Dušková, Libuše (1975) „On some semantic and FSP aspects of the subject in Czech and in English.“ *Philologica Pragensia* 18: 199-221. Reprinted in Dušková, Libuše (1999) *Studies in the English Language*. Part 2. Prague: Karolinum - Charles University Press. Chapter 32.
- Dušková, Libuše (2002) “Constancy of syntactic function across languages.“ In Hladký, Josef (ed.) *Language and Function: To the Memory of Jan Firbas*. Studies in Functional and Structural Linguistics 49. Amsterdam/ Philadelphia: John Benjamins. Pp 127-144.
- Dušková, Libuše (2015) “Constancy of the syntactic and FSP function of the subject.“ In *From Syntax to Text: The Janus Face of Functional Sentence Perspective*. Praha: Karolinum. Pp 14-29.
- Dušková, Libuše et al., (2012), original translation. *Mluvnice současné angličtiny na pozadí češtiny* [online]. URL: <http://emsa.ff.cuni.cz/>. Ústav anglického jazyka a didaktiky, FF UK.
- Goldberg, Yoav (2017) *Neural Network Methods for Natural Language Processing*. London: Morgan and Claypool.
- Hutchins, W. John, ed., (2000) “Early years in machine translation: memoirs and biographies of pioneers“ [online]. Amsterdam/Philadelphia: John Benjamins B.V [cit. 2018-01-15].
- Isabelle, Pierre and George Foster (2006) “Machine Translation: Overview” In: Brown, Keith, and Anne Anderson (2006) *Encyclopedia of language & linguistics*. 2nd ed. Amsterdam: Elsevier. Print.
- Jelínek, Milan (2017) “Kondenzace.” In: Petr Karlík, Marek Nekula, Jana Pleskalová (eds.), *CzechEncy - Nový encyklopedický slovník češtiny* [online]. URL: <https://www.czechency.org/slovník/KONDENZACE>. [cit. 2019-05-13].
- Knittlová, Dagmar, Bronislava Grygová, and Jitka Zehnalová (2010) *Překlad a překládání*. Olomouc: Univerzita Palackého v Olomouci.
- Mathesius, Vilém (1947) „0 funkci podmětu.“ In: *Čeština a obecný jazykozpyt*. Praha: Melantrich. Pp 277-285.
- Mikolov, T & Yih, W.-T & Zweig, G. (2013). “Linguistic regularities in continuous space word representations.” *Proceedings of NAACL-HLT*. 746-751. [online]. URL: <https://www.aclweb.org/anthology/N13-1090> [cit. 2019-02-21]
- Quirk, Randolph et al. *A Comprehensive Grammar of the English Language*. London: Longman. 1985.
- Turovsky, Barak (2016) “Found in translation: More accurate, fluent sentences in Google Translate” [online]. URL: <https://blog.google/products/translate/found-translation-more-accurate-fluent-sentences-google-translate/>. [cit. 2019-05-24]
- Vilar, David, Jia Xu, Luis Fernando D’ Haro, and Hermann Ney (2006) „Error Analysis of Machine Translation Output.“ In: *International Conference on Language Resources and Evaluation*, Pp 697-702, Genoa, Italy.
- Vinay, Jean-Paul, Jean Darbelnet (1995). Translation by Juan C. Sager and M.-J. Hamel. *Comparative Stylistics of French and English*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Wilks, Yorick (2006) “Computational Linguistics: History” In: Brown, Keith, and Anne Anderson (2006) *Encyclopedia of language & linguistics*. 2nd ed. Amsterdam: Elsevier. Print.
- Wu, Yonghui, et al. (2016) “Google’s Neural Machine Translation System: Bridging the Gap between Human and Machine Translation” [online]. [cit. 2018-01-15].

8. Appendix

Text	No.	Source text	GT output	Human Translation
A	1	V kontrastu k této výzdobě stěn je na zemi nádherný turecký koberec, vpravo kulatý stůl, pohovka, kožená klubovní křesla a knihovna, v níž místo knih stojí láhve s vínem a kořalkami.	In contrast to this wall decoration, there is a beautiful Turkish carpet on the ground, a round table on the right, a sofa, leather club chairs and a bookcase with wine and spirits bottles instead of books.	In contrast with the content of the walls, the floor is covered with a magnificent Turkish carpet, stage right is a round armchair, settee, sumptuous leather armchair, bookshelves containing not books but bottles of wines and spirits.
A	2	DOMIN: Čertova výroba! Ale zajisté, slečno Gloryová, všechno uvidíte. Prosím posaďte se. Zajímala by vás historie vynálezu?	DOMIN: Devil's production! But certainly, Miss Glory, you will see everything. Please sit down. Are you interested in the history of the invention?	Domin: Ah, the damned factory! But of course, Miss Glory, you can see everything. Do please sit down. Would you be interested in hearing the history of our invention?
A	3	DOMIN: Tak tedy. (Sedne si na psací stůl, pozoruje Helenu uchvácen a odříkává rychle:)...	DOMIN: Well, then. (She sits down on her desk, watches Helen captivated and chants fast :)	Domin: Well this is what happened. (sits at desk, seems captivated by Helena and speaks quickly)
A	4	DOMIN (slavnostně): A tehdy, slečno, starý Rossum napsal mezi své chemické vzorce tohleto: [...]	DOMIN (solemnly) And then, Miss Rossum wrote among his chemical formulas, ...	Domin:(triumphant) And then, Miss Glory, this is what he wrote down in his chemical notes:
A	5	...Je však jiný způsob, jednodušší, tvárnější a rychlejší, na nějž příroda vůbec nenarazila...	But there is another way, simpler, more dexterous, and faster, which nature has never encountered.	There is however another way which is simpler, easier to mould, and quicker to produce than Nature ever stumbled across.
A	6	[...] Tuto druhou cestu, po které se mohl brát vývoj života, jsem dnešního dne objevil.“ Představte si, slečno, že tahle veliká slova [on] psal nad chrchlem jakéhosi koloidálního rosolu, který by ani pes nesežral. [...]	I have discovered this second way in which the evolution of life could be taken today. "Imagine, Miss, this great word was written over the sputum of some colloidal jelly that even a dog would not eat. ...	think: he wrote these words about a blob of some kind of colloidal jelly that not even a dog would eat."

A	7	Představte si ho, že sedí nad zkumavkou a myslí na to,...	Imagine him sitting over a tube and thinking...	Imagine him sitting with a test tube and thinking
A	8	jak z ní vyroste celý strom života, jak z ní budou vycházet všechna zvířata, počínajíc nějakým vířníkem a končíc – končíc samotným člověkem. jak z ní vyroste celý strom života, jak z ní budou vycházet všechna zvířata, počínajíc nějakým vířníkem a končíc – končíc samotným člověkem.	and thinking of how the whole tree of life will grow out of it, all the animals coming out of it, starting with some rotifers and ending - ending with the man himself.	Thinking about how it could grow out into an entire tree of life made of all the animals starting with a tiny coil of life and ending with . . . ending with man himself.
A	9	Člověkem z jiné látky, než jsme my. Slečno Gloryová, to byl ohromný okamžik.	Man from a substance other than us. Miss Glory, that was a great moment.	Man made of different material than we are
A	10	DOMIN: Škoda. (Seskočí ze stolu a usedne vedle Heleny.) Ale víte, co v čítankách není? (Ťuká si na čelo.) Že byl starý Rossum úžasný blázen. Vážně, slečno Gloryová, ale tohle nechte pro sebe. Ten starý výstředník chtěl opravdu dělat lidi.	DOMIN: Too bad. (She jumps off the table and sits next to Helena.) But do you know what's not in the reading room? (She taps her forehead.) That old Rossum was an amazing fool. Seriously, Miss Glory, but leave this to yourself. The old eccentric really wanted to make people.	Domin: That is a pity. (jumps off desk and sits beside Helena) But do you know what's not in all the papers? (taps his forehead) That old Rossum was completely mad. Seriously. But keep that to yourself. He was quite mad. He seriously wanted to make a human being.
A	11	DOMIN: Přibližně, slečno Heleno. Ale starý Rossum to mínil doslovně. Víte, chtěl jaksi vědecky sesadit boha. Byl strašlivý materialista, a proto to všechno dělal. Nešlo mu o nic víc než podat důkaz, ...	DOMIN: Approximately, Miss Helen. But old Rossum meant it literally. You know, he wanted to scientifically depose God. He was a terrible materialist, and he did it all. He was no more than giving evidence that	Something like that, yes, but old Rossum meant it entirely literally. He wanted, in some scientific way, to take the place of God. He was a convinced materialist, and that's why he wanted to do everything simply to prove

A	12	... že nebylo žádného pánaboha zapotřebí. Proto si uťmanul udělat člověka navlas, jako jsme my. Znáte trochu anatomii?	... that no god was needed. That's why he made himself make a man like us. Do you know a little anatomy?	that there was no God needed. That's how he had had the idea of making a human being, just like you or me down to the smallest hair. Do you know anything about anatomy, Miss Glory?
A	13	DOMIN: Já také. Představte si, že si [on] vzal do hlavy vyrobit všechno do poslední žlázy jako v lidském těle. Slepé střevo, mandle, pupek, samé zbytečnosti. Dokonce i – hm – i pohlavní žlázy.	DOMIN: Me too. Imagine taking a head to make everything into the last gland as in a human body. Blindness, almonds, navel, all useless. Even i - um - even gonads.	Domin:No, nor do I. But just think of how old Rossum got it into his head to make everything, every gland, every organ, just as they are in the human body. The Appendix. The tonsils. The belly-button. Even the things with no function and even, er, even the sexual organs.
A	14	DOMIN: – nejsou zbytečné, já vím. Ale mají-li se lidé uměle vyrábět, pak není – hm – nijak třeba –	DOMIN: - They're not useless, I know. But if people are to artificially produce, then there is no need -	Domin:They do have a function, I realise that. But if people are going to be made artificially then, er, then there's not really much need for them.

A	15	<p>DOMIN: Ukážu vám v muzeu, co spackal za deset let dohromady. Měl to být muž, žilo to celé tři dny. Starý Rossum neměl trochy vkusu. Bylo to hrozné, co udělal. Ale mělo to uvnitř všechno, co má člověk. Skutečně, úžasně piplavá práce. A tehdy sem přišel inženýr Rossum, synovec starého. Geniální hlava, slečno Gloryová. Jak uviděl, co tropí starý, řekl: „To je nesmysl, vyrábět člověka deset let. Nebudeš-li ho vyrábět rychleji než příroda, pak na ten celý krám nakašlat.“ A pustil se sám do anatomie.</p>	<p>DOMIN: I'll show you at the museum what he's been messing with in ten years. It was supposed to be a man, lived for three days. Old Rossum had little taste. It was terrible what he did. But it had everything inside the man. Really, amazingly hard work. And that was when Rossum, an old nephew, came here. Brilliant head, Miss Glory. As he saw the old man, he said, "That's nonsense to make a man for ten years." If you don't make it faster than nature, then cough the whole thing. "And he went into the anatomy himself.</p>	<p>Domin: In the museum I'll show you the monstrosity he created over the ten years he was working. It was supposed to be a man, but it lived for a total of three days. Old Rossum had no taste whatsoever. This thing is horrible, just horrible what he did. But on the inside it's got all the things that a man's supposed to have. Really! The detail of the work is quite amazing. And then Rossum's nephew came out here. Now this man, Miss Glory, he was a genius. As soon as he saw what the old man was doing he said, 'This is ridiculous, to spend ten years making a man; if you can't do it quicker than Nature then you might as well give up on it'. And then he began to study anatomy himself.</p>
A	16	<p>Starý Rossum ho doslovně proklel a do své smrti usmolil ještě dvě fyziologické obludy, až ho nakonec našli v laboratoři mrtvého. To je celá historie.</p>	<p>Old Rossum cursed him literally, and he killed two more physiological monsters until his death, until he was found dead in the lab. That's history.</p>	<p>Old Rossum literally cursed him. He carried on in his laboratory, producing two more physiological monstrosities, until one day they found him there dead. And that's the whole story.</p>

A	17	DOMIN: Mladý Rossum, slečno, to byl nový věk. Věk výroby po věku poznání. Když si okoukl anatomii člověka, viděl hned, že je to příliš složité a že by to dobrý inženýr udělal jednodušeji. Začal tedy předělávat anatomii a zkoušel, co se dá vynechat nebo zjednodušit – Zkrátka – slečno Gloryová, nenudí vás to?	DOMIN: Young Rossum, miss, that was a new age. Age of production after age of knowledge. Seeing the human anatomy, he saw that it was too complicated and that a good engineer would make it easier. So he began to remodel the anatomy and tried what could be omitted or simplified - In short - Miss Glory, wasn't it bored?	Domin: Ah now, young Rossum; that was the start of a new age. After the age of research came the age of production. He took a good look at the human body and he saw straight away that it was much too complicated, any good engineer would design it much more simply. So he began to re-design the whole anatomy, seeing what he could leave out or simplify. In short, Miss Glory . . . I'm not boring you, am I?
A	18	DOMIN: Počkejte. Které jsou zbytečné, když [stroj] má třeba tkát nebo sčítat. Naftový motor nemá mít třapce a ornamenty, slečno Gloryová.	DOMIN: Wait. Which is unnecessary when you need to weave or add up. The diesel engine is not supposed to have cones and ornaments, Miss Glory.	Domin: No, wait. Which are simply not needed for activities such as weaving or calculating. A petrol engine doesn't have any ornaments or tassels on it,
A	19	...Výroba má být co nejjednodušší a výrobek prakticky nejlepší. Co myslíte, jaký dělník je prakticky nejlepší?	Production should be as simple as possible and the product is the best. What do you think is practically the best?	and making an artificial worker is just like making a petrol engine. The simpler you make production the better you make the product. What sort of worker do you think is the best?
A	20	DOMIN: Ano. Z ničeho nic jim praskla noha nebo něco. Naše planeta je patrně trochu malá pro obry. Ted' děláme jen Roboty přirozené velikosti a velmi slušné lidské úpravy.	DOMIN: Yes. Nothing suddenly burst their feet or something. Our planet is probably a little small for giants. Now we are only doing the Robots of natural size and very decent human adjustment.	Domin: Yes. All of a sudden, for no reason, a leg or an arm would break. This planet just seems too small for monsters like that. So now we just make them normal size and normal proportions.
A	21	HELENA: Viděla jsem první Roboty u nás. Obec je koupila... chci říci vzala do práce – DOMIN: Koupila, drahá slečno. Roboti se kupují.	HELENA: I saw the first Robots here. The village bought them ... I want to say to work - DOMIN: She bought, dear Miss. Robots are buying.	Helena: I saw my first robot in our village. They'd bought him so that. . . . that's to say they'd employed him to . . . Domin: Bought it, Miss Glory. Robots are bought and sold.

A	22	HELENA (vstane a podává jí ruku): Těší mne. Je vám asi hrozně smutno tak daleko od světa, vidíte?	HELENA (gets up and gives her a hand): I'm pleased. You may feel so sad from the world, don't you?	Helena:(stands and offers her hand) Pleased to meet you. It must be very hard for you out here, cut off from the rest of the world.
A	23	DOMIN: Nepoznala byste, že je z jiné látky než my. Prosím, má i typické chmýří blondýnek. Jen oči jsou drobátko – Ale zato vlasy! Obrat' se, Sullo!	DOMIN: You wouldn't know he's from a substance other than us. Please have a typical blonde fluff. Just the eyes are a crumb - But the hair! Turn around, Sullo!	Domin:It feels just the same as human skin. Sulla even has the sort of down on her face that you'd expect on a blonde. Perhaps her eyes are a bit small, but look at that hair. Turn around, Sulla.
A	24	HELENA (obejme Sullo): Nebojte se, Sullo, já vás nedám! Řekněte, drahoušku, jsou k vám všichni tak suroví? To si nesmíte dát líbit, slyšíte? Nesmíte, Sullo!	HELENA (hugs Sulla): Don't worry, Sullo, I won't give you! Say, sweetheart, are all of you so raw? You must not like it, do you hear? You can't, Sullo!	Helena:(arms around Sulla) Don't worry, Sulla, I won't let them take you. Do they always treat you like this? You shouldn't put up with it, do you hear, you shouldn't put up with it.
A	25	(Vtom spustí tovární píšťaly a sirény.)	(Then he starts the factory pipes and sirens.)	(just then, factory whistles and sirens sound)
B	26	Američané, kteří v roce 1944 padli v Normandii, byli urostlí chlapani a měřili v průměru 173 cm, a kdyby se nakladli jeden za druhého, chodidla k temenům hlav, měřili by [oni] dohromady 38 kilometrů.	The Americans who died in Normandy in 1944 were tall boys and measured an average of 173 centimeters, and if they were stacked one after the other, their feet would be 38 kilometers in total.	The Americans who fell in Normandy in 1944 were tall men measuring 173 centimeters on average, and if they were laid head to foot they would measure 38 kilometers
B	27	Němci byli také urostlí chlapani, a nejurostlejší ze všech byli senegalští střelci v první světové válce, ti měřili 176 cm, a tak je [oni] posílali do prvních řad...	The Germans were also tall boys, and Senegalese shooters in World War I were the oldest of all, and they were 176 cm tall, sending them to the front row...	The Germans were tall too, while the tallest of all were the Senegalese fusiliers in the First World War who measured 176 centimeters, and so they were sent into battle on the front line
B	28	...aby se jich Němci polekali.	... to scare the Germans.	in order to scare the Germans.
B	29	O první světové válce se říkalo,...	The First World War was said...	It was said of the First World War
B	30	O první světové válce se říkalo, že v ní lidé padali jako semena, ...	to have fallen as seeds in it, ...	that people in it fell like seeds

B	31	A Angličané vymysleli tanky a Němci plyn, kterému se říkalo yperit,	And the English invented the tanks and the Germans the gas called mustard ...	And the English invented the tank and the Germans invented gas, which was known as yperite because the Germans first used it near the town of Ypres
B	32	a také se mu říkalo hořčice, protože štípal do nosu jako hořčice z Dijonu, a to prý byla pravda, a někteří vojáci, kteří se po válce vrátili domů, už nikdy nechtěli jíst dijonskou hořčici.	and it was also called mustard, because it pinched his nose like Dijon mustard, and that was true, and some of the soldiers who returned home after the war never wanted	and it was also called mustard because it stung the nose like Dijon mustard, and that was apparently true, and some soldiers who returned home after the war did not want to eat Dijon mustard again.
B	33	A chudí lidé se těšili, že pojedou vlakem,...	And the poor people were looking forward to take the train,...	And poor people looked forward to riding in a train
B	34	a venkované se těšili, že poznají velká města a že zatelefonují na okresní poštu a nadiktují telegram manželce MÁM SE DOBŘE, DOUFÁM, ŽE SE MÁŠ TAKÉ DOBŘE.	and the villagers were looking forward to meeting the big cities and calling the county post and dictating a telegram to my wife I HAVE GOOD, Hope YOU HAVE ALSO GOOD.	country folk looked forward to seeing big cities and phoning the district post office to dictate a telegram to their wives, I'M FINE, HOPE YOU ARE TOO.
B	35	Generálové se těšili, že o nich bude v novinách,	The generals were looking forward to being in the newspapers,...	The generals looked forward to being in the newspapers,
B	36	...a lidé z národnostních menšin měli radost, že budou sdílet válku s těmi, kteří mluví bez přízvuku, a že s nimi budou zpívat pochodové zpěvy a veselé odrhovačky.	...and the people of national minorities were delighted to share the war with those who had no accent and to sing with them marching songs and cheerful bouncers.	, and people from national minorities were pleased that they would be sharing the war with people who spoke without an accent and that they would be singing marching songs and jolly popular ditties with them.
B	37	a ve které létaly vzducholodě a aeroplány a bombardovaly zázemí a města a civilní obyvatelstvo a ponorky potápěly lodě a děla střílela padáčkem deset nebo dvanáct kilometrů daleko.	... and [so many people] flew airplanes and bombed the background and cities and the civilian population and submarines sank ships and cannons fired a parachute ten or twelve kilometers away.	and in which airships and airplanes flew and bombarded the rear and towns and civilians, and submarines sunk ships and artillery could lob shells ten or twelve kilometers.

B	38	A když senegalští střelci viděli poprvé letadlo, mysleli, že je to ochočený pták, a jeden senegalský voják vyřezával z mrtvých koní kusy masa a házel je co nejdál od sebe, aby letadla odlákal.	And when Senegalese gunners saw the plane for the first time, they thought it was a tame bird, and one Senegalese soldier carved pieces of meat from dead horses and tossed them away from each other to lure the planes away.	And when the Senegalese fusiliers first saw an airplane they thought it was a tame bird and one of the Senegalese soldiers cut a lump of flesh from a dead horse and threw it as far as he could in order to lure it away.
B	39	A vojáci nosili zelené a maskovací uniformy, protože nechtěli, aby je nepřítel viděl,...	And the soldiers wore green and camouflage uniforms because they didn't want the enemy to see them,	And the soldiers wore green and camouflage uniforms because they did not want the enemy to see them
B	40	...což bylo v té době moderní, protože v předchozích válkách nosili vojáci hodně barevné uniformy, aby byli vidět zdaleka.	...which was modern at the time, because in the previous wars the soldiers wore a lot of colorful uniforms to be seen from afar.	, which was modern at the time because in previous wars soldiers had worn brightly-colored uniforms in order to be visible from afar
B	41	A vojáci nosili na krku nebo na zápěstí známku se svým jménem a číslem pluku, aby se vědělo, kdo je kdo a kam poslat soustrastný telegram,	And the soldiers wore a sign with their name and regiment number on their neck or wrist, to know who was who and where to send a condolence telegram	And the soldiers wore around their neck or wrist a tag with their name and the number of their regiment to indicate who was who, and where to send a telegram of condolences
B	42	ale když jim výbuch utrhl hlavu nebo ruku a známka se ztratila, vojenský štáb vyhlásil, že to byli neznámí vojíni, a ve většině hlavních měst jim zřídili věčný plamen,	but when the explosion plucked their head or hand and the stamp was lost, the military staff said it was unknown Private soldiers, and in most capitals set up an eternal flame	, but if the explosion tore off their head or arm and the tag was lost, the military command would announce that they were unknown soldiers, and in most capital cities they instituted an eternal flame preserves the memory of something long past.
B	43	aby se na ně nezapomnělo, protože oheň udržuje paměť něčeho dávného.	... to keep them from forgetting, because the fire keeps the memory of something ancient.	lest they be forgotten, because fire preserves the memory of something long past.

B	44	A padlých Francouzů bylo celkem 2 681 km a padlých Angličanů 1 547 km a padlých Němců 3 010 km při průměrné délce jedné mrtvoly 172 cm.	And the fallen French were a total of 2,681 km and the fallen English were 1,547 km and the fallen Germans 3,010 km with an average length of 172 cm.	And the fallen French measured 2,681 kilometers, the fallen English, 1,547 kilometers, and the fallen Germans, 3,010 kilometers, taking the average length of a corpse as 172 centimeters.
B	45	A dohromady na celém světě bylo padlých vojáků 15 508 km.	And all over the world, fallen soldiers were 15,508 km.	And a total of 15,508 kilometers of soldiers fell worldwide.
B	46	...ale epidemiologové říkali, že chřipka zabila víc lidí v zemích, kde nebyla válka, na tichomořských ostrovech, v Indii nebo Spojených státech, a anarchisté říkali, že je to tak v pořádku, protože svět je prohnílý a spěje ku zkáze.	...but epidemiologists said that the flu killed more people in countries where there was no war, the Pacific islands, India or the United States. and the anarchists said it was all right, because the world is rotten and goes to ruin.	but the epidemiologists said that the disease killed more people in countries where there was no war, such as in Oceania, India or the United States, and the Anarchists said that it was a good thing because the world was corrupt and heading for destruction.
B	47	Ale jiní historikové říkali, že dvacáté století začalo ve skutečnosti dřív, že na jeho počátku stála průmyslová revoluce, která narušila tradiční svět, a že za všechno můžou lokomotivy a parníky.	But other historians have said that the twentieth century had actually begun before, that there was an industrial revolution at its beginning, which broke the traditional world, and that locomotives and steamers could be worth it.	But other historians said that the twentieth century actually started earlier, that it began with the industrial revolution that disrupted the traditional world and that all this was the fault of locomotives and steamships
B	48	A v roce 1918 se rozšířila po celém světě chřipka, které se říkalo španělská,	And in 1918, the flu known as the Spanish spread throughout the world.	And in 1918 an influenza known as Spanish Flu spread throughout the world
B	49	a zabila přes dvacet miliónů lidí.	killing over twenty million people.	killing over twenty million people.

B	50	A Rusové říkali, že celá Evropa je v úpadku a že katolíci a protestanti Evropu úplně zkazili, a navrhovali, že vyženou Turky z Cařihradu a pak že připojí Evropu k Rusku, aby byla zachráněna víra.	And the Russians said the whole of Europe was in decline and that the Catholics and Protestants had completely spoiled Europe, and suggested they drive the Turks out of Constantinople and then join Europe to Russia to save the faith.	And the Russians said that the whole of Europe was decadent and that the Catholics and Protestants had completely ruined Europe and they proposed to throw the Turks out of Constantinople and then annex Europe to Russia so as to preserve the faith.
C	51	Dole z něho trčel nahrubo opracovaný kolík a v podstavci, mnohem pozdějším a konvenčně přezdobeném, byl otvor, zúžený zašlým těsněním, aby odpovídal průměru čepu.	A roughly polished pin protruded from it, and in the base, much later and conventionally trimmed, the hole tapered through the faded seal to match the pin diameter.	There was a crudely fashioned peg protruding from the bottom of the statue, and in the pedestal—which was much newer, and covered to excess with the conventional ornamentation of a later time—was a hole, partially stuffed with worn-out packing so the peg in the statue would fit it snugly.
C	52	Sošku samu poznamenal -- rukou pana Řerichy -- obrodný proces, a znamení dosud odolávala času i jeho staronovým osvícením.	The statue itself - Mr. Rericha's - revived the process, and the signs still resisted the time and its old-age enlightenment.	The statue itself had a new coat of paint applied by Mr. Rericha's hand—a mark of the political revival of 1968—and so far it had managed to with-stand the subsequent reversal of fortune as an earlier version of enlighten-ment once again established its sway
C	53	Jak jsem mohl tušit, že právě jeho si Pán Bůh vybere, aby na něm několika kosteleckým babičkám demonstroval, jak podivné jsou Jeho cesty.	How could I have imagined that his Lord would choose his Lord to demonstrate to him several church grandmothers how strange His ways were.	How could I know the Lord would choose him, of all His servants, to show a handful of Hronov grannies the mystery of His ways?

C	54	Jejich zkřivené tváře a ručičky vztažené k šikmému světcí měly patrně naznačit extatickou zbožnost: v rámci celkové kompozice daly se však také vyložit jako přirozená reakce lidí, jimž padá na hlavu kolos rhodský.	Their distorted faces and hands relative to the oblique saint were likely to indicate ecstatic piety: but within the overall composition they could also be interpreted as the natural reaction of the people falling upon the head of the Rhodes.	Their twisted faces and their tiny hands held up in supplication were presumably meant to suggest religious ecstasy; unfortunately, given the composition, their pose looked more like the natural reaction of people upon whose heads the Colossus of Rhodes was about to topple.
C	55	Nechtěnou aktuálnost nerušila ani dosti středověká písmena, jež z pekelně černého pozadí ohnivě hlásala:	Unwanted topicality did not disturb even the rather medieval letters, which from the hellish black background were praising fire:	Such unintentional up-to-dateness was not diminished in the least by the fiery declaration written in would-be medieval lettering against a background of hellish black:
C	56	Otočil jsem se zády k oltáři, abych se rozhlédl po zbývajících oknech.	I turned my back to the altar to look around the remaining windows.	I turned my back to the altar and looked around at the remaining windows
C	57	Na ně však už panu Řeřichovi nezbyl čas.	But there was no time left for Mr. Rerich.	but time had run out on Mr. Rericha.
C	58	Zato bylo otevřenými dveřmi vidět daleko do krajiny, na šťavnaté louky a temný les, jenž odděloval Písečnici od Kostelce a světa.	On the other hand, through the open door, he could see far into the countryside, the succulent meadows and the dark forest that separated the Sandbox from Kostelec and the world.	Through the open door I could look far out into the countryside and see the succulent meadows and dark forest that separated Písečnice from Hronov and the world.
C	59	Ušklíbl jsem se, zvedl jsem svatého Josefa k jeho podstavci, a vtom, jak byl hladce natřený, vyklouzl mi z ruky a spadl na podlahu.	I grimaced, picking up Saint Joseph's pedestal, and slipping out of my hand as he smoothly painted and dropping to the floor.	I frowned and tilted Saint Joseph up to put him back on his pedestal, but with his glossy coat of fresh paint he slipped from my hands and fell to the floor.
C	60	Něžně jsem se sehnul k tomu divotvorci, jenž v ústraní kaple pod Kobylí hlavou přežil víc než dvě století, a poznal jsem, že je konečně po něm.	I tenderly bent over to the whimper who had been in the seclusion of the chapel below Kobyle's head for more than two centuries, and I knew he was finally there.	Tenderly I knelt down to the miracle-worker who had survived, in the solitude of this chapel under Mare's Head Hill, for more than two hundred years, and I saw that his end had finally come.

C	61	Zaviklal jsem kolíkem, a napůl se vylomil.	I screamed through the peg, and broke out halfway.	I worked the peg out, breaking part of it off as I did so.
C	62	Kdyby nebyla soška tak stará, řekl bych, že obyčejné kancelářské lepicí pasty.	If it wasn't so old, I'd say ordinary office paste.	If the statue hadn't been so old, I would have sworn the glue was ordinary office paste.
C	63	ne symbolickým, ale mým vlastním, zle postiženým gonorrhoeou, jež nemohla pocházet od nikoho jiného než od profesorky ruštiny Ireny Znenáhlíkové.	not symbolic but my own, badly afflicted gonorrhoea, who could not come from anyone other than Irena Znenáhlíková, a Russian professor.	not a symbolic one, but my own, afflicted with a bad case of gonorrhea that I could only have got from a Russian teacher called Irena Znenahlikova
C	64	Moje gonokoky obcestovaly tedy celou zeměkoulí, což snad vysvětlovalo jejich jedinečnou odolnost, málem nezničitelnost.	So my gonococci traveled all over the globe, perhaps explaining their unique resilience, almost indestructible.	So my gonococci had circumnavigated the globe—which may help to explain their special resilience, because they turned out to be all but indestructible.
C	65	Předkožkou, podobnou napuchlé pneumatice, proklouzávaly šedivé kapičky, popsané ve všech příručkách.	The puffy tire-like foreskin, slipped through the gray droplets described in all manuals.	and my foreskin resembled an over-inflated inner tube oozing grey drops—exactly as described in the clinical textbooks.
C	66	Mohu si dovolit představit se tomuhle malému, šeríky obrostlému městečku tím, že hned první den navštívím venerologickou ordinaci?	Can I afford to imagine this small, lilac-grown town by visiting the vener's office the very first day?	Could I afford to begin my stay in this small, lilac-bowered town with a visit to the venereological ward of the local hospital on my very first day?
C	67	Bude skutečně v jejich silách pomlčet o tom, že nový profesor jejich almy mater se do Kostelce přijel léčit na kapavku?	Will he really be silent about the fact that the new professor of their alma mater came to Kostelec to heal for gonorrhoea?	Was it realistic to hope they would keep quiet about a new teacher who came all the way to Hronov seeking a cure for the dap?
C	68	Čím víc jsem o tom přemýšlel, tím víc se ze mě vytrácela odvaha odhalit strašlivě zdevastovanou část těla očím nějaké absolventky.	The more I thought about it, the more I lost the courage to reveal the horribly devastated part of my body to the eyes of a graduate.	The more I thought about it, the more I felt my courage ebbing away. How could I bring myself to expose this ghastly, misshapen organ to the eyes of recent graduates?

C	69	Nezáživně jsem blábolil o Stalinově spisku o jazykovědě, jak to bylo předepsáno v osnovách, a můj duch se zabýval otázkou, zda ty všude popsané kapičky neprosáknou kalhotama.	I gossiped stupidly about Stalin's book of linguistics, as prescribed in the curriculum, and my ghost was wondering if the droplets described here would be soaked in their pants.	While I muttered some dry and indigestible remarks about Stalin's contribution to the sci-ence of language, as the syllabus required, my mind wrestled with the larger question of whether the discharge, also described in the textbooks, would eventually soak through my trousers and betray me.
C	70	Trochu mě to vyvedlo z míry, ale pohled na úplně všemi mastmi mazanou liščí tvář mě okamžitě poučil, že jde o provokaci.	It made me a little bit annoyed, but the look of all the ointment-smeared fox face taught me immediately that it was a provocation.	and at first I didn't know what to think. But one more look at that sly, foxy face told me that here was another test of my endurance
C	71	Na stranu se nedovolovaly ani desetileté holčičky na polické měšťance, kde jsem se živil před tím, než jsem [já] marxistickým žargonem na hradištském školení zapůsobil na téměř všemocného ředitele kostelecké sociálky a byl uprostřed školního roku přeložen do jeho dívčího ráje.	Not even the ten-year-old girls at the polish townfolk were allowed to go where I made a living before the Marxist jargon of Hradiště training on the almost omnipotent director of Kostelec Social and was transferred to his girl's paradise in the middle of the school year.	My only previous experience teaching girls had been with ten-year-olds at the municipal school in Police; that had been before I attended the fateful political schooling in Hradiste, where my command of Marxist jargon made such an impression on the all-powerful principal of the Hronov Health and Social Workers' School that I found myself transferred, in mid-term, to his paradise for young maidens

C	72	Prosím, pravil jsem ledově, a Liška -- ta přezdívka byla něco jako láska na první pohled -- v úzké tmavomodré sukni a v bílém tričku, pod nímž bylo vidět velmi dobře vyplněnou podprsenku, se krokem profesionální kurvy vzdálila z místnosti.	Please, I said icily, and Fox - the nickname was something like love at first glance - in a narrow dark blue skirt, and in a white T-shirt under which a very well filled bra was visible, she stepped away from the room with a professional whore.	"If you must 'powder your nose'," I said, with all the icy deliberation I could muster, "then you may go," and Vixi (the nickname occurred to me in a rush of feeling that was something like love at first sight) swaggered out of the room in her tight navy-blue skirt and the translucent white T-shirt, beneath which I could see the outline of an amply filled brassiere, swinging her hips like a seasoned streetwalker.
C	73	Třídou zašumělo.	There was a noise in the classroom.	A buzz went through the classroom.
C	74	Sotva slyšitelné chichotání pokračovalo a ustalo náraz, teprve když vrzly dveře, kaštanová Liška se, opět krokem profesionální kurvy, vrátila, a já si všiml, že na tričku došlo ke změně.	Hardly audible giggles continued and stopped suddenly, only when the door creaked, the Chestnut Fox came back, again a professional whore, and I noticed a change in the shirt.	The tittering continued, then came to a sudden stop as the door squeaked open and the chestnut-haired Vixi swung back into the room. There was something different about her, but it took me a moment to see what it was.
C	75	Pokus o záchranu návratem k maršálovým jazykovědným spekulacím se proměnil v debakl a já měl pocit, že se mi vzadu kouří z propocené košile.	The rescue attempt by returning to marshal linguistic speculation turned into a debacle, and I felt like I was smoking a sweaty shirt in the back.	My efforts to save the situation by coming back to Marshal Stalin's linguistic ruminations ended in utter failure, and I could feel the steam rising from the back of my sweat-soaked shirt.
D	76	Jmenoval se Hrdý Budžes, byl velice statečnější a vytrval, i když měl všelijaký potíže.	His name was Proud Budesha, he was very brave and persevering, even though he had all kinds of trouble.	about a certain gentleman called B. Proudew
D	77	Jinak se mi včera stala taková nemilá věc.	Otherwise, yesterday I had such an unfortunate thing.	Yesterday I had this rather unpleasant matter to deal with as well.
D	78	Lebek a koster se [já] bojím ze všeho nejvíc.	Skulls and skeletons are most afraid of everything.	. I'm afraid of skulls and bones more than anything else.

D	79	Taky se dost bojím čertů a taky psů, protože mě jednou jeden kousnul do nohy, ale to bylo v Zákopech u dědy a u babičky.	Also, I'm scared of the devils and dogs too, because once bitten my leg, but it was in Zákopě with grandpa and.	I'm quite scared of devils and dogs too, cos I was once bitten on the leg by one, but that was in Zákopy at Grandma's and Grandad's.
D	80	Ale nemilá věc se mi líbí, to se prostě tak říká.	But I like the unpleasant thing, it just says so.	But an unpleasant matter— I do like that. That is how they say it.
D	81	Včera, když jsme přišli do školy, tak nám pani učitelka Koláčková řekla, vlastně ne když jsme přišli do školy, ale když škola skončila, tak nám řekla, ať zůstaneme ještě chvíli potichu sedět, ...	Yesterday, when we came to school, Mrs. Koláčková told us, not really when we came to school, but when the school was over, she told us to stay quiet for a while,	Yesterday when we got to school, our teacher, Mrs Koláčková, told us... well, not actually when we got to school, but when school finished, she told us to remain quietly seated,
D	82	tak nám řekla, ať zůstaneme ještě chvíli potichu sedět, že nám musí něco říct.	she told us to stay quiet for a while, to say something.	as she had something to tell us.
D	83	[Ona] má takový krátký černý vlasy a hrozně hezky kreslí.	He has such a short black hair and draws very nicely.	She has short black hair and she draws awfully well.
D	84	Ted'ka umřela, tak už ji asi neuvidím, ale ty výkresy může pani učitelka klidně ukazovat pořád.	She died now, so I won't see her anymore, but the drawings can still be shown by the teacher.	Now she's died, so I don't suppose I'll be seeing her again, but our teacher might well keep showing us her drawings.
D	85	Doma jsem to hned vyprávěla a ptala jsem se, jakto, že umřela, když to je malá holčička, protože já už vím, že někdo umře a pak už [někdo] nikdy nepřejde domu ani nikam, ale hlavně, když je starej.	I told it at home and asked how she died when she was a little girl, because I already know someone would die and then she would never come home again, but especially when she was old.	I told them at home straight away, and I asked how it was that she died when she's a little girl, cos I already know that some people die and don't go back home or anywhere, but mostly when they're old.
D	86	Jak má někdo srdce, tak skoro určitě umře.	As someone has a heart, they will almost certainly die.	If somebody has a weak heart then they will almost certainly die.

D	87	Měla jsem radost, protože jsem dostala na vysvědčení jedničky a šli jsme na dorty, já jinak na dorty nesmím, protože jsem tlustá, tak jsem měla radost a taky, že byly prázdniny, ale byla jsem smutná, protože nám řekli, že umřela naše pani družinářka Olga Jeřábkova.	I was happy because I got the number one on the school report and we went to the cakes, I can't make the cakes otherwise, because I'm fat, so I was happy and there were holidays, but I was sad because they told us that our lady died teammate Olga Jeřábkova.	The same thing happened to me before the holidays too. I was happy cos I got an A in my report, so we went out for some cakes. Other times I'm not allowed to have cakes cos I'm fat, so I was happy, and also because it was the holidays, but then I was sad, cos they told us that our after-school helper Olga Jeřábková had died.
D	88	Ona totiž se otrávila plynem a vybuchla a vybuchnul i ten dům, kde bydlela a tak ještě umřeli nějaký cizí lidi.	She was poisoned by the gas and exploded, and the house where she lived and some strangers died.	You see, she poisoned herself with gas and she exploded, and the building where she lived exploded too, so some other people died as well.
D	89	Tátovi potom v kuchyni říkala, že ji uštvaly ty svině.	Then Dad said in the kitchen that she had been bitten by that bitch.	. Then in the kitchen she said to Daddy that those swines had hounded her to death.
D	90	To asi myslela Rusové nebo možná komunisti, protože Rusové a komunisti jsou svině, ale nesmí se to říkat.	Perhaps the Russians or perhaps the Communists thought that, because the Russians and the Communists are swine, but it must not be said.	I think she meant the Russians or maybe the Communists, cos the Russians and Communists are swines, though you aren't allowed to say so.
D	91	Tak já nevím, z naší třídy tam chodí celá třída a já bych tam taky chtěla chodit.	So I don't know, there is a whole class coming out of our class and I want to go there too.	So I don't know, out of our class the whole class goes and I'd like to go too.
D	92	Do baletu chodí hodně dětí, hlavně holčiček, ale ty jsou všechny moc hezký a nikdo tam není tlustý, jenom já a tak se mi smějou, když cvičím a pořád.	There are a lot of kids going to ballet, especially girls, but they are all very nice and nobody is fat, just me and they laugh at me when I practice and always.	Lots of children go to ballet, mostly girls, but they're all very pretty and nobody there is fat except me, so they laugh at me when I practise, all the time.
D	93	Pak se to vypálí v peci a je socha.	There are a lot of kids going to ballet, especially girls, but they are all very nice and nobody is fat, just me and they laugh at me when I practice and always.	I fire things in the kiln and make them into statues

D	94	Taky tam chodily jiný děti,...	There were other children,...	Other kids went there too,
D	95	...jenže panu Peckovi se líbilo, co maluju a co mluvím,....	...but Mr. Peck liked what I was painting and what I was talking about....	but Mr Pecka liked what I said and painted,
D	96	...a přemluvil Kačenku, aby mi dovolila chodit večer	...and persuaded Kačenka to let me go out in the evening...	and he persuaded Kačenka to let me go in the evening
D	97	, když choděj dospělý lidi.	...when she was going to adult people.	when the adults go
D	98	Pan Pecka má gramofon, a když se maluje, tak pouští Mozarta,	Mr. Pecka has a gramophone, and when he paints, he plays Mozart,	Mr Pecka has a record player and when we're painting, he plays Mozart.
D	99	to prej je nejlepší skladatel hudby ze všech, co jsou na celým světě, něco jako Bedřich Smetana, ale ještě mnohem lepší.	he is the best music composer of all the world, something like Frederick Smetana, but much better.	They say he's the best music composer in the whole world. Something like Bedřich Smetana, but much better even than him.
D	100	Taky si tam povídaj zajímavý věci a já myslím, že já budu sochař, až budu velká.	Also, talk interesting things there and I think I'll be a sculptor when I'm big.	They talk about interesting things there and I think I'll be a sculptor when I grow up.