Translation Competence in Machines: A Study of Adjectives in English-Swedish Translation.

Lars Ahrenberg
Department of Computer and Information Science, Linköping University
lars.ahrenberg@liu.se

Abstract

Recent improvements in neural machine translation calls for increased efforts on qualitative evaluations so as to get a better understanding of differences in translation competence between human and machine. This paper reports the results of a study of 1170 adjectives in translation from English to Swedish, using the Parallel Universal Dependencies Treebanks for these languages. The comparison covers two dimensions: the types of solutions employed and the incidence of debatable or incorrect translations. It is found that the machine translation uses all of the solution types that the human translation does, but in different proportions and less competently.

1 Introduction

The performance of today's machine translation systems is sometimes characterized as 'human-level' or achieving 'human parity' (Hassan et al., 2018; Bojar et al., 2018). While claims of this kind have been criticized for not being based on proper evaluations, e.g. by (Graham et al., 2019; Läubli et al., 2020), it is nevertheless a fact that the quality of machine-translated text have improved considerably in recent years, due to new neural models such as the Transformer (Vaswani et al., 2017).

These developments do not only motivate the need for better quantitative evaluations but also call for qualitative evaluations that can pinpoint the differences between different translators, whether human or digital. In this paper my interest is with comparing state-of-the art online translation with human translation of the same source data.

1.1 An example

Comparisons of machine translations with human translations tend to focus either on errors or on general quality criteria such as accuracy and fluency. With the advances of neural machine translation such criteria just seem too blunt to be useful. Neural MT is both accurate and fluent most of the time so any search for differences requires something more fine-grained.

Consider the following English sentence and three possible Swedish translations:

1. He is a bad lier
2a. Han är en dålig lögnare
'He is a bad lier'
2b. Han ljuger dåligt
'He lies badly'
2c. Han är dålig på att ljuga
'He is bad at lying'

All three translations are accurate, fluent and understandable. Still, the first is an example of interference or what (Katourgi, 2020) calls *översättningssvenska*, which we can translate as 'translational Swedish' or 'Swedish translationese'. For this example it means that the translation has the same structure, word by word, as the source sentence, while other, more natural or idiomatic alternatives exist.

Translation (2b) has turned a copulative sentence with a noun phrase predicative into a verb phrase, where the verb translates the noun and an adverb translates the adjective. Translation (2c) is again copulative but involves a head switch; the adjective is translated by an adjective, but this adjective is now the only predicative, having a verb in the infinitive as dependent, this verb translating the English noun.

Katourgi does not say that translations of the type 2a are bad, nor that those of 2b and 2c are better. However, as the title of his book reveals, he claims that they can be too 'noticeable' especially if there are too many of them. The point is thus that a translator should know all available alternatives and not least those that are natural and more common in an indigenous Swedish context.

The aim of this study is to assess the quality of the translations produced by a state-of-the-art online NMT system at a particular point in time for one language pair, English-Swedish. Compared to Chinese or German, Swedish is a small language, but it still has high-quality MT systems available. The study has two quality aspects in focus: the types of solution the system can produce and to what extent it can apply those solutions accurately.

As for the first part, it is related to taxonomies of what has variously been termed *translation procedures* (Vinay and Darbelnet, 1958/77) or *translation relations* (van Leuven-Zwart, 1989), though in this paper we will call them *solution types*. The second part is an analysis of what we can call debatable solutions or issues (Lommel et al., 2015). This analysis uses a linguistically-based taxonomy of issues, and has been done by the author only. It is however supported with evidence from the human translation.

(Ahrenberg, 2017) concluded for a study on the same language pair and direction that ”... the MT is in many ways, such as length, information flow, and structure more similar to the source than the HT. More importantly, it exhibits a much more restricted repertoir of procedures, and its output is estimated to require about three edits per sentence”. Here, an edit is caused by an issue that was judged to require alteration. A specific aim of this paper is to see whether these conclusions are still valid.

The data used for this study comes from the English and Swedish Parallel Universal Dependencies treebanks (Zeman et al., 2017). We wished also to investigate the possibilities of using this data set for translation studies, although they were not collected for this purpose.

2 Translation competence

The notion of translation competence has been approached in different ways. One type of analysis seeks to identify all possible properties that are required of a translator. A major proponent of this approach is the PACTE group at the University of Barcelona, who have elaborated a model based on subcomponents in several works, e.g. in (PACTE, 2011). Others, like (Malmkjær, 2009), have argued that a characterization of translation competence should focus on those factors that distinguish it from any other profession, including any other type of bilingualism. From this perspective the only subcomponent of the PACTE model that Malmkjær finds relevant is the transfer competence:

> ”... the ability to complete the transfer process from the ST (source text) to the TT (target text), i.e. to understand the ST and re-express it in the TL (target language), taking into account the translation’s function and the characteristics of the receptor”.

Nevertheless, when PACTE studies how translation competence is acquired they put translation problems in focus, in particular what they call Rich Points, i.e., passages that may be challenging to translate. In this work I take translation competence to mean the ability to find an appropriate solution for all words in the source text, where a solution may of course be to leave it untranslated. Thus, the set of ’problems’ includes not just difficult ones, but all words or constructions meeting a given criterion. This prevents the selection of rich points from being skewed and allows for quantitative analysis. All in all this should give a better picture of the abilities of a translator. The chosen construction is adjectives in relation to a head.

We can interpret produced translations in terms of skills that we ascribe to the translators, whether human or digital. Necessary skills are specified in many text books on translation. Here, I mention a few of them from a text book by (Ingo, 2007)[1]

- a robust sense of style in the target language
- active and creative language skills in the target language
- familiarity with target language genre conventions
- ability to express oneself naturally in the target language
- ability to change style in accordance with the style of the source text
- possession of an imaginative mind so as not be bound by the patterns of the source text and hindered from finding the natural and idiomatic expressions of the target language

We can observe that this list demands of a translator to be able to strike a delicate balance between stylistic and genre-related conventions on the one hand, and creativeness and imaginative abilities on the other.

[1] Translations by the author.

2.1 Translation problems

NMT models make mistakes of different kinds. Even (Hassan et al., 2018) who claims human parity performed an error analysis of the English output and found most of the errors in the categories Incorrect words, Ungrammatical, Missing words, and Named Entities. Thus, both accuracy and fluency are affected. Looking at the same sentences (Läubli et al., 2020) observe that fluency mistakes (word order, ungrammaticality, ...) are still common in the machine translations, somewhat contrary to the expectations that NMT systems have specifically improved as regards fluency. They also observe that cross-sentential constraints affect machine translations more often than human translations. This means that a sentence translation can appear fluent in isolation, but be judged as inappropriate in the document context.

(Ahrenberg, 2017) found that the most frequent errors in his data related to the accuracy of word translations. In that study, close to 50% of the errors noted were of this kind. The next error type in frequency concerned morphological form, slightly less than 25% of all.

2.2 Comparing machine translation and human translation

There have been quite a number of studies trying to distinguish machine translations from human translations by automatic means e.g., (Aharoni et al., 2014). These studies usually favour features that can be detected automatically as well, such as the occurrence of common function words and part-of-speech ngrams. (Nguyen-Son et al., 2017), found other features, such as word distributions, complex phrase constructions, and the occurrence of phrasal verbs to be helpful, in particular when combined with coherence features across sentences or within whole paragraphs. The data used in this study consist of isolated sentences, so we cannot employ coherence features. However, we can compare our approach to what can be gained from other features.

There have also been studies aimed at automating, at least partly, the recognition of solution types, or divergencies as they are often called, for example (Deng and Xue, 2017; Zhai et al., 2019). This was an initial aim also of this study, but was abandoned for reasons that will be explained below.

3 Adjectives in English-Swedish translation

Adjectives have very much the same behaviour in English and Swedish. They can be modifiers/attributes, predicatives, heads of arguments such as subjects and objects, conjuncts and be part of lexicalized phrases. All of these functions are actually found in the English source data. The distribution of these functions in the source data is shown in Table 1 together with simple examples to show what is meant.

Function	Frequency	Example
modifier	956	a *red* shirt
predicative	122	it is *red*
conjunct	42	white and *red*
argument head	34	help the *poor*
lexphrase	16	*at best*
Total	1170	

Table 1: Adjectival functions in the source data.

Given that the vast majority of adjectives can be translated straightforwardly by an identical syntactic construction, it can be expected that this pattern will be over-used by inexperienced translators and by machine translations that tend to prefer frequent patterns over rarer ones. Thus, when the adjective and head noun are independent semantic units that form a complex that can be interpreted compositionally, the unmarked translation is a word-for-word translation, in particular if both items are part of the core vocabulary of the language. This applies to the first four examples of Table 1. The lexphrase also has a standard translation, though one which is not compositional: *i bästa fall*, 'in (the) best case'.

For some English adjectives a common alternative translation is to form a compound. This happens with *wooden – trä-*, *main*, where *huvud-* is a common choice, and *special* with the translations *speciell*, *särskild* or *special-*. Examples are: *wooden table – träbord*, *main purpose – huvudsyfte*, *special unit – specialenhet*. This solution type is actually quite common in the studied data set.

Another possibility is that the adjective and the noun form a single designation of some referent, which acts as a term or name for the referent. This requires that the translator knows this and also is able to find out the term or name used in the target language. Common results then are compounds, *red herring – avledningsmanöver*, ’distraction action’, transfers, such as *British Council* or *American Express*, or a lexicalized phrase, where the translation of the adjective may also be an adjective, but one that does not occur outside of that phrase, as in *common sense – sunt förnuft*, ’sound sense’.

Swedish has a greater propensity than English for using adjectives as heads of nominal phrases. Thus, a nominal head in an English source text is sometimes not translated. *white people* may be translated by *vita människor* but just *vita* ’whites’ would do just as well, if not better. The word *one* is often not translated when it is used instead of repeating a mentioned noun, or when the referent is understood from the context: *the only one – den enda*, *they will build a new road and tear* **the old one** *up. – de ska bygga en ny väg och riva upp* **den gamla**.

Yet another possibility is that the pair of adjective and noun are part of a larger construction that acts as a unit in the translation. It may simply be that a preceding preposition gives the whole an adverbial function and the possibility to translate the whole thing with an adverb. Examples:

in *early* morning
tidigt på morgonen
’early in the_morning’

She was killed *in cold blood*
Hon mördades *kallblodigt*
’She was_murdered coldblooded-ly’

The relevant embedding construction may also be larger:

at your *earliest* convenience
så fort du kan
’as fast you can’

If the embedding construction is found superfluous for the target audience, it may not be translated at all, and this will then affect the adjective-noun pair in the same way. With this as background we now proceed to the study.

4 Data

The source sentences for the study are taken from the English part of the Parallel Universal Dependencies treebanks (PUD)[2]. These treebanks were created for a shared task on multilingual parsing from raw text (Zeman et al., 2017). The sentences are taken from news and Wikipedia articles, but only a few from each article. Thus, there may be lexical overlaps but no coherent paragraphs. This means that we cannot study discourse phenomena such as cohesion.

PUD-segments were translated into Swedish outside of the shared tasks. The Swedish translations follow the same directions as for other PUD treebanks, namely that ”Translators were instructed to prefer translations closer to original grammatical structure, provided it is still a fluent sentence in the target language” (*ibid. p.4*). This requirement is one that we could also ask of a machine translation system.

Only those sentences where English is the source language have been used. They amount to 750 segments. We define an adjective as any token assigned the UD part of speech ADJ in the English PUD treebank. There are 1170 of them.

The machine translations were produced by Google Translate on 25-26th of February, 2021. They were then tagged and parsed with the UDPipe tools (Straka, 2018) using a model for UD_Swedish-Talbanken.

Basic statistics for the data can be found in Table 2. The figures follow a standard pattern for English-to-Swedish translations. It can be noted that the Type-Token-Ration for the machine translation is much closer to the human translation than to the English source text.

Dataset	Types	Tokens	TTR
English PUD	4714	15840	0.297
Swedish PUD	5125	14432	0.355
MT-translated PUD	4949	14129	0.350

Table 2: Statistics of the datasets.

The human translations have earlier been provided with manual word alignments by the author. We hoped that the structural properties of the image of an adjectival relation could be determined automatically from the word alignment. This approach, however, turned out to be problematic as the annotations for part-of-speech and dependencies are not harmonized across the two languages. The translation of many words, such as ’many’ and ’same’ that were tagged ADJ in the source treebank, were translated in the expected, standard fashion, but had a different tag (PRON

[2]https://github.com/UniversalDependencies/UD_English-PUD/

and DET for these words), causing the automatic analysis to suggest a part-of-speech shift. Similarly, a reference such as the 'Metropolitan Club' has been translated verbatim and is thus word-to-word. However, where the English annotates 'Metropolitan' as an adjective modifying a noun, the Swedish sees two proper nouns, where the second is a dependent of the first via the UD relation *flat*. The automatic analysis thus suggests a shift of parts-of-speech and a reversal of the dependency. Cases of this kind abound, and for this reason the sorting and the analysis have required more manual effort than anticipated. Thus, all data points have had a manual review and the same holds for the machine translations.

5 Analysis

We compare a machine translation with human translations of 750 English sentences which are part of the Parallel Universal Dependencies (PUD) dataset. We analyse translations along two dimensions, solution types based on structural properties and issues.

5.1 Solution types

The solution types are divided into two major classes, **Isomorphisms** and **Restructurings**

A translation is an isomorphism if the following properties hold: (1) the adjective is translated by a single token, a; (2) the head token is translated by a separate single token, h; (3) h is the head of a in the translation; (4) a and h have the same part-of-speech as their source tokens and the dependency relation and the order between them is also the same. It may be the case that the distance between a and h is different than the distance between the corresponding source words. These differences are not directly caused by the adjective and its head and so are not considered relevant.

Restructuring is an umbrella term for all other situations. We sub-classify restructurings according to the structural effect. Table 3 gives examples of each category from the corpus.

A *shift* occurs when the first three clauses above hold, but there is a change in part-of-speech and/or relation. Using the dependency relations of the UD framework, a change in part-of-speech will almost always involve a shift of dependency relation as well, so we will note a relation shift only when there is no change in part-of-speech. An example is when the dependency of an adjective is changed from 'xcomp' (head of a subject-less verb phrase) to 'ccomp' (head of a finite clause with subject).

An *omission* occurs when the adjective in the source sentence lacks a corresponding target token. This means as well that there is no corresponding dependency either. In case the head has not been translated we use the label *head-omission*.

A *convergence* occurs when the adjective and its head are mapped onto the same target token, or the same set of target tokens. The opposite situation, a *divergence*, happens when either the adjective or its head is aligned with two or more target tokens, so that the single edge of the source tree is mapped on some subgraph with two or more edges in the target tree.

A *head-shift* occurs when both the adjective and its head are aligned with single tokens, but the dependency relation is reversed, i.e., h will be a dependent of a. This category is different from the category of *head changes*, which means that both the source adjective and its head have been translated, but they are no longer related as a dependent to a head. Finally, an *order-reversal* means that the order between a and h is reversed in comparison with the order between their source words.

From Table 4 we see that the human translation is more prone to restructure than the machine translation. In fact, this difference is consistent across all of the five adjectival functions shown in Table 1. However, the difference is not so great as to be statistically significant at a 0.05 critical level using a Chi-Square test with one degree of freedom.

Also, for some 43% of all instances (506 out of 1170) HT and MT have produced identical translations, see Table 5. The large majority of these cases are isomorphisms and the tokens concerned are common lexical items with more or less standard translations, such as *first – första, many – många, , new – nya, other – andra, possible – möjliga, whole – hela*. Another set of adjectives for which translations are shared are words with a common historical root, or words that Swedish has borrowed from English, such as *artificial – artificiell, civil – civil, international – internationell, military – militär, popular – populära*. In 85% of the instances (1003 out of 1170) the two translations agree on the broad type of solution, and in most of these (941 out of 1170) they also agree on the sub-type.

Category	English	Swedish
	Isomorphisms	
modifier	the *peaceful* transition	den *fredliga* övergången
predicative	this will be a little *different*	kommer detta bli lite *annorlunda*
	Restructurings	
convergence	The *South Korean* company initially thought...	Det *sydkoreanska* företaget trodde ...
divergence	over 70 % are *alive*	mer än 70 % var *vid liv*
omission	*provincial* police surveillance operations	polisens övervakningsoperationer
headomission	of new ideas with *old ones*	som nya idéer bildade med *gamla*
head shift	*preferential*$_1$ *access*$_2$ to government	*företräde*$_1$ i regeringens *tillgänglighet*$_2$
head change	*Much*$_{nsubj:5}$ ... has been about *identity*	*Mycket*$_{nsubj:3}$ ... har handlat om *identitet*
shift of POS	the protein ... that's *responsible*$_{ADJ}$	det protein ... som *ansvarar*$_{VERB}$ för
shift of deprel	I'd be *amazed*$_{root}$ if	Jag skulle bli förbluffad$_{xcomp}$
order-reversal	in the realm of the *unimaginable*	i det *ofattbaras* rike

Table 3: Different types of solutions.

System	Isom	Restr	Total
MT	943	227	1170
HT	878	292	1170

Table 4: Distribution of isomorphic and restructured solutions for MT and HT.

Criterion	Isom	Restr	Total
Token identical	455	51	506
Type identical	827	176	1003
Sub-type identical	825	116	941

Table 5: Number of identical translations between MT and HT.

The largest differences between the two systems concern the use of restructurings, as shown in Table 6. When we look at these more fine-grained sub-types of restructurings, there are several cases where the two translations choose the same type of solution, convergence being the most common. Examples are found with geographical adjectives such as *South Korean – sydkoreansk, northern Sami – nordsamiska*

While the two systems agree to a large extent in the use of convergences, the case is quite different with divergences. The human translation employs this solution type four times as often as the machine translation. The same is true, though to a lesser degree, of part-of-speech shifts and head-shifts. A possible explanation is that the human translator has a better sense of what fluency or naturalness means for the target language.

The system, on the other hand, has a greater use of omissions, although for quite a small percentage of the full dataset. It also produces more of head changes where the direct connection between dependent and head in the source is broken up in the translation.

Type	MT	HT
convergence	120	110
divergence	17	70
headchange	15	6
headshift	5	12
headomission	3	6
omission	22	8
posshift	46	75
reversal	0	2
relshift	2	3
Total:	230	292

Table 6: Distribution of different types of restructurings in MT and HT.

5.2 Issues

For issue classification we use the taxonomy shown in Table 7. It is basically structured according to which linguistic level is affected. The label *Meaning* means that one can debate the accuracy of the choice. It includes cases that (Hassan et al., 2018) label Incorrect words, but also what they call Unknown words, a category which is only rarely found in their translations. However, in our machine translations they are quite common, to be further discussed below. *Word choice* means that we may discuss whether the chosen word in the translation is the best choice. It is less serious than the previous category. The label *Morphol-*

ogy means that there is a lack of congruence between the adjective and its head, or of any one of them in relation to another related token such as a determiner. *Grammar* means that the translation has produced an ungrammatical substring that includes the adjective or its head. *Style* is similar to Word choice but in relation to grammar. Thus, the grammar is ok, but there are perhaps better solutions, i.e., a different type of solution could be preferred. Finally, *Orthography* concerns spellings and the use of capital letters, etc.

Issue	Frequency	Same as HT
Meaning	54	2
Word choice	118	25
Morphology	26	4
Grammar	8	1
Style	30	5
Orthographic	3	0
Total:	239	40

Table 7: System solutions that may be debated according to linguistic levels.

First it should be said that the table shows that issue classification is a subjective process. At least one person, i.e., the translator responsible for Swedish PUD, can be assumed to accept the system translations as they coincide with those of her own. However, we can note with some relief that the issue type where the differences are most pronounced concerns Meaning. For this reason we look at this category in more detail.

5.3 Problems with accuracy

Looking further at the issues pertaining to Meaning, they can largely be divided into three classes: (i) innovations, where the system seems to make up words, probably based on its models of subwords; they may sometimes be understood nevertheless; (ii) mistranslations, where the translation may mislead the reader but is perfectly fluent; and (iii) odd mistranslations that affect both accuracy and fluency and probably will cause the reader to stop for a while and try to infer what is meant.

The innovative solutions produces words that either don't exist in Swedish, or have alternatives that are vastly more common. The following are a few examples:

- 'villainous' is translated as *skurkig*, 'crooky' instead of *skurkaktig*, 'like a crook'.
- 'skerry-protected waterway' is translated as *skärvägsskyddad*, where the human translator found herself forced to rewrite as *av skär skyddad*, 'by skerries protected'.
- 'isthmus' is translated as *landmus*, 'land mouse' instead of the correct *näs*.
- 'zodiacal' is translated as *zodiakal*, a word which exists but is uncommon.

To this list we may add a few cases where the English words are copied into the Swedish translation: 'glitchy, twitchy Odi' is left as such where the human translator provides normal Swedish words. We can observe that the source adjectives in these cases are quite rare; in fact none of them can be found, even at the C2 level, in the English Vocabulary Profile. This means that even as a proficient speaker of English as a second language you are not expected to know them.

The second subset is made up of plain mistranslations, sometimes yielding the opposite of what was in the source as when 'uncooperative' is translated as *samarbetsvillig*, 'cooperative'. References to centuries are a problem; the system sometimes gets it right as with '16th century' becoming *1500-talet*, but mostly gets it wrong; for example with the '6th', '8th' and '14th' centuries.

Inconsistencies are found also with other adjectives of nationality, so that 'Macedonian' is translated either as *makedonisk*, as in the HT, or *makedonsk*.

In the third type of situation the system's choice is just odd, making you wonder what is actually meant. Some examples of this kind are:

- 'the dress code was too stuffy': the HT says *stel*, 'stiff', which is correct, whereas the system says *täppt*, which would be appropriate if you were talking about someone's nose.
- 'skilled jobs' is rendered as *skickliga arbeten*, with an adjective that is appropriate for a 'skilled worker'. The HT has the correct *kvalificerade*.
- 'lower forgone earnings' was translated as *nedre förlorade inkomster*, where *nedre* is appropriate for positions and geography but cannot be applied to earnings.

5.4 Problems with compounds

The system is very happy at producing convergencies and normally does so quite accurately. But sometimes it is overdoing it, producing clumsy compounds such as *Obama-*

specialassistent, 'Obama special assistant', *lagstiftningsförlamning*, 'legislative paralysis', *Post-Classic-perioden*, 'the Post Classic period'. It may also pick an unfortunate translations for one of the parts of a compound, as in *södersamiska* instead of *sydsamiska* for 'south Sami'.

It also occassionally separates the parts of a compound which results in a breach of grammar: *storstads kommun*, 'a city's municipality' instead of *storstadskommun* for 'metropolitan municipality' or *ras tolerans*, 'a race's tolerance' instead of *rastolerans* for 'racial tolerance'.

6 Conclusions

A general conclusion is that the system seems to have improved, for example compared to (Ahrenberg, 2017). It has gained in the type of solutions it has available, and in creativity, but this has come with a price. As regards translation competence with respect to the translation of adjectives, it can be summarized as follows:

- The system is more prone than the human translator to choose an isomorphic solution. The tendency is consistent across grammatical functions;
- The system uses the same types of restructurings as the human translation, but to different degrees;
- In particular, the human translation employs divergencies, part-of-speech shifts, and head shifts to a much larger extent than the system;
- On the other hand, the system shows more of head changes and omissions than the human translations;
- As shown in Table 4, the system produces some 200 debatable translations including about 50 (4.3% of all) that can be considered errors of accuracy.
- (Not surprisingly) the system has the greatest problems with uncommon words. For these words the system often produces innovative solutions, probably on the basis of its sub-word models. However, this means that the system essentially lacks the competence to distinguish words from non-words.

It is interesting to note that the large restructurings that were illustrated in the introduction are rare. There is one example on the model of sentence (1), where someone is described as 'a keen guitarist'. Both human and machine chose a word-by-word translation in spite of the fact that there is no Swedish word that exactly corresponds to 'keen'. The human translator chose *flitig*, 'diligent, hard-working', and the system chose *skicklig*, 'skilled', none of which is optimal. A more idiomatic way of expressing the meaning of 'keen' in Swedish would be to use a verb such as *gilla* or *tycka om*, both meaning 'like'. The human translator was instructed to stay close to the source, so that may be an explanation for not chosing a major rewriting; the system, however, has no awareness of the directive.

As for the use of PUD treebanks to study differences between human and machine translations there are both pro's and con's. On the positive side, the sentences contain both common and uncommon words and thus provides a nice sample of problems for translation across frequency ranges. The same is actually true of grammar so there are a good number of 'Rich Points' that can be selected. On the downside from the point of view translation studies is the fact that the resource consists of isolated sentences, so that discourse effects cannot be studied. Another drawback is that the annotations of the English and Swedish treebanks are not harmonised. This can partly be explained by differences in annotation practices, and partly by parsing errors that have not been corrected. Even though I had made a complete alignment at the word level for all sentences, attempts to automate the categorisation of solution types failed because of the inconsistencies. Similarly, UDpipe was helpful for tagging the machine translations, but also makes many parsing errors.

In future work, the study can be extended to dependencies of nouns and verbs using a similar approach. And the study of adjectives can be repeated at a future date.

References

Roee Aharoni, Moshe Koppel, and Yoav Goldberg. 2014. Automatic detection och machine translated text and translation quality estimation. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Short Papers), June 23-25 2014*, pages 289—295, Baltimore, ML, USA. Association for Computational Linguistics.

Lars Ahrenberg. 2017. Comparing machine translation and human translation: A case study. In *Proceedings of The First Workshop on Human-Informed Translation and Interpreting Technology (HiT-IT)*, pages 21–28, Varna, Bulgaria.

Ondřej Bojar, Christian Federmann, Mark Fishel, Yvette Graham, Barry Haddow, Philipp Koehn, and Christof Monz. 2018. Findings of the 2018 conference on machine translation (WMT18). In *Proceedings of the Third Conference on Machine Translation: Shared Task Papers*, pages 272–303, Belgium, Brussels. Association for Computational Linguistics.

Dun Deng and Nianwen Xue. 2017. Translation divergences in chinese–english machine translation: An empirical investigation. *Computational Linguistics*, 43(3:521—565.

Yvette Graham, Barry Haddow, and Philipp Koehn. 2019. Translationese in machine translation evaluation.

Hany Hassan, Anthony Aue, Chang Chen, Vishal Chowdhary, Jonathan Clark, Christian Federmann, Xuedong Huang, Marcin Junczys-Dowmunt, William Lewis, Mu Li, Shujie Liu, Tie-Yan Liu, Renqian Luo, Arul Menezes, Tao Qin, Frank Seide, Xu Tan, Fei Tian, Lijun Wu, Shuangzhi Wu, Yingce Xia, Dongdong Zhang, Zhirui Zhang, and Ming Zhou. 2018. Achieving human parity on automatic chinese to english news translation.

Rune Ingo. 2007. *Konsten att översätta: översättningens praktik och didaktik*. Studentlitteratur.

Alexander Katourgi. 2020. *Svenskan går bananer: En bok om översättningar som syns*. Lys förlag.

Kitty van Leuven-Zwart. 1989. The relation of translation to original. *Target*, 1(2):100–101.

Arle Lommel, Aljoscha Burchardt, and Hans Uszkoreit. 2015. Multidimensional quality metrics. `http://www.qt21.eu/mqm-definition/definition-2015-12-30.html`.

Samuel Läubli, Sheila Castilho, Graham Neubig, Rico Sennrich, Qinlan Shen, and Antonio Toral. 2020. A set of recommendations for assessing human–machine parity in language translation. *JAIR*, 67:653–672.

Kirsten Malmkjær. 2009. What is translation competence? *Revue française de linguistique appliquée*, XIV:121–134.

Hoang-Quoc Nguyen-Son, Ngoc-Dung T. Tieu, Huy H. Nguyen, Junichi Yamagishi, and Isao Echi Zen. 2017. Identifying computer-generated text using statistical analysis. In *2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pages 1504–1511.

PACTE. 2011. Results of the validation of the pacte translation competence model: Translation problems and translation competence. In *Methods and Strategies of Process Research: Integrative Approaches to Translation Studies*, pages 317–343. John Benjamins.

Milan Straka. 2018. UDPipe 2.0 prototype at CoNLL 2018 UD shared task. In *Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, pages 197–207, Brussels, Belgium. Association for Computational Linguistics.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need.

J.-P. Vinay and J. Darbelnet. 1958/77. *Stylistique comparée du français et de l'anglais*. Paris, Didier.

Daniel Zeman, Martin Popel, Milan Straka, Jan Hajič, Joakim Nivre, Filip Ginter, Juhani Luotolahti, Sampo Pyysalo, Slav Petrov, Martin Potthast, Francis Tyers, Elena Badmaeva, Memduh Gokirmak, Anna Nedoluzhko, Silvie Cinková, Jan Hajič jr., Jaroslava Hlaváčová, Václava Kettnerová, Zdeňka Urešová, Jenna Kanerva, Stina Ojala, Anna Missilä, Christopher D. Manning, Sebastian Schuster, Siva Reddy, Dima Taji, Nizar Habash, Herman Leung, Marie-Catherine de Marneffe, Manuela Sanguinetti, Maria Simi, Hiroshi Kanayama, Valeria de Paiva, Kira Droganova, Héctor Martínez Alonso, Çağrı Çöltekin, Umut Sulubacak, Hans Uszkoreit, Vivien Macketanz, Aljoscha Burchardt, Kim Harris, Katrin Marheinecke, Georg Rehm, Tolga Kayadelen, Mohammed Attia, Ali Elkahky, Zhuoran Yu, Emily Pitler, Saran Lertpradit, Michael Mandl, Jesse Kirchner, Hector Fernandez Alcalde, Jana Strnadová, Esha Banerjee, Ruli Manurung, Antonio Stella, Atsuko Shimada, Sookyoung Kwak, Gustavo Mendonça, Tatiana Lando, Rattima Nitisaroj, and Josie Li. 2017. CoNLL 2017 shared task: Multilingual parsing from raw text to Universal Dependencies. In *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, pages 1–19, Vancouver, Canada. Association for Computational Linguistics.

Yuming Zhai, Pooyan Safari, Gabriel Illouz, Alexandre Allauzen, and Anne Vilnat. 2019. Towards recognizing phrase translation processes: Experiments on english-french.