

Survey of Common Errors of English to Vietnamese Google Translator in Business Contract

Trần Lê Tâm Linh

Foreign Language Center, University of Science, Viet Nam National University, Hochiminh City, Vietnam
Email: tranletamlinh@yahoo.com.vn

Received April 1st, 2013; revised May 2nd, 2013; accepted May 10th, 2013

Copyright © 2013 Trần Lê Tâm Linh. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Machine translation is a prolonged and difficult problem, but it becomes more and more attractive due to its huge benefits for our society. Nowadays, Google Translator (GT) is one of the most common machine translation softwares because of its relatively high accuracy. However, some language pairs with their different typology and some specialized texts which are translated by GT have not issued really good translation results yet. In this paper, we survey the common errors of machine translation from the point of view of comparative linguistics in business contracts when these legal documents are translated from English into Vietnamese by using GT. Hopefully, these results can be used to intend the effective ways improving translation models of computer in the future.

Keywords: Machine Translation (MT); Lexical Error; Semantic Error; Word Order Error; Missing Error; Redundant Errors; Google Translator

Introduction

Since Vietnam has intergrated into World Trade Organisation, most of the global relationships such as commercial, import, export transactions are involved in business contracts, so these kinds of legal documents become more and more essential. As a result, drafting contracts considered the most vital. For studying languages in the contracts, the role of translation is remarkable. Nowadays, the volume of English-Vietnamese bilingual contracts becomes so huge that hand-made translation cannot catch up with the needs of booming information in business. To solve this problem, machine translation can meet the pressing and necessary requirements of business contract translation. That is the main reason why we have studied about common errors in English-Vietnamese business contracts translated by Google translation software.

Programs of machine translation (MT) for natural language have been built since early 1950s, but their success has still limited within pairs of languages which have nearly the same grammar, structures as well as vocabulary such as English-French, English-Russian... Among them, the kinds of scientific or legal documents are translated remarkably successfully due to their clear grammar and their simple single meaning. On the other hand, pairs of completely differently typological languages which are translated by MT have translating results still badly. For instance, because English belongs to an inflectional language, but Vietnamese is a kind of isolating language. As a result, there are too many errors in business contracts when they are translated by MT from English into Vietnamese vice versa. The main reason to carry out a survey of common errors of machine translating is to find the methods correcting them. These works need to combine between linguists and profes-

sional computer experts because everything can be remedied if its causes are found.

Overview

Related Works and General Aspects of Linguistics

There are many works which are related to studying machine translation, contrastive languages and general aspects of linguistics. For example, firstly, *Error classification for MT evaluation* was studied by Mary A. Flanagan. She presents a system for classifying errors in MT output as a means of evaluating output quality. Classification of errors provides a basis for comparing translations produced by different machine translation (MT) systems and formalizes the process of error counting. Error classification can provide a descriptive framework that reveals relationships between errors. For example, if subject and verb do not agree in person or number, the error can be classified as one of agreement, rather than an incorrect noun inflection, or verb inflection or both. Error categorization can also help the evaluator to map the extent of the effect in chains of errors, allowing comparison among MT systems (Mary, 1996: p. 66). Secondly, *Evaluation of automatic translation output* is a difficult task. Several performance measures like Word Error Rate, Position Independent Word Error Rate and the BLEU and NIST scores are widely use and provide a useful tool for comparing different systems and to evaluate improvements within a system. However the interpretation of all of these measures is not at all clear, and the identification of the most prominent source of errors in a given system using these measures alone is not possible. Therefore some analysis of the generated translations is needed in order to identify the main problems and to focus the research efforts. This area is however mostly unex-

plored and few works have dealt with it until now. In this paper we will present a framework for classification of the errors of a machine translation system and we will carry out an error analysis of the system used by the RWTH in the first TC-STAR evaluation (David, 2006: p. 1). Thirdly, machine translation evaluation is a difficult task, since there is not only one correct translation of a sentence, but many equally good translation options. Often, machine translation systems are only evaluated quantitatively, e.g. by the use of automatic metrics, which is fast and cheap, but does not give any indication of the specific problems of a MT system. Besides, *Error analysis of statistical machine translation output* has been researched by David Vilar, Jia Xu, Luis Fernando D'Haro, and Hermann Ney (David, 2006). Fourthly, Association for Computational Linguistics has said that some analysis of the generated output is needed in order to identify the main problems and to focus the research efforts. On the other hand, human evaluation is a time consuming and expensive task. In their paper, they investigate methods for using of morpho-syntactic information for automatic evaluation: standard error measures WER and PER are calculated on distinct word classes and forms in order to get a better idea about the nature of translation errors and possibilities for improvements (Maja, 2006: p. 1). Fifthly, according to *Morpho-syntactic information for automatic error analysis of statistical machine translation output* from proceedings of the workshop on statistical machine translation (Maja June, 2006, New York City, pages 1-6). Sixthly, *Word error rates: decomposition over POS classes and applications for error analysis* has been said that the obtained results are shown to correspond to the results of a human error analysis. The results obtained on the European Parliament Plenary Session corpus in Spanish and English give a better overview of the nature of translation errors as well as ideas of where to put efforts for possible improvements of the translation system (Maja, 2007: p. 47). Seventhly, *a preliminary study of the length of sentence in legal English* (Duong, 2008). Eighthly, there are many works of studying contrastive languages in contracts as well as law documents such as *editing techniques of law normative act and legal languages* (Nguyen, 2010). Tenthly, there are many successful works for errors of machine translation such as *A tool for error analysis of machine translation output*, according to Sara S. (2011).

Besides, there should be various ideas to study languages in these kinds of law documents to help those who use contracts to avoid misunderstanding ambiguous words which can cause problems due to them. Whenever the quality of translation is said, its common errors will be interested most because they can cause any serious problems without finding and correcting in time. So, there should be the discriminating criterion of common errors in order to suggest the ways to correct them.

On analysing any errors of any languages, researchers should find them belonging to grammatical errors, lexical meaning errors or pragmatical errors. In this paper, we survey the common errors of machine translation from the point of view of comparative linguistics in business contracts when these legal documents are translated from English into Vietnamese by using GT. Hopefully, these results can be used to intend the effective ways improving translation models of computer in the future.

Software Supports to Research

Introduction of Blast Software

Blast (the Bi-Lingual Annotator/Annotation/Analysis Support

Tool) is an error annotation tool for machine translation output. It came from a Swedish author, Sara Stymne, Linköping University, Linköping, Sweden.

Blast, which is considered as a tool for error analysis of machine translation output, can aid the user by highlighting similarities with a reference sentence. Blast is flexible in that it can be used with output from any MT system, and with any hierarchical error typology. It has a modular design, allowing easy extension with new modules. To the best of our knowledge, there is no other publicly available tool for MT error annotation. Since we believe that error analysis is a vital complement to MT evaluation, we think that Blast can be useful for many other MT researchers and developers. 2 MT Evaluation and Error Analysis Hovy et al. (2002) discussed the complexity of MT evaluation, and stressed the importance of adjusting evaluation to the purpose and context of the translation. However, MT is very often only evaluated quantitatively using a single metric, especially in research papers. Quantitative evaluations can be automatic, using metrics such as Bleu (Papineni et al., 2002) or Meteor (Denkowski & Lavie, 2010), where the MT output is compared to one or more human reference translations. Besides, Blast is also accepted to the Association for Computational Linguistics (ACL'11), demonstration session. Portland, Oregon, USA. July 2011.

Application of Blast into Analyzing Vietnamese Common Machine Translation Errors

The material is studied for this paper is "Legal documents on labour and economic contracts, settlement of labour and economic disputes (Vietnamese-English)" [18]. This bilingual book concludes 733 pages. After typing the whole books, we extracted 2947 language pairs of English and Vietnamese. Then they are processed by our special software to delete the repeated language pairs. As a result, there are 2068 remaining pairs to survey.

Blast has three different working modes: annotation, edit and search. The main mode is annotation, which allows the user to add new error annotations. The edit mode allows the user to edit and remove error annotations. The search mode allows the user to search for errors of different types. It can also create support annotations, that can later be updated by the user, and calculate and print statistics of an annotation project.

From **Figure 1**, we can see a screenshot of Blast. The MT output is shown to the annotator one segment at a time, in the upper part of the screen. A segment normally consists of a sentence and the MT output (Vietnamese) can be accompanied by a source sentence (English), a reference sentence (Vietnamese), or both. Error annotations are marked in the segments by bold, underlined, colored text, and support annotations are marked by light background colors. The bottom part of the tool, contains the error typology, and controls for updating annotations and navigation. The error typology is shown using a menu structure, where submenus are activated by the user clicking on higher levels.

Results for Error Analysis of Blast on Machine Translation Output

Overall Results

There are total 2068 language pairs containing about 60,017 words. The data have 4529 errors processing. As a result, the

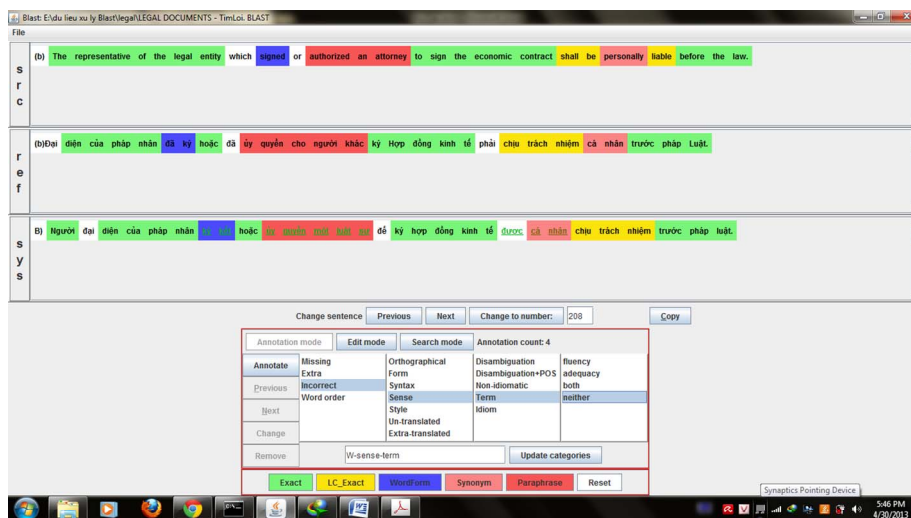


Figure 1.
Example of blast model.

average error per sentence is 2.97, but the average error per sentence with errors is 3.015. Besides, it also shows the average length of a sentence is 13.252 words (**Table 1**).

Number of Sentences with a Certain Number of Errors

The results for 2068 sentences (S) with a certain number of errors (E) illustrate that the best result from this table is 566 of 2068 sentences keeping their meanings in the context. On the other hand, the maximum errors per sentence is 13, but there are only 5 sentences in this case. There are 431 sentences having 1 error; 322 sentences containing 2 errors; 272 sentences with 3 errors; 161 sentences with 4 errors; 135 sentences with 5 errors; 81 sentences with 6 errors; 34 sentences with 7 errors; 30 sentences with 8 errors; 19 sentences with 9 errors; 7 sentences with 10 errors; and 2 sentences having 11 or 12 errors (**Table 2**).

Main Classification

According to main classification, there are 4 basic categories

Table 1.
Table of result from blast.

Number of sentences: 2068
Number of words: 60,017
Number of errors: 4529
Average error per sentence: 2.97
Average error per sentence with errors: 3.015
Average of words per sentence: 13.252

Table 2.
Number of errors per sentence.

(E)	0	1	2	3	4	5	6
(S)	566	431	322	273	161	135	81
(E)	7	8	9	10	11	12	13
(S)	34	30	19	7	2	2	5

of errors such as errors extra word errors (E), missing word errors (M), wrong word order (O) and incorrect word (W). The results of these including incorrect words (W) have the most error rate which is 58.40% with 2645 errors; wrong word order has the lowest rate including 504 errors occupying 11.13%; whereas missing errors having a little higher rate is 11.70% with 530 errors; and the second highest occupying 18.77% with 850 errors (**Figure 2**).

Cross Classifications

Based on cross classifications, there are three of four ranges of errors considerable because they still remain good context meanings with levels of adequacy, fluency, and both of them (**Figure 3**).

Results and number of errors for cross classifications, but their context meanings are acceptable. The adequacy has the highest rate with 524 errors (11.57% of total number identified errors); the fluency has unremarkable with only 0.02%; whereas there are 9 errors, but they still remain both adequacy and fluency level.

All Classifications

All classifications mean that there are 4 basic categories of errors including (M), (E), (O) and (W), each of which has sub-categories such as orthographical, form, syntax, sense, style, untranslated and extra-translated errors following.

Missing Words: 530 Errors (Approximately 11.7%)

There are 2 types of lack of words to make errors such as content missing and grammar missing, each of which also divided into 2 sub-types. Although they are considered errors, they don't affect their meanings in the context. According to the diagram there are 11 errors of content missing (M-cont-ade = 0.24%) and 5 ones of grammar missing (M-gram-ade = 0.11%) keeping good meanings like human translation. On the other hand, 350 errors of content missing (M-cont-neither = 7.73%) and 69 errors of grammar missing (M-gram-neither = 1.52%), which create bad MT sentences (**Figure 4**).

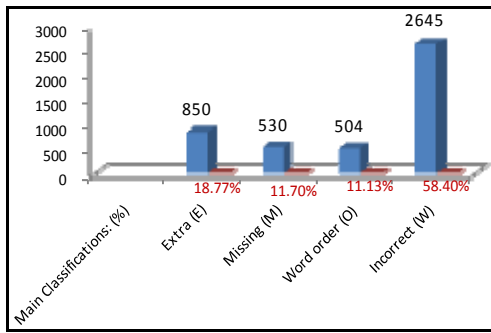


Figure 2.
Results for the four common categories of errors.

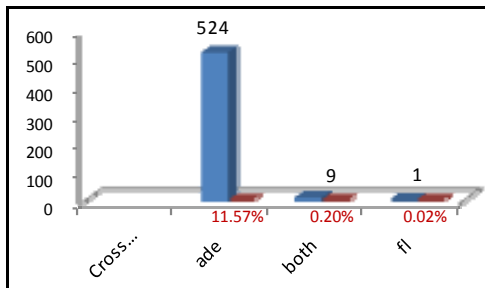


Figure 3.
Result of cross classification.

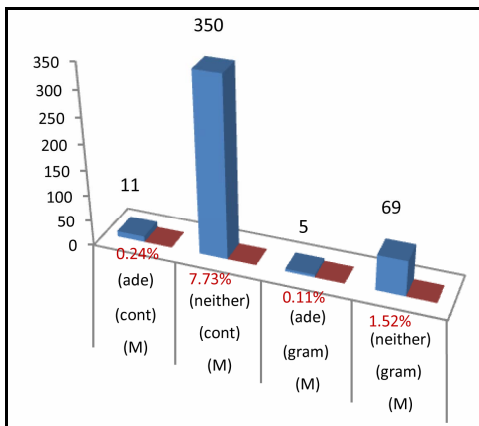


Figure 4.
Errors of missing words.

Extra Words (Or Redundant Errors): 850 Errors (About 18.77%)

This figure shows that there are 850 errors of the extra word category divided into 6 sub-categories. The first ones are considered errors, but they still remain adequate (E-cont-ade = 0.29%), fluent (E-gram-fl = 0.02%), even both adequate and fluent (E-gram-both = 0.02%) or extra grammar words (E-gram-ade = 0.22%) meanings in the context, which means that these MT output are acceptable. On the contrary, the other ones are real errors causing big problems for MT. For example, there are 433 errors due to machine translating extra words (E-cont-neither = 9.56% of total errors) to create redundant meanings compared to source sentences; and 235 errors (E-gram-neither = 5.19%) translated by MT in the wrong grammar aspect (Figure 5).

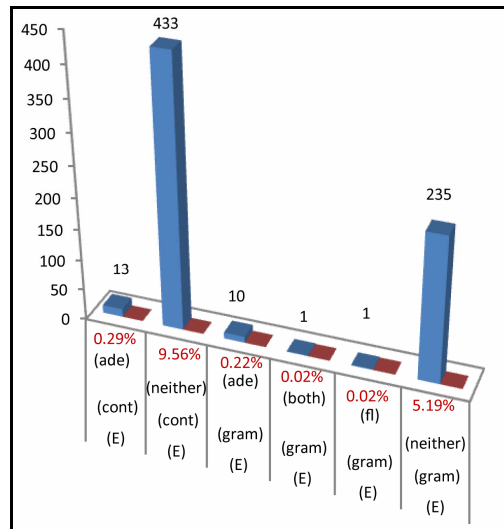


Figure 5.
Errors of extra words.

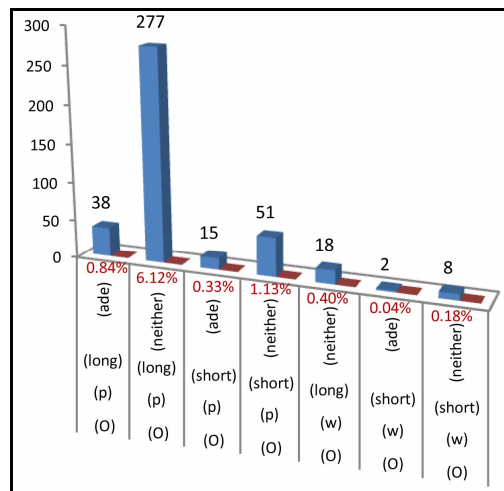


Figure 6.
Word order errors.

Wrong Word Order: 504 Errors (Approximately 11.13%)

This category contains 504 wrong word order, but they are divided into 7 sub-categories (instead of 8 ones). There are 38 phrases are errors of word orders in the long distance (p-long-ade = 0.84%) and 15 phrases are errors of word orders in the short distance (p-short-ade = 0.33%), but both of them still have good meanings in the context. By contrast, 277 phrases are errors of word orders in the long distance (p-long-neither = 6.12%) and 51 phrases are errors of word orders in the short distance (p-short-neither = 1.13%), which are really errors to make the bad results for MT (Figure 6).

Besides, the same situation also happens when there are 2 cases for single word order error in the short distance that the first one contains 2 errors (w-short-ade = 0.04%) but it still remains meanings in the context; and the other includes 8 errors (w-short-neither = 0.18%) with completely wrong meanings; finally, single word order errors in the long distance with completely wrong meanings have 18 errors (w-long-neither = 0.40%).

The Other Incorrect Words: 2.645 Errors (About 58.4%)

This category is named as incorrect/wrong words (W) which have sub-categories such as orthographical, form, syntax, sense, style, un-translated and extra-translated errors.

1) Orthographical errors belonging to lexical meanings: 216 errors (4.77%)

Errors belong to orthographical ones including punctuation (punct), capitalization (casing), number forming (number) and the others (other). Among these errors, punctuation errors have the highest rate with 70 errors, 54 of which (1.19%) create wrong meanings (W-orth-punct-neither), but the 16 other errors don't make those sentences have wrong meanings (W-orth-punct-ade). Then, the second highest rate is capitalization ones with 89 errors (1.96%) in which 39 words are really errors (W-orth-casing-neither), but other 50 errors still remain their meanings in the context (W-orth-casing-ade). Besides, the number forming errors also have the same result such as adequate meanings (W-orth-number-ade) in the context (0.13%) and real errors without suitable meanings with the source language (W-orth-number-neither). Moreover, we have named the other errors (W-orth-other-neither) because sometimes the source sentences typed wrong spelling (for example: Instead of typing "price", "prince" is typed with the meaning completely wrong), the other case such as wrong spelling in the MT sentences (**Figure 7**).

2) Form and style errors belonging to pragmatical meanings: 423 errors (9.34%)

Results for errors of style and form, which belong to pragmatical meanings. Firstly, there are 197 errors of style (W-style-ade = 3.95%), but they still good meanings in the context. Moreover, 3 errors of style have fluent and adequate meanings (W-style-both = 0.07%). On the other hand, the 134 others of style lead to wrong meanings (W-style-neither = 2.96%). Secondly, the form errors divided into 4 sub-categories such as agreement (Incorrect agreement between subject-verb, noun-adjective, past participle agreement with preceding direct object, etc.), co-reference, source mismatch, and "other" which is signed (W-form-other-ade) or (W-form-other-ade). Although these form errors have an unmarkable rate, they also show us the detailly various errors of MT. As a result, there are the only

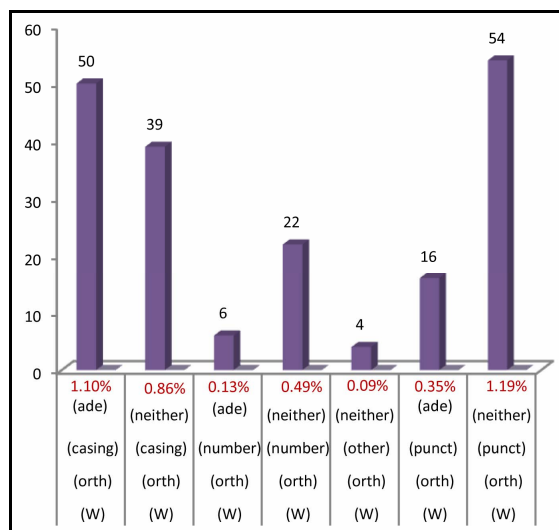


Figure 7.
Orthographical errors.

one agreement form errors with adequate meaning (W-form-agree-ade = 0.02%), 3 agreement form errors with wrong meanings (W-form-agree-neither = 0.07%), 2 co-reference form errors with their wrong meanings (W-form-coref-neither = 0.04%), 8 source mismatch form errors with their adequate meanings (W-form-mismatch-ade = 8%), 17 source mismatch form errors with their wrong meanings (W-formmismatch-neither = 0.38%), and wrong form of target/system sentences (W-form-other-neither = 0.15%) (**Figure 8**).

3) Sense errors belong to semantic meanings: 1718 errors (37.93%)

Especially, there are 156 errors of business contract terms which are chosen common entries of dictionaries but they still remain adequate meanings in the context (W-sense-term-ade = 3.44%) and 5 others having both fluent and adequate meanings (W-sense-term-both = 0.11%). Moreover, there are some kinds of sense errors in this diagram having good meanings in the context such as errors of disambiguation due to chosen wrong entries by MT (W-sense-dis-ade = 0.18%), 2 non-idiomatic sense errors (W-sense-nondiom-ade = 0.04%). On the other hand, the most errors in this diagram are 176 errors of disambiguation sense with wrong meanings (W-sense-dis-neither = 3.89%). Besides, GT cannot translate idiom, non-idiomatic words well, and they become common errors (such as W-sense-idiom-neither = 0.04% and W-sense-nondiom-neither = 0.09%) (**Figure 9**).

4) Syntax errors, extra-translated and un-translated errors: 287 errors (6.24%)

This diagram shows that errors of syntagmatic meanings which belong to a range of clause (W-syntax-clause-neither = 0.09%), wrong function (W-syntax-function-neither = 0.46%), disambiguation because of wrong part of speech (W-syntax-pos-neither = 0.49%), errors due to MT output having extra words compared to source language but unchanging meaning in the context (W-syntax-exTrans-ade = 0.02%), and the highest rate of errors due to keeping the same foreign language of source sentence (W-syntax-foreign-neither = 5.26% with 238 errors) (**Figure 10**).

Discussion

According to the criteria of identifying errors above, they can be divided into 4 basic categories of machine translation errors such as missing words (M), redundant/extra words (E), wrong word order (O), and incorrect words (W). In this paper, we only survey the common errors of machine translation from the point of view of comparative linguistics in business contracts when these legal documents are translated from English into Vietnamese by using Google Translator owing to the Blast software which analyzes them automatically and systematically in order to find an effective way to study more deeply. Hopefully, the next papers, we will describe more detailly about all kinds of MT errors such as finding them, describing them, explaining them, categorizing them, evaluating them, and suggesting latest trends to improve English-Vietnamese machine translation the most effectively.

Conclusion

In summary, the results of this study showed several important things about machine translation in Vietnam. First, machine translation has not given the good results yet. Then,

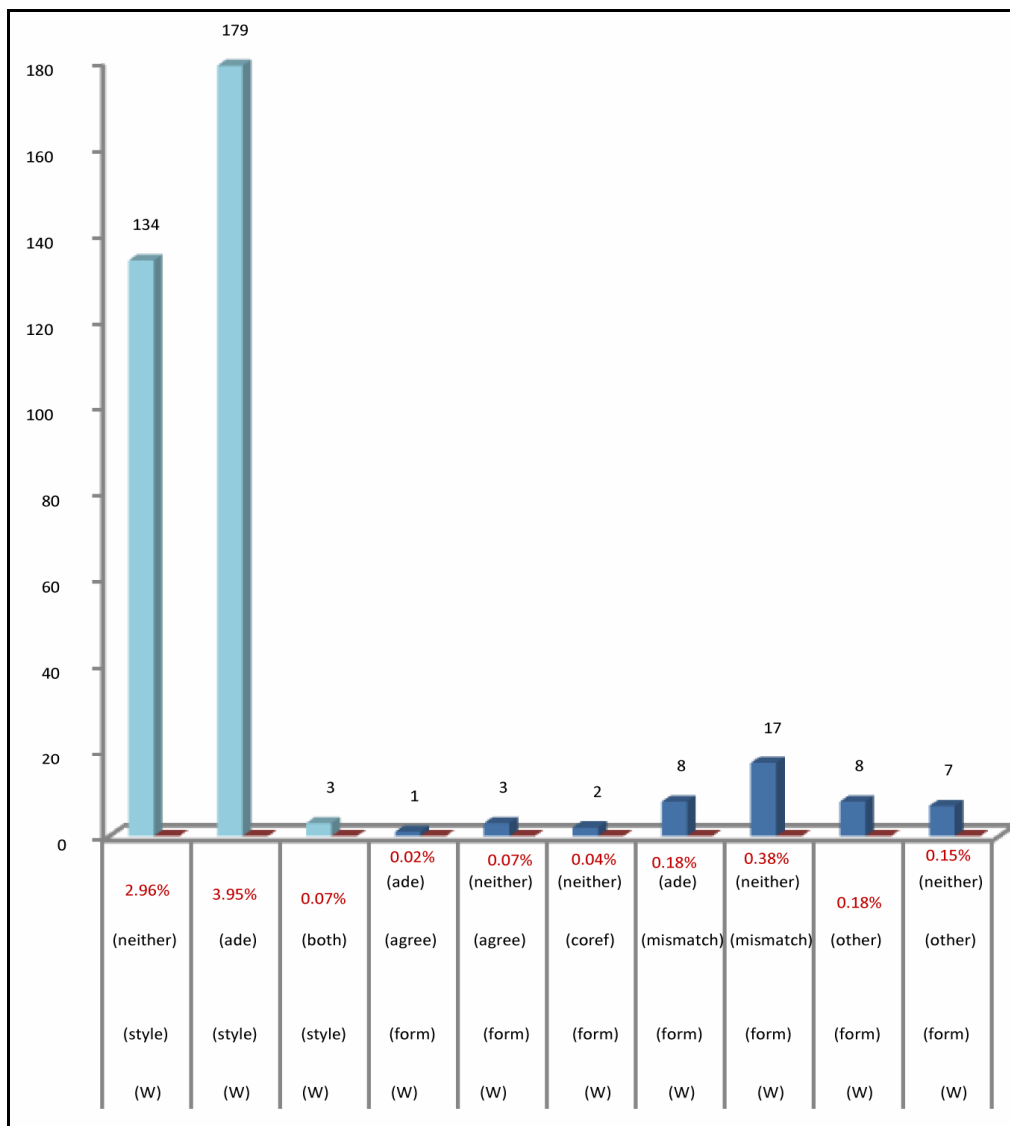


Figure 8.
Errors of form and style.

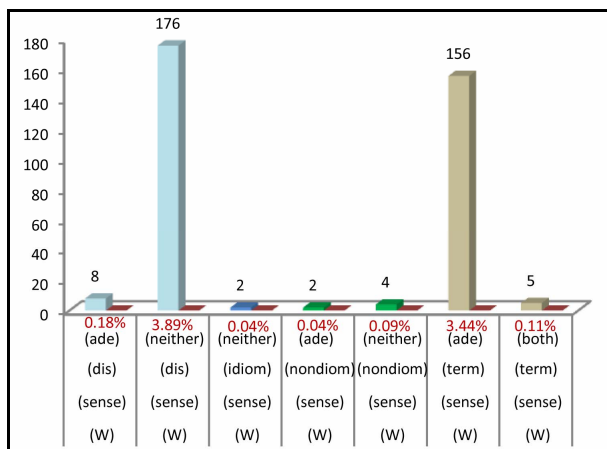


Figure 9.
Semantic errors.

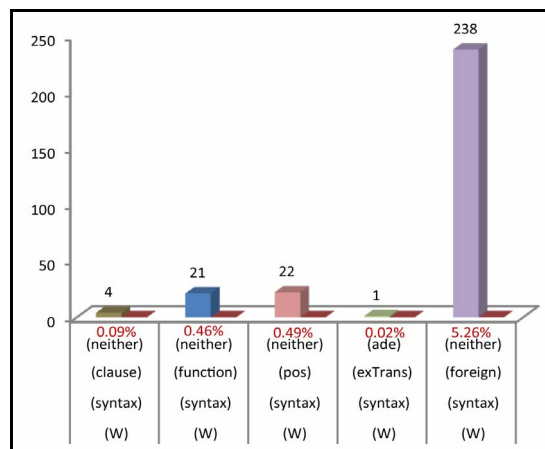


Figure 10.
Syntax errors, extra-translated and un-translated errors.

there are too little studying about machine translating from English into Vietnamese in general as well as business contracts. A part of this reason is that most of machine translation software using the method of statistical machine translation which requires the more bilingual corpus there are, the more exact results are given. However, to build a kind of corpus in English-Vietnamese costs too much. So, this problem becomes more and more difficult. On the other hand, although statistical machine translation has been confirmed its strength due to a huge corpus, its results are still suspected because there are too many errors after translating. So, there should have been the works studying more deeply for this field. It is important that linguists and computer programmers coordinate more closely to find solutions in order to limit those common errors.

REFERENCES

- David, V., Jia, X., Luis, F. D., & Hermann, N. (2006). Error analysis of statistical machine translation output. www.lrec-conf.org/proceedings/lrec2006/pdf/413_pdf.pdf
- Duong, T. N. (2008). Thử tìm hiểu độ dài của câu tiếng Anh chuyên ngành Luật (A preliminary study of the length of sentence in legal English). ĐHQGHN, NN 24: 207-215. http://tapchi.vnu.edu.vn/4_208_NN/2.pdf
- Maja, P., & Hermann, N. (2006). Morpho-syntactic information for automatic error analysis of statistical machine translation output. *Proceedings of the workshop on statistical machine translation*, New York City, June 2006, c2006 Association for Computational Linguistics, 1-6.
- Maja, P., & Hermann, N. (2007). Word error rates: Decomposition over POS classes and applications for error analysis. *Proceedings of the second workshop on statistical machine translation*, Prague, June 2007, c 2007 Association for Computational Linguistics, 48-55.
- Mary, A. F. (1996). Error classification for MT evaluation. CompuServe 1000 Massachusetts Ave. Cambridge. www.mt-archive.info/AMTA-1994-Flanagan.pdf
- Nguyen Ngoc Hung (2010). Kỹ thuật soạn thảo văn bản quy phạm pháp luật, ngôn ngữ pháp lý (Editing techniques of law normative act and legal languages). TaiLieu.VN: June 15, 2010. <http://www.tech24.vn/.../6025-Ky-thuat-soan-thao-van-ban-quy-pham-phap-m-pha...>
- Sara, S. (2011). *Blast: A tool for error analysis of machine translation output*. Linköping: Department of Computer and Information Science, Linköping University. <http://www.aclweb.org/anthology-new/P/P11/P11-4010.pdf>