

MFSH(7)

036
WANG
C-3



澳門大學
UNIVERSIDADE DE MACAU

Thesis for Master of Arts in Translation
FACULTY OF SOCIAL SCIENCES AND HUMANITIES

機器翻譯預處理可行性研究

– Google 網上翻譯系統中譯英個案研究

Pre-editing for Machine Translation

– *A Case Study of C-E Informative Text Translation with
Google Translate*

Student: Beatriz WONG

Supervisor: Dr. Hari VENKATESAN

二零一二年十一月

November 2012

ABSTRACT

Pre-editing for Machine Translation

– A Case Study of C-E Informative Text Translation with Google Translate

Beatriz Wong

University of Macau

Machine translation plays an important role in instant translation technology today but still faces challenges in terms of its quality and value, as human interventions are still involved in different dimensions such as post-editing and pre-editing. There is a dearth of study investigating the possibility of pre-editing for improved machine translation quality when compared with other aspects of MT such as automated pre-processing and post-editing; therefore, this report is designed to study the practicability of human pre-editing for machine translation.

Twenty 145-175-word Chinese informative texts randomly selected from different electronics and house appliances instruction manuals were used as the raw data in this study (ST₀). The raw source texts were compared with the source texts being pre-edited (ST₁) based on five editing criteria: (a) word segmentation, (b) sentence breaking, (c) reordering of syntactical structures, (d) omission of unnecessary aspect markers and auxiliary words, and (e) other changes. Both source text groups were machine translated into English with Google Translate and were compared and analyzed in terms of their naturalness, grammatical correctness and content clarity. A survey of target texts comparisons was also designed and included two native speakers of English as respondents to compare with the analysis of the author.

Consensus was reached among the informants and the author. Findings show that there is a higher readability in most target texts of pre-edited source texts, albeit not necessarily accompanied by better performance in terms of naturalness, grammatical correctness and content clarity.

Although there are exceptions, it is concluded that pre-editing texts for machine translation is a viable way for a better text output and could be an alternative (besides post-editing) for saving human effort in obtaining a higher quality of text output. However, applying pre-editing does not mean elimination of post-editing but rather an option for user who is familiar with both source and target languages. Pre-editing discussed in this thesis is also related to the specific language pair in concern.

Keywords: Machine translation, Google Translate, pre-editing, informative text

TABLE OF CONTENTS

Abstract	ii
Acknowledgements	iii
Declaration	iv
List of Tables	vi
List of Abbreviations	vii
1. Introduction	1
1.1 Background	1
1.2 Research Questions	4
1.3 Research Limitations	5
2. Literature Review	6
2.1 Initiatives for MT and its Quality Assessment	6
2.2 Post-Editing for MT	7
2.3 Pre-Editing for MT	8
2.4 Text Typology and MT	11
2.5 The Effects of Linguistic Difference on MT	13
2.6 Linguistic Differences between Chinese and English	14
3. Research Methodology	16
3.1 Data Collection	16
3.2 Procedure and Analysis	20
4. Data Analysis and Discussions	22
4.1 Differences between ST_0 and ST_1	24
4.1.1 Pre-editing with Criteria-c	24
4.1.2 Pre-editing with Criteria-a	28
4.1.3 Pre-editing with Criteria-d	30
4.1.4 Pre-editing with Criteria-b	32
4.1.5 Pre-editing with Criteria-e	33
4.2 Differences between TT_0 and TT_1	39
4.2.1 Comparisons of TT_0 and TT_1 with ST	39
4.2.2 Comparisons of TT_0 and TT_1 Readability	43
4.3 Discussion	46
5. Conclusion	50
References.....	52
Appendix I. Data Sets	58
Appendix II. Survey	78