

澳門大學

University of Macau

社會及人文科學學院

Faculty of Social Sciences and Humanities

翻譯碩士論文

Master of Translation Studies Thesis

以 BLEU 評估在線翻譯系統研究- Google 語言工具及 SYSTRANBox

Assessing Online Translation Systems using the BLEU Score: Google Language Tools & SYSTRANBox

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Translation Studies at the University of Macau

學生姓名（Student's Name）：羅美妍 Law Mei In

指導導師（Supervisor's Name）：Dr. Hari Venkatesan

2011 年 7 月

DECLARATION

I declare that this thesis represents my own work, except where due acknowledgment is made, and that is has not been previously included in a thesis, dissertation or report submitted to this University or to any other institution for a degree, diploma or other qualification.



Signed: _____

ACKNOWLEDGEMENT

This paper owes much to the thoughtful and helpful comments from my supervisor. Dr. Hari Venkatesan. Thanks Dr. Hari for being my supervisor and thanks for all the patience, suggestions, guidance and corrections. His suggestions helped me in all the time of research and writing of this thesis. Without him, this thesis could not be successfully completed.

My thanks also go to all the teachers in the Department of English, University of Macau, especially Professor Zhang MeiFang. I have learnt a lot from all of them. Their teaching did not only help me deepen my understanding in translation studies and give me the useful academic competence, but also broaden my way of thinking.

I would like to thank my fellow classmates and my family members for all the care, support and help.

Last but not the least; I would like to thank all those who have directly and indirectly offered their help to this thesis.

ABSTRACT

"MT is a waste of time because you will never make a machine that can translate Shakespeare." (Zhao, 2005)

People generally believe that machine translation is not trustworthy. It would be impossible to expect perfect translation from translation systems. However, at the same time, in some specific situations, machine translation is used in our daily lives to give perfect translation, e.g. METEO. One of the factors affecting the performance of the translation systems would be the form of the source language. When the source language is in controlled language, the performance of the translation system would be better. However, it appears that the result is not so satisfying when it is not in controlled language.

In this study, the performance of the two systems, Google Language Tools and SYSTRANBox, was studied. The data used was a simple form of natural language, user manual. Scholars usually compare translation systems with human evaluation. However, in this way, it takes up a much longer time and a lot of resources and man power. In order to solve this problem, scholars developed another way to evaluate translation – automatic evaluation. In this research, the BLEU score, which was

developed by the professionals of IBM, is used. BLEU is categorized under the N-gram measurement which is one of different ways to evaluate the quality of a piece of translation in a quicker way. In the end of the study, it was found that it might not be a good idea to evaluate the performance of individual systems, as it would be difficult to retrieve the problems of inadequacy of the system. However, in terms of the comparison between two or more systems, the BLEU score is still an expeditious way to gain an overview. With this study, it was found that Google Language Tools performed better than SYSTRANBox.

KEYWORDS

BLEU score, N-gram, Google Language Tools, SYSTRANBox

TABLE OF CONTENTS

DECLARATION ..... **i**

ACKNOWLEDGEMENT ..... **ii**

ABSTRACT ..... **iii**

KEYWORDS ..... **iv**

TABLE OF CONTENTS ..... **v**

TABLE OF FIGURES AND TABLES ..... **vii**

LIST OF ABBREVIATONS ..... **viii**

1. Introduction **1**

1.1. Background 1

1.2. Problem Description 3

1.3. Chapters and Sections 4

2. Literature Review **5**

2.1. Systems Introduction 5

2.1.1. SYSTRANBox 5

2.1.2. Google Language Tools 7

2.2. Rule-based Machine Translation System and Statistic-based Machine Translation System 8

2.2.1. Rule-based Machine Translation System, RBMT system 8

2.2.2. Statistic-based Machine Translation System, SBMT system 9

2.3. Problems in Evaluation 11

2.4. Different ways of Evaluation 12

2.4.1. Human Evaluation 12

2.4.2. Automatic Evaluation 14

2.5. Related Studies 18

3. Data and Research Method **22**

3.1. Data 22

3.2. Research Method 25

3.3. Other Works Related to The Usage of BLEU Score 32

4. Analysis 35

4.1. Method of Calculation 35

4.2. BLEU Score of the Candidates 38

4.3. General Analysis of the Results 39

5. Comparison between 2 Systems 43

5.1. A Check up of the BLEU result 45

6. Conclusion 65

Bibliography 69

Appendix A- User Manual of iPod in Chinese (Source Text) 71

Appendix B- User Manual of iPod in English (Reference) 85

Appendix C – Translation Provided by Google Language Tools (Candidate). 101

Appendix D – Translation provided by SYSTRANBox (Candidate) 119

Appendix E- Precision Score and Brevity Penalty of Individual Sentence (Google Language Tools) 138

Appendix F- Precision Score and Brevity Penalty of Individual Sentence (SYSTRANBox) 150

TABLE OF FIGURES AND TABLES

fig. 1: Modified Precision Score of Google Language Tools and SYSTRANBox 1
fig.2: Slopes Showing 50% Decrease in Value 1

Table 1. Results 38
Table 2. Percentage Decrease of the Precision Scores 45
Table 3. Check-up Table 46
Table 4: Precision Score and Brevity Penalty of Individual Sentence (Google) 138
Table5. Precision Score and Brevity Penalty of Individual Sentence (SYSTRANBox) 150

UNIVERSIDADE DE MACAU
仁義禮知信
澳門大學

LIST OF ABBREVIATONS

ALPAC	Automatic Language Processing Advisory Committee
ARPA	Advanced Research Projects Agency, ARPA
BLEU	Bilingual Evaluation Understudy
GLT	Google Language Tools
METEOR	Metric for Evaluation of Translation with Explicit Ordering
MT	Machine Translation
NIST	National Institute of Standard and Technology
NLPR	National Laboratory of Pattern Recognition
RBMT	Rule-based Machine Translation
SBMT	Statistic-based Machine Translation
WER	Word Error Rate

UNIVERSIDADE DE MACAU
仁義禮知信
澳門大學