

A Critical Review of Current E-to-C Machine Translation of Academic Abstracts

CHEN Yuanyuan, Eric (M-A7-5516-3)

Thesis Supervisor: Dr. Hari VENKATESAN



A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF MASTER OF ARTS AT THE UNIVERSITY OF MACAU.

AUGUST 29, 2011

Abstract of thesis entitled A Critical Review of Current E-to-C Machine Translation of Academic Abstracts. Submitted by CHEN Yuanyuan (M-A7-5516-3) for the degree of Master of Arts at the University of Macau in August 2011.

1. ABSTRACT

The purpose of this study is to look at possible strategies of evaluating machine translations and to review the capabilities of current MT systems. It also aims to measure the level of equivalence that could be supported by the machine translation systems and to compare the translation quality of different types of machine translation systems. After reviewing the background of MTs and the main-flow evaluation metrics, five latest machine translation systems (Systran, Google, PT2008, LEC Trans DotNet, and Kingsoft AIT) have been tested by assessing their translations of five abstracts adopted from *the Financial Analysts Journal*. By analyzing the translation quality from the perspective of equivalence proposed by Mona Baker (1992), this research has found that machine translation can only partially achieve lower levels of equivalence for preset domains, such as equivalence at word level and equivalence at above word level. Although most systems claim to be full-text capable, they could barely deal with equivalence at contextual level and could only make appropriate translations to a limited extent. Based on the performances of different systems, each layer of translation equivalence is discussed further. Several procedures in pre-editing are recommended for further exploration as they may possibly reduce the ambiguity of the original text.

CONTENTS

1. ABSTRACT	3
2. ACKNOWLEDGEMENT	4
3. INTRODUCTION	6
3.1. Machine Translation and Its Developments	6
3.2. Different types of MT systems and their competencies	9
4. LITERATURE REVIEW	12
4.1. Manual Evaluations	12
4.2. Automatic Evaluations (Deterministic Metrics and Learned Metrics)	17
4.2.1. String-based metrics	17
4.2.2. Dependency-based Metrics	19
4.2.3. Learned Metrics	20
4.3. Evaluations of E-to-C MT systems	21
4.4. Summaries and Discussions	22
4.5. Machine Translation and Translation Theories	23
4.6. Towards a Theoretical Framework of Equivalence	26
5. RESEARCH METHODOLOGY	33
5.1. Source Texts	33
5.2. Target MT systems	33
5.3. Testing Procedure	36
6. DISCUSSIONS	39
6.1. Equivalence at word level	39
6.2. Equivalence at above Word-level	50
6.3. Equivalence at Grammatical Level	58
6.4. Equivalence at Pragmatic Level and Contextual Equivalence	68
7. CONCLUSIONS AND RECOMMENDATIONS	71
8. REFERENCES	76
9. APPENDIX	81