

**Portuguese – Chinese Machine Translation Based on Translation
Corresponding Tree (TCT) Annotation Scheme**

by

Tang, Chi Wai (Kevin)

M-A1-6244-6

Master of Science in E-Commerce Technology

2006



Faculty of Science and Technology
University of Macau

TABLE OF CONTENTS

ABSTRACT	III
TABLE OF CONTENTS	V
LIST OF FIGURES	IX
LIST OF TABLES	XII
ACKNOWLEDGEMENT	XIV
CHAPTER 1. INTRODUCTION.....	1
1.1 LITERATURE REVIEW OF THE HISTORY IN MACHINE TRANSLATION	2
1.2 REVIEW OF MT METHODOLOGIES.....	4
1.2.1 Rule-Based Machine Translation	4
1.2.2 Statistic-Based Machine Translation	5
1.2.3 Corpus Based Machine Translation	6
1.3 MT ARCHITECTURE	7
1.3.1 Direct Translation	7
1.3.2 Interlingua Approach	9
1.3.3 Transfer approach	11
1.3.4 Interlingua vs. Transfer	13
1.4 DIFFERENT TYPES OF EBMT.....	14
1.4.1 Surface Based EBMT.....	17
1.4.2 Structure Based EBMT.....	18
1.5 REVIEW OF EXISTING MT SYSTEM.....	19
1.6 LINGUISTIC CHARACTERISTICS BETWEEN PORTUGUESE AND CHINESE	20
1.7 PROBLEM AND MOTIVATION	22
1.7.1 Existing MT Problems.....	22
1.7.2 Language deviation problem.....	24
1.7.3 Motivation.....	25
CHAPTER 2. TCT REPRESENTATION	30
2.1 TCT NOTATION	30

2.1.1	TCT Scheme	31
2.1.2	TCT notation explanation.....	32
2.2	ADVANTAGES OF TCT	34
2.2.1	Maintain grammatical consistency	35
2.2.2	Consist of Multi-level alignment	35
2.2.3	Extension flexibility	36
2.3	CASE ANALYSIS OF TCT	37
2.3.1	Total Mapping.....	37
2.3.2	Partial Mapping	38
2.3.3	Un-mapping	39
2.3	CHAPTER CONCLUSION	41
CHAPTER 3.	EXAMPLE BASE CONSTRUCTION	42
3.1	THE TCT CONSTRUCTION PROCESSES	42
3.2	CONSTRUCTING AN TCT STRUCTURE FOR A TRANSLATION EXAMPLE	44
3.2.1	Phase 1: Pre-processing of example pair.....	44
3.2.2	Phase 2: Correspondences establishment.....	46
3.2.3	Phase 3: Post-editing validation.....	50
3.3	BILINGUAL LANGUAGE PAIR CONSTRUCTION	51
3.4	CHAPTER CONCLUSION	52
CHAPTER 4.	PORTUGUESE TO CHINESE TRANSLATION BASED ON TCT	53
4.1	TRANSLATION PROCESS	53
4.2	PRE-PROCESSING STAGE	54
4.2.1	Morphological analysis	54
4.2.2	Part of speech tagging.....	55
4.2.3	Syntax parsing	56
4.3	TRANSLATION STAGES	58
4.3.1	Phase 1: Linguistic analysis.....	58
4.3.2	Phase 2: Example retrieval.....	58
4.3.3	Phase 3: Recombination and Target Generation	59
4.4	TCT TRANSLATION EXAMPLE	60
4.5	TCT GENERATION	65
4.6	CHAPTER CONCLUSION	66

CHAPTER 5:	APPLICATION TO CHINESE TO PORTUGUESE TRANSLATION	67
5.1	PROBLEMS IN CHINESE TO PORTUGUESE ANNOTATION	67
5.1.1	<i>Chinese segmentation</i>	67
5.1.2	<i>Syntax analysis</i>	69
5.2	TCT CONSTRUCTION AND TRANSLATION FOR CHINESE TO PORTUGUESE	70
5.2.1	<i>Construction with new language pair</i>	70
5.2.2	<i>Chinese to Portuguese Translation</i>	72
5.3	CONVERTING EXISTING PORTUGUESE TO CHINESE EXAMPLE TCT TREES INTO CHINESE TO PORTUGUESE TCT TREES	72
5.3.1	<i>Conversion Algorithm</i>	73
5.3.2	<i>Conversion demonstration</i>	73
5.4	CHAPTER CONCLUSION	78
CHAPTER 6.	PORTUGUESE - CHINESE TRANSLATION SYSTEM	79
6.1	SYSTEM ARCHITECTURE	79
6.1.1	<i>Communication Module</i>	80
6.1.2	<i>Translation Module</i>	81
6.1.3	<i>Clause format</i>	82
6.1.4	<i>Translation Sequence</i>	84
6.1.5	<i>Database Structure</i>	85
6.2	APPLICATION USER INTERFACE	87
6.3	TRANSLATION OPERATION	88
6.4	CHAPTER CONCLUSION	90
CHAPTER 7.	EVALUATION AND CONCLUSION	92
7.1	EVALUATION	92
7.1.1	<i>Evaluation of the Portuguese to Chinese Translation System</i>	92
7.1.2	<i>Evaluation of the Chinese to Portuguese Translation System</i>	95
7.1.3	<i>Problem analysis</i>	97
7.1.4	<i>Evaluation Conclusion</i>	103
7.2	RESEARCH CONCLUSION	104
7.3	FUTURE DEVELOPMENT	105
BIBLIOGRAPHY		107

APPENDIX I	PORTUGUESE TO CHINESE TRANSLATION EVALUATION	114
APPENDIX II	CHINESE TO PORTUGUESE TRANSLATION EVALUATION	122