

Translation and Technology

Palgrave Textbooks in Translating and Interpreting

Series Editors: **Gunilla Anderman** and **Margaret Rogers**, The Centre for Translation Studies, University of Surrey, UK

Palgrave Textbooks in Translating and Interpreting bring together the most important strands of thinking in a fast-developing field. Volumes in the series are designed for Masters students in Translation Studies and Interpreting, as well as for upper-level undergraduates considering a career in this area. Researchers and practitioners keen to refresh their knowledge of practice across the field will also find new material not readily accessible elsewhere. The series will include a core theory book and a number of individual volumes on specific topics covered in many MA courses.

Titles include:

C.K. Quah
TRANSLATION AND TECHNOLOGY

Forthcoming titles include:

Ann Corsells
PUBLIC SERVICE INTERPRETING AND TRANSLATING

Palgrave Textbooks on Translating and Interpreting

Series Standing Order ISBN 1-4039-0393-X hardback 1-4039-0394-8 paperback
(*outside North America only*)

You can receive future titles in this series as they are published by placing a standing order. Please contact your bookseller or, in case of difficulty, write to us at the address below with your name and address, the title of the series and an ISBN quoted above.

Customer Services Department, Macmillan Distribution Ltd, Houndmills, Basingstoke, Hampshire RG21 6XS, England

Translation and Technology

C.K. Quah

palgrave
macmillan



© C.K. Quah 2006

Preface © Gunilla Anderman and Margaret Rogers 2006

Softcover reprint of the hardcover 1st edition 2006

All rights reserved. No reproduction, copy or transmission of this publication may be made without written permission.

No paragraph of this publication may be reproduced, copied or transmitted save with written permission or in accordance with the provisions of the Copyright, Designs and Patents Act 1988, or under the terms of any licence permitting limited copying issued by the Copyright Licensing Agency, 90 Tottenham Court Road, London W1T 4LP.

Any person who does any unauthorised act in relation to this publication may be liable to criminal prosecution and civil claims for damages.

The author has asserted her right to be identified as the author of this work in accordance with the Copyright, Designs and Patents Act 1988.

First published 2006 by

PALGRAVE MACMILLAN

Houndmills, Basingstoke, Hampshire RG21 6XS and

175 Fifth Avenue, New York, N.Y. 10010

Companies and representatives throughout the world

PALGRAVE MACMILLAN is the global academic imprint of the Palgrave Macmillan division of St. Martin's Press, LLC and of Palgrave Macmillan Ltd. Macmillan® is a registered trademark in the United States, United Kingdom and other countries. Palgrave is a registered trademark in the European Union and other countries.

ISBN 978-1-4039-1832-1 ISBN 978-0-230-28710-5 (eBook)

DOI 10.1057/9780230287105

This book is printed on paper suitable for recycling and made from fully managed and sustained forest sources.

A catalogue record for this book is available from the British Library.

A catalog record for this book is available from the Library of Congress.

10 9 8 7 6 5 4 3 2 1

15 14 13 12 11 10 09 08 07 06

For Paul

This page intentionally left blank

Contents

<i>List of Figures, Tables and Boxes</i>	ix
<i>Series Editors' Preface</i>	xii
<i>Acknowledgements</i>	xiv
<i>List of Abbreviations</i>	xvi
Introduction	1
1 Definition of Terms	6
Machine translation	8
Human-aided machine translation	11
Machine-aided human translation	13
Human translation	14
The localization industry	19
Conclusion	20
2 Translation Studies and Translation Technology	22
Translation theory	22
Academic and professional groups in translation	26
Linguistic theories in machine translation systems	30
Translation studies	36
The translation process	43
Conclusion	55
3 Machine Translation Systems	57
Major historical developments	58
Architectures	66
Hybrid and interactive machine translation systems	84
Online machine translation systems	85
Commercial machine translation systems	87
Reasons for using machine translation systems	89
Conclusion	91

4 Computer-Aided Translation Tools and Resources	93
Workbenches	93
Translation support tools and resources	106
Localization tools	113
Commercial computer-aided translation tools	117
Standards for data interchange	119
Conclusion	128
5 Evaluating Translation Tools	129
Machine translation systems	129
Computer-aided translation tools	131
Stakeholders	133
Evaluation methods	135
General frameworks for evaluating translation tools	139
Conclusion	151
6 Recent Developments and Future Directions	152
Machine translation systems	152
Computer-aided translation tools	156
Translation systems with speech technology	157
Translation systems for minority languages	162
Translation on the web	164
Machine translation systems and the semantic web	166
The localization industry	170
Conclusion	171
7 Translation Types Revisited	172
Relationships between topics and translation types	173
Machine translation systems	191
Computer-aided translation tools	193
Conclusion	195
<i>Appendices</i>	197
<i>References</i>	204
<i>Index</i>	218

List of Figures, Tables and Boxes

Figures

1.1	Classification of translation types	7
1.2	Machine translation model	9
1.3	Machine translation system based on usage	10
1.4	Human-aided machine translation model	12
1.5	Machine-aided human translation model	13
2.1	Chronology of translation theories	23
2.2	Translation process model	29
2.3	Example of sentence representations	31
2.4	Holmes' schema of translation studies	37
2.5	A schema of applied translation studies	42
2.6	A model of the translation process including pre- and post-editing tasks	43
2.7	Example of an English SL text and its pre-edited version	44
2.8	Unedited and post-edited Spanish machine translation output	46
2.9	Example of natural and controlled languages	48
2.10	Example of original English text and its AECMA simplified English version	50
2.11	Example of natural English, simplified English and simplified Arabic texts	51
2.12	Example of an English controlled language text and its translations	53
2.13	Illustration of the translation process using a machine translation system	54
3.1	Chronology of machine translation development	58
3.2	Example of structural representations	68
3.3	Machine translation architectures	68
3.4	Direct translation model	70
3.5	Interlingua model	72
3.6	Interlingua multilingual machine translation system model	72
3.7	Transfer model	74
3.8	Transfer using tree-to-tree parsing	75
3.9	Transfer multilingual machine translation system model	76

3.10	Statistical-based model	78
3.11	Probabilities workflow in the statistical-based approach	80
3.12	Example-based model	81
3.13	Translations by online machine translation systems	87
4.1	Example of HTML code in a web page	99
4.2	Example of the web page without HTML code	99
4.3	Example of a translation workflow using a translation memory system	102
4.4	Example of an English source text	102
4.5	Pre-translation 1	103
4.6	Database model in translation memory systems	103
4.7	Reference model in translation memory systems	104
4.8	Flowchart to illustrate how to build a parallel corpus	109
4.9	Example of a text header in a corpus	110
4.10	Example of part-of-speech tagging	111
4.11	Example of a concordance for the word 'round'	112
4.12	Types of tool used in a localization project	114
4.13	Example of the translation process using a machine translation system, a translation database and a terminology database	117
4.14	Example of TMX data-sharing	120
4.15	Example of a header in TMX	121
4.16	Example of a body in TMX	122
4.17	Example of a header in TBX	124
4.18	Example of a body in TBX	125
4.19	Example of XLIFF in the localization process	126
4.20	Example of a header in XLIFF	127
4.21	Example of a body in XLIFF	127
4.22	Example of an alternate translation element in XLIFF	127
5.1	Example of a glass-box evaluation	138
5.2	Example of a black-box evaluation	139
5.3	Example of an evaluation process	141
5.4	Standardization projects for evaluating machine translation systems	142
5.5	EAGLES general evaluation framework	145
6.1	Future-use model of translation technology	154
6.2	Speech technology in translation	158

Tables

1.1	An example of a table for describing translation types	8
3.1	Example of a word entry in KAMI	67

3.2	Imitation in the example-based approach	82
3.3	Semantic similarity in the example-based approach	82
3.4	Classification of commercial machine translation systems	88
4.1	Example of perfect matching	95
4.2	Examples of fuzzy matching	96
4.3	Higher and lower threshold percentages for fuzzy matching	97
4.4	Examples of matching suggestions for 'bow'	98
4.5	Example of segments	100
4.6	Example of translation units	101
4.7	Example of English-French translation units from a database	102
4.8	Classification of commercial computer-aided translation tools	118
7.1	Degree of automation	174
7.2	Human intervention	175
7.3	Integrated tools	175
7.4	Application of theory	176
7.5	Application of theory in machine translation systems	177
7.6	Source-language texts	178
7.7	Target-language texts	180
7.8	Stages of the translation process	181
7.9	Types of text	182
7.10	Language dependency	185
7.11	Types of source language	186
7.12	Data interchange standards in translation	187
7.13	Translation groups and data interchange standards	188
7.14	Levels of evaluation	189
7.15	Methods of evaluation	190
7.16	Features in a machine translation system	191
7.17	Language coverage in machine translation systems	193
7.18	Texts and computer-aided translation tools	194
7.19	Language dependency in computer-aided translation tools	194
7.20	Number of languages in computer-aided translation tools	195

Boxes

1.1	A translator at work	14
5.1	FEMTI evaluation framework	147

Series Editors' Preface

Recent years have witnessed momentous changes in the study of Modern Languages, globally as well as nationally. On the one hand, the rapid growth of English as a universal *lingua franca* has rendered the command of other languages a less compelling commodity. On the other hand, the demand for intercultural mediators including translators and interpreters has grown as a result of many recent social, political and economic developments; these include legislative changes, the emergence of supranational organisations, the ease of travel, telecommunications, commercial pressures raising awareness of local needs, migration and employment mobility, and a heightened awareness of linguistic and human rights. Today, linguistically oriented students wishing to pursue a career in which they are able to further their interest in languages and cultures would be more inclined to choose vocationally relevant courses in which translation and interpreting play an important part rather than traditional Modern Language degrees.

Thus the possibilities for professional work in translation and interpreting have been extended, particularly as a result of developments in technology, whether as facilitating the translation process or as a means of dissemination and broadening access to communications in a range of media. The role of translation is, for example, becoming increasingly important in the context of modern media such as television and cinema, whether for documentary or entertainment purposes. And the technological possibilities for providing interpreting services, whether to the police officer on the beat or to the businessperson on a different continent, have extended the previously physically confined nature of mediating the spoken word.

Not only do these new vistas open up opportunities for the professional linguist, they also point to expanding areas of research in Translation and Interpreting Studies. Practice and theory are of mutual benefit, especially in the case of a relatively young discipline such as Translation Studies. As a result, the first aim of this series, written primarily for the MA and advanced undergraduate student, is to highlight contemporary issues and concerns in order to provide informed, theoretically based, accounts of developments in translation and interpretation. The second aim is to provide ready access for students interested in the study and pursuit of Modern Languages to

vocational issues which are of relevance to the contemporary world of translating and interpreting. The final aim is to offer informed updates to practising professionals on recent developments in the field impacting on their discipline.

Linguistic, Culture and Translation Studies
University of Surrey
Guildford
UK

GUNILLA ANDERMAN
MARGARET ROGERS

Acknowledgements

I am indebted to three individuals for their contributions. This book would have taken more time to complete if it had not been for Chooi Tsien Yeo who researched background information for me. Words cannot express my gratitude to Stephen Moore, in between translation deadlines, for putting his experiences as a professional translator into writing. I am extremely indebted to Paul Marriott for his comments and suggestions, particularly on helping to visualize a new way to depict the multidimensional classification of translation types in Chapter 7.

I would like to acknowledge especially the Duke University Libraries and Institute of Statistics and Decision Science at Duke University in providing me with the environment and research facilities where most of this book was written. Also my thanks to the National University of Singapore Libraries, George Edward Library at the University of Surrey, and the Department of Statistics and Actuarial Science at the University of Waterloo for their help.

I would also like to acknowledge the following authors, publishers and organizations for allowing the use of copyright material in this book: John Hutchins, Harold Somers and Elsevier (Academic Press Ltd) for the classification of translation types in Chapter 1; Eugene Nida and the Linguistic Society of America for the translation process in Chapter 2; John Smart and Smart Communications, Inc. for the controlled and simplified English samples in Chapter 2; Francis Bond and Takefumi Yamazaki for the KAMI Malay–English dictionary entry in Chapter 3; Paolo Dongilli and Johann Gamper for the building of a parallel corpus in Chapter 4; Tony Jewtushenko and Peter Reynolds of OASIS for XLIFF in Chapter 4; Enrique de Argaez at Internet World Stats for the statistical figure on the Internet population in Chapter 6; Michael Carl, Reinhard Schäler, Andy Way, Springer Science and Business Media, and Kluwer Academic Publishers for the model of the future use of translation technology in Chapter 6. To António Ribeiro, Tessadit Lagab, Margaret Rogers and Chooi Tsien Yeo, my most sincere thanks for translating from English into Portuguese, French, German and Chinese respectively. I am solely responsible for any translation errors that occurred. A special thank you goes to Elsie Lee, Shaun Yeo, Angeliki Petrits, Mirko Plitt and Ken Seng Tan for answering some of my queries.

To Caroline, Elizabeth, Gillian and Lyndsay, thank you for helping out with keying in corrections on the earlier drafts. Lastly, to my '*sifu*' and friend Peter Newmark, a big thank-you for all the translation discussions we had during our coffee-biscuit sessions years ago.

If it had not been for the series editors, Gunilla Anderman and Margaret Rogers, this book would not have been written. I am forever grateful to both of them for their feedback and comments. Thanks to Jill Lake of Palgrave Macmillan for her patience and understanding due to my 'country-hopping' from Southeast Asia to North America during the writing of this book.

Waterloo, Canada

CHIEW KIN QUAH

List of Abbreviations

ACRoTERMITE	Terminology of Telecommunications
AECMA	European Association of Aerospace Industries
AIA	Aerospace Industries Association of America
ALPAC	Automatic Language Processing Advisory Committee
ALPS	Automatic Language Processing System
ALT-J/C	Automatic Language Translator Japanese to Chinese
ALT-J/E	Automatic Language Translator Japanese to English
ALT-J/M	Automatic Language Translator Japanese to Malay
AMTA	Association of Machine Translation in the Americas
ASCC	Automatic Spelling Checker Checker
ASD	AeroSpace and Defence
ATA	American Translators Association
BASIC	British American Scientific International, Commercial
BLEU	Bilingual Evaluation Understudy
BSO	Buro voor Systeemontwikkeling
CAT	Computer-Aided Translation
CAT2	Constructors, Atoms and Translators
CESTA	Campagne d'Evaluation de Systèmes de Traduction Automatique
CFE	Caterpillar Fundamental English
CIA	Central Intelligence Agency
CICC	Center of International Cooperation for Computerization
CRATER	Corpus Resources and Terminology Extraction
CTE	Caterpillar Technical English
CULT	Chinese University Language Translator
DARPA	Defense Advanced Research Projects Agency
DBMT	Dialogue-based Machine Translation
DIPLOMAT	Distributed Intelligent Processing of Language for Operational Machine Aided Translation
DLT	Distributed Language Translation
DTS	Descriptive Translation Studies
EAGLES	Expert Advisory Group on Language Engineering Standards
EARS	Effective, Affordable Reusable Speech-to-Text
EDIG	European Defence Industries Group

ELDA	Evaluations and Language resources Distribution Agency
ELRA	European Language Resources Association
ENSPAN	English Spanish Machine Translation System
ENIAC	Electronic Numerical Integrator and Computer
EURODICAUTUM	European Terminology Database
EUROSPACE	Aerospace and Defence Industries Association of Europe
EUOTRA	European Translation
EVALDA	Infrastructure d'ÉVALUation à ELDA
EWG	Evaluation Working Group
FAHQT/FAHQMT	Fully Automatic High Quality (Machine) Translation
FEMTI	A Framework for the Evaluation of Machine Translation in ISLE
GENETER	Generic Model for Terminology
GETA	Groupe d'Étude pour la Traduction Automatique
HAMT	Human-Aided/Assisted Machine Translation
HICATS	Hitachi Computer Aided Translation System
HT	Human Translation
HTML	HyperText Markup Language
IAMT	International Association of Machine Translation
IATE	Inter-Agency Terminology Exchange
INTERSECT	International Sample of English Contrastive Texts
ISI	International Statistical Institute
ISLE	International Standards for Language Engineering
ISO	International Organization for Standardization
JEIDA	Japan Electronic Industry Development Association
JEITA	Japan Electronics and Information Technology Association
JICST-E	Japan Information Center of Science and Technology
KAMI	Kamus Melayu-Inggeris (Malay-English Dictionary)
KANT	Knowledge-based Accurate Translation
KGB	Komitet Gosudarstvennoi Bezopasnosti
LDC	Linguistic Data Consortium
LISA	Localisation Industry and Standards Association
LMT	Logic-based Machine Translation
LTC	Language Technology Centre
LTRAC	Language Translation Resources Automatic Console
MAHT	Machine-Aided/Assisted Human Translation
MANTRA	Machine Assisted Translation
MARTIF	Machine Readable Terminology Interchange Format

MASTOR	Multilingual Automatic Speech-to-Speech Translator
MAT	Machine-Aided/Assisted Translation
METAL	Mechanical Translation and Analysis of Language
METU	Middle East Technical University
MLIR	MultiLingual Information Retrieval
MT	Machine Translation
NAATI	National Accreditation Authority for Translators and Interpreters Ltd.
NIST	National Institute of Standards and Technology
OASIS	Organization for the Advancement of Structured Information Standards
OCP	Oxford Concordance Programme
OCR	Optical Character Recognition
OLIF	Open Lexicon Interchange Format
OS	Operating System
OSCAR	Open Standards for Container/Content Allowing Re-use
PaTrans	Patent Translation
PAHO	Pan-American Health Organization
PDA	Personal Digital Assistant
PESA	Portuguese-English Sentence Alignment
RDF	Resource Description Framework
RFC	Request for Comments
SALT	Standards-based Access to Lexicographical & Terminological Multilingual Resources
SGML	Standard Generalised Markup Language
SPANAM	Spanish American Machine Translation System
SUSY	Saarbrücker ÜbersetzungsSYstem
SYSTRAN	System Translation
TAP	Think-Aloud Protocols
TAUM	Traduction automatique à l'Université de Montréal
TBX	TermBase eXchange
TEMAA	Testbed Study of Evaluation Methodologies: Authoring Aids
TGT-1	Text-into-Gesture Translator
THETOS	Text into Sign Language Automatic Translator for Polish
TMF	Terminological Markup Framework
TMX	Translation Memory eXchange
TOLL	Thai On-Line Library
TONGUES	Act II Audio Voice Translation Guide Systems
TS	Translation Studies
TTS	Theoretical Translation Studies

WebDIPLOMAT	Web Distributed Intelligent Processing of Language for Operational Machine Aided Translation
WebOnt	Web Ontology
WWW	World Wide Web
W3C	WWW Consortium
XLIFF	XML Localisation Interchange File Format
XLT	XML Representation of Lexicons and Terminologies
XML	Extensible or Extensive Markup Language