

Advances in Empirical Translation Studies

Empirical translation studies is a rapidly evolving research area. This volume, written by world-leading researchers, demonstrates the integration of two new research paradigms: socially oriented and data-driven approaches to empirical translation studies. These two models expand current translation studies and stimulate reader debates around how development of quantitative research methods and integration with advances in translation technologies would significantly increase the research capacities of translation studies. Highly engaging, the volume pioneers the development of socially oriented innovative research methods to enhance the current research capacities of theoretical (descriptive) translation studies in order to tackle real-life research issues, such as environmental protection and multicultural health promotion. Illustrative case studies are used, bringing insight into advanced research methodologies of designing, developing and analysing large-scale digital databases for multilingual and/or translation research.

MENG JI is Professor of Translation Studies and Chinese Studies at the University of Sydney, Australia. She has published extensively on corpus translation studies, contrastive linguistics and quantitative translation methodologies.

MICHAEL OAKES is a reader in Computational Linguistics in the Research Institute in Information and Language Processing at the University of Wolverhampton, UK. His research interests are corpus linguistics, information retrieval and studies of disputed authorship.

Cambridge University Press
978-1-108-42327-4 — Advances in Empirical Translation Studies
Edited by Meng Ji , Michael Oakes
Frontmatter
[More Information](#)

Cambridge University Press
978-1-108-42327-4 — Advances in Empirical Translation Studies
Edited by Meng Ji, Michael Oakes
Frontmatter
[More Information](#)

Advances in Empirical Translation Studies

Developing Translation Resources and Technologies

Edited by

Meng Ji

University of Sydney

Michael Oakes

University of Wolverhampton



CAMBRIDGE
UNIVERSITY PRESS

Cambridge University Press
 978-1-108-42327-4 — Advances in Empirical Translation Studies
 Edited by Meng Ji, Michael Oakes
 Frontmatter
[More Information](#)

CAMBRIDGE UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre,
New Delhi – 110025, India

79 Anson Road, #06–04/06, Singapore 079906

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning, and research at the highest international levels of excellence.

www.cambridge.org

Information on this title: www.cambridge.org/9781108423274

DOI: 10.1017/9781108525695

© Cambridge University Press 2019

This publication is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published 2019

Printed and bound in Great Britain by Clays Ltd, Elcograf S.p.A.

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

Library of Congress Cataloging-in-Publication Data

Names: Ji, Meng, 1982– editor. | Oakes, Michael P., editor.

Title: Advances in empirical translation studies : developing translation resources and technologies / edited by Meng Ji, Michael Oakes.

Description: New York, NY : Cambridge University Press, 2019. | Includes bibliographical references and index.

Identifiers: LCCN 2019008010 | ISBN 9781108423274 (hardback)

Subjects: LCSH: Translating and interpreting – Research – Methodology. | Machine translating – Research – Methodology. | Computational linguistics. |

BISAC: LANGUAGE ARTS & DISCIPLINES / Linguistics / General.

Classification: LCC P306.5 .A278 2019 | DDC 418/.02–dc23

LC record available at <https://lccn.loc.gov/2019008010>

ISBN 978-1-108-42327-4 Hardback

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for external or third-party internet websites referred to in this publication and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

Contents

<i>List of Figures</i>	<i>page</i> vii
<i>List of Tables</i>	ix
<i>List of Contributors</i>	xi
<i>Preface</i>	xiii
MENG JI AND MICHAEL OAKES	
1 Advances in Empirical Translation Studies	1
MENG JI	
2 Development of Empirical Multilingual Analytical Instruments	13
MENG JI	
3 Statistics for Corpus-Based and Corpus-Driven Approaches to Empirical Translation Studies	28
MICHAEL OAKES	
4 The Evolving Treatment of Semantics in Machine Translation	53
MARK SELIGMAN	
5 Translating and Disseminating World Health Organization Drinking-Water-Quality Guidelines in Japan	77
MENG JI, GLENN HOOK AND FUKUMOTO FUMIYO	
6 Developing Multilingual Automatic Semantic Annotation Systems	94
LAURA LÖFBERG AND PAUL RAYSON	
7 Leveraging Large Corpora for Translation Using Sketch Engine	110
SARA MOZE AND SIMON KREK	
8 Developing Computerised Health Translation Readability Evaluation Tools	145
MENG JI AND ZHAOMING GAO	

vi	<i>Contents</i>	
9	Reordering Techniques in Japanese and English Machine Translation MASAAKI NAGATA	164
10	Audiovisual Translation in Mercurial Mediascapes JORGE DÍAZ-CINTAS	177
11	Exploiting Data-Driven Hybrid Approaches to Translation in the EXPERT Project CONSTANTIN ORĂSAN, CARLA PARRA ESCARTÍN, LIANET SEPÚLVEDA TORRES AND EDUARD BARBU	198
12	Advances in Speech-to-Speech Translation Technologies MARK SELIGMAN AND ALEX WAIBEL	217
13	Challenges and Opportunities of Empirical Translation Studies MENG JI AND MICHAEL OAKES	252
	<i>Index</i>	265

Figures

2.1	Introducing multi-sectoral interaction to advance environmental health risk (EHR) transition	page 15
3.1	Regression line showing the relation between the percentage frequencies of the words ‘that’ and ‘of’ in English texts written by non-native speakers	34
3.2	Confidence limits for the use of ‘that’ in English texts by native speakers of various language pairs	37
3.3	Boxplot showing how the number of types per 1,000 words depends on the interaction between mode and language	39
3.4	PCA for native-speaker-authored (n), non-native-speaker-authored (u) translated (t) texts in the ENNTT corpus	44
3.5	PCA for native-speaker-authored (n) and non-native-speaker-authored (u) texts in the ENNTT corpus	45
3.6	PCA for texts translated into English from Germanic- and Romance-language-family originals	46
3.7	PCA where translated texts are grouped according to their original languages	49
4.1	Contrasting syntactic and semantic intermediate structures	55
4.2	The Vauquois Triangle	56
4.3	A hybrid intermediate structure from the ASURA system	60
4.4	A sentence representation in the UNL interlingua	61
4.5	Sentence representations in the IF interlingua	62
4.6	Part of a phrase table for statistical machine translation	63
4.7	Two vector spaces for English, with corresponding Spanish spaces	64
4.8	Connections among rules forming a network	66
4.9	A neural network showing encoding, decoding, and attention	67
4.10	A fragment of Google’s Knowledge Graph	71
5.1	Path analysis of the translation and dissemination of WHO Guidelines in Japan	83
7.1	Building a query for the idiom <i>grasp the nettle</i> with the CQL builder	113
		vii

viii *List of Figures*

7.2	Parallel query using the Slovene and English Eur-Lex judgements subcorpora	114
7.3	Slovene–English parallel concordance for ‘postopek soodločanja’ in the ‘Sentence view’	115
7.4	Generating a multi-level frequency list in Sketch Engine	118
7.5	Frequency list featuring candidate translations with ‘framework’	118
7.6	The word-sketch functionality	120
7.7	Word sketch for ‘effective’	121
7.8	Word sketch for ‘efficient’	122
7.9	Sketch difference comparing the adjectives ‘effective’ and ‘efficient’	125
7.10	Extract from the Eur-Lex-generated sketch difference comparing ‘security’ and ‘safety’	126
7.11	Multi-word sketch for ‘legal framework’	128
7.12	The Sketch Engine thesaurus in action	130
7.13	Slovene–English bilingual word sketch for ‘učinkovit’ and ‘effective’	132
7.14	Short excerpt from the Italian–English bilingual word sketch for ‘regolamento’ and ‘regulation’	134
7.15	The ‘Change extraction options’ box in Sketch Engine	137
7.16	English term candidates extracted from the specialised parallel corpus	138
7.17	Slovene term candidates extracted from the specialised parallel corpus	139
7.18	Defining basic parameters for the new corpus in Sketch Engine	140
7.19	Step 1 of the corpus-building procedure	140
7.20	Corpus files and the ‘Compile corpus’ command	140
7.21	Single- and multi-word Slovene terms extracted from the Lisbon Treaty text	142
9.1	Outline of statistical machine translation	165
9.2	Pre-ordering method based on head finalization applied to an English binary constituent tree	166
9.3	Pre-ordering method based on head finalization applied to English and Chinese dependency structures	167
9.4	Pre-ordering method-based maximization of rank correlation coefficient applied to Japanese binary constituent tree	170
9.5	Pre-ordering method-based maximization of rank correlation coefficient applied to word-based typed dependency for Japanese	171
9.6	Correlation between human evaluation and BLEU	173
11.1	Main components of ActivaTM	206

Tables

2.1	Word frequencies of occurrence by category in FACTIVA (Australia)	<i>page</i> 21
2.2	Cross-sectoral correlation scores: Australia	23
2.3	Comparison between EPI and Multi-Sectoral Interaction (MSI)	24
2.4	Spearman’s correlation test	25
5.1	Regression weights (targets and goals)	84
5.2	Regression weights (standards and principles)	85
5.3	Regression weights (approaches and methods)	86
5.4	Regression weights (actions)	86
5.5	Regression weights between the media and industrial sectors	88
5.6	Spearman’s correlation test of effects from direct sources of information and the mass media	90
6.1	USAS semantic taxonomy	96
8.1	Exploratory factor analysis model	150
8.2	Rotated component matrix of the two-dimensional factor analysis model	151
8.3	Model fit: Discriminant analysis	154
8.4	Discriminant analysis (stepwise)	155
8.5	Standardised canonical discriminant function coefficients	156
8.6	Structure matrix of the five-item model	156
8.7	Classification result	156
8.8	Model fit	158
8.9	Wilks’ lambda	158
8.10	Classification results	159
9.1	Word segmentation, part-of-speech tagging and dependency parsing accuracies for patent sentences	169
9.2	Translation accuracy in BLEU for baseline (phrase-based SMT) and head finalization (phrase-based SMT with pre-ordering)	169

Cambridge University Press
978-1-108-42327-4 — Advances in Empirical Translation Studies
Edited by Meng Ji , Michael Oakes
Frontmatter
[More Information](#)

x	<i>List of Tables</i>	
9.3	Translation accuracy in BLEU for baseline (phrase-based SMT) and bunsetsu-based pre-ordering and word-based pre-ordering	172
12.1	Features combined in current speech-translation systems	236

Contributors

ALEX WAIBEL, Professor of Computer Science, Karlsruhe Institute of Technology (KIT), Germany

CARLA PARRA ESCARTÍN, Research Fellow, ADAPT Centre, Dublin City University, Ireland

CONSTANTIN ORĂSAN, Associate Professor, University of Wolverhampton, UK

EDUARD BARBU, Research Fellow, University of Tartu, Estonia

FUKUMOTO FUMIYO, Professor of Computer Science, Yamanashi University, Japan

GLENN HOOK, Professor Emeritus of Japanese Studies, University of Sheffield, UK

JORGE DÍAZ-CINTAS, Professor of Translation Studies, University College London, UK

LAURA LÖFBERG, Research Fellow, University of Lancaster, UK

LIANET SEPÚLVEDA TORRES, Pangeanic, Spain

MARK SELIGMAN, President, Spoken Translation, Inc., Berkeley, USA

MASAAKI NAGATA, Senior Research Fellow, NTT Communication Science Laboratories, NTT Corporation, Japan

MENG JI, Associate Professor of Translation Studies, University of Sydney, Australia

MICHAEL OAKES, Reader in Computational Linguistics, University of Wolverhampton, UK

PAUL RAYSON, Associate Professor of Computer Science, University of Lancaster, UK

SARA MOZE, Lecturer in Corpus Linguistics, University of Wolverhampton, UK

Cambridge University Press
978-1-108-42327-4 — Advances in Empirical Translation Studies
Edited by Meng Ji, Michael Oakes
Frontmatter
[More Information](#)

xii *List of Contributors*

SIMON KREK, Research Fellow, Centre for Language Resources and Technologies, University of Ljubljana; Jozef Stefan Institute, Ljubljana, Slovenia

ZHAOMING GAO, Associate Professor of Translation Studies, National Taiwan University

Preface

This book introduces socially oriented and data-driven approaches to empirical translation studies. Using case studies, it illustrates the interaction and growing interdependency between theoretical and applied translation studies in recent times, in an effort to mitigate the long-standing divide between the two, which has hampered the growth of translation studies in general.

The first four chapters of the book come under the umbrella of understanding empirical translation studies. The book opens with Ji's account in Chapter 1 of the advances which have been made in empirical translation studies, which have occurred in four main waves: pure and applied research; product- and process-oriented translation research; data concerning written translations and interpreting written translations and interpreting data; and now the development of data- and technology-intensive subfields that have important social and practical applications. In Chapter 2, Ji expands on the ideas of data intensiveness and social innovation in the field. In Chapter 3, Oakes then distinguishes between data-based and data-driven approaches, both of which are data-intensive approaches to empirical translation studies, by showing how it is possible to distinguish the original languages of translations from various European languages into English. In Chapter 4, a leading practitioner in the field, Mark Seligman, describes the evolving treatment of semantics in machine translation. The theme of the next four chapters is the development of multilingual resources and data-driven analytical tools. In keeping with the social turn in empirical translation studies, Ji, Hook and Fukumoto look at the translation and dissemination of WHO drinking-water-quality guidelines in Japan in Chapter 5. In Chapter 6, Löfberg and Rayson describe their work on the USAS semantic-tagging system, which assigns a meaning code to each word in the input text. They describe how the tool has been made to work for many languages, and look at its social usefulness. In Chapter 7, Moze and Krek then write about Sketch Engine, which is a powerful suite of corpus tools for cross-linguistic analysis. It can be used with large web-based corpora for computer-aided translation. In Chapter 8, Ji and Gao describe digital analytical instruments that they have developed for assessing the readability of Chinese health translations. The following group of four chapters is concerned with the

development of practical and industrial applications for end users. In Chapter 9, Nagata describes his experience of how machine-translation techniques can be enhanced by reordering the words of the text, when the source and target languages have very different word orders. In Chapter 10, Díaz-Cintas gives a comprehensive overview of advances in audiovisual translation and subtitling. Orăsan et al. in Chapter 11 follow the data-driven paradigm by describing the EU-funded EXPERT project, which exploited empirical approaches to translation, including the evaluation of machine-translated outputs. Seligman again draws on his practical experience in Chapter 12 to describe advances in speech-to-speech translation technologies. In Chapter 13, Ji and Oakes conclude by looking at the opportunities and challenges ahead for empirical translation studies.

We hope that this book will bring together theoretical (descriptive) and applied translation studies, so as to expand the horizon of the field of empirical translation studies as a whole, and that we have shown how data-intensive approaches will enable the tackling of new and significant social and research issues.

MENG JI AND MICHAEL OAKES
Sydney and Wolverhampton
March 2018