

ROUTLEDGE STUDIES IN TRANSLATION TECHNOLOGY

The Future of Translation Technology

Towards a World without Babel

Chan Sin-wai



The Future of Translation Technology

Towards a World without Babel

Chan Sin-wai

First published 2017
by Routledge
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

and by Routledge
711 Third Avenue, New York, NY 10017

*Routledge is an imprint of the Taylor & Francis Group,
an informa business*

© 2017 Chan Sin-wai

The right of Chan Sin-wai to be identified as author of this work
has been asserted by him in accordance with sections 77 and 78
of the Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this book may be reprinted or
reproduced or utilized in any form or by any electronic,
mechanical, or other means, now known or hereafter invented,
including photocopying and recording, or in any information
storage or retrieval system, without permission in writing from the
publishers.

Trademark notice: Product or corporate names may be trademarks
or registered trademarks, and are used only for identification and
explanation without intent to infringe.

British Library Cataloguing in Publication Data

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

Library of Congress Cataloging-in-Publication Data

Names: Chan, Sin-wai, author.

Title: The future of translation technology : towards a world without
Babel / by Sin-wai Chan.

Description: Milton Park, Abingdon, Oxon ; New York, NY : Routledge,
[2017] | Series: Routledge Studies in Translation Technology; 1 |
Includes bibliographical references and index.

Identifiers: LCCN 2016015991 | ISBN 9781138842045 (hardback) |
ISBN 9781315731865 (ebook)

Subjects: LCSH: Translating and interpreting—Technological innovations.

Classification: LCC P306.97.T73 C53 2017 | DDC 418/.020285—dc23

LC record available at <https://lcn.loc.gov/2016015991>

ISBN: 978-1-138-84204-5 (hbk)

ISBN: 978-1-315-73186-5 (ebk)

Typeset in Galliard
by Apex CoVantage, LLC



Printed and bound in Great Britain by
TJ International Ltd, Padstow, Cornwall

The Future of Translation Technology

Technology has revolutionized the field of translation, bringing drastic changes to the way translation is studied and done. To an average user, technology is simply about clicking buttons and storing data. What we need to do is to look beyond a system's interface to see what is at work and what should be done to make it work more efficiently. This book is both macroscopic and microscopic in approach: macroscopic as it adopts a holistic orientation when outlining the development of translation technology in the last forty years, organizing concepts in a coherent and logical way within a theoretical framework, and predicting what is to come in the years ahead; microscopic as it examines in detail the five stages of technology-oriented translation procedure and the strengths and weaknesses of the free and paid systems available to users. *The Future of Translation Technology* studies, among other issues:

- The Development of Translation Technology
- Major Concepts in Computer-aided Translation
- Functions in Computer-aided Translation Systems
- A Theoretical Framework for Computer-aided Translation Studies
- The Future of Translation Technology

This book is an essential read for scholars and researchers of translational studies and computational linguistics, and a guide to system users and professionals.

Chan Sin-wai, Professor at the School of Humanities and Social Science, The Chinese University of Hong Kong, Shenzhen, has taught and conducted research on translation technology, translation studies, and Chinese-English translation for many years. He has so far authored, edited, compiled, and translated forty-eight academic books, in sixty-two volumes. His book publications in 2016 include *A New Comprehensive Chinese-English Dictionary* (in three volumes), *Routledge Encyclopedia of Translation Technology* and *Routledge Encyclopedia of the Chinese Language*.

Routledge studies in translation technology

Series Editor

Chan Sin-wai

For a full list of titles in this series, please visit <https://www.routledge.com/series/RSITT>

1 The Future of Translation Technology

Towards a World without Babel

Chan Sin-wai

Illustrations

Figures

2.1	Screenshot of Dashboard of SDL-Trados 2015	58
2.2	Screenshot of SDL-Trados: 'Progress of Individual Projects'	58
2.3	Screenshot of SDL-Trados: 'Completion of Projects: Today'	59
2.4	Screenshot of SDL-Trados: 'Progress on the Translation of Files'	59
3.1	Editing Interface of Wordfast Anywhere	123
3.2	MadCap Lingo V9	124
4.1	Main Window of Anaphraseus	173
4.2	User Data of ESteam	178
4.3	Translation Editor of Isometry	184
4.4	User Interface of LogiTerm Web	187
4.5	Project Management Page of Memsource Cloud	191
4.6	Editing Interface of MetaTaxis	192
4.7	Translation Editor of MT2007	193
4.8	Editor of OmegaT	197
4.9	SDL TMS 2013	201
4.10	SDL WorldServer	203
4.11	Similis Manager	204
4.12	Similis: Statistics of Translation Tasks	205
4.13	User Interface of Snowman	206
4.14	Screenshot of TMs and Glossaries Management in Wordfast Anywhere	212
4.15	Screenshot of the Login Window in Wordfast Anywhere	213
4.16	Working Document of WordFisher	214
4.17	XTM Cloud	215
4.18	Apertium Translation Interface	223
4.19	Lucy LT KWIK Translator	228
4.20	SDL Automated Translation Solutions	231

Tables

2.1	Statistics of Languages Support by Seven CAT Systems	51
3.1	Project Wizards of Nine CAT Systems	71
3.2	Designations of Terminology Databases of Nine CAT Systems	80
3.3	Designations of Translation Memory Databases of Nine CAT Systems	84
3.4	Editing Environments of Nine CAT Systems	121
4.1	Comparison between the Costs of Translation Performed by Lingotek Community Translation and Industry Standards (Adapted from Vandenberg 2009: 21)	186
4.2	Most Popular TM tools with Non-TM-users (Adapted from Table 3 in Lagoudaki 2006: 15)	195
4.3	Top 10 Most Widely Used TM Tools (Adapted from Table 7 in Lagoudaki 2006: 20)	195
4.4	Top 10 Most Widely Used TM Tools by other than Windows OS Users (Adapted from Figure 24 in Lagoudaki 2006: 21)	196
4.5	CAT Systems and their MT Links	221
4.6	Online Systems Linked to CAT Systems	222
4.7	Bi-directional Language Pairs Supported in AppTek's TranSphere® MT	224
4.8	Language Pairs for which Uni-directional Translation is Supported in AppTek's TranSphere® MT	224
5.1	A Framework for Computer-aided Translation Studies	247
6.1	Divisions of Translation	259

Preface

This book studies the development of translation technology in the last four decades from a historical, conceptual, functional, and theoretical perspective, as well as how translation technology will move ahead in the years to come, eventually getting to the stage where the Tower of Babel will no longer exist and the entire world will be free from language barriers.

Before we discuss the role of translation technology in the future world of translation, an explanation of the term may be in order. According to Lynne Bowker, author of *Computer-Aided Translation Technology: A Practical Introduction*, translation technology refers to different types of computerized tools used in the translation process. This definition covers the general tools used in computing, such as word processors and electronic resources, and the specific tools used in translating, such as corpus-analysis tools and terminology management systems (Bowker 2002: 5–9). A broader definition of the term is given in *A Dictionary of Translation Technology*, which describes translation technology as ‘a branch of translation studies that specializes in the issues and skills related to the computerization of translation’ (Chan 2004: 258). In this book, the latter definition is followed, which means that translation technology is considered to cover both computer-aided translation and machine translation tools, and is regarded academically as an integral part of translation studies and professionally as an important component of vocational training for translators.

It is true that the major subject of this book is translation technology, but it should be emphasized that the human factor is also important in all translating activities, either manual or mechanical. In this digital age, all forms of translation are essentially computer-aided translation, with varying degrees of human intervention. Machine translation, which is fully automatic in text generation, needs human intervention at the stage of post-editing and, preferably, also at the pre-translation stage in the form of pre-editing. Computer-aided translation, which is interactive, needs human intervention at all stages of translation. Finally, human translation, which is manual, needs the computer as a means to produce the target text.

As the title of this book suggests, we work towards a world without Babel. Babel, as we know, refers to a story in Genesis of the Bible, which says:

And the Lord said, ‘Behold, they are one people, and they have all one language; and this is only the beginning of what they will do; and nothing

that they propose to do will now be impossible for them. Come, let us go down, and there confuse their language, that they may not understand one another's speech.' So the Lord scattered them abroad from there over the face of all the earth, and they left off building the city. Therefore its name was called Babel, because there the Lord confused the language of all the earth; and from there the Lord scattered them abroad over the face of all the earth.

(*The English Standard Bible*, Genesis, 11: 3)

The Tower of Babel can perhaps be regarded as a happy beginning for translators, as it was due to the lack of linguistic sameness that translation was needed to facilitate communication. As we move into the second decade of the twenty-first century, we begin to realize that technology has helped to remove the language barriers that have separated different language communities in the world for a very long time. With the rapid advances in computer science and translation technology, it is quite likely that in the span of a few decades, the Tower of Babel will collapse and we will be able to communicate without any kind of language and cultural barriers.

This book, which explores the issues mentioned above, is divided into six chapters. Chapter 1 provides a chronological and critical analysis of the development of translation technology, mainly computer-aided translation, in different countries and regions since its inception more than forty years ago. Chapter 2 discusses the seven major concepts in translation technology, which have shaped the development of functions in the systems, and the realization of the concepts through these functions. Chapter 3 examines the systems and the functions that are used during the five stages of data management: the initiating, data creation, data processing, data editing, and finalizing stages. Chapter 4 explores various aspects of free and paid computer-aided translation systems. Chapter 5 presents a framework of computer-aided translation studies proposed by the author to show that translation technology has come of age, academically and professionally. Lastly, Chapter 6 looks into the future of translation technology based on the main trends that have emerged in recent decades.

It is hoped that this study on translation technology will shed new light on the nature and application of computer-aided translation and machine translation, thus opening up a new vista for this area.

References

- Bowker, Lynne (2002) *Computer-aided Translation Technology: A Practical Introduction*, Ottawa: University of Ottawa Press.
- Chan, Sin-wai (2004) *A Dictionary of Translation Technology*, Hong Kong: The Chinese University Press.

Acknowledgements

In completing this book, my gratitude goes to scholars and institutions which invited me to give talks on translation technology in recent years and colleagues at the Centre for Translation Technology, The Chinese University of Hong Kong.

The actual writing of this book and the formation of concepts have been greatly helped by talks and lectures on the various aspects of computer-aided translation that I gave at educational institutions in Taiwan, China, and Hong Kong in the last nine years. These include lectures and speeches delivered at the Department of English Language and Literature of Soochow University in Taiwan, the Department of Foreign Languages of Beihang University, Institute of Computational Linguistics of Peking University, School of Microelectronics and Software of Peking University, the School of Interpreting and Translation Studies of Guangdong University of Foreign Studies, the Faculty of the Humanities and Social Sciences of the Open University of Hong Kong, the School of Chinese of the University of Hong Kong, and the School of Foreign Languages of the University of Shanghai for Science and Technology. I would like to thank them all for their invitations, which helped me to complete this book project sooner than I had expected.

Special mention must be made of the lecture series on computer-aided translation given at Beihang University in Beijing, China. At the invitation of Professor Qian Duoxiu, Chairperson of the Department of Translation and Interpreting at Beihang, I went there in October 2007 for two weeks to deliver a series of lectures on 'Perspectives on Computer-aided Translation'. These lectures form the basis of this book. The finalization of the contents of this book was greatly helped by a public lecture I gave at the Central Library of Hong Kong on 6 December 2008. I would like to express my gratitude to Professor Wong Kwok-pun, former Professor of Translation and Chairman of the Department of Translation of The Chinese University of Hong Kong, who initiated and organized the Central Library talks.

For data collection, data analysis, text translation, and proofreading, my thanks go to my colleagues at the Centre for Translation Technology at The Chinese University of Hong Kong, in particular Miss Florence Li Wing Yee, who devoted so much of her time to this publication project, and Miss Sara Roman Galdran,

who read through the manuscript and made many useful suggestions. For the publication of this book and creation of the Routledge Translation Technology Studies, my gratitude goes to Miss Christina Low, Commissioning Editor of Routledge.

Lastly, I dedicate this book to Proefessor Kenneth Young, Master of C.W. Chu College, The Chinese University of Hong Kong, for his encouragement and support in my academic development.

The author and publishers would like to thank the following copyright holders for permission to reproduce the following material:

“The Development of Translation Technology: 1967–2013” (pp. 3–31) from Chan Sin-wai (ed.) *The Routledge Encyclopedia of Translation Technology*. Published with the permission of Routledge.

“Computer-aided Translation: Major Concepts” (pp. 32–67) from Chan Sin-wai (ed.) *The Routledge Encyclopedia of Translation Technology*. Published with the permission of Routledge.

“A Framework for Computer-aided Translation Studies” (pp. 57–74) from Laurence K.P. Wong and Chan Sin-wai (ed.) *The Dancer and the Dance: Essays in Translation Studies* as Chapter 5: “A Theoretical Framework for Computer-aided Translation Studies”. Published with the permission of Cambridge Scholars Publishing.

Figures “Screenshot of Dashboard of SDL-Trados 2015”, “Screenshot of SDL-Trados: ‘Progress of Individual Projects’”, “Screenshot of SDL-Trados: ‘Completion of Projects: Today’”, “Screenshot of SDL-Trados: ‘Progress on the Translation of Files’”, “SDL TMS 2013”, “SDL WorldServer”, and “SDL Automated Translation Solutions” featured in Chapters 2 and 4. Used with kind permission of SDL PLC.

Figures “Editing Interface of Wordfast Anywhere”, “Screenshot of TMs and Glossaries Management in Wordfast Anywhere”, and “Screenshot of the Login Window in Wordfast Anywhere” featured in Chapters 3 and 4. Used with kind permission of Wordfast LLC.

Figure “MadCap Lingo V9”, featured in Chapter 3. Used with kind permission of MadCap Software, Inc.

Figure “Main Window of Anaphraseus”, featured in Chapter 4. Used with kind permission of Anaphraseus (Oleg Tsygany).

Figure “User Data of ESteam”, featured in Chapter 4. Used with kind permission of ESteam AB.

Figure “Translation Editor of Isometry”, featured in Chapter 4. Used with kind permission of Finite Field (Toshiya Kazuyoshi).

Figure “User Interface of LogiTerm Web”, featured in Chapter 4. Used with kind permission of Terminotix Inc.

Figure “Project Management Page of Memsource Cloud”, featured in Chapter 4. Used with kind permission of Memsource.

Figure “Editing Interface of MetaTaxis”, featured in Chapter 4. Used with kind permission of MetaTaxis Software and Services.

Figure “Translation Editor of MT2007”, featured in Chapter 4. Used with kind permission of Andrew Manson.

Figure “Editor of OmegaT”, featured in Chapter 4. Used with kind permission of OmegaT (Marc Prior).

Figures “Similis Manager” and “Similis: Statistics of Translation Tasks” featured in Chapter 4. Used with kind permission of Lingua et Machina.

Figure “User Interface of Snowman”, featured in Chapter 4. Used with kind permission of Foshan Snowman Computer Co. Ltd.

Figure “Working Document of WordFisher”, featured in Chapter 4. Used with kind permission of Tibor Környei.

Figure “XTM Cloud”, featured in Chapter 4. Used with kind permission of XTM International Ltd.

Figure “Apertium Translation Interface”, featured in Chapter 4. Used with kind permission of Apertium (Mikel L. Forcada).

Figure “Lucy LT KWIK Translator”, featured in Chapter 4. Used with kind permission of Lucy Software and Services GmbH.

While we have made every effort to contact copyright holders of material used in this volume, we would be grateful to hear from any we were unable to contact.

Contents

<i>List of illustrations</i>	vii
<i>Preface</i>	ix
<i>Acknowledgements</i>	xi
1 The development of translation technology: 1967–2014	1
2 Major concepts in computer-aided translation	30
3 Functions in computer-aided translation systems	68
4 Computer-aided translation: Free and paid systems	167
5 A theoretical framework for computer-aided translation studies	244
6 The future of translation technology	258
<i>Index</i>	279

1 The development of translation technology

1967–2014

Introduction

The history of translation technology, or more specifically computer-aided translation (CAT), is short, but its development has been fast (Chan 2015: 3). It is generally recognized that the failure of machine translation in the 1960s as a result of the infamous ALPAC report (1966) led to the emergence of computer-aided translation. The development of computer-aided translation in the course of the last forty-seven years, from its beginning in 1967 to 2014, can be divided into four periods. The first period, which goes from 1967 to 1983, is a period of germination. The second period, covering the years between 1984 and 1992, is a period of steady growth. The third period, from 1993 to 2002, is a decade of rapid growth. The last period, which extends from 2003 to 2014, is a period of global development.

1967–1983: A period of germination

Computer-aided translation, as mentioned above, originated from machine translation, which, in turn, resulted from the invention of the computer. By 1966, when the ALPAC report was published, machine translation had made considerable progress in a number of countries since the invention of the first computer, ENIAC, in 1946. Several events that took place over these two decades are worth noting. In 1947, merely one year after the advent of the computer, Warren Weaver, President of the Rockefeller Foundation, and Andrew D. Booth of Birkbeck College, London University, proposed to make use of the newly invented computer to translate natural languages, becoming the first two scholars who discussed the possibility of incorporating computers into the translation process (Chan 2004: 290–291). In 1949, Warren Weaver wrote a memorandum for peer review outlining the prospects of machine translation, which went down in history as ‘Weaver’s Memorandum’. In 1952, Yehoshua Bar-Hillel held the first conference on machine translation at the Massachusetts Institute of Technology. Some of the papers that were presented in the conference were compiled by William N. Locke and Andrew D. Booth into an anthology entitled *Machine Translation of Languages: Fourteen Essays*, the first

2 The development of translation technology

book on machine translation (Locke and Booth 1955). In 1954, Leon Dostert of Georgetown University and Peter Sheridan of IBM used the IBM701 machine to make a public demonstration of the translation of Russian sentences into English, which marked a milestone in machine translation (Chan 2004: 125–126; Hutchins 1999). Later that year, the inaugural issue of *Mechanical Translation*, the first journal in the field of machine translation, was published by the Massachusetts Institute of Technology (Yngve 2000: 50–51). In 1962, the Association for Computational Linguistics was founded in the United States, and the journal of the association, *Computational Linguistics*, began to be published. It was roughly estimated that by 1965, there were sixteen countries or research institutions engaged in studies on machine translation, including the United States, the former Soviet Union, the United Kingdom, Japan, France, West Germany, Italy, the former Czechoslovakia, the former Yugoslavia, East Germany, Mexico, Hungary, Canada, Holland, Romania, and Belgium (Zhang 2006: 30–34).

The development of machine translation in the United States since the late 1940s, however, fell short of expectations. In 1963, the Georgetown machine translation project was terminated, signifying the end of the largest machine translation project in the United States (Chan 2004: 303). In 1964, the government of the United States set up the Automatic Language Processing Advisory Committee (ALPAC), which comprised seven experts in the field, to enquire into the state of machine translation (ALPAC 1966; Warwick 1987: 22–37). In 1966, the report of the Committee, entitled *Languages and Machines: Computers in Translation and Linguistics*, pointed out that ‘there is no immediate or predictable prospect of useful machine translation’ (ALPAC 1966: 32) and, as machine translation was twice as expensive as human translation, it failed to meet people’s expectations. The Committee thus recommended that resources to support machine translation should be terminated. The report also mentions that ‘as it becomes increasingly evident that fully automatic high-quality machine translation was not going to be realized for a long time, interest began to be shown in machine-aided translation’ (ALPAC 1966: 25). Therefore, the focus on machine translation shifted to machine-aided translation that was ‘aimed at improving human translation, with an appropriate use of machine aids’ (ALPAC 1966: iii), and they concluded that ‘machine-aided translation may be an important avenue toward better, quicker, and cheaper translation’ (ALPAC 1966: 32). The ALPAC report dealt a serious blow to machine translation in the United States, which was to remain stagnant for more than a decade, and it also had a negative impact on the research on machine translation in Europe and Russia. However, this provided an opportunity for machine-aided translation to come into being. All these show that the birth of machine-aided translation is closely related to the development of machine translation.

Computer-aided translation, nevertheless, would not have been possible without the support of related concepts and software. It was no mere coincidence that the idea of a translation memory, which is one of the major concepts and functions of computer-aided translation, emerged during this period. According

to John Hutchins, the concept of translation memory can be traced back to the period between the 1960s and the 1980s (Hutchins 1998). In 1978, when Alan Melby of the Translation Research Group of Brigham Young University conducted research on machine translation and developed an interactive translation system ALPS (Automated Language Processing Systems), he incorporated the idea of translation memory into a tool known as 'Repetitions Processing', which aimed at finding matched strings (Kingscott 1984: 27–29; Melby 1978; Melby and Warner 1995: 187). In the following year, Peter Arthern, in his paper on the issue of whether machine translation should be used in a conference organized by the European Commission, proposed the method of 'translation by text-retrieval' (Arthern 1979: 93). According to Arthern:

This information would have to be stored in such a way that any given portion of text in any of the languages involved can be located immediately . . . together with its translation into any or all of the other languages which the organization employs.

(Arthern 1979: 95)

In October 1980, Martin Kay published an article 'The Proper Place of Men and Machines in Language Translation' at the Palo Alto Research Center of Xerox. He proposed to create a machine translation system in which the display on the screen was divided into two windows. The text to be translated would appear in the upper window, while the translation would be composed in the bottom one to allow the translator to edit the translation with the help of simple facilities peculiar to translation, such as aids for word selection and dictionary consultation, which are labeled by Kay as a 'translator amanuensis' (Kay 1980: 9–18). In view of the level of word-processing capacities at that time, his proposal was inspiring to the development of computer-aided translation and exerted a huge impact on its research later on. Kay is generally considered as a pioneer in proposing an interactive translation system.

It can be seen that the idea of translation memory was established in the late 1970s and early 1980s (Bruderer 1975: 258–261, 1977: 529–556). Hutchins believes that the first person to propose the concept of translation memory was Arthern. However, as Melby and Arthern proposed the idea almost at the same time, both could be considered as forerunners. In addition, it should be acknowledged that Arthern, Melby, and Kay made a great contribution to the growth of computer-aided translation in its early days.

The first attempt to deploy the idea of translation memory in a machine translation system was made by Alan Melby and his co-researchers at Brigham Young University, who jointly developed the Automated Language Processing System, or ALPS for short. This system provided access to previously translated segments which were identical (Hutchins 1998: 291). Some scholars classify this type of full match as a function of first-generation translation memory systems (Elita and Gavrila 2006; Gotti, Langlais, Macklovitch, Bourigault, Robichaud, and Coulombe 2005; Kavak 2009). One of the major shortcomings

4 *The development of translation technology*

of this generation of computer-aided translation systems was that sentences with full matching were very small in number, minimizing the reusability of the translation memory and the role of the translation memory database (Wang 2011: 141).

Some researchers around 1980 began to collect and store translation samples with the intention of redeploying and sharing their translation resources. Constrained by the limitations of computer hardware (such as limited storage space), the cost of building a bilingual database was high, and with the immaturity in the algorithms for bilingual data alignment, translation memory technology was forced to remain in a stage of exploration. As a result, a truly commercial computer-aided translation system did not emerge during the sixteen years of this period, and, therefore, translation technology did not have an impact on the translation practice and translation industry (Zachary 1979: 13–28).

1984–1992: A period of steady growth

The eight-year period between 1984 and 1992 is characterized by a steady growth of computer-aided translation and by some developments that took place: corporate operation, in 1984; system commercialization, in 1988; and regional expansion, in 1992 (Marčuk 1989: 682–688).

Company operation

It was during this period that the first computer-aided translation companies, Trados in Germany and Star Group in Switzerland, were founded. These two companies later had a great impact on the development of computer-aided translation.

The German company was founded by Jochen Hummel and Iko Knyphausen in Stuttgart, Germany, in 1984. The name Trados GmbH stood for ‘TRANslation and DOcumentation Software’. This company was set up initially as a language service provider (LSP) to work on a translation project that they had received from IBM. As the company later developed computer-aided translation to help complete the project, the establishment of Trados GmbH is regarded as the starting point of the period of steady growth in computer-aided translation (Garcia and Stevenson 2005: 18–31; <http://www.lspzone.com>).

Of equal significance was the founding of the Swiss company STAR AG in the same year. STAR, an acronym of ‘Software, Translation, Artwork, and Recording’, provided manual technical editing and translations with information technology and automation. Two years later, STAR opened its first foreign office in Germany in order to serve the increasingly important software localization market, and started developing two software products, namely GRIPS and Transit, for information management and translation memory, respectively. At the same time, client demand and growing export markets led to the establishment of additional overseas locations in Japan and China. The STAR Group still plays an important role in the translation technology industry (<http://www.star-group.net>).