BENJAMINS TRANSLATION LIBRARY

Computers and Translation

A translator's guide

EDITED BY Harold Somers

Computers and Translation

A translator's guide

Edited by

Harold Somers

UMIST

John Benjamins Publishing Company

Amsterdam/Philadelphia

™ The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences – Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

Library of Congress Cataloging-in-Publication Data

Computers and translation : a translator's guide / edited by Harold Somers.
p. cm. (Benjamins Translations Library, ISSN 0929–7316 ; v. 35)
Includes bibliographical references and indexes.
1. Machine translating. I. Somers, H.L. II. Series.

P308. C667 2003
418'.02'0285-dc21 2003048079
ISBN 90 272 1640 1 (Eur.) / 1 58811 377 9 (US) (Hb; alk. paper)

© 2003 – John Benjamins B.V.
No part of this book may be reproduced in any form, by print, photoprint, microfilm, or any other means, without written permission from the publisher.

John Benjamins Publishing Co. · P.O. Box 36224 · 1020 ME Amsterdam · The Netherlands
John Benjamins North America · P.O. Box 27519 · Philadelphia PA 19118-0519 · USA

Computers and Translation

Benjamins Translation Library

The Benjamins Translation Library aims to stimulate research and training in translation and interpreting studies. The Library provides a forum for a variety of approaches (which may sometimes be conflicting) in a socio-cultural, historical, theoretical, applied and pedagogical context. The Library includes scholarly works, reference works, post-graduate text books and readers in the English language.

General editor
Gideon Toury
Tel Aviv University

Associate editor
Miriam Shlesinger
Bar Ilan University

Advisory board
Marilyn Gaddis Rose
Binghamton University
Yves Gambier
Turku University
Daniel Gile
Université Lumière Lyon 2 and ISIT Paris
Ulrich Heid
University of Stuttgart
Eva Hung
Chinese University of Hong Kong
W. John Hutchins
University of East Anglia
Zuzana Jettmarová
Charles University of Prague
Werner Koller
Bergen University
Alet Kruger
UNISA
José Lambert
Catholic University of Leuven
Franz Pöchhacker
University of Vienna
Rosa Rabadán
University of León
Roda Roberts
University of Ottawa
Juan C. Sager
UMIST Manchester
Mary Snell-Hornby
University of Vienna
Sonja Tirkkonen-Condit
University of Joensuu
Lawrence Venuti
Temple University
Wolfram Wilss
University of Saarbrücken
Judith Woodsworth
Mt. Saint Vincent University Halifax
Sue Ellen Wright
Kent State University

Volume 35

Computers and Translation: A translator's guide
Edited by Harold Somers

试读结束：需要全本请在线购买： www.ertongbook.com

For Mum, Happy 81st

and Dad, the first linguist I ever met,
who (I hope) would have found all this fascinating

and for Nathan and Joe, the next generation

List of contributors

Jeffrey Allen, Mycom France, Paris, France.
Postediting@aol.com

Doug Arnold, Department of Language and Linguistics, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, England. doug@essex.ac.uk

Paul Bennett, Centre for Computational Linguistics, UMIST, PO Box 88, Manchester M60 1QD, England. paul.bennett@umist.ac.uk

Scott Bennett, 43 West Shore Road, Denville, NJ 07834, USA.
three.bennetts@verizon.net

Lynne Bowker, School of Translation and Interpretation, University of Ottawa, 70 Laurier Ave E., PO Box 450, Station A, Ottawa ON K1N 6N5, Canada. lbowker@uottawa.ca

Bert Esselink, L10nbridge, Overschiestraat 55, 1062 HN Amsterdam, The Netherlands. bert@locguide.com

Laurie Gerber, Language Technology Broker, 4774 Del Mar Avenue, San Diego CA, USA. lgerber@gerbersite.com

Willem-Olaf Huijsen, Institute for Linguistics OTS, Utrecht University, Trans 10, 3512 JK, Utrecht, The Netherlands. willem-olaf.huijsen@let.ruu.nl

John Hutchins, University of East Anglia, Norwich NR4 7TJ, England.
wjhutchins@compuserve.com

Elke Lange, SYSTRAN Software, Inc., 9333 Genesee Avenue, San Diego CA 92121, USA. elange@systransoft.com

Sara Laviosa, Università degli Studi di Bari, Italy. SaraLaviosa@hotmail.com

Teruko Mitamura, Language Technologies Institute, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh PA 15213, USA. teruko+@cs.cmu.edu

Eric Nyberg, Language Technologies Institute, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh PA 15213, USA. ehn@cs.cmu.edu

Harold Somers, Centre for Computational Linguistics, UMIST, PO Box 88, Manchester M60 1QD, England. harold.somers@umist.ac.uk

John S. White, PRC Northrop Grumman Information Technology, MacLean VA, USA. white_john@prc.com

Jin Yang, SYSTRAN Software, Inc., 9333 Genesee Avenue, San Diego CA 92121, USA. jyang@systransoft.com

Table of contents

List of figures

试读结束：需要全本请在线购买： www.ertongbook.com

List of tables

CHAPTER 1

Introduction*

Harold Somers
UMIST, Manchester, England

1. Preliminary remarks

This book is, broadly speaking, and as the title suggests, about computers and translators. It is not, however, a Computer Science book, nor does it have *much* to say about Translation Theory. Rather it is a book for translators and other professional linguists (technical writers, bilingual secretaries, language teachers even), which aims at clarifying, explaining and exemplifying the impact that computers have had and are having on their profession. It is about Machine Translation (MT), but it is also about Computer-Aided (or -Assisted) Translation (CAT), computer-based resources for translators, the past, present and future of translation and the computer.

Actually, there is a healthy discussion in the field just now about the appropriateness or otherwise of terms like the ones just used. The most widespread term, "Machine Translation", is felt by many to be misleading (who calls a computer a "machine" these days?) and unhelpful. But no really good alternative has presented itself. Terms like "translation technology" or "translation software" are perhaps more helpful in indicating that we are talking about computers, the latter term emphasising that we are more interested in computer programs than computer hardware as such. Replacing the word "translation" by something like "translator's" helps to take the focus away from translation as the end product and towards translation as a process[1] carried out by a human (the translator) using various tools, among which we are interested in only those that have something to do with computers.

We hope that this book will show you how the computer can help you, and in doing so we hope to show also what the computer *cannot* do, and thereby reassure you that the computer, far from being a threat to your livelihood, can become an essential tool which will make your job easier and more satisfying.

1.1 Who are we?

This book has been put together by academics (teachers and researchers in language and linguistics, especially computational linguistics, translation theory), employees of software companies, and — yes — even translators. All the contributors have an interest in the various aspects of translation and computers, and between them have several hundred years' worth of experience in the field. All are committed to telling a true story about computers and translation, what they can and cannot do, what they are good for, and what they are not. We are *not* trying to sell you some product. But what we *are* aiming to do is to dispel some of the myths and prejudices that we see and hear on translators' forums on the Internet, in the popular press, even in books about translation whose authors should know better!

1.2 Who are you?

We assume that you are someone who knows about and is interested in languages and translation. Perhaps you are a professional linguist, or would like to be. Or perhaps you are just a keen observer. In particular, you are interested in the topic of computers and translation and not too hostile, though perhaps healthily sceptical. The fact you have got hold of this book (perhaps you have already bought it, or are browsing in a bookshop, or a colleague has passed it on to you) is taken to mean that you have not dismissed the idea that computers can play a part in the translation process, and are open to some new ideas.

You are probably *not* a computer buff: if you are looking for lots of stuff about bits and bytes, integer float memory and peripheral devices then this is not the book for you. On the other hand, you are probably a regular computer-*user*, perhaps at the level of word-processing and surfing the World Wide Web. You know, roughly, the difference between "software" and "hardware", you know about windows and desktops, files and folders. You may occasionally use the computer to play games, and you may even have used some software that involves a kind of programming or authoring. But by enlarge that's not really your area of expertise.

On the other hand, you *do* know about language. We don't need to tell you about how different languages say things differently, about how words don't always neatly correspond in meaning and use, and how there's almost never an easy answer to the question "How do you say X in language Y?" (though we may remind you from time to time). We assume that you are familiar with traditional grammatical terminology (noun, verb, gender, tense, etc.) though you may not have studied linguistics as such. Above all, we don't need to remind you that translation is an art, not a science, that there's no such thing as a single "correct" translation, that a

translator's work is often under-valued, that translation is a human skill — one of the oldest known to humankind[2] — not a mechanical one. Something else you already know is that almost no one earns their living translating literary works and poetry: translation is mostly technical, often nonetheless demanding, but just as often routine and sometimes — dare we admit it? — banal and boring. Whatever the case, the computer has a role to play in your work.

1.3 Conventions in this book

This is a technical book, and as such will, we hope, open avenues of interest for the reader. For that reason, we give references to the literature to support our arguments, in the usual academic fashion. Where specific points are made, we use footnotes so as to avoid cluttering the text with unwieldy references. We also want to direct the reader to further sources of information, which are gathered together at the end of each chapter. Technical terms are introduced in bold font. Software product names are given in italics, and are thus distinguished typographically from the (often identical) names of the company which produce them.

Often it is necessary to give language examples to illustrate the point being made. We follow the convention of linguistics books as follows: **cited forms** are always given in italics, regardless of language. Meanings or **glosses** are given in single quotes. Cited forms in languages other than English are always accompanied by a **literal gloss** and/or a translation, as appropriate, unless the meaning is obvious from the text. Thus, we might write that *key-ring* is rendered in Portuguese as *porta-chave* lit. 'carry-key', or that in German the plural of *Hund* 'dog' is *Hünde*. Longer examples (phrases and sentences) are usually separated from the text and referred to by a number in brackets, as in (1). Foreign-language examples are accompanied by an aligned literal gloss as well as a translation (2a), though either may be omitted if the English follows the structure of the original closely enough (2b).

(1) This is an example of an English sentence.

(2) a. *Ein Lehrbuchbeispiel in deutscher Sprache ist auch zu geben.*
a text-book-example in German language is also to give
'A German-language example from a text-book can also be given.'

b. *Voici une phrase en français.*
this-is a sentence in French

We follow the usual convention from linguistics of indicating with an asterisk that a sentence or phrase is **ungrammatical** or otherwise **anomalous** (3a), and a question-mark if the sentence is dubious (3b).

(3) a. *This sentence are wrong.
b. ?Up with this we will not put.

2. Historical sketch

A mechanical translation tool has been the stuff of dreams for many years. Often found in modern science fiction (the universal decoder in *Star Trek*, for example), the idea predates the invention of computers by a few centuries. Translation has been a suggested use of computers ever since they were invented (and even before, curiously). Universal languages in the form of numerical codes were proposed by several philosophers in the 17th Century, most notably Leibniz, Descartes and John Wilkins.

In 1933 two patents had been independently issued for "translation machines", one to Georges Artsrouni in France, and the other to Petr Petrovich Smirnov-Troyanskii in the Soviet Union. However, the history of MT is usually said to date from a period just after the Second World War during which computers had been used for code-breaking. The idea that translation might be in some sense similar at least from the point of view of computation is attributed to Warren Weaver, at that time vice-president of the Rockefeller Foundation. Between 1947 and 1949, Weaver made contact with a number of colleagues in the USA and abroad, trying to raise interest in the question of using the new digital computers (or "electronic brains" as they were popularly known) for translation; Weaver particularly made a link between translation and cryptography, though from the early days most researchers recognised that it was a more difficult problem.

2.1 Early research

There was a mixed reaction to Weaver's ideas, and significantly MIT decided to appoint Yehoshua Bar-Hillel to a full-time research post in 1951. A year later MIT hosted a conference on MT, attended by 18 individuals interested in the subject. Over the next ten to fifteen years, MT research groups started work in a number of countries: notably in the USA, where increasingly large grants from government, military and private sources were awarded, but also in the USSR, Great Britain, Canada, and elsewhere. In the USA alone at least $12 million and perhaps as much as $20 million was invested in MT research.

In 1964, the US government decided to see if its money had been well spent, and set up the Automated Language Processing Advisory Committee (ALPAC). Their report, published in 1966, was highly negative about MT with very damaging consequences. Focussing on Russian–English MT in the USA, it concluded that MT was slower, less accurate and twice as expensive as human translation, for which

试读结束：需要全本请在线购买： www.ertongbook.com