MACHINE TRANSLATION

An Introductory Guide

Douglas Arnold

Lorna Balkan

Siety Meijer

R. Lee Humphreys

Louisa Sadler

This book was originally published by NCC Blackwell Ltd. as *Machine Translation: an Introductory Guide*, NCC Blackwell, London, 1994, ISBN: 1855542-17x. However, it is now out of print, so we are making this copy available. Please continue to cite the original publication details.

First published 1994

First published in the USA 1994

Originally published as D.J. Arnold, Lorna Balkan, Siety Meijer, R. Lee. Humphreys and Louisa Sadler, 1994, *Machine Translation: an Introductory Guide*, ISBN: 1855542-17x. Originally published in the UK by NCC Blackwell Ltd., 108 Cowley Rd, Oxford OX4 IJF, and in the USA by Blackwell Publishers, 238 Main St. Cambridge, Mass. 02142. Blackwells-NCC, London.

This copy differs from the published version only in that some cartoons are omitted (there is blank space where the cartoons appear), and it has been reduced to print two pages on one A4 page (the original is slightly larger).

March 6, 2001

Preface

Automatic translation between human languages ('Machine Translation') is a Science Fiction staple, and a long-term scientific dream of enormous social, political, and scientific importance. It was one of the earliest applications suggested for digital computers, but turning this dream into reality has turned out to be a much harder, and in many ways a much more interesting task than at first appeared. Nevertheless, though there remain many outstanding problems, some degree of automatic translation is now a daily reality, and it is likely that during the next decade the bulk of routine technical and business translation will be done with some kind of automatic translation tool, from humble databases containing canned translations of technical terms to genuine Machine Translation Systems that can produce reasonable draft translations (provided the input observes certain restrictions on subject matter, style, and vocabulary).

Unfortunately, how this is possible or what it really means is hard to appreciate for those without the time, patience, or training to read the relevant academic research papers, which in any case do not give a very good picture of what is involved in practice. It was for this reason that we decided to try to write a book which would be genuinely introductory (in the sense of not presupposing a background in any relevant discipline), but which would look at all aspects of Machine Translation: covering questions of what it is like to use a modern Machine Translation system, through questions about how it is done, to questions of evaluating systems, and what developments can be foreseen in the near to medium future.

We would like to express our thanks to various people. First, we would like to thank each other. The process of writing this book has been slower than we originally hoped (five authors is five pairs of hands, but also five sets of opinions). However, we think that our extensive discussions and revisions have in the end produced a better book in terms of content, style, presentation, and so on. We think we deserve no little credit for maintaining a pleasant working atmosphere while expending this level of effort and commitment while under pressure caused by other academic responsibilities.

We would also like to thank our colleagues at the Computational Linguistics and Machine Translation (CL/MT) group at the University of Essex for suggestions and practical support, especially Lisa Hamilton, Kerry Maxwell, Dave Moffat, Tim Nicholas, Melissa Parker, Martin Rondell and Andy Way.

For proofreading and constructive criticism we would like to thank John Roberts of the Department of Language and Linguistics at the University of Essex, and John Roberts and Karen Woods of NCC Blackwell. We are also grateful to those people who have helped us by checking the examples which are in languages other than English and Dutch, especially Laurence Danlos (French), and Nicola Jörn (German).

Of course, none of them is responsible for the errors of content, style or presentation that remain.

D.J. Arnold
L. Balkan
R. Lee Humphreys
S. Meijer
L. Sadler

Colchester, August 1993.

Contents