

MACHINE TRANSLATION REVIEW

The Periodical
of the
Natural Language Translation Specialist Group
of the
British Computer Society
Issue No. 15
December 2005

The *Machine Translation Review* incorporates the Newsletter of the Natural Language Translation Specialist Group of the British Computer Society. Publication has now ceased and this is the last issue.

The Review has always welcomed contributions, articles, book reviews, advertisements, and all items of information relating to the processing and translation of natural language and thanks all those who have provided material over the several years during which the Review has been published. It has, however, been decided to discontinue further publication.

However, if further contributions are received, the Group is happy to consider these for publication on its website. Notices and reviews of books received will continue to be published in this way and should be sent in the first instance to the editor at the following address.

Derek Lewis
The Editor
Machine Translation Review
School of Modern Languages
Queen's Building
University of Exeter
Exeter
EX4 4QH
United Kingdom

Tel: +44 (0)1392 264264
E-mail: D.R.Lewis@exeter.ac.uk

The *Machine Translation Review* is published by the Natural Language Translation Specialist Group of the British Computer Society. All published items are subject to the usual laws of Copyright and may not be reproduced without the permission of the publishers.

All opinions expressed in this Review are those of the respective writers and are not necessarily shared by the BCS or the Group.

ISSN 1358-8346

Please note: From April 2000 this Review will be published electronically and will be available on our web site at the British Computer Society (see page 5). The format will be in HTML, in the same way as some back copies have already been stored electronically, so it will be easy for readers to print copies if they wish. Each section will be separate so readers may print selected parts only.

Some copies will be printed for the Copyright libraries and for purchase at a modest price plus postage and packing for those without electronic access. Members of the Natural Language Translation Specialist Group of the British Computer Society will be advised of each issue.



Contents

Group News and Information	4
The Committee	4
BCS Library	5
Website	5
Books Received	
<i>Derek Lewis</i>	6
Conferences and Workshops	10
Membership	11

Group News and Information

The British Computer Society

Charity No. 292786

THE NATURAL LANGUAGE TRANSLATION SPECIALIST GROUP

Please reply to:

72 Brattle Wood

Sevenoaks

Kent, TN13 1QU

Tel: 01732 455446

E-mail: wiggjd@bcs.org.uk

The Committee

The telephone numbers and e-mail addresses of the Committee are as follows:

David Wigg (Chair & Treasurer)

Tel.: +44 (0)1732 455446 (H)

Tel.: +44 (0)207 815 7472 (W)

E-mail: wiggjd@bcs.org.uk

Monique L'Huillier (Secretary)

Tel.: +44 (0)1273 474107 (H)

E-mail: monique@elda.demon.co.uk

Derek Lewis (Editor)

Tel.: +44 (0)1404 814186 (H)

E-mail: d.r.lewis@exeter.ac.uk

Roger Harris (Webmaster)

Tel.: +44 (0)208 800 2903 (H)

E-mail: rwsh@nationalfinder.com

Douglas Clarke

E-mail: dougclarke7@hotmail.com

Heather Fulford

E-mail: h.fulford@lboro.ac.uk

John Hutchins

E-mail: wjhutchins@compuserve.com

Ian Kelly

E-mail: idkk@idkk.com

Veronica Lawson

E-mail: veronica_1@compuserve.com

Ruslan Mitkov

E-mail: r.mitkov@wlv.ac.uk

Mark Stevenson

E-mail: m.stevenson@dcs.shef.ac.uk

BCS Library

Books kindly donated by members are passed to the BCS library at the IEE, Savoy Place, London, WC2R 0BL, UK (tel: +44 (0)207 240 1871; fax: +44 (0)207 497 3557). Members of the BCS may borrow books from this library either in person or by post. All they have to provide is their membership number. The library is open Monday to Friday, 9.00 am - 5.00 pm.

Website

The website address of the BCS-NLTSG is: <http://www.bcs-mt.org.uk>

BOOKS RECEIVED

The following books have been received from publishers.

Véronis, Jean (ed.) (2000). *Parallel Text Processing. Alignment and Use of Translation Corpora*. Dordrecht/Boston/London: Kluwer Academic Publishers. 402 pp. ISBN 0-7923-6546-1 (HB)

This book evolved from the ARCADE evaluation exercise for assessing alignment systems for parallel texts (i.e. those with a translation) that began in 1995. After briefly surveying the development of parallel text processing, the volume presents chapters on alignment methodology, applications, and resources and evaluation.

Carl, Michael and Andy Way (eds) (2003). *Recent Advances in Example-Based Machine Translation*. Dordrecht/Boston/London: Kluwer Academic Publishers. 483 pp. ISBN 1-4020-1400-7 (HB)

Originating from a workshop held at the MT Summit VIII in Santiago in 2001, this volume includes sections on the foundations of example-based machine translation (EBMT), run-time approaches to EBMT, template-driven EBMT, and EBMT and derivation trees. It is the first major attempt to assemble a representative number of EBMT approaches in a single book (there are contributions from 26 different authors) and to provide an up-to-date overview of the field.

Damper, R.I. (ed.) (2001) *Data-Driven Techniques in Speech Synthesis*. Dordrecht/Boston/London: Kluwer Academic Publishers. 316 pp. ISBN 0-412-81750-0 (HB)

This book begins with a helpful introduction to text-to-speech (TTS) systems that describes early work in developing rule systems for the automatic generation of speech from phonetic symbols during the 1960s. It then moves on to focus on data-driven techniques based on neural networks that emerged in the 1980s. Four chapters are devoted to techniques for converting text into phonetic transcript, while others describe a framework for integrating various sources of linguistic knowledge, the role and importance of diphones (which cover transitions between phonemes), models of intonations, and experiments on accent and phrasing. The volume thus provides an interesting review of the variety of data-driven approaches employed in speech synthesis.

Harald Baayen, R. (2001). *Word Frequency Distributions*. Dordrecht/Boston/London: Kluwer Academic Publishers. 333 pp. ISBN 0-7923-7017-1 (HB)

This volume is intended as an introduction to the statistical analysis of word frequency distributions. These are characterised by very large numbers of rare words, which leads to strange results, including the feature that mean frequencies change as the number of observations increases. The declared aim of the book is to make techniques for the analysis of such distributions more accessible to non-specialists by presenting basic concepts and notation and describing various models with examples of applications. However, the volume is not an introduction to statistics and probability theory and non-specialist linguists will find the material largely incomprehensible without a familiarity with the methods and language of

these fields. A CD-ROM provides the Linux-based LEXSTATS suite of programs for carrying out statistical analyses.

Junqua, Jean-Claude and Gertjan van Noord (2001). *Robustness in Language and Speech Technology*. Dordrecht/Boston/London: Kluwer Academic Publishers. 269 pp. ISBN 0-7923-67090-1 (HB)

This book brings together selected contributions from the ELSNET-sponsored 6th European Summer School on Language and Speech Communication held in Barcelona in 1998. It considers the most promising research directions to be robust feature extraction, rapid speaker adaptation, separate modelling of phonetic and speaker-dependent information, environment-adaptive automatic speech recognition and language/task adaptation, and the combination of syntactic knowledge, global semantic analysis and pragmatic information in order to produce better language models and robust parsing algorithms. The specific topics include acoustic features and distance measure, speaker compensation in automatic speech recognition, robustness in statistical language modelling, the modelling of spontaneous speech events, regular approximation of context-free grammars, weighted grammar tools, robust parsing of word graphs, and balancing robustness and efficiency.

Minker, Wolfgang, Mariani, Joseph and Alex Waibel (1999). *Stochastically-Based Semantic Analysis*. Dordrecht/Boston/London: Kluwer Academic Publishers. 221 pp. ISBN 0-7923-8571-3 (HB)

This volume investigates the problem of automatic language understanding (NLU) for spoken language systems and proposes a parsing method that is designed to be sufficiently general and flexible so as to be easily ported to different applications, domains and human languages. After reviewing the background and research in NLU, including existing systems, a chapter introduces the applications, subject areas and natural language corpora used in the authors' system, while others describe rule-based and stochastic NLU components. The conclusion analyses the performance of a specifically developed stochastically based frame parser using a semantic case grammar formalism that was applied to the American ATIS (Air Travel Information Services) task and ported to the French MASK (Multimodal-Multimedia Automated Service Kiosk) application and the ESST (English Spontaneous Speech Task).

Stede, Manfred (1999). *Lexical Semantics and Knowledge Representation in Multilingual Text Generation*. Dordrecht/Boston/London: Kluwer Academic Publishers. 219 pp. ISBN. (HB)

After reviewing the nature of lexical items and their representation in natural language processing, this book considers how to classify lexical variation, build domain models for natural language generation (NLG), develop levels of representation, and analyses so-called verb alternations and extensions. After describing the prototype of a system architecture for NLG, the author demonstrates how the representations in the domain model and in the lexicon can work together in the system architecture to derive different verbalisations and lexical paraphrases at sentence level. He also investigates the possibilities for extending the system into a text generator for paragraphs. The volume is an extended and revised version of a dissertation completed at Toronto and Ulm (Germany).

Nirenberg, Sergei, Harold Somers and Yorick Wilks (eds) (2003). *Readings in Machine Translation*. Cambridge, Massachusetts: The MIT Press. i-xv and 406 pp. ISBN0-262-14074-8 (HB)

This handsomely produced volume is primarily a collection of 36 hard-to-find but often-cited papers on Machine Translation. The editors describe them as ‘the “classical” MT papers that researchers and students want, or should be persuaded, to read’ and which make up the ‘communal inheritance’ of MT (for readers’ convenience a full list of the papers is included at the end of this short review). The three main sections are historical, theoretical and methodological issues, and system design. Each section is prefaced with a brief introduction. The first section contains ‘historical papers’ from the beginnings of MT (including Warren Weaver’s memorandum, Bar-Hillel’s ‘box in the pen’ article, and excerpts from the ALPAC report) up until the late 1960s. The second section looks at sublanguage, controlled input, the role of human beings in MT, the impact of linguistic approaches, the transfer/interlingua issue, and representation of meaning and knowledge. The third section is devoted to system design and includes papers on knowledge-based, statistical and example-based translation. Given the historical approach and the uncertainty of knowing what usefully to include beyond the early 1990s, this is when coverage ends. All the papers appeared originally in English and the authors admit with regret that material from Russia/USSR, Japan and France and other languages could not be included. Individual systems are not described, although references to such descriptions, and also to major textbooks, are provided. There is no doubt that reading the papers provides a fascinating and absorbing tour through several decades of MT activity. The introductions alone give crisp and clear overviews of the key developments. A particularly positive feature of this well produced, lengthy and large format book is its reasonable price (just under £36).

Section 1: HISTORICAL

Translation: Warren Weaver

Mechanical Translation: A.D. Booth

The Mechanical Determination of Meaning: Erwin Reifler

Stochastic Methods of Mechanical Translation: Gilbert W. King

A Framework for Syntactic Translation: Victor H. Yngve

The Present Status of Automatic Translation of Languages: Yehoshua Bar-Hillel

A New Approach to the Mechanical Syntactic Analysis of Russian: Ida Rhodes

A Preliminary Approach to Japanese-English Automatic Translation: Susumu Kuno

On the Mechanization of Syntactic Analysis: Sydney M. Lamb

Research Procedures in Machine Translation: David G. Hays

ALPAC: The (In)Famous Report: John Hutchins

Correlational Analysis and Mechanical Translation: Silvio Ceccato

Automatic Translation: Some Theoretical Aspects and the Design of a Translation System:

O.S. Kulagina and I.A. Mel'čuk

Mechanical Pidgin Translation: Margaret Masterman

English-Japanese Machine Translation: S. Takahashi, H. Wada, R. Tadenuma, and S.

Watanabe

Section 2: THEORETICAL AND METHODOLOGICAL ISSUES

Automatic Translation and the Concept of Sublanguage: J. Lehrberger

The Proper Place of Men and Machines in Language Translation: Martin Kay

Machine Translation as an Expert Task: Roderick L. Johnson and Peter Whitelock
 Montague Grammar and Machine Translation: Jan Landsbergen
 Dialogue Translation vs. Text Translation - Interpretation Based Approach: Jun-ichi Tsujii and Makoto Nagao
 Translation by Structural Correspondences: Ronald M. Kaplan, Klaus Netter, Jürgen Wedekind, and Annie Zaenen
 Pros and Cons of the Pivot and Transfer Approaches in Multilingual Machine Translation: Christian Boitet
 Treatment of Meaning in MT Systems: Sergei Nirenburg and Kenneth Goodman
 Where Am I Coming From: The Reversibility of Analysis and Generation in Natural Language Processing: Yorick Wilks
 The Place of Heuristics in the Fulcrum Approach to Machine Translation: Paul L. Garvin
 Computer Aided Translation: A Business Viewpoint: John S.G. Elliston

Section 3: SYSTEM DESIGN

Three Levels of Linguistic Analysis in Machine Translation: Michael Zarechnak
 Automatic Translation - A Survey of Different Approaches: B. Vauquois
 Multi-level Translation Aids: Alan K. Melby
 EUROTRA: Computational Techniques: Rod Johnson, Maghi King, and Louis des Tombe
 A Framework of a Mechanical Translation between Japanese and English by Analogy Principle: Makoto Nagao
 A Statistical Approach to Machine Translation: Peter F. Brown, John Cocke, Stephen A. Della Pietra, Vincent J. Della Pietra, Fredrick Jelinek, John D. Lafferty, Robert L. Mercer, and Paul S. Roossin
 Automatic Speech Translation at ATR: Tsuyoshi Morimoto and Akira Kurematsu
 The Stanford Machine Translation Project: Yorick Wilks
 The Textual Knowledge Bank: Design, Construction, Applications: Victor Sadler
 Machine Translation Without a Source Text: Harold L. Somers, Jun-ichi Tsujii, and Danny Jones

Conferences and Workshops

The following is a list of forthcoming or recent conferences and workshops. E-mail addresses and websites are given where known.

The online Language and Speech Calendar maintained by ELSNET (European Network in Human Language Technologies) contains announcements of events in the area of Natural Language and Speech, and related areas. This is an extensive website which provides links to a wide range of conferences, meetings and workshops. The site may be accessed at:
<http://www.elsnet.org>.

A list of 2004 Computational Linguistics Conferences is also accessible at:
<http://www.cs.rochester.edu/u/tetreaul/conferences.html>

21-23 September 2005
RANLP 2005 (Recent Advances in Natural Language Processing)
Borovets, Bulgaria
<http://www.lml.bas.bg/ranlp2005>

24-25 November 2005
ASLIB: Translating and the Computer 27
New Connaught Rooms, London, WC2
<http://www.aslib.com/conferences/>

24-26 May 2006
LREC 2006
Magazzini del Cotone Conference Center, Genoa, Italy

19-20 June 2006
11th Annual Conference of the European Association of Machine Translation (EAMT)
Oslo, Norway
<http://www.eamt.org/events.html>

17-21 July
COLING/ACL 2006
Joint 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics
Sydney, Australia
<http://www.aclweb.org/acl2005/acl2006.pdf>

Autumn 2006
Annual Conference of AMTA (Association for Machine Translation in the Americas)
Boston, Massachusetts (USA)
<http://www.amtaweb.org/>

MEMBERSHIP: CHANGE OF ADDRESS

If you change your address, please advise us on this form, or a copy, and send it to the following (this form can also be used to join the Group):

Mr. J.D.Wigg
BCS-NLTSG
72 Brattle Wood
Sevenoaks, Kent TN13 1QU
U.K.

Date:/...../.....

Name:

Address:

Postal Code: Country:

E-mail: Tel.No:

Fax.No:

Note for non-members of the BCS: your name and address will be recorded on the central computer records of the British Computer Society.

Questionnaire

We would like to know more about you and your interests and would be pleased if you would complete as much of the following questionnaire as you wish (please delete any unwanted words).

- 1.
 - a. I am mainly interested in the computing/linguistic/user/all aspects of MT.
 - b. What is/was your professional subject?
 - c. What is your native language?
 - d. What other languages are you interested in?
 - e. Which computer languages (if any) have you used?

- 2. What information in this Review or any previous Review have you found:
 - a. interesting? Date
 -
 -
 - b. useful (i.e. some action was taken on it)? Date
 -
 -

3. Is there anything else you would like to hear about or think we should publish in the *MT Review*?

.....

.....

.....

- 4. Would you be interested in contributing to the Group by,
 - a. Reviewing MT books and/or MT/multilingual software
 - b. Researching/listing/reviewing public domain MT and MNLP software
 - c. Designing/writing/reviewing MT/MNLP application software
 - d. Designing/writing/reviewing general purpose (non-application specific) MNLP procedures/functions for use in MT and MNLP programming
 - e. Any other suggestions?
 -
 -
 -

Thank you for your time and assistance.