

**Association for Computational Linguistics**

**EACL 2003**

**10<sup>th</sup> Conference of The European Chapter**

**Proceedings of the 7<sup>th</sup> International  
EAMT workshop on  
MT and other language technology tools  
Improving MT through other language  
technology tools Resources and  
tools for building MT**

**April 13<sup>th</sup> 2003**

**Agro Hotel, Budapest, Hungary**

**Association for Computational Linguistics**

**EACL 2003**

**10<sup>th</sup> Conference of The European Chapter**

**Proceedings of the 7<sup>th</sup> International  
EAMT workshop on  
MT and other language technology tools  
Improving MT through other language  
technology tools Resources and  
tools for building MT**

**April 13<sup>th</sup> 2003**

**Agro Hotel, Budapest, Hungary**

The conference, the workshops and the tutorials are sponsored by:

Chief Patron of the Conference:

Dr. Ferenc Baja

Political State Secretary

Office of Government Information Technology and Civil Relations

Prime Minister's Office



Linguistic Systems BV

Leo Konst (Managing director)

Postbus 1186, 6501 BD Nijmegen, Nederland

tel: +31 24 322 63 02

fax: +31 24 324 21 16

e-mail: [info@euroglot.nl](mailto:info@euroglot.nl), [leokonst@telebyte.nl](mailto:leokonst@telebyte.nl),

<http://www.euroglot.nl>

Xerox Research Centre Europe

Irene Maxwell

6 chemin de Maupertuis

38240 Meylan, France

Tel: +33 (0)4.76.61.50.83

Fax: +33 (0)4.76.61.50.99

email: [info@xrce.xerox.com](mailto:info@xrce.xerox.com)

website: [www.xrce.xerox.com](http://www.xrce.xerox.com)



ATALA

Jean Veronis

[Jean.Veronis@up.univ-mrs.fr](mailto:Jean.Veronis@up.univ-mrs.fr)

45 rue d'Ulm

75230 Paris Cedex 5, France

<http://www.atala.org>



ELRA/ELDA

Khalid Choukri

[choukri@elda.fr](mailto:choukri@elda.fr)

55-57 rue Brillat Savarin

75013 Paris, France

Tel: (+33 1) 43 13 33 33,

Fax: (+33 1) 43 13 33 30

<http://www.elda.fr>



©April 2003, Association for Computational Linguistics

Order copies of ACL proceedings from: Priscilla Rasmussen, Association for Computational Linguistics, 3 Landmark Center, East Stroudsburg, PA 18301 USA, Phone +1-570-476-8006, Fax +1-570-476-0860, URL <http://www.acl-web.org>.

## INTRODUCTION:

There is an ever-growing need for tools for translation, and Europe is envisaging a particular challenge with the enlargement of the European Union which will add 10-12 new languages.

One of the important trends emerging from the 2001 MT Summit in Santiago de Compostela was that MT is going more and more towards an integration or combination with other tools.

In the planning of this workshop we wanted to follow up on these two trends by focussing on how MT and other language technology tools can be combined in order to produce translation faster and better, and secondly on how language technology tools can support faster production of MT systems.

Various ways of combining 'proper' MT with other types of language technology tools in order to improve performance and efficiency of translation have been discussed. This may include pre-editing tools, taggers, post-editing tools, access to bilingual concordances, term extraction tools, categorisation tools, semantic clustering etc. In the programme we see a selection of such proposals for new directions for improving MT by embedding it in an environment of other tools.

Embedding is a very popular term, and it is worth noticing that we are using the concept in a slight different way here. Embedded MT normally refers to MT embedded in another application, e.g. information retrieval, document production. But the embedding discussed here is an embedding or an enrichment of MT by combining it with other tools, namely tools which may improve the quality, even if the tools themselves are solving problems that are much simpler than the MT problem.

The other theme for the workshop was chosen because the growing demand for MT for new languages and new language pairs makes it necessary to find ways of supporting the production of new language pairs. Both providers and researchers are using resources, statistics, and language technology to make progress in this field. Unfortunately, this theme is not strongly represented in the papers we received.

This is the first time an EAMT workshop is organised at EACL and it has been a very good experience which can certainly be repeated in the future.

Finally, I should like to thank my collaborators in the Programme Committee: Gáb or Prószéky, Jörg Schütz and Harold Somers for their contributions and support.

Copenhagen, March 2003

Bente Maegaard, programme chair

**SPONSOR:**

European Association for Machine Translation (EAMT)



**WORKSHOP PROGRAMME COMMITTEE:**

Bente Maegaard  
Center for Sprogteknologi  
Njalsgade 80, DK-2300 Copenhagen, Denmark  
bente@cst.dk

Gábor Prószéky  
Morphologic  
MorphoLogic  
1118 Késmárki u. 8, Budapest, Hungary  
proszeky@morphologic.hu

Jörg Schütz  
IAI  
Universität des Saarlandes,  
Martin-Luther-Str. 14, D-66111 Saarbrücken, Germany  
joerg@iai.uni-sb.de

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

# CONFERENCE PROGRAM

## Sunday, April 13

- 9:30-9:45 Welcome and introduction  
Bente Maegaard (Center for Sprogteknologi, CST)
- 9:45-10:30 *Improving Machine Translation Quality with Automatic Named Entity Recognition*  
Bogdan Babych, Anthony Hartley (University of Leeds)
- 10:30-11:00 Coffe Break
- 11:00-11:45 *Two Approaches to Aspect Assignment in an English-Polish Machine Translation System*  
Anna Kupsc (Polish Academy of Sciences & Carnegie Mellon University)
- 11:45-12:30 *Multi-language Machine Translation through Interactive Document Normalization*  
Aurélien Max (GETA & Xerox Research Centre Europe)
- 12:30-14:00 Lunch
- 14:00-14:45 *Computer-based Support for Patients with Limited English*  
Harold Somers (UMIST), Hermione Lovel (University of Manchester)
- 14:45-15:30 Panel discussion on improving MT
- 15:30-16:00 Coffee Break
- 16:00-16:45 *Parallel Corpora Segmentation Using Anchor Words*  
Francisco Nevado, Francisco Casacuberta, Enrique Vidal (Universidad Politécnica de Valencia)
- 16:45-17:30 *An Evaluation of a Lexicographer's Workbench: building lexicons for Machine Translation*  
Rob Koeling (University of Sussex), Adam Kilgarriff, David Tugwell,  
Roger Evans (University of Brighton)
- 17:30-18:00 Panel discussion on resources and tools for building MT



## Table of Contents

<i>Introduction</i>	
Bente Maegaard .....	iii
<i>Improving Machine Translation Quality with Automatic Named Entity Recognition</i>	
Bogdan Babych, Anthony Hartley .....	1
<i>An Evaluation of a Lexicographer's Workbench: building lexicons for Machine Translation</i>	
Rob Koeling, Adam Kilgarriff, David Tugwell, Roger Evans .....	9
<i>Two Approaches to Aspect Assignment in an English-Polish Machine Translation System</i>	
Anna Kupsc .....	17
<i>Multi-language Machine Translation through Interactive Document Normalization</i>	
Aurélien Max .....	25
<i>Parallel Corpora Segmentation Using Anchor Words</i>	
Francisco Nevado, Francisco Casacuberta, Enrique Vidal .....	33
<i>Computer-based Support for Patients with Limited English</i>	
Harold Somers, Hermione Lovel .....	41