

**Proceedings of the MT Summit XI Workshop
Using Corpora for Natural Language Generation:
Language Generation and Machine Translation (UCNLG+MT)**

Anja Belz and Sebastian Varges (eds)

Contents

| | |
|---|----|
| Preface | iv |
| Organisation | iv |
| KEYNOTE PAPER: | |
| “Automatic Language Translation Generation Help Needs Badly” Kevin Knight | 1 |
| REGULAR PAPERS: | |
| <i>Automatic Evaluation of Generation and Parsing for Machine Translation with Automatically Acquired Transfer Rules</i> Yvette Graham, Deirdre Hogan and Josef van Genabith | 5 |
| <i>Generalizing Syntactic Collocates for Creative Language Generation</i> David Hardcastle | 13 |
| <i>Towards Broad Coverage Surface Realization with CCG</i> Michael White, Rajakrishnan Rajkumar, Scott Martin | 22 |
| <i>Method of Selecting Training Sets to Build Compact and Efficient Statistical Language Model</i> Keiji Yasuda, Hirofumi Yamamoto and Eiichiro Sumita | 31 |
| WORK-IN-PROGRESS PAPERS: | |
| <i>The Induction and Evaluation of Word Order Rules Using Corpora Based on the Two Concepts of Topological Models</i> Bernd Bohnet | 38 |
| <i>A Probabilistic Approach to Linguistic Analysis in Machine Translation Output Evaluation</i> Olivier Gouirand | 46 |
| <i>Declarative Syntactic Processing of Natural Language Using Concurrent Constraint Programming and Probabilistic Dependency Modeling</i> Irene Langkilde-Geary | 55 |
| EXTENDED ABSTRACTS FOR THE DISCUSSION SESSION ON LANGUAGE GENERATION AND MACHINE TRANSLATION: | |
| <i>NLG is STILL Relevant to MT</i> Nizar Habash | 64 |
| <i>Evaluation of NLG: Some Analogies and Differences with Machine Translation and Reference Resolution</i> Andrei Popescu-Belis | 66 |
| <i>Generation Issues in Machine Translation</i> Gregor Thurmair | 69 |
| <i>One-Way Translation: An Opportunity for NLG and MT Research to Interact</i> Sebastian Varges | 71 |

Contents (continued)

THE ATTRIBUTE SELECTION FOR GRE CHALLENGE:

Introduction

Ehud Reiter, Anja Belz, Albert Gatt and Jette Viethen 74

The Attribute Selection for GRE Challenge: Overview and Evaluation Results

Anja Belz and Albert Gatt 75

IS-FBN, IS-FBS, IS-IAC: The Adaptation of Two Classic Algorithms for the Generation of Referring Expressions in Order to Produce Expressions Like Humans Do

Bernd Bohnet 84

NIL: Attribute Selection for Matching the Task Corpus Using Relative Attribute Groupings Obtained from the Test Data

Raquel Hervás and Pablo Gervás 87

DIT: Frequency Based Incremental Attribute Selection for GRE

John Kelleher 90

Evaluating an Open Domain GRE Algorithm on Closed Domains (System IDs: CAM-B, CAM-T, CAM-BU and CAM-TU)

Advaith Siddharthan and Ann Copestake 92

Cost-based attribute selection for GRE (GRAPH-SC/GRAPH-FP)

Mariet Theune, Pascal Touset, Jette Viethen and Emiel Kraemer 95

TITCH: Attribute Selection Based on Discrimination Power and Frequency

Philipp Spanger, Takahiro Kurosawa and Takenobu Tokunaga 98

EVALUATION METHODS TRACK:

Content Determination in GRE: Evaluating the Evaluator

Kees van Deemter and Albert Gatt 101

Preface

The Workshop on Language Generation and Machine Translation (UCNLG+MT) took place in Copenhagen on 11th September 2007, in conjunction with MT Summit XI. It was the second of the UCNLG workshops which have the general aims

1. to provide a forum for reporting and discussing data-based methods for generating language;
2. to open up the NLG research field to neighbouring areas in NLP and to foster cross-fertilisation between these research fields; and
3. to promote the sharing of data in NLG and the use of data-based methods of evaluation.

The special theme of this second UCNLG workshop was Language Generation and MT. Our aim in having this special theme was to explore to what extent research issues are similar, and resources and techniques can be shared, between the NLG and MT fields. We were honoured to have as our keynote speaker Kevin Knight of ISI, University of Southern California, whose paper was aptly titled “*Automatic Language Translation Generation Help Needs Badly*”. Several of the full workshop papers spanned both NLG and MT. There was a discussion session on Language Generation and MT chaired by Robert Dale, with contributions by Nizar Habash, Andrei Popescu-Belis, Gregor Thurmair, and Sebastian Varges.

We are pleased to say that UCNLG+MT was also host to the first Shared-Task Challenge in NLG: the *Attribute Selection for GRE Challenge*, organised by Anja Belz, Albert Gatt, Ehud Reiter and Jette Viethen. This Challenge was met with enthusiasm by the NLG community, attracting 19 individual registrations. 13 of the registered participants formed 6 teams and submitted a total of 22 systems by the deadline.

The UCNLG+MT Workshop was endorsed by the Association for Computational Linguistics Special Interest Group on Generation (SIGGEN). This volume presents extended abstracts for the keynote paper by Kevin Knight and the presentations in the discussion session on language generation and MT, as well as the seven full-length workshop papers which were selected in a rigorous double-blind reviewing process from submissions from France, Germany, Japan, Spain, the United Kingdom and the United States. Also included are the Shared-Task reports describing the participating systems, and an overview paper reporting the evaluation results; these papers were not peer-reviewed.

We would like to thank all the authors who submitted papers and extended abstracts for the workshop, our programme committee who helped us put together an excellent workshop programme, the organisers of the MT Summit XI conference (in particular the workshop chairs), and everybody else who helped us — against the odds — to make this workshop happen.

September 2007

Anja Belz and Sebastian Varges

Organisation

Workshop Co-chairs and Organisers

Anja Belz, University of Brighton, UK
Sebastian Varges, University of Trento, Italy

Programme Committee

Srinivas Bangalore, AT&T, USA
Stephan Busemann, DFKI, Germany
Robert Dale, Macquarie University, Australia
Kevin Knight, ISI, University of Southern California, USA
Irene Langkilde-Geary, University of Brighton, UK
Chris Mellish, University of Aberdeen, UK
Jon Oberlander, University of Edinburgh, UK
Richard Power, Open University, UK
Flo Reeder, The Mitre Organisation, USA
Ehud Reiter, University of Aberdeen, UK
Amanda Stent, SUNY, USA
Michael Strube, EML Research, Germany
Kees van Deemter, University of Aberdeen, UK
Bonnie Webber, University of Edinburgh, UK
Mike White, Ohio State University, USA