



MT Marathon

26th – 30th January 2009

Prague, Czech Republic

Lectures, Talks, Labs

	Morning Lecture	Research Talks	Afternoon Lab Session	Evening
	9.00 – 10.30	11.00 – 12.30	14.00 – 17.00	17.00 – ?
Monday	Introduction to MT and MT Evaluation (Adam Lopez)	Projects: Introduction Intro to Stat-XFER (Alon Lavie) Intro to TectoMT (Zdeněk Žabokrtský, Ondřej Bojar)	Manual judgement of MT quality	
Tuesday	Word Alignment (Barry Haddow)	PostCAT, apertium-cy, MBMT	Implementing IBM model 1	Projects: Update
Wednesday	Phrase-Based Models and Decoding (Chris Dyer)	Joshua, SAMT, MERT+	Installing and running Moses (Hieu Hoang and Josh Schroeder)	
Thursday	TectoMT: Processing Trees (Zdeněk Žabokrtský and others)	RIA, Sub-Tree Aligner	TectoMT hands-on experience: Installation and tutorial (Jana Kravalová)	Projects: Update
Friday	Richer Models and Optimization (Philipp Koehn)	Projects: Final Short Presentations	Using factored models and MERT in Moses (Hieu Hoang, Barry Haddow, Abhishek Arun)	

Accepted Contributions, Research Talks

The following contributions will be presented during late mornings:

1. apertium-cy: **F. M. Tyers, K. Donnelly**: apertium-cy - a collaboratively-developed free RBMT system for Welsh to English
2. Joshua: **Z. Li, C. Callison-Burch, W. Thornton, S. Khudanpur**: Joshua: an Open-source Decoder for Parsing-based Machine Translation
3. MBMT: **A. van den Bosch and P. Berck**: Memory-Based Machine Translation and Language Modeling
4. MERT+: **N. Bertoldi, B. Haddow, J.-B. Fouet**: Improved Minimum Error Rate Training in Moses
5. PostCAT: **J. Graça, K. Ganchev, B. Taskar**: PostCAT - Posterior Constrained Alignment Toolkit
6. RIA: **Y. Graham, J. van Genabith**: An Open Source Rule Induction Tool for Transfer-Based SMT
7. SAMT: **A. Venugopal, A. Zollmann**: Grammar based statistical MT on Hadoop. An end-to-end toolkit for large scale PSCFG based statistical machine translation
8. Sub-Tree Aligner: **V. Zhechev**: Unsupervised Generation of Parallel Treebanks through Sub-Tree Alignment
9. Z-MERT: **O. Zaidan**: Z-MERT: A Fully Configurable Open Source Tool for Minimum Error Rate Training of Machine Translation Systems (There is no presentation for this paper.)

Research talk presentations should be 20 to 25 minutes long with additional 5 minutes for a discussion.

Contents

Monday	4
Introduction to MT	4
MT Evaluation	13
Intro to Stat-XFER	19
Stat-XFER: Czech to English Translation Project	29
TectoMT for Plaintext Freaks	33
Tuesday	35
Word Alignment	35
Wednesday	42
Phrase-Based Models and Decoding	42
Thursday	51
Introduction to the Software Framework	51
TectoMT: Alignment	57
Bad news, NLP Hacking and Feature Fishing	59
TectoMT: Tutorial	62
Friday	69
Discriminative Training and Factored Translation Models	69