

DMTW 2015

**Proceedings of the
1st Deep Machine
Translation Workshop**

3 – 4 September 2015
Charles University in Prague,
Faculty of Mathematics and Physics,
Institute of Formal and Applied Linguistics
Prague, Czech Republic

Published by:

Charles University in Prague,
Faculty of Mathematics and Physics,
Institute of Formal and Applied Linguistics
Malostranské náměstí 25
118 00 Praha 1
Czech Republic

ISBN 978-80-904571-7-1

Preface

This volume contains the papers presented at DMTW-2015: Deep Machine Translation Workshop held on September 3-4, 2015 in Prague.

Each submission was reviewed by at least 2, most by three, and on the average 2.8, program committee members. The committee decided to accept 13 papers. The program also includes 1 invited talk. The submission and proceedings creation has been handled by the EasyChair system.

We thank the QTLep EU project for providing local funding for the workshop organization, and to Rudolf Rosa for handling local arrangements and for his technical help with submissions, web page and proceedings.

August 28, 2015
Prague, Czech Republic

Jan Hajič
António Branco

Program Committee

Eneko Agirre	University of the Basque Country
António Branco	University of Lisbon
Aljoscha Burchardt	Deutsches Forschungszentrum für Künstliche Intelligenz GmbH
Rosa Del Gaudio	Higher Functions LDA
Jan Hajič	Charles University in Prague
Eva Hajičová	Charles University in Prague
Petya Osenova	Sofia University and Bulgarian Academy of Sciences
Martin Popel	Charles University in Prague
Khalil Sima'an	University of Amsterdam
Kiril Simov	Bulgarian Academy of Sciences
Gertjan van Noord	University of Groningen
Deyi Xiong	Soochow University

Table of Contents

Linguistic Linked Open Data: What’s in for Machine Translation?	1
<i>Christian Chiarcos</i>	
Modelling the Adjunct/Argument Distinction in Hierarchical Phrase-Based SMT.....	2
<i>Sophie Arnoult and Khalil Sima’An</i>	
Towards Deeper MT - A Hybrid System for German	12
<i>Eleftherios Avramidis, Aljoscha Burchardt, Maja Popovic and Hans Uszkoreit</i>	
Splitting Compounds by Semantic Analogy.....	20
<i>Joachim Daiber, Lautaro Quiroz, Roger Wechsler and Stella Frank</i>	
Delimiting Morphosyntactic Search Space with Source-Side Reordering Models	29
<i>Joachim Daiber and Khalil Sima’An</i>	
Evaluating a Machine Translation System in a Technical Support Scenario	39
<i>Rosa Del Gaudio, Aljoscha Burchardt and Arle Lommel</i>	
Development of Telugu-Tamil Transfer-Based Machine Translation system: With Special reference to Divergence Index.....	48
<i>Parameswari Krishnamurthy</i>	
Deep-syntax TectoMT for English-Spanish MT	55
<i>Gorka Labaka, Oneka Jauregi, Arantza Díaz de Ilarraza, Michael Ustaszewski, Nora Aran- berri and Eneko Agirre</i>	
First Steps in Using Word Senses as Contextual Features in Maxent Models for Machine Translation.....	64
<i>Steven Neale, Luís Gomes and António Branco</i>	
Lexical choice in Abstract Dependency Trees	73
<i>Dieke Oele and Gertjan van Noord</i>	
Large Scale Translation Quality Estimation	81
<i>Miguel Angel Rios Gaona and Serge Sharoff</i>	
Translation Model Interpolation for Domain Adaptation in TectoMT	89
<i>Rudolf Rosa, Ondrej Dusek, Michal Novák and Martin Popel</i>	
Factored models for Deep Machine Translation	97
<i>Kiril Simov, Iliana Simova, Velislava Todorova and Petya Osenova</i>	
Machine Translation for Multilingual Troubleshooting in the IT Domain: A Comparison of Different Strategies	106
<i>Sanja Štajner, João Rodrigues, Luís Gomes and António Branco</i>	

Linguistic Linked Open Data: What's in for Machine Translation?

Christian Chiarcos

Angewandte Computerlinguistik

Fachbereich für Informatik und Mathematik

Goethe-Universität Frankfurt am Main

Robert-Mayer-Str. 10

D-60325 Frankfurt am Main

chiarcos@informatik.uni-frankfurt.de

During the past years, the notion of Linked (Open) Data has gained considerable reception in different communities working with language resources, ranging from academic and applied linguistics over lexicography to natural language processing and information technology. In this context, the Open Linguistics Working Group of the Open Knowledge Foundation (OWLG, <http://linguistics.okfn.org/>), founded in 2010 in Berlin, Germany, is playing an important integrative role, by reaching out to a broad band-width of disciplines, by facilitating interdisciplinary information exchange through meetings, workshops, datathons and joint publications, but most notably by introducing and maintaining the Linguistic Linked Open Data (LLOD) cloud diagram. Being deeply involved in this emerging community at the intersection between the different disciplines mentioned above, I will introduce the basic concepts of Linked Open Data for linguistics/NLP, summarize motivations and history of Linguistic Linked Open Data so far. Since creating the first instantiation of the LLOD cloud diagram in 2012, LLOD has attracted a lot of activity, we have reached an agreement on vocabularies for many aspects of language resources and the number of resources included is continuously on the rise. This growth is documented, for example, by declaring LLOD “the new hot topic in our (= language resource) community” (Nicoletta Calzolari, LREC-2014 closing session). But with substantial amounts of data being available, the focus of activity in the LLOD community is slowly shifting from resource creation to applications of Linguistic Linked Open Data. The primary promise of providing open, but heterogeneously structured and scattered language resources in a more interoperable way has been fulfilled, and it facilitates using and re-using existing language resources in novel contexts. Beyond this, innovative LLOD-based applications for common problems in Natural Language Processing, Digital Humanities and linguistics are on the horizon. The second part of the talk will give a glimpse on these prospects by discussing use cases and potential applications of LLOD for (Deep) Machine Translation.