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
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 QTLeap 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

<table>
<thead>
<tr>
<th>Name</th>
<th>Institution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eneko Agirre</td>
<td>University of the Basque Country</td>
</tr>
<tr>
<td>António Branco</td>
<td>University of Lisbon</td>
</tr>
<tr>
<td>Aljoscha Burchardt</td>
<td>Deutsches Forschungszentrum für Künstliche Intelligenz GmbH</td>
</tr>
<tr>
<td>Rosa Del Gaudio</td>
<td>Higher Functions LDA</td>
</tr>
<tr>
<td>Jan Hajíč</td>
<td>Charles University in Prague</td>
</tr>
<tr>
<td>Eva Hajíčová</td>
<td>Charles University in Prague</td>
</tr>
<tr>
<td>Petya Osenova</td>
<td>Sofia University and Bulgarian Academy of Sciences</td>
</tr>
<tr>
<td>Martin Popel</td>
<td>Charles University in Prague</td>
</tr>
<tr>
<td>Khalil Sima’An</td>
<td>University of Amsterdam</td>
</tr>
<tr>
<td>Kiril Simov</td>
<td>Bulgarian Academy of Sciences</td>
</tr>
<tr>
<td>Gertjan van Noord</td>
<td>University of Groningen</td>
</tr>
<tr>
<td>Deyi Xiong</td>
<td>Soochow University</td>
</tr>
</tbody>
</table>
# Table of Contents

Linguistic Linked Open Data: What’s in for Machine Translation? .......................... 1  
   Christian Chiarcos

Modelling the Adjunct/Argument Distinction in Hierarchical Phrase-Based SMT .......... 2  
   Sophie Arnoult and Khalil Sima’an

Towards Deeper MT - A Hybrid System for German ........................................... 12  
   Eleftherios Avramidis, Aljoscha Burchardt, Maja Popovic and Hans Uszkoreit

Splitting Compounds by Semantic Analogy ......................................................... 20  
   Joachim Daiber, Lautaro Quiroz, Roger Wechsler and Stella Frank

Delimiting Morphosyntactic Search Space with Source-Side Reordering Models .......... 29  
   Joachim Daiber and Khalil Sima’an

Evaluating a Machine Translation System in a Technical Support Scenario ............... 39  
   Rosa Del Gaudio, Aljoscha Burchardt and Arle Lommel

Development of Telugu-Tamil Transfer-Based Machine Translation system: With Special  
   reference to Divergence Index ............................................................................ 48  
   Parameswari Krishnamurthy

Deep-syntax TectoMT for English-Spanish MT ..................................................... 55  
   Gorka Labaka, Oneka Jauregi, Arantza Díaz de Ilarraza, Michael Ustaszewski, Nora Aran-  
   berri and Eneko Agirre

First Steps in Using Word Senses as Contextual Features in Maxent Models for Machine  
   Translation ............................................................................................................ 64  
   Steven Neale, Luís Gomes and António Branco

Lexical choice in Abstract Dependency Trees ....................................................... 73  
   Dieke Oele and Gertjan van Noord

Large Scale Translation Quality Estimation ........................................................... 81  
   Miguel Angel Ríos Gaona and Serge Sharoff

Translation Model Interpolation for Domain Adaptation in TectoMT ......................... 89  
   Rudolf Rosa, Ondřej Dusek, Michal Novák and Martin Popel

Factored models for Deep Machine Translation .................................................... 97  
   Kiril Simov, Iliana Simova, Velislava Todorova and Petya Osenova

Machine Translation for Multilingual Troubleshooting in the IT Domain: A Comparison  
   of Different Strategies ......................................................................................... 106  
   Sanja Štajner, João Rodrigues, Luis Gomes and António Branco
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.