



Abu-MaTran

AUTOMATIC BUILDING OF MACHINE TRANSLATION

PIAP- GA-2012-324414

D2.4 SaLTMiL Workshop

Dissemination level	Public
Delivery date	2015/12/31
Status and version	Final, v1.0
Authors and affiliation	Mikel L. Forcada (UA), Nikola Ljubešić (UZ) and Antonio Toral (DCU)



Project funded by the European Community under the Seventh Framework Programme for Research and Technological Development



Contents

Executive Summary	2
1 Introduction	3
2 Workshop Report	3
2.1 Regular papers	4
2.2 Discussion	4

Executive Summary

This deliverable corresponds to task T2.4 (scientific Workshop) within work package 2 (Dissemination and Outreach) aiming at the dissemination of the project for different audiences.

In this respect, the Abu-MaTran consortium co-organised in May 2014 the SaLTMiL workshop, an activity of the ISCA Special Interest Group on Speech and Language Technology for Minority Languages (SaLTMiL). This workshop was strategically co-located with the 2014 edition one of the main conferences of our area, the Language Resources and Evaluation Conference (LREC), held in Reykjavík, Iceland in May 2014.

The SaLTMiL workshop was about “free/open-source language resources for the machine translation of less-resourced languages”, a topic which is very relevant to our project as the languages of our case study (Croatian, Serbian, etc.) are considered under-resourced indeed (cf. e.g. META-NET White Papers).

One researcher of the project (Mikel L. Forcada, UA) was an organiser of the workshop and two researchers of the project (Antonio Toral, DCU and Nikola Ljubešić, UZ) were part of the Programme Committee. The workshop was well attended with 31 registered attendees and about 40 attendees in total.

1 Introduction

The initial plan, as described in Annex 1, was to organise a workshop of the FreeRBMT (Free/Open-source Rule-based Machine Translation) series during month 34 of the project (October 2015). However, the consortium decided to co-organise instead the “SaLTMiL workshop” in month 17 (May 2014), co-located with the 2014 edition of one of the main conferences of our area, the Language Resources and Evaluation Conference (LREC), held in Reykjavík, Iceland, in May 2014.¹

The FreeRBMT series of workshops² aimed to “bring together the experience of researchers and developers in the field of rule-based machine translation who have decided to get on board the free/open-source train”. The 2014 SaLTMiL workshop is related in that it is about “free/open-source language resources for the machine translation of less-resourced languages”. However, we thought that SaLTMiL suited the project better than FreeRBMT as SaLTMiL also put emphasis on less-resourced languages, and the languages of our case study (Croatian, Serbian, etc.) are indeed considered under resourced (cf. e.g. META-NET White Papers).

One researcher of the project (Mikel L. Forcada, UA) was an organiser of the workshop and two researchers of the project (Antonio Toral, DCU and Nikola Ljubešić, UZ) were part of the Programme Committee.

2 Workshop Report

The workshop was very well attended. About 40 people were present, more than the 31 attendees registered as of 22nd May, 2014. After a brief welcoming address by Mikel L. Forcada, there were two oral sessions where regular papers (peer-reviewed by the programme committee) were presented (cf. Section 2.1). This was followed by a discussion session (cf. Section 2.2).

¹<http://lrec2014.lrec-conf.org/en/>

²2009: <http://xixona.dlsi.ua.es/freerbmt09/>; 2011: www.uoc.edu/freerbmt11/; 2012: <https://www.chalmers.se/hosted/freerbmt12-en>.

2.1 Regular papers

Six papers were presented at the workshop. Details including the author names and titles follow. The proceedings are publicly available.³

1. **Iñaki Alegria, Unai Cabezón, Unai Fernández de Betoño, Gorka Labaka, Aingeru Mayor, Kepa Sarasola and Arkaitz Zubiaga.** Wikipedia and Machine Translation: killing two birds with one stone
2. **Gideon Kotzé and Friedel Wolff.** Experiments with syllable-based English-Zulu alignment In the second session, chaired by Trond Trosterud, the paper by Matthew Marting and Kevin Unhammer was presented by Francis Tyers as the authors could not make it to Iceland.
3. **Inari Listenmaa and Kaarel Kaljurand.** Computational Estonian Grammar in Grammatical Framework
4. **Matthew Marting and Kevin Unhammer.** FST Trimming: Ending Dictionary Redundancy in Apertium
5. **Hrvoje Peradin, Filip Petkovski and Francis Tyers.** Shallow-transfer rule-based machine translation for the Western group of South Slavic languages
6. **Alex Rudnick, Annette Rios Gonzales and Michael Gasser.** Enhancing a Rule-Based MT System with Cross-Lingual WSD

2.2 Discussion

The discussion session concerned two questions that were posed to the audience:

1. Is research in minority-language machine translation already mainstream?
2. What are the main difficulties in building or putting together free/open-source language resources for small languages, and how should they be addressed? Are we pooling these resources correctly?

³http://ixa2.si.ehu.es/%7Ejipsagak/SALTMIL/LREC_2014_Workshop_Proceedings_Saltmil.pdf

The discussion on the first question quickly turned into a discussion on rule-based MT (RBMT) systems for under-resourced languages: how hard it is to publish RBMT systems, the importance of linguistic research and training in computational linguistics, etc. The use of rich morphologies in statistical MT (SMT), was also heavily discussed, as well also the role of computational descriptions of the morphology of languages as a way to actually document the under-resourced languages. The pooling and licensing of linguistic resources for under-resourced languages was also discussed in detail. A more detailed account of the discussion — which, as has been seen, touched upon many topics which are important to the Abu-Matran project— can be found in SaLTMiL’s website.⁴

⁴<http://ixa2.si.ehu.es/saltnil/index.php/en/%3Fview=article&catid=13:home&id=104:report-on-the-9th-saltnil-workshop-reykjavik-2014&format=pdf>