



# ACCURAT

Analysis and Evaluation of Comparable Corpora  
for Under Resourced Areas of Machine Translation

[www accurat-project.eu](http://www accurat-project.eu)

**Project no. 248347**



**Deliverable D6.8**  
**Final leaflet and poster**

**Version No. 1.0**  
**29/06/2012**

**Document Information**

Deliverable number:	D6.8
Deliverable title:	Final leaflet and poster
Due date of deliverable:	29/06/2012
Actual submission date of deliverable:	29/06/2012
Main Author(s):	Marko Tadić
Participants:	Božo Bekavac, Željko Agić, Nikola Ljubešić
Internal reviewer:	Tilde
Workpackage:	WP6
Workpackage title:	Dissemination
Workpackage leader:	FFZG
Dissemination Level:	PU
Version:	V1.0
Keywords:	Poster, leaflet, dissemination materials

**History of Versions**

<b>Version</b>	<b>Date</b>	<b>Status</b>	<b>Name of the Author (Partner)</b>	<b>Contributions</b>	<b>Description/ Approval Level</b>
V0.5	05/06/2012	Draft	FFZG	Marko Tadić	First draft
V0.8	25/06/2012	Draft	FFZG	Marko Tadić, Božo Bekavac, Željko Agić, Nikola Ljubešić	A presentation of ACCURAT poster added
V1.0	29/06/2012	Final	FFZG	Corrected errors in mixing terms “flyer” and “leaflet”	Submitted

**EXECUTIVE SUMMARY**

With this deliverable a description of the final leaflet and poster is presented.

## Table of Contents

<b>1 Introduction</b> .....	<b>4</b>
<b>2. Dissemination activities</b> .....	<b>5</b>
2.1. Flyer and poster .....	5
2.2. T-shirts.....	9
<b>2 Conclusion</b> .....	<b>11</b>

## 1 Introduction

The overall goal of WP6 is to **disseminate** project results and to transfer the project knowledge, technologies, lessons learned and best practices to interested communities and thus to ensure their worldwide impact and long-term sustainability. Also, one of the goals of WP6 is to provide insight into exploitation plans of the tools developed within the ACCURAT project.

This deliverable D6.8 is the report on production of the final ACCURAT leaflet and final ACCURAT poster that were used for dissemination activities in order to spread the information about the ACCURAT project during the last four months of the project at the most important events in the field.

## 2. Dissemination activities

According to the implementation of dissemination activities described in the Dissemination and Exploitation Plan, the Final leaflet and the Final poster were scheduled for M30. In order to maximise the impact of our dissemination activities, we have produced the Final leaflet and the Final poster earlier, i.e. in April 2012. In this way these dissemination instruments were at our disposal for the main events in the field (e.g. LREC2012, EAMT2012, META-FORUM2012, TKE2012, etc. The detailed list of all dissemination activities is available at the D6.9).

In addition to the Final leaflet and Final poster, also 100 Final t-shirts were produced and distributed at these events.

The dissemination instruments described, have significantly raised the profile and visibility of the project covering classic dissemination channels (web-site, appearance at the conferences and papers in journals and proceedings).

### 2.1. *Leaflet and poster*

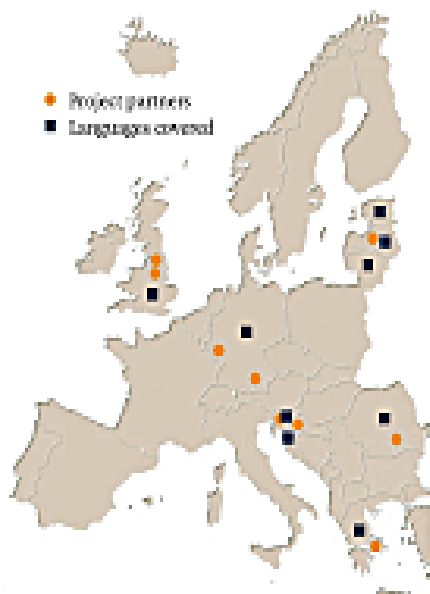
Following the Plan for dissemination and exploitation (D6.1) at M27 the Final leaflet and the Final poster were produced. They present the current results and achievements of the project partners, particularly stressing the availability of the ACCURAT Toolkit for processing comparable corpora.

The leaflets are A4 twofold full colour leaflet that yields effectively six pages. Posters are 70x100 cm in size and in full colour also.

The whole print run of the Final leaflet was produced and distributed between the project partners for further dissemination activities.

The poster and leaflet also demonstrate the cooperation of the ACCURAT project with META-NET alliance. The initial design and layout of the poster was adapted to accommodate the META-NET and META-SHARE logos and to denote the alignment of ACCURAT with META-NET activities.

Analysis and Evaluation of Comparable Corpora  
for Under-Resourced Areas of Machine Translation



A true cross-European project connecting under-resourced languages, that are either EU official languages or are in the process of becoming one soon, with well-resourced languages such as English and German. Languages covered by the project: Latvian, Lithuanian, Estonian, Romanian, Greek, Croatian, Slovenian, English and German.

[www accurat-project.eu](http://www accurat-project.eu)



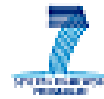
**Project partners**

- TTU** TTU, Riga, Latvia
- USFS** University of Sheffield, Computer Science Department, ML Group, Sheffield, UK
- CRU** University of Leeds, Centre for Translation Studies, Leeds, UK
- ILIP** Athena Research and Innovation Center in Information Communication & Knowledge Technologies, Institute for Language and Speech Processing, Athens, Greece
- IRIA** University of Zagreb, Faculty of Humanities and Social Sciences, Department of Linguistics, Zagreb, Croatia
- DFKI** Deutsches Forschungszentrum für Künstliche Intelligenz, IT Lab, Saarbrücken, Germany
- IRIT** IRIT, French Academy, Research Institute for Artificial Intelligence, Toulouse, France
- UW** Linguistik GmbH, Munich, Germany
- ZEMANTIA** Zemanoloco, Ljubljana, Slovenia

**Contact**

TTU, SL  
75a Vindabas Gatve  
LV-1004 Riga  
LATVIA

Contact person  
Ariana Fitzpatrick  
Project manager  
T: +371 67004001  
F: +371 67005250  
M: +371 29200344  
E: ariana.fitzpatrick@ttu.lv  
W: <http://www.ttu.lv>



The project has received funding from the European Commission's Horizon Research and Innovation Programme under Grant Agreement 101019124

META<sup>3</sup>NET

Analysis and Evaluation of Comparable Corpora  
for Under-Resourced Areas of Machine Translation

Is your translation  
**ACCURAT**  
enough?

[www accurat-project.eu](http://www accurat-project.eu)

Analysis and Evaluation of Comparable Corpora for Under-Resourced Areas of Machine Translation

**Main goal**

To develop methods and techniques to overcome one of the central problems of Machine Translation (MT) – the lack of linguistic resources for under-resourced areas of machine translation. The main goal is to find, analyse and evaluate novel methods that exploit comparable corpora in order to compensate for the shortage of linguistic resources, and ultimately to significantly improve MT quality for under-resourced languages and narrow domains.

**Using comparable corpora**

The applicability of current data-driven methods directly depends on the availability of large quantities of parallel corpus data.

For this reason the translation quality of current data-driven MT systems varies dramatically from quite good for language pairs with large corpora available (e.g. English and French) to almost unusable for under-resourced languages and narrow

[www.accurat-project.eu](http://www.accurat-project.eu)

domains where little data is available (e.g. Latvian and Croatian).

The goal of ACCURAT is to improve the machine translation translation quality for under-resourced languages and narrow domains.

- for under-resourced languages covering Latvian, Lithuanian, Bosnian, Greek, Croatian, Romanian and Slovenian;
- narrow domains (e.g. web page energy data processing, automotive engineering etc.) by finding out the novel approaches how comparable corpora can compensate for a shortage of parallel linguistic resources.

**Important results**

- **Comparability metrics** - developed methodology and identified features that are used to measure comparability of source and target language documents in comparable corpora;
- **Automated methods and tools for collecting comparable corpora** from the web;

- Several ACCURAT multilingual comparable and parallel corpora have been gathered from the Web;
- Research methods for alignment and extraction of lexical, terminological and other linguistic data from comparable corpora;
- Improved SMT and RBMT systems from applying acquired data against developed baseline results;
- ACCURAT Toolkit for collecting and processing comparable corpora – developed, published and freely downloadable from the project web site.

**Tested use cases**

Evaluation and validation of the ACCURAT project results in practical applications by adjusting and enhancing MT systems based on data collected from comparable corpora:

- **translation solutions in narrow domains** (renewable energy, sports news, political and financial news, ICT news, news on disasters, automotive engineering etc.);
- application in web authoring: blog authoring;
- software localisation: increasing efficiency in localization, integration with CAT tools.

ACCURAT Toolkit 2.0  
ACCURAT results will also be available through

Figure 1 ACCURAT Final leaflet

# Is your translation?

Analysis and Evaluation of Comparable Corpora for Under-Resourced Areas of Machine Translation

# ACCURAT

### Main goal

To develop methods and software to increase use of the central principle of Machine Translation (MT) - The lack of linguistic resources for under-resourced areas of machine translation. The main goal is to find suitable and find new (or) methods that exploit available corpora to solve the problem for the shortage of linguistic resources, and ultimately to improve MT quality for under-resourced languages and narrow domains.



### Key innovation

The creation of new methods and tools to increase use of the central principle of Machine Translation (MT) - The lack of linguistic resources for under-resourced areas of machine translation. The main goal is to find suitable and find new (or) methods that exploit available corpora to solve the problem for the shortage of linguistic resources, and ultimately to improve MT quality for under-resourced languages and narrow domains.

### Using comparable corpora

The main goal is to find suitable and find new (or) methods that exploit available corpora to solve the problem for the shortage of linguistic resources, and ultimately to improve MT quality for under-resourced languages and narrow domains.



### Results

• Comparable corpora - developed methods and software to increase use of the central principle of Machine Translation (MT) - The lack of linguistic resources for under-resourced areas of machine translation. The main goal is to find suitable and find new (or) methods that exploit available corpora to solve the problem for the shortage of linguistic resources, and ultimately to improve MT quality for under-resourced languages and narrow domains.



### A Cross-European Project

Cooperating under various languages that are under-resourced (languages) in the project of Machine Translation (MT) - The lack of linguistic resources for under-resourced areas of machine translation. The main goal is to find suitable and find new (or) methods that exploit available corpora to solve the problem for the shortage of linguistic resources, and ultimately to improve MT quality for under-resourced languages and narrow domains.

### ACCURAT Toolkit

Toolkit for collecting and processing comparable corpora from distributed data from the project website.

### Contact

For more information, please contact the project website.



META-SHARE ACCURAT Final poster META-NET  
**enough?**  
www.accurat-project.eu

Figure 2 ACCURAT Final poster



## 2.2. T-shirts

After surprisingly successful presentation of ACCURAT project by T-shirts at LREC2010, we decided that we will continue this type of disseminating activity. After the mid-term t-shirts we introduced and produced also the final t-shirts, using the same slogan as at the Final leaflet and the Final poster. They were distributed at conferences and other occasions.





Figure 3 ACCURAT Final t-shirt

## 2 Conclusion

In this deliverable a detailed description of dissemination instruments the Final leaflet, the Final poster and the Final t-shirt in the ACCURAT project is being provided. It describes how these instrument have been produced to convey the information about the project, to raise the awareness of its achievements and to present its results to the intended audience: research community, media professionals and general public.