

Abu-MaTran seeks to enhance **industry-academia cooperation** as a key aspect to tackle one of Europe's biggest challenges: multilingualism. We aim to increase the hitherto low **industrial adoption of machine translation** by identifying crucial cutting-edge research techniques that can accelerate the deployment of Machine Translation systems, preparing them to be suitable for commercial exploitation and finally transferring this knowledge to industry. The main tools developed in the project in this regard so far include:

- Tools to **accelerate the development of statistical machine translation systems**: crawlers of parallel and monolingual corpora from the web. These include *Bitextor* and *ILSP Focused Crawler* to crawl parallel data, *TweetCaT* to crawl tweet corpora and *Spidextor* to crawl jointly parallel and monolingual data from top-level domains.
- Tools to **accelerate the development of rule-based machine translation** systems:
  - A paradigm annotation tool, that allows users to extend morphological dictionaries with new entries.
  - Tools to assist with the extension of dictionaries, addressed both at experts (paradigm annotation tool) and non-experts (dynamically generated yes/no questions).
  - A tool for the automatic acquisition of transfer rules from small parallel corpora.
- Techniques to **improve the translation quality for morphologically rich languages**: morphological segmentation, both linguistically-motivated and unsupervised.

We have demonstrated the usefulness of these components by participating in **shared tasks**. Namely, we have taken part in WMT 2014 translation task (top rank, English-to-French), WMT 2015 translation (top rank, English-to-Finnish) and quality estimation (top rank, word-level sub-task) and in TweetMT 2015 tweet translation (top rank, Spanish from/to Catalan and Spanish-to-Basque).

While the tools developed in the project are to a large extent language independent, we have a specific **case study**, selected due to its strategic interest for Europe; we provide machine translation for the language of a new member state (Croatian) and then extend to related languages of candidate member states.

Abu-MaTran also works on **knowledge transfer from industry to academia**. In this regard we aim to transfer to academia the know-how of industry regarding management, processes, etc. to make research products more robust. In this regard, the consortium has organised workshops devoted to the topics of efficient software management and linguistic data creation.

The project exploits the **open-source business model**: all the resources produced are released as free/open-source software, resulting in effective knowledge transfer beyond the consortium. All the resources are available on the project's website.

Finally, we have an ambitious outreach plan, whose flagship activity is the establishment of a national **Linguistic Olympiad** in Spain (OLE). This has attracted so far the participation of around 400 students in its 2014 edition and over 600 in 2015.

More info [abumatran.eu](http://abumatran.eu) · [ole.abumatran.eu](http://ole.abumatran.eu) · [paradigm.abumatran.eu](http://paradigm.abumatran.eu) · [translator.abumatran.eu](http://translator.abumatran.eu)