Publishable summary

Context

The diversity of languages in Europe makes translation vitally important to the economic, cultural and social lives of Europeans. Machine translation (MT) provides a way of fully or partially automating the translation process, and hence reducing the costs and enabling more text and speech to be translated.

Research in MT has progressed at a rapid rate in recent years, including the recent EU funded projects EuroMatrix (Framework 6) and EuroMatrixPlus (Framework 7). Statistical machine translation (SMT) has become the dominant paradigm, a technique that requires a large quantity of parallel text in the desired language pair to build a system. Machine translation, however, is still a complex field and presents substantial barriers for entry to potential researchers, and to potential users of the technologies. The principal aim of the MosesCore project is to reduce these barriers, making it easier to join and participate in the MT research community, and easier to become an MT user. Through the coordination of freely available shared software and data, and by organising appropriate networking events we aim to stimulate the research and adoption of MT in Europe.

The MosesCore project addresses the needs of both machine translation researchers, seeking to improve the state-of-the-art in MT by building on top of existing technologies, and also users of machine translation. The latter may be based in commercial organisations or public bodies, and will be interested in using MT technology to build translation systems which are tailored to their needs, principally using open-source software. We also bring these two groups of people together in shared events, so that researchers can be better informed about the needs of the users of the technologies, and so that users can learn how best to apply the latest research.

We are trying to create an environment where machine translation technology is open to all, not just the big research laboratories and large corporations. In other words, to make sure that there is a place for research teams of all size, from "one grad-student" upwards, and where small and medium sized enterprises can easily create their own customised translation systems. This requires the existence of high quality, stable and well-supported open-source implementations of existing translation technology, which researchers can use to implement their new ideas. We coordinate the development and maintenance of such software, and also organise the release of appropriate standard data sets.

In this respect, the MosesCore project builds upon the successful strategy of the EuroMatrix and EuroMatrixPlus projects of supporting the open-source Moses translation toolkit, as it expands and develops in its role as the open-source SMT platform of choice for MT researchers and users. In the MosesCore project we provide stewardship for Moses development, ensuring that it retains its position by incorporating the latest research and also improving the stability, ease-of-use and flexibility of the toolkit.

In order to foster collaboration in the development and use of machine translation technology, the MosesCore project organises appropriate meetings for researchers and users. These

meetings take several different forms, serving different combinations of participants, and offering different styles of event, but all with the objective of making machine translation research more productive, and driving the industry take-up of this technology. For researchers we offer both the traditional style of researcher meeting consisting of seminars and discussions (the Workshop on Machine Translation), as well as the more collaborative and practical "Machine Translation Marathons". where participants work together on useful projects in small groups. These marathons have been broadened in scope to attract a larger attendance from the translation technology user and developer community, who are currently trying to build systems using open-source translation tools. We also seek to encourage the commercial acceptance of Moses (and other open-source tools) through "Industrial Outreach" events, enabling users and developers to share experiences about the use of open-source machine translation.

Objectives

The major objectives of the MosesCore project are
- To provide coordination and stewardship of the development of open source software for machine translation, notably the Moses statistical MT toolkit. This will result in at least three major releases of Moses, one in each year of the project.
- Outreach to the research community through academic workshops, evaluation campaigns and the machine translation marathons.
- Outreach to current and potential users of MT by providing a well maintained web presence, an active newsletter, and three annual outreach events for knowledge sharing and tutorials .
- Improved interaction between academic and industrial MT stakeholders through both the outreach events and tutorials , and the marathons.

Work Packages

The MosesCore project is made up of four main work packages.

WP1.   Moses Coordination and Integration
WP2.   Machine Translation Marathons
WP3.   Workshop and Evaluation Campaign
WP4.   Industrial Outreach

Partners

The MosesCore consortium consists of the following partners

UEDIN	University of Edinburgh (Coordinator)	United Kingdom
TAUS	TAUS	Netherlands
FBK	Fondazione Bruno Kessler	Italy

| CUNI | Charles University, Prague | Czech Republic |
| CTI | Capita Translation and Interpreting | United Kingdom |

Note the CTI were formerly known as ALS (Applied Language Solutions).

Impact

We believe that MosesCore can impact on everyone who comes into contact with machine translation, whether they are researchers, developers or users, as well as impacting on the wider society.

For machine translation researchers, MosesCore supplies them with both standard tools and data-sets, and coordinates meeting for the to network and share ideas. The Moses toolkit provides them with both a standard baseline and a standard platform on which to implement their ideas. MosesCore enables the continued development of this toolkit, as it both incorporates the latest research, and improves in stability and usability. The MosesCore project also supplies researchers with standard, freely available, data sets, released each year for the shared task.

This shared task consists of a small number of related sub-tasks, each of which allows researchers to compare their machine translation techniques with others in a controlled setting. We also organise extensive human evaluation of the results of machine translation. The results of the shared task are reported at the annual Workshop in Machine Translation (WMT), organised by MosesCore, a major academic workshop which also includes MT research papers.

Machine translation researchers also benefit from the MT marathons, which bring together MT researchers, as well as developers and some users, to collaborate on developing new implementations of MT technology. The MT marathon includes a summer school to help introduce new researchers and developers to the field.

The industrial outreach events organised by MosesCore benefit new and potential users and developers of MT, by introducing them to the technology (and open-source MT in particular) and providing tutorials and knowledge sharing. These events also serve to gather information on users' MT requirements, which can be passed back to developers and researchers.

Ultimately, the wider society should benefit from the improvements to MT which MosesCore will help to enable.

Progress Summary

In the first year of MosesCore the main accomplishments were:

- A release v1.0 of the Moses SMT toolkit, with a significantly expanded testing procedures, and incorporating many new features and improvements.

- Well-attended industrial outreach (MT Showcase) events organised in Asia, Europe and North America.

- A successful MT Marathon, held in Edinburgh, Scotland in September 2012.

- Preparations for the first MosesCore workshop and shared task, to be held in Sofia, Bulgaria in August 2013.

For the first release of Moses, we have expanded the testing regime to include complete end-to-end testing of the training pipeline, and unit-testing of individual classes and methods, as well as increasing the coverage and extendibility of the existing component-wise regression testing. The first year has seen a significant amount of new functionality incorporated into Moses (see the release notes at http://www.statmt.org/moses/RELEASE-1.0/release-notes.pdf) as well as many improvements and bug-fixes.

The MT showcase events, organised by TAUS, took place in Monaco, Beijing and Seattle in 2012, and the attendance met or exceeded our targets. These events were supplemented by a bimonthly newsletter, and an outreach website featuring uses cases and reports, and an online MT tutorial.

The week-long MT Marathon in Edinburgh was attended by around 70 participants from Academia and Industry, and featured a "summer" school, invited talks, MT hacking projects, labs and papers describing new tools. Planning has already started for the 2013 edition in Prague.

The progress on WP3 (shared task and evaluation) has been slower, since the first MosesCore-sponsored workshop is not due to take place until August 2013. We have been investigating and piloting improved methods of human evaluation, as well as gathering new training and test data.

Website and Contact

More information about MosesCore can be found on the website (www.mosescore.eu) or by contacting the project email address (info@mosescore.eu).