

Impact case study (REF3b)

Institution: The University of Edinburgh						
Unit of Assessment: B11 — Computer Science and Informatics						
Title of case study: The Moses Machine Translation Toolkit						
<p>1. Summary of the impact</p> <p>The research on machine translation carried out at the University of Edinburgh has led to the development of Moses, the dominant open source toolkit for building machine translation (MT) systems. The toolkit has found wide adoption in academic research worldwide: the Moses paper was the most cited paper in all of the Association for Computational Linguistics conferences in 2011. Moses has also been widely used by commercial concerns such as Adobe, Symantec and Sybase, and agencies such as the European Commission and the World Trade Organisation. The research contribution of the School of Informatics in the University of Edinburgh has significantly increased the commercial viability and availability of machine translation.</p> <p>The toolkit has been one of the main drivers in lowering the barrier to entry to machine translation, making MT available to small and medium-size companies and opening up new markets and opportunities.</p> <p>Today, Moses is one of the most widely adopted MT systems in the translation industry, dominating the open-source space for MT. Its maturity and quality, as well as its liberal open-source license, means that it is often preferred over proprietary systems.</p>						
<p>2. Underpinning research</p> <p>Key researchers at the University of Edinburgh:</p> <table border="1"> <tr> <td>Philipp Koehn, Professor. 2005–present.</td> <td>Miles Osborne, Lecturer 2000–2005. Reader. 2005–present.</td> </tr> <tr> <td>Hieu Hoang, PhD candidate 2005–2010, supervised by P. Koehn. Researcher 2010–present.</td> <td>Barry Haddow, Researcher 2008–present.</td> </tr> <tr> <td>Phil Williams, PhD candidate 2007–2012, supervised by P. Koehn.</td> <td>Abhishek Arun, PhD candidate 2005–2010, supervised by P. Koehn.</td> </tr> </table> <p>Machine translation is a research field that investigates the use of a computer to translate from one natural language to another. It has obvious practical benefits in enabling people to communicate with others who do not share a common language.</p> <p>Most modern MT research focuses on the use of machine learning and statistical techniques to create translation systems. The creation of these systems usually requires a large amount of language-dependent data while the software algorithms and designs remain language-agnostic.</p> <p>In 2005, the Edinburgh MT group originated the Moses toolkit. This was used as the basis of an internationally competitive Human Language Processing workshop at Johns Hopkins University in the summer of 2006. The US National Science Foundation funded the workshop. This workshop yielded the first public release of Moses. (Ref: http://www.clsp.jhu.edu/workshops/archive/ws06/).</p> <p>Moses is an implementation of the statistical (or data-driven) approach to machine translation. An efficient search algorithm finds quickly the highest probability translation among the exponential number of choices. This is the dominant approach in the field at the moment, and is employed by the on-line translation systems deployed by Google and Microsoft.</p> <p>The toolkit is continuously being developed to improve its efficient and usability, and to incorporate the latest advances in MT research. University of Edinburgh researchers are at the forefront of the development of the toolkit.</p>	Philipp Koehn , Professor. 2005–present.	Miles Osborne , Lecturer 2000–2005. Reader. 2005–present.	Hieu Hoang , PhD candidate 2005–2010, supervised by P. Koehn. Researcher 2010–present.	Barry Haddow , Researcher 2008–present.	Phil Williams , PhD candidate 2007–2012, supervised by P. Koehn.	Abhishek Arun , PhD candidate 2005–2010, supervised by P. Koehn.
Philipp Koehn , Professor. 2005–present.	Miles Osborne , Lecturer 2000–2005. Reader. 2005–present.					
Hieu Hoang , PhD candidate 2005–2010, supervised by P. Koehn. Researcher 2010–present.	Barry Haddow , Researcher 2008–present.					
Phil Williams , PhD candidate 2007–2012, supervised by P. Koehn.	Abhishek Arun , PhD candidate 2005–2010, supervised by P. Koehn.					

3. References to the research

3.1. Relevant papers

1. **Moses: Open Source Toolkit for Statistical Machine Translation**, *Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondrej Bojar, Alexandra Constantin, Evan Herbst*, ACL 2007, demonstration session.
 - a. WWW <http://www.aclweb.org/anthology/P07-2045>
 - b. PDF <http://www.aclweb.org/anthology-new/P/P07/P07-2045.pdf>
2. **Factored Translation Models**, *Philipp Koehn and Hieu Hoang*, EMNLP 2007.
 - a. WWW <http://www.aclweb.org/anthology/D/D07/D07-1091>
 - b. PDF <http://acl.ldc.upenn.edu/D/D07/D07-1091.pdf>
3. **Agreement Constraints for Statistical Machine Translation into German**, *Philip Williams and Philipp Koehn*, Proceedings of the Sixth Workshop on Statistical Machine Translation (WMT), 2011.
 - a. WWW <http://www.aclweb.org/anthology/W11-2126>
 - b. PDF <http://www.aclweb.org/anthology-new/W/W11/W11-2126.pdf>
4. **Enriching Morphologically Poor Languages for Statistical Machine Translation**, *Eleftherios Avramidis and Philipp Koehn*, ACL 2008.
 - a. WWW <http://www.aclweb.org/anthology/P/P08/P08-1087>
 - b. PDF <http://aclweb.org/anthology-new/P/P08/P08-1087.pdf>
5. **Improving Mid-Range Reordering using Templates of Factors**, *Hieu Hoang and Philipp Koehn*, EACL 2009.
 - a. WWW <http://www.aclweb.org/anthology/E09-1043>
 - b. DOI <http://dx.doi.org/10.3115/1609067.1609108>
6. **Monte Carlo Inference and Maximization for Phrase-based Translation**, *Abhishek Arun, Chris Dyer, Barry Haddow, Phil Blunsom, Adam Lopez and Philipp Koehn*, Conference on Computational Natural Language Learning, 2009.
 - a. WWW <http://www.aclweb.org/anthology/W09-1114>
 - b. WWW <http://dl.acm.org/citation.cfm?id=1596394>

Publications [1], [2], and [4] are most indicative of the quality of the underlying research.

3.2. Grants and scholarship awarded

- MosesCore, <http://www.statmt.org/mosescore/>, 01/02/12–31/01/15. Value: £399,046. European project EC-FP7-288487.
- EuromatrixPlus, <http://www.euromatrixplus.eu>, 01/03/09–30/04/12. Value: £730,345. European project EC-FP7-231720.
- Euromatrix, <http://www.euromatrix.net>, 01/09/06–28/02/09. Value: £442,097. European project FP6-034291.
- LetsMT, <https://www.letsmt.eu/>, 01/03/10–31/08/12. Value: £138,560. European project EC-CIP-ICT-PSP2009-3.

4. Details of the impact

4.1. Commercialization

The development of Moses has led to a significant increase in the understanding and use of machine translation in the translation industry [A]. The free, open-source license has allowed many organizations to access the latest developments in MT research that had once been the preserve of governments and large IT companies. AVB Translations, a leading Netherlands translation firm (turnover in 2011 of €4.6M) with specific complex translation requirements requiring rapid accurate confidential translations between English and Dutch of legal texts reported that Moses produces “usable translations very quickly and at 50%–60% of normal translation cost” [B].

Impact case study (REF3b)

Industry bodies such as the Translation Automaton User Society (TAUS, <http://www.taus.net>) have championed the benefits of MT and Moses, educating industrial users on how to profit from MT as well as the mechanics of using the software.

A commercial ecosystem has formed around the toolkit, consisting of large and small companies, multinational organizations, users and suppliers of Moses-based services [C]. As TAUS has observed, since 2009 when MT became a mainstream industry tool due to the...

commoditization of one of the most significant and far-reaching innovations in translation technology...

*We've moved from a handful of commercially usable MT providers to a few dozen in a few short years. Many of these newcomers are using the open source statistical MT solution, **Moses**...*

A wide variety of commercial Moses-based MT solutions have emerged, ranging from self-service solutions to full service customization.

The Moses toolkit has also spawned other open-source projects such as 'Moses 4 Localization' [D] and 'Do Moses Yourself' [E] which fills in the gap between research-led MT and the commercial world.

4.2. Users of Moses

The following is a list of organisations that are known to be using the Moses toolkit:

International Organisations:

European Commission [G], World Trade Organisation, World Intellectual Property Organisation

Companies:

Autodesk [I], Adobe, Xerox Research Centre Europe, Symantec, Sybase

Translation Agencies and Technology Providers:

Applied Language Solutions (ALS), AVB Translations, CrossLang, Duolingo, Digital Silk Road, Hunnect, Logrus International, Lucy Software, Precision Translation Tools, Systran, myGengo, Moravia, Pangeanic, Safaba, Simple Shift, Tauyou, Tilde, Translated, Trusted Translations

4.3. Impact benefits delivered by Moses

Translation cost, information privacy protection, and translation quality are three of the main advantages of using Moses for the translation industry.

- Considering translation cost, Google Translate charges \$20 per million characters for bulk users. The Premium Translation software by Systran, a well-known translation software developer, cost £619. Licenses for Language Weaver's software vary in price between \$5,000 and \$125,000. In contrast, Moses is free.
- Considering information privacy, a problem arises because translation services like Google Translate cannot be used by organisations such as the UN who handle commercially and politically sensitive documents. In contrast, Moses can be run on a user's own internal computer system, creating no concerns about sensitive information leaving the host's domain.
- Concerning translation quality, Moses has been shown to significantly outperform general-purpose translation engines such as Google Translate when the system is trained with in-domain data for specific clients. The paper "Let's MT! — A Platform for Sharing SMT Training Data" by Tiedemann and Weijnitz (<http://bit.ly/19f7jr5>) compares Google Translate and Moses, trained with in-domain, user-specific data for English-Swedish translation and uses the well-known BLEU (Bilingual Evaluation Understudy) metric to compare the results. (BLEU is a metric that is highly correlated with human judgements of translation quality.) Tiedemann and Weijnitz report that Moses has a 20% higher BLEU score on English-to-Swedish translation than Google Translate and a 40% higher BLEU score on Swedish-to-English translation.

4.4. Opening up new translation fields

The availability of high quality, cheap, fast machine translation offered by the Moses toolkit has made machine translation more affordable and accessible to more companies. As well as productivity gains and lower prices, new markets have also opened up for machine translation. We describe three examples below.

1. Services such as the EC's Europe Media Monitor (EMM, <http://emm.newsbrief.eu>) translate over a hundred thousand articles a day in 50 languages for dissemination within the organisation. The EMM have used Moses since 2009 and are one of the most high profile users [G].
2. Electronic discovery (e-discovery) [H] is the digital forensic analysis of vast amounts of information during litigation and commercial transactions, such as company takeovers, in order to find interesting and relevant information. The global economy has increased the need for high-speed bulk translation of foreign-language documents and emails during this process. Technology providers such as Simple Shift (simple-shift.com) use Moses as the underlying technology to build translation systems for this market.
3. Translation of real-time interactive chat and near real-time translation of user reviews, public forums, and bulletin boards have been demonstrated. Multilingual interaction cannot afford the luxury of expert human post-revision but they must also be of sufficient quality. In contrast to general-purpose translation services such as Google and Bing, systems built on Moses can be trained on user-specific and domain-specific data, resulting in better translation quality [J].

4.5. Teaching machine translation to graduate-level students

The Moses toolkit has also significantly increased the awareness and teaching of machine translation. In addition to the educational workshops run by TAUS aimed at industrial users, academic-led workshops ('MT Marathon') [F] are organised annually by participants of the Euromatrix and EuromatrixPlus projects where machine translation is taught to graduate students for a week. It not only provides hands-on experience for students but also allows them meet and interact other students and teachers in an informal setting. It is usually one of the most intensive and in-depth exposures these students have had to MT.

Most students come from institutions other than the University of Edinburgh, often from institutions with small machine translation groups. The MT Marathon is a rare chance for them to meet other MT researchers. Building the community of machine translation experts in this way is another impact of the research done at Edinburgh in machine translation.

5. Sources to corroborate the impact

Evidence for the impact of the Moses MT toolkit can be found at the following web addresses.

- A. <https://www.taus.net/articles/six-moses-machine-translation-use-cases>
- B. <http://www.slideshare.net/TAUS/1545-avb-legal-moses-engine>
- C. <http://www.tausdata.org/blog/2010/10/doing-business-with-moses-open-source-translation/>
- D. <http://groups.google.com/group/m4loc>
- E. <http://www.precisiontranslationtools.com/>
- F. www.statmt.org/mtm12
- G. <http://www.mt-archive.info/MTMarathon-2011-Turchi.pdf>
- H. <http://www.catalystsecure.com/blog/2012/05/was-samsung-deal-a-watershed-for-use-of-machine-translation-in-ftc-second-requests/>
- I. <http://translate.autodesk.com/productivity.html>
- J. <http://www.translationautomation.com/technology/a-snapshot-of-real-time-multilingual-chat.html#comments>

Archive copies of these webpages are available from <http://ref2014.inf.ed.ac.uk/impact/>