INTEGRATING MACHINE TRANSLATION INTO MOOCS

Conference Paper · March 2017

DOI: 10.21125/edulearn.2017.0765

4 authors:

Sheila Castilho
Dublin City University
19 PUBLICATIONS **89** CITATIONS
SEE PROFILE

Federico Gaspari
Unversity of Bologna at Forlì, Italy
26 PUBLICATIONS **113** CITATIONS
SEE PROFILE

Joss Moorkens
Dublin City University
26 PUBLICATIONS **56** CITATIONS
SEE PROFILE

Andy Way
Dublin City University
263 PUBLICATIONS **2,703** CITATIONS
SEE PROFILE

INTEGRATING MACHINE TRANSLATION INTO MOOCS

Sheila Castilho, Federico Gaspari, Joss Moorkens, Andy Way

ADAPT Centre, Dublin City University (IRELAND)

Abstract

This paper presents TraMOOC (Translation for Massive Open Online Courses), a European research project developed with the intention of empowering international learners in the digital multilingual world by providing reliable machine translation (MT) specifically tailored to MOOCs from English into 11 languages (Bulgarian, Chinese, Croatian, Czech, Dutch, German, Greek, Italian, Polish, Portuguese, and Russian). The paper describes how the project is addressing the challenges involved in developing an innovative, high-quality MT service for producing accurate translations of heterogeneous multi-genre MOOC materials, encompassing subtitles of video lectures, assignments, tutorials, and social web text posted on student blogs and fora. Based on the results of a large-scale and multi-method evaluation conducted as part of the TraMOOC project, we offer a reflection on how to best integrate state-of-the-art MT into MOOC platforms. The conclusion summarizes the key lessons learned, that can be applied by the wider community of international professionals with an interest in the multilingual aspects of innovative education and new learning technologies.

Keywords: MOOCs, machine translation (MT), translation, e-learning, distance learning.

1 INTRODUCTION

1.1 Background and motivation of the study

Massive Open Online Courses (MOOCs) offer valuable learning opportunities in several disciplines to many students, to a large extent regardless of their background, location, and personal circumstances [1]. Views about the actual potential of MOOCs inevitably vary, mostly depending on the subjects being taught and on the pedagogic attitudes of the instructors ([2], [3], [4] and [5]), but MOOCs are gradually starting to have an impact on teaching practice, at least for some disciplines (see, e.g., [6]). One widely held view is that MOOCs may represent effective means of disseminating knowledge and training to disadvantaged communities or individual students living in remote areas, with limited or no access to traditional teaching and learning facilities, such as colleges, public libraries, qualified teaching staff, technical equipment or laboratories [7]. However, rather surprisingly, there is growing evidence that MOOC participants are in fact predominantly already qualified professionals from privileged backgrounds mostly based in high-income, industrialized countries (e.g. [8], [9] and [10]).

One explanation of this seeming failure of MOOCs' original intended mission of broadening access to education and training is that this disappointing situation hinges significantly on language-related limitations. MOOCs are typically available in one language that is shared between tutors and students, which has the added bonus of enabling interactions on social platforms and fora accompanying formal instruction [11]. However, language barriers impede broad use of high-quality MOOC materials across national and language boundaries, severely limiting peer-to-peer as well as student-instructor interactions alongside the more formal components of MOOC-based instruction: English is often chosen as the common language of MOOCs with international reach; this, however, is far from ideal, especially because it prevents large groups of potential users from fully engaging in a fulfilling MOOC experience, thus wasting precious learning opportunities for innumerable motivated students around the world. In an increasingly globalized and mobile society, in which academic institutions as well as individual trainers are under growing pressure to seize the opportunities offered by internationalization, there is a strong need for high-quality digital teaching and learning resources to be distributed across linguistic and cultural boundaries [12].

Against this background, the paper reports the experience of the international research project TraMOOC (Translation for Massive Open Online Courses, whose official website can be visited at http://tramooc.eu/). The paper is structured as follows: after these introductory remarks on the background and motivation of the study, Section 1.2 provides more detail on the project, emphasizing its aims and expected outcomes. Section 2 describes the evolution of the main approaches to MT system design, from the traditional rule-based architecture to the more recent statistical and neural

paradigms, that are now competing to be recognized as the state-of-the-art. Section 3 discusses the application of MT for MOOCs, highlighting the difficulties inherent in the types of texts that form a MOOC, and Section 4 details our development and evaluation of MT systems within the TraMOOC project. Finally, Section 5 concludes by summarizing the key lessons learned from this work that can be useful to the wider community of instructors and institutions interested in delivering innovative and effective education opportunities via MOOCs to multilingual students, also outlining some possibilities for future work in the rapidly evolving area at the crossroads of MOOCs and MT.

1.2 The TraMOOC project: aims and expected outcomes

One issue that cuts across all MOOCs with significant impact on their uptake and effectiveness is that of the language(s) of instruction: this, in itself, is a crucial factor in restricting or, on the contrary, widening access to education and training delivered via MOOCs [13]. Making MOOC contents available in multiple languages has obvious benefits, and there have already been attempts to support language diversity within MOOCs with a European focus [14]. In the ambitious attempt to address the numerous and complex challenges entailed by this endeavour, TraMOOC aims at developing high-quality MT of the multifarious text genres typically included in MOOCs from English into 9 European (i.e. Bulgarian, Croatian, Czech, Dutch, German, Greek, Italian, Polish and Portuguese) and 2 so-called BRIC languages (namely, Chinese and Russian). While these diverse target languages constitute strong use cases in the MOOC space, some of them have been proven difficult to translate into, which is further compounded by the weak or fragmentary support in terms of language resources and processing tools that are required to build some of the relevant MT systems. This scenario poses significant research and development challenges to the TraMOOC project consortium.

The main outcome of the project lies in the development of a high-quality semi-automated MT platform for all types of textual data normally encountered in MOOCs, which typically range from subtitles of video lectures to instructions for completing assignments, presentation slides, posts shared on student blogs and comments sent to course fora. The core of the final service will be open-source and some premium add-on services are expected to be commercialized, including MT support for additional target languages of interest to the users, MT post-editing, transcription and subtitling of video-based course contents, as well as professional translation. The ultimate goal is to turn the MOOC translation service into a platform enabling the integration of any MT system chosen by the users, for any desired language, for the educational domain.

2 THE EVOLUTION OF MACHINE TRANSLATION SYSTEM DESIGN

MT has made substantial progress over the course of its history. Until the mid-1990s, rule-based MT systems were the norm: these required significant investments and huge resources to be built, including skilled computational linguists and programmers. This meant that MT systems were available only for a limited number of well-resourced languages with substantial commercial interest. In the late 1990s, a new data-driven paradigm emerged in MT system development, namely statistical MT (SMT), which quickly became the dominant approach in both research and market-oriented commercial applications. The principle underlying this approach is to do away with explicit linguistic rules altogether. In contrast, translation patterns (i.e. correspondences between phrases in the source and in the target languages) are inferred automatically from the analysis of parallel corpora, i.e. huge collections of sentence-aligned professional (i.e. human-quality) translations. SMT systems estimate the degree of probability for the correspondence of short bilingual chunks of text extracted from the analysis of the parallel corpora, and subsequently generate the output in the target language based on complex statistical calculations.

SMT systems can be built much faster and at a fraction of the cost of traditional rule-based ones, for many more language pairs, using open-source development toolkits, such as Moses [15]. In addition, SMT systems can be customized much more effectively than rule-based ones to different domains and text types. More recently, the neural approach has emerged as a promising further development in MT system design, attracting interest not only from academic researchers, but also from players in the language, translation and localization industry, because neural MT (NMT) systems have outperformed SMT systems for a number of language pairs in recent comparative evaluations. Simply put, NMT exploits neural networks and deep learning techniques drawn from artificial intelligence to map entire sentences from the source to the target language all at once, instead of breaking them down into smaller units (typically individual words, or fixed sequences of a few words), as is the case in SMT. This offers some advantages, although it is still debated whether NMT is superior to SMT.

3 MACHINE TRANSLATION FOR MOOCS

Several recent studies address the crucial issue of evaluating and improving the quality, effectiveness and success of MOOCs (see, e.g., [16] and [17]), and research has also been devoted to evaluating the level of engagement afforded by MOOCs (e.g. [18]). This body of work provides, either implicitly or explicitly, indications concerning good practice [19]. What is conspicuously absent from this substantial body of work is the language dimension of MOOC-based instruction, especially when, as is often the case, MOOCs have the ambition of being delivered internationally, to course takers with different linguistic and cultural backgrounds: this necessarily raises the issue of how to effectively translate these digital teaching and learning resources, so that their eventual multilingual nature contributes to their overall value for students, rather than detracting from it.

We regard this as a major gap in the MOOC literature, and contend that the language used to impart knowledge and support interactions associated with MOOCs is a key factor in the quality, effectiveness and success of learning experiences for international students, which should receive more attention, and this paper wishes to represent a first step in this direction. The broad questions addressed in the work reported here are whether the time is ripe for the integration of MT into MOOCs, and how to best go about selecting the most effective MT solution for this purpose.

A particularly interesting application domain that has recently emerged for MT concerns user-generated content (UGC) [20]. Successful techniques have been developed, for example, for the domain adaption of MT systems to deal with user comments in the e-commerce scenario [21], with several experiments showing the feasibility of this rather challenging task, even though it is certainly hard to obtain high-quality MT output in this area. UGC is also found in typical MOOC data, and the TraMOOC project aims at providing reliable MT for it, too, which is extremely challenging, because UGC is often poorly formulated, with relatively frequent spelling mistakes and grammatical inaccuracies, and more generally sub-standard, or non-conventional, language.

4 MACHINE TRANSLATION FOR TRAMOOC

For the TraMOOC project, we undertook to evaluate which of the two leading approaches to MT system design competing to be the state-of-the-art in the field, namely SMT or NMT (see Section 2), is better suited to be integrated into a MOOC platform to effectively deliver digital learning resources multilingually. The overall study is reported in more detail in [22].

The SMT and the NMT systems used for this evaluation were built using state-of-the-art procedures, aimed at guaranteeing the highest possible quality; in particular, for the statistical approach, a phrase-based architecture was used, while the NMT systems generally followed the settings of [23]. All the systems were trained on a variable mix of general, i.e. out-of-domain, and in-domain educational data, due to the different resources available for each language combination. The general training data ranged from a minimum of 21.30 million sentence pairs for EN-RU, to a maximum of almost 32 million sentence pairs for EN-PT; the much smaller in-domain training data sets consisted of a minimum of approximately 140000 sentence pairs for EN-EL, going up to 2.31 million sentence pairs for EN-RU.

Four sets of 250 English sentences each were translated into German, Greek, Portuguese and Russian using the SMT and NMT systems. Our evaluation is based on a set of four widely used automatic MT quality evaluation metrics: HTER (Translation Error Rate) [24], BLEU [25], METEOR [26] and chrF [27]. For the human assessment, we have selected the following state-of-the-art metrics: fluency and adequacy, post-editing, error annotation and ranking. Professional translators were asked to rate the translations according to those metrics and to post-edit the sentences. These procedures are widely used in the MT field in order to assess the quality of a given MT system.

The results of this large-scale evaluation, which are reported in full in [22], show that NMT receives higher scores than SMT with all four automatic evaluation metrics (even though improvements for Portuguese are very limited), and side-by-side ranking also shows a clear preference for NMT output across the board, for all the language pairs and MOOC domains covered in this comparative study. We can also conclude that NMT offers improvements in terms of fluency and word order errors over SMT, mostly due to its better handling of word reordering. In addition, fewer sentences translated with the NMT systems include errors, and NMT seems to perform better than SMT on morphologically rich and highly inflected target languages.

In contrast, however, adequacy does not show marked improvements with NMT, and the situation is mixed for errors of omission, addition and mistranslation, so much so that overall NMT does not entail

noticeable reductions in post-editing effort. Moreover, in-depth investigations of automatic MT evaluation metric scores reveal that the performance of NMT tends to degrade for longer sentences (more than 20 tokens), where SMT appears to be more reliable: the sentence length of the MOOC text to be translated is one of the factors to be considered in order to decide which MT system provides the best quality. Based on this evidence, for the final stages of the TraMOOC project the decision was made to favour the NMT approach over SMT for the language pairs under consideration, as there are indications that this approach holds the greatest potential for quality going forward. However, applying MT to new language pairs and other MOOC domains may present different challenges, which is why we are hesitant to make broader conclusive generalizations.

5 CONCLUSIONS AND FUTURE WORK

This paper has discussed the outcomes of a large-scale, multi-method evaluation, comparing the quality of SMT and NMT output for MOOC data in a diverse set of language combinations of interest to the TraMOOC project, i.e. from English into German, Greek, Portuguese and Russian. The evaluation involved four state-of-the-art automatic evaluation metrics (i.e. HTER, BLEU, METEOR, and chrF), as well as a range of more labour-intensive manual methods (fluency and adequacy, post-editing, error annotation and ranking). In conclusion, consistently with other application domains, the large-scale multi-method evaluations based on our MOOC data suggest that the emerging neural approach to MT offers some noticeable advantages over the competing and well-established SMT paradigm.

Our findings also show that, while NMT represents an improvement over SMT in some areas, further work is still required to consolidate the current promising performance before NMT can be recognized as the new state-of-the-art in MT [22]. As far as our own work in TraMOOC is concerned, subsequent planned evaluations include the identification of source-language phenomena that are likely to cause particularly serious errors in the output, depending on the target languages and the MT system type. Another avenue for further research consists in task-based evaluations, which would provide a useful addition to the range of evaluation methods that have already been applied in preparation for this paper: task-based evaluation involves exposing real users to MOOC content machine-translated into one of TraMOOC's target languages, and then assessing their understanding and knowledge of that translated material. Their performance can be judged against the baseline of students using the same original English-language MOOC in preparation for identical tests, to give an indication of how effective and successful the application of the MT system concerned is for the specific MOOC domain.

The application of MT to the various types of texts that incorporate a MOOC is undoubtedly a complex task. At the end of the TraMOOC project we hope to provide a roadmap for using automatic translation and user post-editing of MOOC materials, as well as a platform via which this work may be carried out using state-of-the-art MT technology, as part of the ultimate aim of making MOOC resources more accessible to non-English-speaking users. The results of our evaluations of NMT quality using MOOC texts have so far been promising. This work is continuing at a larger scale, with results feeding back to the MT development team, in the hope of facilitating multilingual MOOC resources that are comprehensible and beneficial to global end users.

ACKNOWLEDGEMENTS

The TraMOOC project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement № 644333. The ADAPT Centre for Digital Content Technology at Dublin City University is funded under the Science Foundation Ireland Research Centres Programme (Grant 13/RC/2106) and is co-funded under the European Regional Development Fund.

REFERENCES

[1] M. Nanfito, MOOCs: *Opportunities, Impacts, and Challenges. Massive Open Online Courses in Colleges and Universities*. CreateSpace Independent Publishing Platform, 2014.

[2] S.D. Krause, and C. Lowe (eds), *Invasion of the MOOCs: The Promises and Perils of Massive Open Online Courses*. Parlor Press, 2014.

[3] D.G. Glance, M. Forsey, and M. Riley, “The pedagogical foundations of massive open online courses,” in *First Monday*, vol. 18, no. 5, 2013. Retrieved from http://firstmonday.org/article/view/4350/3673.

[4] R. Kop, H. Fournier, and J. Mak, “A pedagogy of abundance or a pedagogy to support human beings? Participant support on massive open online courses,” in *International Review of Research in Open and Distance Learning*, vol. 12, no. 7, pp.74–93, 2011.

[5] J. Mackness, M. Waite, G. Roberts, and E. Lovegrove, “Learning in a small, task-oriented, connectivist MOOC: Pedagogical issues and implications for higher education,” in *The International Review of Research in Open and Distance Learning*, vol. 14, no. 4, 2013. Retrieved from www.irrodl.org/index.php/irrodl/article/view/1548.

[6] E.A. Monske, and K.L. Blair (eds), *Handbook of Research on Writing and Composing in the Age of MOOCs*. IGI Global, 2017.

[7] B. Wildavsky, “MOOCs in the Developing World: Hope or Hype?,” in *International Higher Education*, no. 80, pp. 23–25, 2015. Retrieved from https://ejournals.bc.edu/ojs/index.php/ihe/article/view/6154/5392.

[8] G. Christensen, A. Steinmetz, B, Alcorn, A. Bennett, D. Woods, and E.J. Emanuel, “The MOOC Phenomenon: Who Takes Massive Open Online Courses and Why?,” in *Social Science Research Network*, 2013. Retrieved from https://ssrn.com/abstract=2350964.

[9] T. Liyanagunawardena, S. Williams, and A. Adams, “The impact and reach of MOOCs: a developing country’s perspective,” in *eLearning Papers*, no. 33, 2013. Retrieved from www.openeducationeuropa.eu/sites/default/files/asset/In-depth_33_1.pdf.

[10] D. Laurillard, “The educational problem that MOOCs could solve: professional development for teachers of disadvantaged students,” in *Research in Learning Technology*, vol. 24, no. 1, 2016. Retrieved from www.tandfonline.com/doi/full/10.3402/rlt.v24.29369.

[11] S. Mak, R. Williams, and J. Mackness, “Blogs and forums as communication and learning tools in a MOOC,” in *Proceedings of the 7th International Conference on Networked Learning 2010* (L. Dirckinck-Holmfeld, V. Hodgson, C. Jones, M. de Laat, D. McConnell, and T. Ryberg, eds.), pp. 275–284, 2010. Retrieved from www.lancaster.ac.uk/fss/organisations/netlc/past/nlc2010/abstracts/PDFs/Mak.pdf.

[12] C. Yeager, B. Hurley-Dasgupta, and C.A. Bliss, “cMOOCs and Global Learning: An Authentic Alternative,” in *Journal of Asynchronous Learning Networks*, vol. 17, no. 2, pp. 133–147, 2013. Retrieved from http://files.eric.ed.gov/fulltext/EJ1018269.pdf.

[13] T. Beaven, A. Cormas-Quinn, M. Hauck, B. de los Arcos, and T. Lewis, “The Open Translation MOOC: creating online communities to transcend linguistic barriers,” in *Journal of Interactive Media in Education*, 2013. Retrieved from http://jime.open.ac.uk/articles/10.5334/2013-18/.

[14] F. Brouns, N. Serrano Martínez-Santos, J. Civera, M. Kalz, and A. Juan, “Supporting language diversity of European MOOCs with the EMMA platform,” in *Proceedings of the European MOOC Stakeholder Summit 2015 – Research Track*, pp. 157–165, 2015. Retrieved from http://dspace.ou.nl/bitstream/1820/6026/1/emoocs2015-emma-final.pdf.

[15] P. Koehn, H. Hoang, A. Birch, C. Callison-Burch, M. Federico, N. Bertoldi, B. Cowan, W. Shen, C. Moran, R. Zens, C. Dyer, O. Bojar, A. Constantin, and E. Herbst, “Moses: Open Source Toolkit for Statistical Machine Translation,” in *Proceedings of the ACL 2007 Demo and Poster Sessions*, pp. 177–180, 2008.

[16] M.J. Israel, “Effectiveness of Integrating MOOCs in Traditional Classrooms for Undergraduate Students,” in *The International Review of Research in Open and Distributed Learning*, vol. 16, no. 5, 2015. Retrieved from www.irrodl.org/index.php/irrodl/article/view/2222/3402.

[17] D. Gamage, I. Perera, and S. Fernando, “A framework to analyze effectiveness of eLearning in MOOC: Learners’ perspective,” in *Proceedings of the 8th International Conference on Ubi-Media Computing (UMEDIA),* Colombo, Sri Lanka: IEEE, 2015.

[18] C. Milligan, A. Littlejohn, and A. Margaryan, “Patterns of Engagement in Connectivist MOOCs,” in *Journal of Online Learning and Teaching*, vol. 9, no. 2, pp. 149–159, 2013. Retrieved from http://jolt.merlot.org/vol9no2/milligan_0613.pdf.

[19] M. Bali, “MOOC Pedagogy: Gleaning Good Practice from Existing MOOCs,” in *Journal of Online Learning and Teaching*, vol. 10, no. 1, pp. 44–57, 2014. Retrieved from http://jolt.merlot.org/vol10no1/bali_0314.pdf.

[20] A. Way, “Traditional and Emerging Use-Cases for Machine Translation,” in *Proceedings of Translating and the Computer 35*, 2013.

[21] M. Fernández-Barrera, V. Popescu, A. Toral, F. Gaspari, and K. Choukri, “Enhancing Cross-border EU e-commerce through Machine Translation: Needed Language Resources, Challenges and Opportunities,” in *Proceedings of the 10th Language Resources and Evaluation Conference*, pp. 4550–4556, Paris: European Language Resources Association, 2016.

[22] S. Castilho, J. Moorkens, F. Gaspari, I. Calixto, J. Tinsley, and A. Way, “Is Neural Machine Translation the New State of the Art?,” in *The Prague Bulletin of Mathematical Linguistics,* vol. 108, 2017.

[23] R. Sennrich, B. Haddow, and A. Birch, “Edinburgh Neural Machine Translation Systems for WMT 16,” in *Proceedings of the First Conferenc e on Machine Translation*, vol. 2, *Shared Task Papers*, pp. 371–376, 2016.

[24] M. Snover, B. Dorr, R. Schwartz, L. Micciulla, and J. Makhoul, “A study of translation edit rate with targeted human annotation,” in *Proceedings of the Conference of the Association for Machine Translation in the Americas*, pp. 233–231, 2006.

[25] K. Papineni, S. Roukos, T. Ward, and W. Zhu, “BLEU: A Method for Automatic Evaluation of Machine Translation,” in *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pp. 311–318, 2002.

[26] A. Lavie, and A. Agarwal, “METEOR: An Automatic Metric for MT Evaluation with High Levels of Correlation with Human Judgments,” in *Proceedings of the Workshop on Statistical Machine Translation,* pp. 228–231, 2007.

[27] M. Popović, “chrF: character n-gram F-score for automatic MT evaluation,” in *Proceedings of the 10th Workshop on Statistical Machine Translation*, pages 392–395, 2015.