Procesamiento del Lenguaje Natural, Revista nº 51, septiembre de 2013, pp 209-212 recibido 30-04-2013 revisado 16-06-2013 aceptado 21-06-2013

Bologna Translation Service: Improving Access To Educational Courses Via Machine Translation (system demonstration)

Bologna Translation Service: mejorando el acceso a los planes de estudios universitarios mediante la traducción automática (demostración del sistema)

Justyna Pietrzak, Elena García Berasategi, Amaia Jauregi Carrera
Eleka Ingeniaritza Linguistikoa S.L. (Elhuyar Group)
Zelai Haundi kalea 3, Osinalde industrialdea,
20170 Usurbil, Gipuzkoa, Spain
{j.pietrzak, e.garcia, a.jauregi}@elhuyar.com

Resumen: Se presenta una demostración del trabajo realizado en el proyecto Bologna Translation Service (BTS), un proyecto cofinanciado por la Unión Europea en el marco de "Information and Communications Technology Policy Support Programme" (ICT PSP) que ofrece la traducción automática de planes de estudios desde 7 idiomas diferentes (alemán, español, finés, francés, holandés, portugués y turco) al inglés, y desde inglés al chino mandarín. BTS ofrece traducción automática accesible on-line. La estructura completa del sistema incluye motores de traducción automática basada en reglas, basada en estadística, ambas combinadas y también un sistema de post-edición automática y manual.
Palabras clave: traducción automática, traducción automática estadística (SMT), traducción automática basada en reglas (RBMT), post-edición automática, post-edición manual

Abstract: This is a demonstration of the Bologna Translation Service (BTS), an EU-funded project (in the framework of Information and Communications Technology Policy Support Programme - ICT PSP) which specialises in the automatic translation of study programmes from 7 languages (German, Spanish, Finnish, French, Dutch, Portuguese, and Turkish) to English, and from English to Mandarin Chinese. At the core of the BTS framework there are several machine translation (MT) engines through which web-based translation services are offered. The fully integrated BTS architecture groups rule-based and statistical MT, their combination, and automatic and manual post-editing modules.
Keywords: machine translation, statistical machine translation (SMT), rule-based machine translation (RBMT), automatic post-editing, manual post-editing

1 Introduction

There is a continuing increasing need for universities and other higher educational institutions to provide course syllabi and other educational information in English. Access to this content plays a crucial role in the degree to which these institutions effectively attract students and, more importantly, has an impact on international profiling of universities, helping them to operate in a globalized environment.

The regulatory environments in the context of the Bologna Accords combined with budget constraints and limited human resources make it very difficult for higher educational institutions to deliver English documentation, which affects their capacity to promote their services internationally. Confronted with the European Credit Transfer System (ECTS) requirements, many of them now spend vast amounts of money and time providing traditional human translated documents.

2 Objectives

As European higher education and research are two pillars of the knowledge-based society,

ISSN 1135-5948

BTS[1] aims to provide a solution to this problem by offering a low-cost, web-based, high-quality machine translation (MT) service geared towards this specific use case. The project makes use of existing rule-based (RBMT) and statistical (SMT) technologies and tailors them in order to produce the best possible quality for syllabus translations.

The first phase of the project includes the automatic translation of syllabi from 7 languages (German, Spanish, Finnish, French, Dutch, Portuguese, and Turkish) to English and from English to Mandarin Chinese.

The BTS approach is to integrate existing MT components into a web-based collaboration framework, in which users with different roles (requester, reviewer, manager, post-editor...) participate on-line at different stages of the translation workflow. The basis is a SMT engine for all language pairs, which were further refined by adding in data from the educational domain and applying domain adaptations and automated and human post-editing. For a selected number of language pairs, systems were combined (SMT and RBMT combination) in order to further improve translation quality.

3 System description

BTS integrates different MT components in order to improve the quality of translation.

3.1 Statistical Machine Translation

For training and tuning the SMT systems, the freely available Moses (Koehn *et al.*, 2007) software tools and relevant wrapper scripts included in OpenMaTrEx (Dandapat *et al.*, 2010) were used. The phrase-based translation models were trained using an out of domain corpus of nearly 6 million segments and a 2 million segment in-domain one. GIZA++ (Och and Ney, 2003) was used for word-aligment with the default number of iterations for the implementations of IBM Models.

To build the language models (LM), we used the state-of-the-art open-source IRSTLM toolkit (Federico and Cettolo, 2007). The LMs were five- to seven-gram models, applying Kneser-Ney discounting and using word forms. Separate LMs were built for in-domain and out-of-domain corpora.

3.2 Rule-Based Machine Translation

Rule-based systems were developed for 4 language pairs. Only the Turkish to English system was developed entirely for the BTS project. The remainder were modifications and further development of machines that were already in use by Convertus and Eleka.

The state-of-the-art open-source platform for developing RBMT systems *Apertium* (Forcada *et al.*, 2011) was used for Spanish and Portuguese, and in-house RB system made by one of the project's partners, the Convertus company, was applied for Turkish to English and Finnish to English. RBMT models included the adaptation of linguistic data to the environment of the Bologna Translation Service: enrichment of dictionaries (by adding the terminology used in the academic and education fields), modification of bilingual dictionaries (in order to promote domain-specific translations), and adaptation of some transfer rules to syllabi writing style.

3.3 Automated post-editing

In order to keep on improving the quality of translation, the next developed modules were fine-tuned MT systems built upon the advanced SMT and RBMT systems.

SMT fine-tuning was achieved by means of rule-based automated post-editing (rbAPE); rbAPE modules included a pre-processing string-based layer focusing mostly on lexical and syntactic issues. The rbAPE rules were developed for every language pair, with an average of 200 transfer rules, and 1000 generation rules.

For those language pairs for which RBMT systems are used, statistical APE (sAPE) was developed, using the Moses tool. The translation in these cases was carried out from rule-based machine translation to the reference target translation, so to speak, from "not perfect SMT-produced English" into "better post-edited English". In every case, statistically significant improvements were achieved; nevertheless RBMT+sAPE systems

[1] "Bologna Translation Service" project has received funding from the European Community (ICT-PSP 4th Call) under Grant Agreement nº 270915. The official website of the project: http://www.bologna-translation.eu

obtained in general worse automatic evaluation than their respective SMT+rbAPE systems.

3.4 System combination

Automatic evaluation lead us to the conclusion that the output from the SMT systems is considerably better than that from the RBMT systems, but we expected further advances to be shown when such system types are combined. Using the CMU system combination toolkit (Hildebrand and Vogel, 2008), we conducted a series of experiments combining the RBMT and SMT engines for Spanish–, Portuguese–, Finnish– and Turkish–English language pairs. Experiments were conducted with different development corpora, LM and phrase-table combinations, searching for the best performing systems. On average, improvements of 2 BLEU points were achieved with statistically significant improvements.

3.5 Front-end, manual post-editing

All system elements were combined into a workflow, offering web-access to all translation models.

The platform was optimized following previous usability surveys conducted among potential users. In addition to previously mentioned components, the translation memory module was added (personalized and confidential for every user institution), as well as the manual post-editing module, to produce one centralized, fully-fledged MT and post-editing environment.

Service is available to be tested at demo.bologna-translation.eu.

4 Evaluation

For translation quality evaluation, we compiled results for the automatic MT evaluation metrics BLEU (Papineni *et al.*, 2002), METEOR (Banerjee and Lavie, 2005) and TER (Snover, 2006), using 2000 sentence test sets. Statistical significance testing was carried out using the tool 'Multeval' (Clark *et al.*, 2011), which implements approximate randomisation normalisation, which has been shown to be more reliable than bootstrap resampling (Koehn, 2004) to test the statistical significance of MT quality measurements.

Apart from automatic evaluation of the translation quality, three different types of human evaluation were carried out, each of which focused on a different aspect of translation. With the *quality evaluation*, the focus was on the linguistic quality of the translations. With these evaluations we try to answer the question 'how good is the translation?'. With the *productivity evaluation*, we investigate to what extent automated translation can speed up human translation. The question we are trying to answer here is 'is post-editing MT output quicker than translating from scratch?'. With the *usability evaluation*, the focus was on the utility of raw MT to end-users, trying to answer the question 'how useful is the MT output for gisting?'.

For the quality and productivity evaluations, the same test set was used. This test set was compiled by taking a random sample of 300 sentences from the test sets that were used for the automatic evaluations. The size of the set was reduced to keep the human evaluation within acceptable limits in terms of cost and time needed to complete them.

5 Results and further work

Good evaluation scores were obtained for every language pair, as reported in previously published paper (Van de Walle *et al.*, 2013), with an average of 20 BLEU point of improvements for MT systems, and 30% of productivity improvements, compared to human translation.

In order to develop the fully functional system, various universities were invited to take part in qualitative and quantitative surveys before the development of the platform, and their opinions were taken into account during and after prototype development.

MT quality is also constantly improving, due to the incorporation of translation memories and periodic retraining of translation models.

Another round of surveys and evaluations will be conducted in 2014 after the first year of working of BTS.

References

Banerjee, S., A. Lavie. 2005. METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human

Judgments. In: *Proceedings of the ACL 2005 Workshop on Intrinsic and Extrinsic Evaluation Measures for MT and/or Summarization*. University of Michigan, Ann Arbor, MI; pp. 65–72.

Clark, J., C. Dyer, A. Lavie, N. Smith. 2011. Better Hypothesis Testing for Statistical Machine Translation: Controlling for Optimizer Instability. In: *49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies. Short papers*. Volume 2, Portland, OR; pp 176–181.

Dandapat, S., M. L. Forcada, D. Groves, S. Penkale, J. Tinsley, A. Way. 2010. OpenMaTrEx: A Free/Open-Source Marker-Driven Example-Based Machine Translation System. In: H. Loftsson (Red.), *Advances in Natural Language Processing: 7th International Conference on NLP. IceTAL 2010* (Reykjavík, 16-18 Aug. 2010) (Vol. Col. Lecture Notes in Artificial Intelligence, vol. 6233, pp. 121–126). Berlin/Heidelberg: Springer.

Federico, M., M. Cettolo. 2007. Efficient Handling of N-gram Language Models for Statistical Machine Translation. In: *Proc. of ACL Workshop on SMT*. Pp. 88–95, Prague, Czech Republic.

Forcada, M. L., M. Ginestí-Rosell, J. Nordfalk, J. O'Reagan, S. Ortiz-Rojas, J. A. Pérez-Ortiz, F. Sánchez-Martínez, G. Ramírez-Sánchez, F. M. Tyers. 2011. Apertium: a free/open-source platform for rule-based machine translation. In: *Machine Translation*. 25:2(2011), pp 127–144.

Hildebrand A.S., S. Vogel. 2008. Combination of machine translation systems via hypothesis selection from combined n-best lists. In: *MT at work: Proceedings of the Eighth Conference of the Association for Machine Translation in the Americas.* Pp. 254–261, Waikiki, Hawaii, October, 2008.

Koehn, P. 2004. Statistical significance tests for machine translation evaluation. In: *EMNLP-2004: Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, 25-26 July 2004, Barcelona, Spain; 8pp.

Koehn, P., H. Hoang, A. Birch, C. Callison-Burch, M. Federico, N. Bertoldi, B. Cowan, W. Shen, Christine Moran, R. Zens, C. Dyer, O. Bojar, A. Constantin, and E. Herbst. 2007. Moses: open source toolkit for statistical machine translation. In: *ACL 2007: proceedings of demo and poster sessions*, Prague, Czech Republic; pp. 177–180.

Och, F. J., H. Ney. 2003. A systematic comparison of various statistical alignment models. *Computational Linguistics*, 29(1): 19–51.

Och, F. J. 2003. Minimum error rate training in statistical machine translation. *ACL-2003: 41st Annual meeting of the Association for Computational Linguistics*, July 7-12, 2003, Sapporo, Japan; pp. 160–167.

Oflazer, K. 2008. Statistical Machine Translation into a Morphologically Complex Language. In: *Proceedings of CICLING 2008: Conference on Intelligent Text Processing and Computational Linguistics*, February 2008, Haifa, Israel; pp. 376–387.

Papineni, K., S. Roukos, T.d Ward & W-J. Zhu. 2002. BLEU: A Method for Automatic Evaluation of Machine Translation. In: *ACL-2002: 40th Annual meeting of the Association for Computational Linguistics*, Philadelphia, PA; pp. 311–318.

Snover, M., B. Dorr, R. Schwartz, L. Micciulla, L. and J. Makhoul. 2006. A study of translation edit rate with targeted human annotation. In: *AMTA 2006: Proceedings of the 7th Conference of the Association for Machine Translation in the Americas, "Visions for the Future of Machine Translation"*, August 8-12, 2006, Cambridge, MA, USA; pp. 223–231.

Van de Walle, J., H. Depraetate, J. Pietrzak. 2013. Improving access to educational courses via machine translation: Bologna Translation Service evaluation. In: *Proceedings of the 5th international Conference on Education and New Learning Technologies*, Barcelona 1 – 3 July.