

What place for technology in the translation of Welsh? Language Technology, Language

Planning and the Professional Translation of Welsh

Abstract

The Welsh translation industry in Wales was estimated in 2007 to be worth £45,000,000 per annum (Prys *et al.* 2009). The industry has also seen massive growth since its early beginnings in the late 1960s (Jones 2005) and is now an attractive and lucrative career option for graduates of Welsh. Scholars have also drawn attention to the importance translation into and out of minority languages plays in the process of linguistic normalization and language planning (Garcia 2005; Kaufman 2010, 2012). Given that translation in modern Wales is so important to language planning and revitalization, a fact the Welsh Government itself acknowledges in its current language policy document (Welsh Government 2012), this paper intends to ask what role language technology can play in translation and as a result in the revitalization of Welsh. The role language technology in particular can play in the process of language planning and revitalization has also been acknowledged by language planners in Wales (Evas 2012; Welsh Government 2014). In light of this, a study is reported on which tested the benefits of using translation technology for the professional translator of Welsh. This was done within a hypothetico-deductive framework, comparing and contrasting the outcomes of a number of variables that are inextricably linked to the professional translation process. A between-groups design is utilized in which cognitive, technical and temporal effort are gauged between translation and the use of Welsh output from a Machine Translation system and Translation Memory software. All participants were members of The Association of Welsh Translators and Interpreters, and all were employed as practicing translators. This empirical translation process study using specialist recording software has not been done before for Welsh, and is also the first process study of its kind for the language. Results show that the use of language technology did speed up the translation process, that translation became cognitively easier and that a number of variables related to text production were also reduced. This has implications for language planners in Wales as currently only a minority of Welsh translators use such technology despite these benefits.

Introduction: Language Planning and Translation in Wales

The translation industry in Wales has grown in parallel with developments in language policy and planning (Jones 2005a, b; Kaufman 2010, 2012; Andrews 2015), and has been said to be part of the very ‘ethos’ of the Welsh brand of bilingualism that we see and experience in modern Wales today (Miguélez-Carballeira *et al.* 2016, p. 125). In the 1960s, when the Welsh Language Society began to cause civil unrest, Welsh nationalist politicians started to make their first breakthroughs and the British government finally began to give a modicum of status

to the language, there were likely to be no more than a handful of Welsh translators in Wales. The Welsh Office, following the recommendations of the Hughes-Parry Committee, established the first ever professional Welsh translation unit which consisted originally of two full time members of staff at its office in Cardiff. There were no companies, few freelancers and precious few opportunities for training. Today, there over 500 translators and at least 5 major companies employing between ten to twenty translators each, as well as a large number of freelance professionals. Universities offer MAs in aspects of Welsh translation and many of those who graduate in Welsh go on to become trainee and junior translators. This link then between language planning and policy should not be underestimated, especially in Wales. Speaking in general about the relationship between translation and linguistic normalization, González (2005, p. 111) states, ‘In the case of minority languages [...] translation is an activity that has to be fostered and activated, as a mechanism to promote the language itself. In other words, translation is no longer an underlying element of communication, but an essential tool in the process of language recovery or preservation’. In its recent policy document on Welsh, the Welsh Government has stated in unambiguous terms the role it believes translators of Welsh have. In that policy document (Iaith, Fyw, Iaith Byw, or ‘A Living Language, A Language for Living), the government notes ‘The demand for professional translators and interpreters that work in Welsh and English must be met if we are to satisfy the need for bilingual documents and simultaneous translation at events and meetings. The industry has developed significantly in recent years and we recognise the need to continue this work, and to ensure that robust accreditation and regulation structures are in place to provide assurances with regard to quality’. Further, the use of technology within the field of professional translation in Wales has also received attention. Again in ‘A Living Language, A Language for Living’, the Welsh Government (2012, t. 49) stated ‘We must also ensure that the translation profession makes the most of the ICT tools that are available to it in order to ensure efficiency,

consistency and value for money'. In 2014 the 'reuse of translations, translation engines and automated translation for post-editing and quality control by humans, so that there can be greater prominence for Welsh' was noted as a priority in an update document of that original 2012 policy (Welsh Government 2014, p. 11). Finally, the language commissioner in Wales responsible for the regulation of the Welsh Language Measure (Wales) (2011) has also advised on the use of translation technology in a professional context. In one of its advice documents regarding translation and interpreting, the Commissioner notes 'Language technology can help to accelerate and facilitate the work of the translator. It can also assist organizations who commission work to ensure the quality of translation work commissioned and to ensure value for money'. Given the importance of translation and the position taken by the Government as well as the advice given by the language commissioner, two experiments were carried out to test the actual benefits of using translation technology to the professional translator of Welsh. The two technologies investigated are briefly described below. The hypotheses tested and the experimental designs are then outlined. Finally, results are discussed.

Translation Memory and Statistical Machine Translation

Translation memory is a translation technology that is able to reproduce sentences already translated to the translator, which are called exact matches. It can also offer the translator segments that are partially similar, called fuzzy matches, and as a result only require minimum human editing to create an acceptable translation. Statistical Machine Translation learns how to translate using statistical probability. Using a bilingual corpus of two languages, the system learns the most likely translation pairs of phrases within a sentence (called the Translation Model) and then using a monolingual corpus of the target language the system puts those most likely phrase constituents in the right order (called the Language Model). Google Translate roughly works like this, but by using the World Wide Web for its huge corpora. When using these technologies, the translator is required to correct their output so as to ensure it is an

acceptable translation. The literature regarding the interface between the professional translation process, efficiency and the correction of the output from these translation technologies is large, and show that translating with these technologies can reduce cognitive effort (O'Brien 2007b; Carl *et al.* 2015), reduce keyboarding (O'Brien 2007b; Carl *et al.* 2015) and as a result increase productivity as compared to translation without them according to Lange & Bennett (2000), O'Brien (2007), Groves & Schmidtke (2009), Guerberof (2009, 2012, 2014), Flourney & Duran (2009), Kanavos & Kartsaklis (2010), Plitt & Masselot (2010), De Sousa *et al.* (2011), Lee & Liao (2011), Federico *et al.* (2012), Green *et al.* (2013), Aranberri *et al.* (2014), Depraetere *et al.* (2014), Moran *et al.* (2014), Silva (2014), Yamada (2014) and Zhechev (2014). This then was tested for Welsh in two experiments, the first investigating whether the use of Translation Memory and Machine Translation could render the translation process cognitively easier and reduce keyboarding effort, and the second to test productivity gains in more depth.

Experiment 1: Effort

In the first experiment, two hypotheses were tested, based in the evidence provided in the relevant literature. These were:

H1: That the correction of the Machine Translation output and Translation Memory output in the form of fuzzy matches would be cognitively easier than manual translation, as measured by the amount of pauses after the source text had been read and after text production came to an end

H2: That the correction of these outputs would reduce the amount of keyboarding needed to produce a translation, as measured by the number of keystrokes recorded.

In order to test this, Translog-II translation recording software was used in a between-groups design. This software can record a screen video of the translation session, measure the amount of time spent pausing and so amount of cognitive effort expended in translation, as well as the analyse the amount of keystrokes struck and which type. Four professional translators of Welsh

translated a document regarding adoption containing 15 sentences while another four corrected the machine-translated version of the same text. In the second part, these same translators were split into two groups again, where four translators translated a document containing 11 sentences about Web2,0 and the other four corrected fuzzy matches from a Translation Memory system based on this same text. Exact matches were not analysed in this first experiment; given that these are already acceptable translations of the source text, it would have been futile to measure the cognitive effort and keystrokes required to process them.

Experiment 2: Productivity

In the second experiment, the main hypothesis was as follows:

H3: The use of Machine Translation and also Translation Memory output in the form of fuzzy and exact matches would speed up translation, as opposed to translating without them.

Seven professional translators of Welsh processed 25 segments of Machine Translated output, 25 fuzzy matches, 25 exact matches and 25 sentences for which no output was offered. The speed with which these segments were then processed were measured in milliseconds within a tool developed by the Translation Automation Users Society. In all, over 15,000 words were processed. All texts were taken from the public sector domain.

Results of Experiment 1

Despite no statistical difference being found between the average pause durations of translators in the Control Group (CG) who translated and the Experimental Group (EG) who corrected the machine translation output according a Mann-Whitney test ($p = 0.98$), the number of keystrokes struck by the EG was indeed much less than that of the CG. On average, 504 keys were struck by the CG but only 173 by the EG. In terms of the use of translation memory and the first hypothesis in relation to cognitive effort as measured by pauses, the process of using this output did render the process of producing a translation cognitively easier. There was a statistically significant difference according a Mann-Whitney test between the average pause durations of

translators in the CG who translated and the EG who corrected the translation memory output ($p = .000$). The same was found for the second hypothesis and the average number of keystrokes struck by both groups; on average, 77 keys were struck by the CG, but as little as 48 by the EG.

What these results collectively show then is that the use of language technology changes the translation process, and can render this process of actually producing the translation practically easier. Having responded to the first two hypotheses, the results of the second experiment will now be discussed.

Results of Experiment 2

The segments for which translation memory output in the form of fuzzy and exact matches were offered, and the segments for which Machine Translation output was offered, were processed much more quickly than the segments that had to be translated from scratch. Taking the average time each translator took in milliseconds to process each batch of 25 segments (25 fuzzy matches, 25 exact matches, 25 Machine translated segments and 25 segments to be translated from scratch), a mixed methods ANOVA (Analysis of Variance) was carried out. This showed that the type of match processed had a significant effect on the time taken to translate where $F = 34.64$, $p = .000$. The average time taken to complete each segment for which no output was given was 62 seconds, but only 38 seconds for the fuzzy matches, 14 seconds for the exact matches and 18 seconds for the segments which required the Machine Translation output to be corrected. Translation productivity was therefore greatly improved.

Conclusion

By using software capable of tapping into the translation process, a number of translation process variables were recorded with a view to analysing what effect the use of language technology, namely Translation Memory and Machine Translation, has on translation efficiency. These experiments reported on here using this type of software are also new to

research in Welsh. The data analysed show that the translation process can in fact become cognitively less effortful for the translator, that the practical and manual process of translation production in terms of keystrokes is also more efficient and that translator productivity can be greatly improved. These results must be contextualised by looking again at the importance of translation to language planning and policy in Wales. If the use of language technology can speed up translation and help translators produce more, whilst also making the translation process ergonomically easier on the practitioners themselves, then given the role translation plays the use of this technology could also play a major role in process of revitalizing Welsh in Wales.

References

- Aranberri, N *et al.* (2014). Comparison of post-editing productivity between professional translators and lay users. In: O'Brien, S, Simard, M & Specia, L. eds. *Proceedings of the Third Workshop on Post-Editing Technology and Practice*. Vancouver, Canada, October 2014: AMTA, pp. 20-33.
- Carl, M. *et al.* 2015. Post-editing Machine Translation: Efficiency, Strategies and Revision Processes in Professional Translation Settings. In: Ferreira, A. & Schwieter, J.W. eds. *Psycholinguistic and Cognitive Inquiries into Translation and Interpreting*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Depraetere, I. & Vackier, T. 2011. Comparing formal translation evaluation and meaning-oriented translation evaluation: or how QA tools can(not) help. In: Depraetere, I. ed. *Perspectives on Translation Quality*. Berlin/Boston: Mouton de Gruyter, pp. 25-51.
- De Sousa, S. *et al.* 2011. Assessing the Post-Editing Effort for Automatic and Semi-Automatic Translations of DVD Subtitles. In: Angelova, G. *et al.* eds. *Proceedings of of the International Conference in Recent Advances in Natural Language Processing*. Hissar, Bulgaria, 12-14 September, 2011, pp. 97-104.
- Evas, J. 2013. *The Welsh Language in the Digital Age – Y Gymraeg yn yr Oes Ddigidol*. META-NET White Paper Series, cyf. 13. Berlin: META-NET.
- Federico, M. *et al.* 2012. Measuring user productivity in machine translation enhanced computer assisted translation [Online]. In: *Proceedings of the Second International Conference of the Association for Machine Translation in the Americas (AMTA)*. Available: <http://amta2012.amtaweb.org/AMTA2012Files/papers/123.pdf> [Seen: 07/03/2016].
- Flournoy, R. & Duran, C. 2009. Machine Translation and Document Localization at Adobe: From Pilot to Production [Online]. In: Pierre, I. *et al.* eds. *Proceedings of the 12th MT Summit*. Ottawa, August 26-30, 2009. Available: http://www.promt.ru/press/pdf/ADOBE_Flournoy2009.pdf [Seen: 03/06/2015].

Gonzalez, G.M. 2005. Translation of Minority Languages in Bilingual and Multilingual Communities. In: Branchadell, A & West, L.M. *Less Translated Languages*. John Benjamins Publishing Company: Amsterdam/Philadelphia, pp. 105-125.

Goves, D a Schmidtke, D. 2009. Identification and analysis of post-editing patterns for MT. Yn: *Proceedings of the 12th Machine Translation Summit*, Awst 26-30, Ottawa, tt. 429-436.

Green, S. et al. 2013. *The Efficacy of Human Post-Editing for Language Translation* [Online]. Paris: CHI Proceedings 2013. Available: <http://jan.stanford.edu/pubs/green+heer+manning.chi13.pdf> [Seen: 13/05/2015].

Guerberof, A.A.2009.Productivity and quality in the post-editing of outputs from translation memories and machine translation.*International Journal of Localization* 7(1), tt. 11-21.

Guerberof, A.A. 2012. *Productivity and Quality in the Post-editing of Outputs from Translation Memories and Machine Translation*. Unpublished PhD Thesis, Universitat Rovira I Virgili.

Guerberof, A. A. 2013. What Do Professional Translators Think about Post-editing? *Journal of Specialised Translation* (19), pp. 75-95.

Jones, B. P. 2005a. *Translation in Wales* [Online]. Available: http://www.cyfieithwycymru.org.uk/content/uploads/documents/anerchiad_%20ITI.pdf [Seen: 20/01/2015].

Kanavos, P & Kartsaklis, D. Integrating Machine Translation with Translation Memory: A Practical Approach. In: Zhechev, V. ed. *Proceedings of the Second Joint EM+/CNGL Workshop 'Bringing MT to the User: Research on Integrating MT in the Translation Industry*. American Machine Translation Association, Denver, Colorado 4 November, 2010. pp. 11-20.

Kaufmann, J. 2010. Cyfieithu a Pholisi Iaith. *Contemporary Wales* 23(1), pp. 171-183.

Kaufmann, J. 2012. The Darkened glass of bilingualism? Translation and interpreting in Welsh language planning. *Translation Studies* 5(3), pp. 327-344.

Lee, J. & Liao, P. 2011. A Comparative Study of Human Translation and Machine Translation with Post-editing. *Compilation and Translation Review* 4(2), pp. 105-149.

Lange, C.A & Bennett, W.S. 2000. Combining Machine Translation with Translation Memory at Baan. In: Sprung, R.C. ed. *Translating into Success: Cutting-edge strategies for going multilingual in a global age*. Amsterdam/Philadelphia: John Benjamin's Publishing Company, pp. 203-219.

Miguélez-Carballeira et al. 2016. Introduction: Translation in Wales: History, theories and approaches. *Translation Studies* 9(2), pp. 125-136.

Moran, J. et al. 2014. Analysis of Post-editing Data: A Productivity Field Test using an Instrumented CAT Tool. Yn: O'Brien, S. et al. eds. *Post-editing of Machine Translation: Processes and Applications*. Newcastle: Cambridge Scholars Publishing, pp. 126-147.

O'Brien, S. 2007a. An Empirical Investigation of Temporal and Technical Post-editing Effort. *Translation and Interpreting Studies* 2(1), pp. 83-136.

O'Brien, S. 2007b. Eye-tracking and Translation Memory Matches. *Perspectives: Studies in Translatology* 14(3), pp. 185-205.

Plitt, M. & Masselot, F. 2010. A Productivity Test of Statistical Machine Translation Post-Editing in a typical localisation context. *The Prague Bulletin of Mathematical Linguistics* 93, pp. 7-16.

Prys, D. *et al.* 2009. *Gwell Offer Technoleg Cyfieithu ar gyfer y Diwydiant Cyfieithu yng Nghymru: Arolwg Dadansoddol* [Online]. Bangor: Unedau Technoleg Iaith, Canolfan Bedwyr. Available:

<http://www.catcymru.org/wordpress/wpcontent/uploads/PDFGwell%20offer%20technoleg%20cyfieithu%20ar%20gyfer%20y%20Diwydiant%20Cyfieithu%20yng%20Nghymru.pdf> [Seen: 15/05/2015].

Silva, R. 2014. Integrating Post-editing MT in a Professional Translation Workflow. In: O'Brien *et al.* eds. *Post-editing Machine Translation: Processes and Applications*. Newcastle: Cambridge Scholars Publishing, pp. 24-51.

Welsh Government. 2012. *A Living Language – A Language for Living, Welsh Language Strategy 2012-2017*. Cardiff: Welsh Government.

Welsh Government. 2014. *A Living Language – A Language for Living, Moving Forward*. Cardiff: Welsh Government.

Yamada, M. 2011. The effect of translation memory databases on productivity. In: Pym, A. ed. *Translation Research Projects 3*. Tarragona: Intercultural Studies Group, pp. 63-73.

Zhechev, V. 2014. Analysing the Post-editing of Machine Translation at Autodesk. In: O'Brien, S. *et al.* eds. *Post-editing of Machine Translation: Processes and Applications*. Newcastle: Cambridge Scholars Publishing, pp. 2-24.