Towards a CAT tool agnostic standard for User Activity Data

John Moran, Dave Lewis
Trinity College Dublin,
Dublin, Ireland
Moranj3@cs.tcd.ie, Dave.lewis@cs.tcd.ie

Abstract

The dominance of the source word pricing model combined with the fact that most translators work as freelancers has led to a scenario in which until recently most buyers (direct and intermediary) who work with freelancers neither knew nor cared how many words per hour the translators they hire translate. However, this situation is beginning to change. Machine translation has shown that it is possible for translation requesters to impact positively on words per hour productivity. In addition to classical full-sentence MT, advances in various typeahead technologies have resulted in a situation in which a number of options are available to impact positively on a translator's working speed and terminological consistency with previous translations. Finally, evidence is beginning to emerge that productivity gains can be achieved where translators use Automatic Speech Recognition to dictate rather that type the target text. In this paper we will provide a brief overview of these technologies and use cases the impact on translator productivity and describe an architecture to gather translation process data to measure their impact from working translators in a maximally unobtrusive way. We propose an open-standard for User Activity Data in CAT tools (CAT-UAD) so that they can work in any CAT tool that implements this standard and outline a technical architecture to gather such data conveniently and a privacy model that respects translator, intermediary and end-client data sharing concerns and discuss various A/B testing scenarios that can be tested using Segment Level A/B testing.

Keywords: *CAT-UAD, iOmegaT, translator productivity, machine translation post-editing, SLAB testing*

1. Introduction

A number of studies have shown that MT can be a productivity aid for human translators, e.g. Plitt & Masselot (2010), Roukos et al. (2012) and (Moran, Saam, et al. 2014). Often the term *post-editing* is used to describe this use case, but the reality is more complex.

In fact, MT can be presented in a number of ways:

> Full-sentence pre-populated MT
>
> This is the typical post-editing scenario in which a target segment is pre-populated using MT. Unless the MT proposal is fit-for-purpose, action is required on the part of the translator to delete or improve it. Usually this can be done quickly using a keystroke combination. A variation on this theme is adaptive MT where post-editing patterns are identified and applied automatically without the need to retrain the underlying language model. A second variation on this theme is the Example Based Machine Translation (EBMT) paradigm where text fragments are glued together with some morphological processing on the edges. Unlike rule-based or statistical MT this is usually carried out in the CAT tool itself. An example of this can be found in DejaVuX' auto-assemble technology (Atril 2015).
>
> MT-as-reference
>
> In this scenario the translator can glance at an MT proposal in a side pane and insert the proposal in the target segment with a keyboard shortcut if useful. However, even if the translator does not consider it useful enough to bootstrap the translation of the segment it may contain terminology that is useful and hence save on research or thinking time. Anecdotally, this workflow works well with Automatic Speech Recognition (ASR). A translator dictates parts of the proposal into the target segment. This may reduce the temptation to compromise on word order and so reduce the impact of MT on style. Unfortunately, very little research has been carried out on this use case.
>
> Type-ahead technologies

As a feature in CAT tools predictive typing has been a common feature in desktop-based CAT tools for some time. For example in Trados Studio (SDL, 2015) the feature is referred to as Auto-suggest and in MemoQ (Kilgray 2015) as the Muse function. Proposals may be statistically generated from bitext using bilingual terminology extraction. It is also possible to cull false positives to create a smaller termbase[1] to reduce the annoyance they cause.

A second approach is to compile a terminology database over a long period of time. This termbase does not necessarily contain terms. It may contain any words or multi-word fragments a translator guesses will arise again. In this case a 30 second investment to save an entry in the termbase may save five minutes of typing or research over a few years. Intuitively it seems likely that this approach saves more time for translators who are specialized than generalists.

A third and more recent approach is to connect to an MT system from the CAT tool for typeahead purposes. Research into type-ahead technologies including Interactive Machine Translation (IMT) dates back to the 1990's (Foster et al. 1997). Proposals appear ahead of the cursor as the translator types and can be accepted using a keystroke. Generally, one problem with IMT is that it is difficult to evaluate in an academic context as traditional automated metrics like BLEU scores (Papineni et al. 2002) do not apply.

Finally, technology is not the only factor governing translator productivity. It is possible to increase translator productivity by requiring that translators ignore stylistic factors in their work and focus on fidelity (light post-editing). Productivity gains can still be achieved in full post-editing where little or no quality degradation is accepted e.g. Plitt & Masselot (2010) and Moran et al. (2014) but they are lower.

Lack of accurate productivity speed ratios can become critical when MT is used as a reason to give a discount (in addition to discounts for translation memory matches). Where a translation requester asks for an unfair discount that overshoots the utility of the MT this may only become obvious after some time. In this case, once a project has been accepted it may be too late for the translator to reverse the discount. However, the translator may decide not to take on future projects that involve MT discounts from that client again (even though they may be fair). Clearly, unfair discounts are not in the interest of any stakeholder. A better approach is that taken at IBM where MT utility is measured over a long period on a large or ongoing project and a discount is negotiated once both parties agree that the utility measure is accurate (Roukos et al. 2012).

2. Automatic Speech Recognition

Though MT and type-ahead technologies can be beneficial from a productivity perspective, it is likely that on average Automatic Speech Recognition (ASR) has a greater impact (outside of light PE contexts). Certainly, financially it is in the interest of the translator. Discounts for post-editing are often requested in a similar manner to discounts for translation memory matches. Where a translator uses dictation software they bring the productivity enhancing technology to the table so discounts are neither requested, nor are they likely to be granted.

Dictation of written translation (or sight translation) is not a new phenomenon. For example, the now infamous Alpac Report (Pierce et al. 1966) described how translators were highly productive when dictating translations to be typed by human transcriptionists. In a 2001 ITI Survey (a UK-based translators union) with 430 respondents approximately 30 used a typist (Aparicio, A., Benis, M., & Cross 2001). More recently ASR software may have begun to replace human typists and to have found new users. In a recent survey (CIOL & ITI 2011 p.4) 10% reported using ASR, of which 94% used Dragon Naturally Speaking (Nuance 2015). Unfortunately, productivity gains reports from ASR are not as well reported as those for MT. In the introduction to an online tutorial Jim Wardell, an experienced professional translator comments how he has been able to double his earnings over his working lifetime using dictation (Wardell, 2014). In a recent survey of ASR use by translators with 47 respondents, the average reported productivity increase was 110.56% (though the median was 35%) (Ciobanu 2014).

However, as there is no means of tracking working speed over long periods of time in any commercial CAT tool the impact of training and practice are unavailable. For example, techniques used to train interpreters may be useful in sight translation. Also the impact of the recognition quality of the ASR system on translator productivity is unknown. This information gap may also help to explain why there is so little take up of dictation software by translators. It may also explain why there is little or no focus on translators by dictation software publishers who according to Reddy et al. (2009) could improve accuracy by 32% using context derived from the translation task. Finally, it is worth noting that the health gains to be had from this mode of text input (e.g. lower risk of Repetitive Strain Injury) means that even if productivity gains were negligible it would still be worth using the technology.

3. Previous work

A number of means of measuring translation speed exist. Web-based testing platforms that like TAUS DQF (TAUS, 2015) and TransCentre (Denkowski & Lavie 2012) do not provide most of the features found in CAT tools (e.g. a concordance function or translation memory matching) so they can only be used to gather small samples. However, unlike most CAT tools they can provide a Segment Level A/B (SLAB) testing scenario where translation speed in segments without MT (A) are compared to segments with MT (B). An overview of other similar systems and approaches is described in Moran, Saam, et al. (2014).

Our approach is most similar to IBM TM2 (Roukos et al. 2012) which gathers translation process data at the segment level from within a well-featured desktop-based CAT tool.

4. CAT-UAD – A standard format to record User Activity Data

In (Moran, Lewis, et al. 2014) we describe how User Activity Data is gathered in iOmegaT and give an example of the data in XML. In future work we plan to publish a formal specification for CAT-UAD but for the purposes of this paper it can be thought of as a format that records how a translator interacted with a CAT tool during the normal course of their work in an XML format that can be replayed and analysed later (which explains the video camera icon in Figure 1). The XML records details of segment editing sessions as events and context. It also records when a translator returns to a segment (thus taking self-review time into account).

Translators generally use CAT tools for many hours per day and though they may use more than one anecdotal evidence suggests they normally they have a preference for the CAT tool they use most. Although it is likely that most translators do not use all the features that sophisticated desktop-based CAT tools provide in their daily work, nonetheless, anecdotally at least, resistance to using new web-based CAT tools expressed on Internet forums and social media indicates familiarity impacts on productivity.

This is mirrored in our experience. For example, asking a freelance translator familiar with Trados to work in an unfamiliar CAT tool called OmegaT (omegat.org 2015) for a few days to carry out an MT productivity tool is possible but it is not viable for longer periods, e.g. weeks, months or indeed years.

Nonetheless, OmegaT is a well-featured CAT tool as evidenced by the fact that it is commonly used. Download statistics from Sourceforge (the code repository from which it is downloaded) indicate that downloads will soon exceed 10,000 per month and over 2000 users are registered on the user support e-mail list. In its ten-year existence downloads have doubled approximately every four years. However, a recent survey of translators by proz.com (a website for translators) indicated that OmegaT was being used by under 10% of respondents. In contrast various versions of Trados make up the majority of translators with WordFast (Wordfast LLC 2015) and MemoQ in second and third place. Thus, to record and report on the utility of machine translation in terms of translation speed or the effectiveness of training translators in the use of dictation software in a CAT tool agnostic manner, a new data standard is required so that CAT tool developers can log the data in a convenient manner. Also, unlike, for example, the current speed report in MemoQ, time series reports can also be reported at a supra project level (i.e. longitudinally).

Figure 1 shows an overview of how this architecture would look.

In terms of the client-side data collection, OmegaT is shown without the “i” prefix (iOmegaT) as we plan to merge our instrumentation code into the main

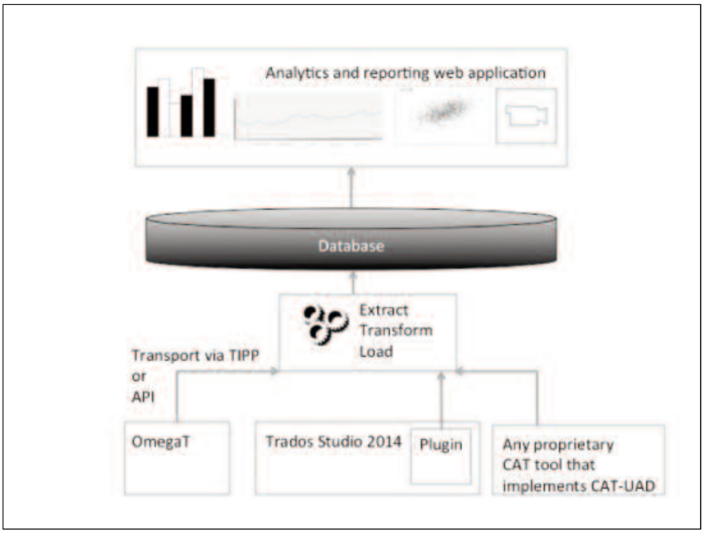

Figure 1: Architectural overview of reporting platform for translation productivity data

OmegaT codebase when the web-based reporting platform has been developed. It is important that a free open-source CAT tool remain central to the platform as this is a maximally flexible option and will make it easier for researchers to carry out reproducible research using SLAB testing in the field, e.g. into various techniques and strategies for interactive MT. Currently, the OmegaT project is the container for the CAT-UAD but this may become a live API. Also, it is possible that it could be added to the TIPP specification (or simply added to the folder structure)[2].

Recent changes to the Trados Studio 2014 Application Programming Interface suggest that a plugin to gather at least some of the data we gather with iOmegaT can be gathered in Trados. However, APIs are not as flexible as open-source applications so it is likely that some A/B testing scenarios that can be implemented in OmegaT will not be possible in Trados Studio.

Finally, conversations with both web-based and desktop based CAT tool publishers suggest there are grounds for cautious optimism that the CAT-UAD standard can be implemented in other proprietary CAT tools once it is formally defined.

In terms of the server-side implementation, the current iOmegaT Translator Productivity Testbench uses console based applications that can be installed locally on a PC. These applications extract, transform and load (ETL) the data gathered from the CAT-UAD files that are stored in the iOmegaT project containers in XML format.

Similarly the web-based reporting platform will be locally installable so all data remains private. In addition a cloud-based option will be available for convenience, albeit with some loss of data privacy.

We have not outlined exact implementation details (e.g. so-called Big Data technologies). However, it is worth noting that recent advances in cloud-computing and data processing provide a number of templates for high volume processing of log data at low cost.

5. Privacy models

Figure 2 shows how the privacy settings could be defined in a CAT tool.

The nature of the translation industry is that translators can be located in almost any jurisdiction so we will use Germany as an example. The recording of User Activity Data in a CAT tool (and in particular translation speed) is a form of workplace monitoring. For translators who are employees pursuant to §87, Subsection 1, No. 1, Works Council Constitution Act (Betriebsverfassungsgesetz - BetrVG) this should be discussed with the relevant works council. For this reason sharing of CAT-UAD should be deactivated by default.

Also, a translator may wish to share translation speed data or other User Activity Data with a third-party (e.g. a company that provides training and support for dictation software). This can be done without infringing a non-disclosure agreement (NDA) with the agency or end client as Words Per Hour and other data identifying the ASR system, MT system or IMT algorithm being used is unrelated to the text being translated. However an option to share linguistic data is required as in some circumstances, e.g. where the reporting application is hosted with the agency it may be useful to include linguistic data and the NDA is not being infringed. Finally, if a translator wishes to remain anonymous or (more likely) an agency wishes to preserve translator anonymity from a client (a larger translation agency or end buyer) requesting a discount for MT post-editing, it should be possible to do so using an anonymous ID in the Username field.

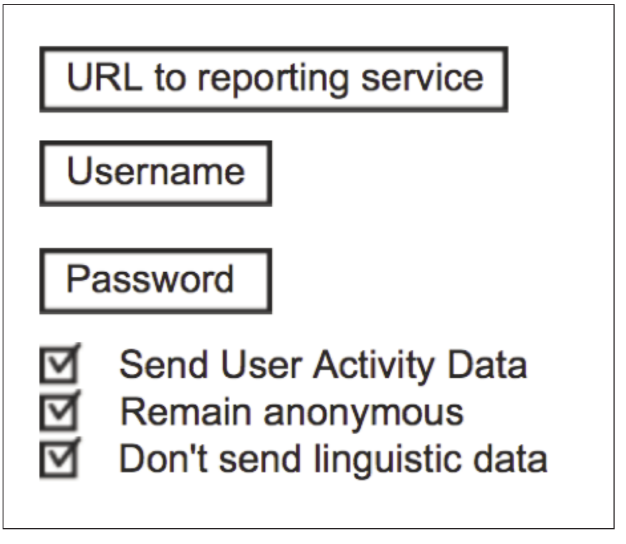

Figure 2: Proposed privacy settings in a CAT tool

6. Future SLAB testing scenarios

In our work to date we have focused on two segment categories, target segments pre-populated with full-sentence MT and empty segments (which we call HT or Human Translation). However, many other SLAB tests are conceivable. For example, dictation with and without MT, two MT systems blind tested against each other (e.g. with two different language models or two different MT providers), two IMT algorithms blind tested against each other, IMT versus HT and even dictating every second segment. Even small improvements should be visible given enough User Activity Data.

7. Summary

In this paper we have presented a number of technologies that can impact on translator productivity. We outlined some means by which translation speed can be measured and showed why a dual strategy of adapting an open-source CAT tool (e.g. to test different IMT scenarios) and instrumenting existing proprietary CAT tools to be maximally unobtrusive to the translators who do not use OmegaT regularly. The latter strategy should make it possible to record translation speed data longitudinally to the benefit of computation linguistics researchers, translators, intermediaries and end buyers.

References

Atril, 2015. Company website [online], available: http://www.atril.com [accessed 1 Mar 2015].

Aparicio, A., Benis, M., & Cross, G., 2001. ITI 2001 Rates & Salaries Survey.

Ciobanu, D., 2014. Of Dragons and Speech Recognition Wizards and Apprentices. *Tradumàtica*, (12). available: http://revistes.uab.cat/tradumatica/article/view/n12-ciobanu/pdf [accessed 1 Mar 2015].

CIOL & ITI, 2011. *2011 Rates and Salaries Survey for Translators and Interpreters*, Denkowski, M. & Lavie, A., 2012. TransCenter: Web-Based Translation Research Suite. In *Workshop on Post-Editing Technology and Practice Demo Session*. San Diego, available: http://www.cs.cmu.edu/~mdenkows/transcenter/ [accessed July 9, 2014].

Elming, J., Winther Balling, L. & Carl, M., 2014. Investigating User Behaviour in Post-editing and Translation using the CASMACAT Workbench. In S. O'Brien et al., eds. *Post-editing of Machine Translation*. Newcastle upon Tyne, pp. 147–169.

Foster, G., Isabelle, P. & Plamondon, P., 1997. Target-text mediated interactive machine translation. *Machine Translation*, 12(1-2), pp.175–194.
Kilgray, 2015. Company website [online], available: http://www.kilgray.com [accessed 1 Mar 2015]

Linport, 2015. website [online], available: http://www. ort.org [accessed 1 Mar 2015]

Moran, J., Lewis, D. & Saam, C., 2014. Analysis of Post-editing Data: A Productivity Field Test using an Instrumented CAT Tool. In S. O'Brien et al., eds. *Post-editing of Machine Translation*. Newcastle upon Tyne, pp. 126–146.

Moran, J., Saam, C. & Lewis, D., 2014. Towards desktop-based CAT tool instrumenta-tion. In Third Workshop on Post-Editing Technology and Practice. Vancouver, p. 99.

Nuance, 2015. Company website [online], available: http://www.nuance.com [accessed 1 Mar 2015].

OmegaT, 2015. Open-source project website [online], available: http://www.omegat.org [accessed 1 Mar 2015].

Papineni, K. et al., 2002. BLEU : a Method for Automatic Evaluation of Machine Translation. *Computational Linguistics*, (July), pp.311–318.

Pierce, J.R. et al., 1966. Computers in translation and linguistics (ALPAC report). report 1416. *National Academy of Sciences/National Research Council*.

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.

Reddy, A. et al., 2009. Incorporating knowledge of source language text in a system for dictation of document translations. *Proceedings of the twelfth Machine Translation Summit*.

Roukos, S., Ittycheriah, A. & Xu, J.-M., 2012. Document-Specific statistical machine translation for improving human translation productivity. In *Computational Linguistics and Intelligent Text Processing*. Springer, pp. 25–39.

SDL, 2015. Company website [online], available: http://www.sdl.com [accessed 1 Mar 2015].

TAUS, 2015. Dynamic Quality Framework [online], available: https://evaluate.taus.net/evaluate/dqf/dynamic-quality-framework [accessed 1 Mar 2015]

Wardell, J., 2014. Using Dragon NaturallySpeaking Speech Recognition Software to Maximize Speed and Quality in memoQ [online], available: https://www.youtube.com/watch?v=VWQOwBUS-kM [accessed 1 Mar 2015].

Wordfast LLC, 2015. Company website [online], available: http://www.wordfast.net [accessed 1 Mar 2015].