# Translation practice in the workplace and Machine Translation

**Nikiforos Karamanis**    **Saturnino Luz**    **Gavin Doherty**

Department of Computer Science
Trinity College Dublin
Ireland
`firstName.lastName@cs.tcd.ie`

## Abstract

Localisation has long been regarded as an appropriate domain for the deployment of MT. This paper reports the results of a qualitative study which investigated the activities of professional translators working in a Language Services Provider. We analyse the observed work practices and highlight issues related to translation efficiency, quality assessment and teamwork. These issues may impact on the successful deployment of MT in a localisation setting but have not been considered extensively by research assessing the appropriateness of MT for localisation as yet.

## 1 Introduction

Localisation has long been regarded as an appropriate domain for the deployment of MT. The "Best Practice Guide" issued by the Localization Industry Standards Association (Dillinger and Lommel, 2004) includes four case studies of successful uses of MT in localisation settings. However, progress has been slow and other reports by localisation experts are still referring to the role of MT as "emerging" (Esselink, 2003; Wittner and Goldschmidt, 2007; Yanishevsky, 2009).

Additional insights have been shared with the MT community by providers of commercial MT services (Groves, 2008) or large organisations (Roturier, 2009) which deploy MT to pre-translate content that is then passed on to professional translators to be post-edited. Although these reports include some high-level feedback from the translators, they do not provide a description of the activities that translators engage in on a day-to-day

basis. Moreover, these accounts mostly express the perspective of other stakeholders such as the client or the MT developer who assume different roles within the localisation workflow. This perspective is reflected on attempts to "educate translators about MT" (Groves, 2008, p.16).

Efforts to deploy MT in localisation are mostly motivated by the expectation that it can help increase throughput. That the output of the MT engine is of substantial quality is seen as a prerequisite to achieving this aim. As a result, a lot of emphasis is placed on assessing MT quality, mainly in terms of fluency and adequacy [as e.g. in Turian et al. (2003)], a practice also adopted in some competitive evaluations of MT engines. Although such judgements of fluency and adequacy are sometimes provided by professional translators, the task that the judges are asked to perform and the conditions under which they perform it for the purposes of such evaluation exercises have little resemblance with their actual working environment and practice.

In industry settings, a reported best practice is to assess MT quality on the basis of judgements provided by members of the organisation which deploys the MT engine rather than by the professional translators who post-edit the MT output. Roturier (2009), for example, reports that linguists who own the MT resources judge whether a machine translated sentence can be understood without reading the source text. It is again unclear how closely this assessment corresponds to the work practice of the translator.

Assessing MT quality has also motivated experimental attempts to measure post-editing effort. Although these experiments seem to have been performed mostly in laboratory conditions, a few attempts to run studies *in situ* have also been re-

ported. The latter are generally regarded as closer to "real operating conditions" (Macklovitch, 2006, p.168). However, the tools that professional translators use on a daily basis are often not part of the experimental set-up. In the study by Macklovitch, a "standalone" tool developed to support an interactive approach to MT [IMT, Barrachina et al. (2009)] was introduced to two Language Services Providers (LSPs). Despite reported gains in post-editing efficiency, translators stated "in no uncertain terms" (Macklovitch, 2006, p.171) that they did not intend to use the tool unless it incorporated functionalities similar to that of a Translation Memory (TM). Koehn and Haddow (2009) also tried to assess the impact of IMT on translation efficiency using an online tool. This study was conducted remotely but the participants were not professional translators.

In another study (O'Brien, 2007), professional translators had difficulty using the dictionary of the tool that was used to log their post-editing actions. As O'Brien observes, the participants' dictionary look-up behaviour may have been influenced by the fact that they normally work with a particular TM and term management tool. Thus, similarly to evaluations of MT quality, attempts to measure post-editing effort do not appear to consider real work practice in significant detail even when the assessment takes place in the actual workplace.

## 1.1 Studying real work practice

MT constitutes the starting point of the investigation in the aforesaid studies, thus defining their methods and purpose. The difficulty of assessing technology without direct reference to real work practice has been acknowledged in certain areas of Computer Science, particularly in the fields of Human-Computer Interaction (HCI) and Computer Supported Cooperative Work (CSCW).

As a result, research in these areas has been moving away from viewing a computer system as "an island" (Norman, 2009) and towards investigating real-life activities and the environment in which these activities take place (typically referred to as "context"). Qualitative frameworks including methods inspired from ethnographic practice have been increasingly employed in support of this kind of investigation (Randall et al., 2007). It is context, not technology, which serves as the starting point of this investigation in order to gain insights for extant and new technologies.

In the study reported in this paper, an LSP serves as the context in which localisation activities are investigated. Instead of treating the LSP as a test-bench for experimenting with MT, we gain insight from activities that have become part of the everyday routine of translators working in this environment but do not necessarily involve MT per se. Our aim is not to educate translators but to learn more about their actual work practices. To our knowledge, this is the first attempt to employ qualitative methods to investigate the translation practice in the LSP and relate the findings of the contextual study to research efforts in MT.

Lagoudaki (2006; 2009) criticised the techno-centric approach that prevails in the development of TM tools for translators. Her online survey focuses on the use of a particular technology (TM) mostly by freelance translators. By contrast, we provide an "eye-witness" account of how translators situated in the LSP combine several tools and sources of information to accomplish their work.

The paper is structured as follows: After providing some background to the study, we discuss how methods commonly used in HCI were deployed for data collection and analysis. Then, we provide examples of real work performed by professional translators that help us understand which problems are particularly challenging, how the translators go about solving them, and which tools support them in doing so. Insights from the investigated environment and the decision making process that underlies the behaviour of the translators in this context are then discussed in relation to research in MT.

## 2 Background

This case study is part of a larger investigation of context in localisation. The investigation includes site visits and engagement with various stakeholders including representatives of a content provider (client), terminologists working for the client and employees in two LSPs. This paper focuses on work which takes place in one of these field sites, a large LSP which employs several in-house translators. The translators are divided into language departments according to the languages that they specialise in. These are relatively small teams typically consisting of 4-6 translators, the most senior of whom has the role of the Team Leader.

Initially, we visited each of the field sites and interviewed 15 employees with a range of roles. These visits lasted for four days and provided us

with a preliminary understanding of the main activities within and across sites and of related roles including that of the translator, the project manager, the systems engineer, quality assurance staff, etc. They were followed up by a second site visit to the LSP aimed at collecting more detailed data about how translators work. A member of our group spent two full working days investigating work activities by six translators. The data collected during this visit constitute the core of this paper. In the following we discuss our methods for collecting and analysing these data.

## 3 Methods

### 3.1 Data gathering

There are several ways in which data can be collected for qualitative research. Methods stemming from ethnographic practice are used in fieldwork in order to inform software engineering (Viller and Sommerville, 2000; Randall et al., 2007). Our approach for data gathering relies on the practice of Contextual Inquiry (Beyer and Holtzblatt, 1998). Contextual Inquiry has been employed in industrial settings and has become a standard reference when discussing techniques for requirements gathering in HCI [see e.g. Sharp et al. (2007, pp.498)].

A Contextual Inquiry is a one-to-one field interview conducted where the work is done. The researcher observes the worker as she performs her everyday tasks and inquires about her actions in order to understand her motivations and strategies. Instances of real activities are surveyed in order to capture details which are hard to elicit with other forms of investigation. The number of people to interview depends on the scope of the investigation but 4-6 interviews are generally considered enough to provide a good idea of how a certain group of people approach their work in a particular setting (Beyer and Holtzblatt, 1998, p.76).

Our researcher interviewed four male and two female translators working in four different language departments (French, Italian, Spanish and Greek). The participants have been working in the LSP for at least one year and had at least three years of experience (including working as freelancers) before joining the company. They all held professional qualifications on translation.

The interviews lasted between 1.5-2.5 hours. Each translator first gave background information about the job she was currently working on (e.g. size of the job, type of text, availability of refer-

ence material, stage of completion, etc). The translators were observed working on various jobs and performing several tasks related to translation.[1]

The translator was asked to explain her actions as she went on with her work. Repetitive activities were discussed in the first couple of instances and then the translator was left to work without interruption "unless something new comes up". In this way, the researcher could observe previously explained activities performed in almost real time.

Each session was audio recorded. The researcher also kept notes and asked the translator to provide screenshots as a visual reference for the recorded activities.

### 3.2 Data analysis

Our fieldwork gave rise to a wealth of data which was analysed following the main principles of Grounded Theory (Glaser and Strauss, 1968).[2] The interviews were transcribed and coded using an open scheme.[3] Themes were then identified in the coded transcripts and used to group together related excerpts. This approach supports the discovery of knowledge from the acquired data instead of relying on prior assumptions.

A narrative was then composed to report examples of real work accomplished by the translators. The narrative consolidates our observations with the participants' actual words into a coherent account. It exemplifies patterns of work in significant detail without overwhelming the reader with everything that occurred in the field.

The narrative was sent to the interviewed translators by email and was discussed with them during a subsequent visit in the LSP. The researcher asked the translators to criticise the narrative and suggest revisions. The translators called the narrative "factually correct" and "characteristic of our daily work" (or "our daily bread" as one person said). Clarifications were also provided and incorporated into the narrative.

In the following section, we present the part of this narrative which illuminates how translators go about solving certain kinds of problems and which tools support them in doing so. We then discuss the relation of these observations to current research efforts in MT.

---

[1] Section 4 provides relevant background information and details of the activities reported in this paper.
[2] See Sharp et al. (2007, pp.389) for an overview.
[3] ELAN (www.lat-mpi.eu/tools/elan/) was used for transcription and coding.

## 4 Findings

### 4.1 Efficient editing behaviours

The main tool used by the interviewed translators is the TM.[4] This is a commercial tool consisting of a database of previous translations as its back end. The user interface includes an editing environment which displays the sentence to be translated (source sentence) and a proposed translation as well as surrounding text. The proposed translation is the translation of a sentence in the TM (TM sentence) which matches the source sentence exactly or approximately (above a certain threshold). The Workbench window shows the source sentence, the TM sentence and the proposed translation. The Concordance window can be used to search the TM for text entered by the translator (Concordance search).

Translators were observed making frequent use of the translations proposed by the TM. Keyboard shortcuts were used often to perform operations such as accepting translations and moving to the next sentence. Text was entered by touch-typing. When translators did not leave the TM's editing environment, revisions took place quickly and editing was normally not interrupted for too long.

Translators interrupted editing in order to perform a Concordance search on several occasions. The text that they entered in the search field almost always consisted of a subpart of the source sentence (a phrase) rather than the complete sentence. Sometimes, the translators would look up a phrase even when a full-sentence translation was already proposed by the TM. On other occasions, there was no proposed translation by the TM and they would have to translate "from scratch". These situations were not always followed up by a Concordance search but when the Concordance search did take place, the translators would again search for a particular phrase.

The look-up action for the Concordance search was performed very quickly since the search function is integrated with the editing environment and can be achieved by simply selecting a phrase and using the "search" shortcut. The Concordance search often returned several results. The translators did not always have to sift through the results but when they did delays were observed.

---

[4]The participants were using the SDL Trados suite.

### 4.2 Evaluating translation quality in situ

In the following incident, the translator was working in a large project. Although it is generally considered better practice to give the whole project to one translator, this project was split between the participant, another translator in the LSP and freelance translators in order to be completed on time. A third translator from the LSP was reviewing their translations to ensure that they were of high quality. While the translator was working, he searched the Concordance for the phrase "stacking options": [Q1] *"so "stacking options" no luck* [in Concordance search] *only "stacking", it's not very easy because "stacking" is the way images go one on top of the other, you can say that in several ways, you can see here that there is no one term, she* [his Team Member] *translated with different wording every time"*

Notably, the translator was not trying to find out what "stacking" meant as if it were an unknown word. Rather, he needed to figure out which of the several translations of the term was the most appropriate for the particular segment: [Q2] *"in most cases the translator is not really stuck as in they don't have a clue about what a term means, I can easily find what "stacking" means e.g. with a dictionary or online, so it would be more helpful for me to know what he* [the Team Leader] *thinks or what my team agrees with rather than starting a debate with a freelancer whom I have never worked with"*

From the various team members, the opinion of the Team Leader is of particular value: [Q3] *"he is the language authority or coordinator, if I do what a freelancer did and he disagrees he has the last word and the responsibility if the client complains"*

The Concordance interface (as well as the Workbench) provides the translator with the required hints. These were the translator's comments during another Concordance search: [Q4] *"it's very important to look at properties, most important is the date it was created, if it was a segment for [Product]-2001 and I am translating [Product]-2009 I don't really care how it was used, it has to be recent,* [. . . ] *to see the name of one of my team-mates it means that they have the same references as I have on the server, they went to the briefing with me so I trust them more, the name and the date are the most important things"*

Freelancers, as already indicated, are trusted

less. These comments were made during another Concordance search:

[Q5] *so again "arrangement" can have several meanings, I know they are talking about the pictures, again I do concordance* [searches Concordance]*, in this case they are all the same but if this was not the same, if here it said something different from this, I would trust this user more because the other one is a freelancer and I know that freelancers do not have all the materials that we have and did not have the briefing, if we had very legal text and the freelancer is specialised in legal terminology then I could trust them more, when they send us the TM of the job that they did at home we import it to our TM but we also put this attribute "Freelance of [Product]-Training", even if I did not remember that she is a freelancer I would know from here"*

Again, the attribute on the interfaces to the TM enabled the translator to inspect the origins of the provided translation in order to assess how trustworthy it is. Although the various translations retrieved by the Concordance are all written by professionals, they are not of equal value. Who provided the translation is a crucial piece of information that the in-house translator is using when evaluating its quality.

The interfaces to the TM make visible whether a translation has been reviewed and revised:

[Q6] *"I did this segment on September 4th at 5pm, it was reviewed by her and she changed something, I have to keep this, I want my presentation to be consistent to have the same translations, I know that this is correct so I accept this"*

This awareness further strengthens the trust that the translator feels in the tool:

[Q7] *"in most cases if there is a difficult term someone researches it and it goes to the TM, after the review it stays in the TM and this is the final decision about it, if I am a new translator and I come across this term I trust the TM"*

### 4.3 Teamwork in the LSP

A Concordance search was normally the first step that the translators took when they were faced with a problem. However, this did not always give the answer. The translators were observed consulting various other reference material (including glossaries and term lists, the localised UI and the source text in the original language when they were available, etc). They had many applications running on their desktops and had to switch their attention from the TM editing environment to another window to access a reference. Some references came as Word documents and the translators were observed searching for terms using the "Find" function. This was typically more time-consuming than searching the Concordance.

Most of the internal references originated from the client but some were also made by the team. E.g. when there were several terms which could be translated in "a few ways", the team collected the inconsistent examples, resolved them collaboratively and recorded the decisions in a spreadsheet about "terminology issues":

[Q8] *"you see this, we did it ourselves, maybe the TM is inconsistent sometimes so in the case of "device" there are a few ways to translate it, we decided to go for this among ourselves"*

Similarly to the "stacking options" example (see [Q1] above), the problem was not that the translator needed help to find out what "device" meant but that he had to choose between the various ways in which this term had been translated in the past. Again, teamwork of this kind occurred when a large project had to be split between several translators in order to meet a tight deadline.

### 4.4 Searching for information online

Sometimes there were no internal references or the references did not provide an answer to the translator's question. In those cases, the translator "did research", mostly by searching online.

In this instance, the translator was challenged by the phrase "dual throttle". He searched an online dictionary which included "forum discussions". After looking at the meanings of the dictionary entries for "throttle", he looked at the links in the forum:

[Q9] *"let's see "full throttle"* [clicks on "full-throttle" link in the forum discussion] *and someone says, this is another window, "blabla martin Scorsese is filming in full throttle", I don't like this, I go back, "full throttle" is about cinema mine is technical,* [pause] *this is by someone else* [clicks on "air-throttle valve" link] [reads in Spanish] *so it's the valve of the accelerator, OK, so next we go to Google to check if it means what I assume that it means* [searches Google]*, let's see how many, "full throttle"* [pause] *this is proz*[5]*,* [pause] *"half*

---

[5]www.proz.com, which was described as an "online portal for translators".

*throttle", "half acceleration" so it's definitely acceleration so no further question"*

Another translator was observed performing a series of online searches using Google. For each search, she commented on the reliability of the websites that the search was directing her to. Websites that were deemed reliable were explored further, the non-reliable ones were ignored.

While the results of online search were scrutinised, this was not the case for information originating from the team, as a third translator stated: [Q10] *"when I do research online e.g. in proz I don't know these people and they don't know my file and I want to know why they chose this term but for a project here I trust them* [his Team Members] *because they have all the information to decide"*

This statement was made in relation to decisions recorded in the "terminology issues" spreadsheet (see [Q8] above). As the translator acknowledged, compiling such resources "requires a bit of work". Searching online was also observed to be more time-consuming than searching the Concordance. When it was done, the translators did not accept the search results in blind faith. Rather, websites and individual entries were inspected to assess the quality of the provided translation.

## 5  Discussion

In this section we summarise the insights that we gained from the contextual study and discuss them in relation to relevant research in MT.

### 5.1  Improving translation efficiency

Our first observation is that the interviewed translators appeared to be working within the editing environment of the TM quickly and efficiently. The main cause for interruption was a Concordance search (section 4.1). Unsuccessful searches were often followed by consulting the reference material (section 4.3). This caused more significant delays, particularly when the translators also had to look for information online (section 4.4).

This observation contrasts with approaches focusing on "saving keystrokes" such as the IMT paradigm evaluated by Macklovitch. All our participants are fast typists and do not seem to require much assistance with editing, especially if this is introduced as a "standalone" application in spite of their familiarity with the TM editing environment.

The results of a more recent exploration by Koehn and Haddow (2009) also suggest that the margin of improvement for typing speed is not large. In their study, delays are mostly caused when participants are inactive (i.e. not typing). The study was done remotely so it is not clear whether the logged inactivity actually corresponds to time that subjects had spent actively looking for solutions to translation problems outside the logged interface and whether these solutions were perceived as improving their translation.[6] Our observations suggest that this is possible although our participants are professional translators working with tools that they are familiar with, unlike participants recruited in Koehn and Haddow's study.

More generally, showing that MT may increase typing speed or reduce inactivity which is caused by the experimental setting e.g. because the participants were using an unfamiliar editing environment or because they could not consult otherwise accessible information (by performing a Concordance search or looking at their references) does not appear to shed much light on the potential benefits of MT in a real working environment.

Work practices defined around MT have resulted in attempts to "educate translators to only correct actual translation errors" and "avoid unnecessary post-editing effort" (Groves, 2008, p.17). As suggested in Offersgaard et al. (2008), this strategy may pay off when the translator cannot benefit much from the TM. However, the causes of interruption identified in our study may be relevant to these cases too, particularly when one aims to support experienced typists.

Hence, more detailed investigations of editing efficiency and inactivity in real settings are required. Given the difficulty of running controlled experiments in such environments [c.f. O'Brien (2006; 2007)], contextual studies like the one presented in this paper can be seen as an alternative methodological choice for investigating these issues further. This can focus research and development efforts on addressing problems actually encountered in the workplace instead of introducing technology based on assumed needs.

### 5.2  Evaluating translation quality

MT typically provides translations for complete sentences or a whole text. Yet, our observations suggest that the real stumbling blocks for translators are finer grained, at the phrase or term level

---

[6]Koehn and Haddow (2009, p.76) tried to evaluate the quality of the translations but the results were inconclusive.

(see the examples in section 4). This is true also when they translate sentences "from scratch".

The attention that is given to customising the MT dictionary together with research in statistical post-editing in the industry (Roturier, 2009) and the suggestion by Koehn and Haddow that the phrase table of statistical MT could serve as a translation aid appear to be focusing on assisting translators with these problems.

However, design decisions which do not involve MT but may still help the translator with these issues are conceivable too. E.g. the Workbench window of the TM enables the translator to look up terms in an integrated glossary. However, most reference material comes in various other forms and the translators in the LSP still need to switch between applications and search Word documents and spreadsheets for terms and phrases (section 4.3). Improving the integration and visualisation of reference material appears to be a more straightforward way to address this problem than feeding a glossary to the MT engine and asking the translator to deal with its output.

Our study also reveals that a particularly crucial question for the translator in the LSP is not "what translation" but "by whom" (section 4.2). The preference of translations by team members over those of freelancers is of particular interest. This is not some arbitrary bias against other professionals. Freelancers are less trusted because related material is not always accessible to them and they are unable to participate in internal briefings. Simply put, freelancers are not full participants to the way the work is done within the LSP and, therefore, they are less trusted.

This suggests that perhaps the ultimate challenge for MT within this ecosystem is to stand at the same level as a trusty team member. What makes the translation of a team member more reliable than that of a freelancer does not have much to do with its quality expressed in terms of fluency and adequacy as done in most human-based evaluations of MT engines. Freelancers are professional translators and their suggestions are arguably both fluent and adequate to a great extent.[7]

However, the freelancer's translation may fall short when compared to a translation made by a team member due to subtle factors which are related to the context of work in the LSP. These factors have not yet been taken into substantial consideration in the human-based evaluations of MT output which tend to happen out of this context.

To accept a translation the translators in the LSP need to be able to evaluate its quality in ways that are meaningful to them. The interface of the TM supports this activity by showing whom the translation comes from and whether it has been reviewed (section 4.2). This increases the trust that the translator places in the tool. This observation together with the collaborative work practices that translators in the LSP develop to resolve translation conflicts (section 4.3) contrasts with the suggestion that they tend to "give into their subjective opinions" (Groves, 2008, p.20). Rather, these findings suggest that the methods used to assess the appropriateness of MT for localisation may benefit from incorporating criteria more directly related to the translators actual work practice.

Accepting a translation from an online search requires extra steps for verification compared to a team member's decision (section 4.4). This is another aspect of the work in the LSP that research in MT does not appear to have captured as yet. Although the specialised researchers who developed an MT engine are probably able to tell why a certain string has been translated in a particular way, for most people who are not working in this domain the MT engine remains a black box.

Hence, providing the translator in the LSP with more information about the origins of a translation is another challenge for the successful integration of MT within the investigated context. Making the MT dictionary or the translation table more transparent (e.g. by highlighting recent updates or by making the effort involved in the preparation of terminological resources more visible across sites) may be helpful steps in this direction.

Overall, one needs to keep in mind that the translators are more likely to ask "why" rather than "what" (see [Q10] in section 4.4). Supporting their effort to answer this question appears to be crucial for the successful introduction of new technology in the investigated setting.

## 5.3 Investigating teamwork in more depth

Previous discussions of TM tools by MT specialists (Somers, 2003) and localisation experts (Esselink, 2003; Wittner and Goldschmidt, 2007) consider re-using previous translations as their main utility. This view is shared by Lagoudaki

---

[7]Arguably, freelance and in-house translations are probably treated as equally valid when corpora are used for the training or evaluation of MT engines.

(2006, p.3) whose online survey elicited information mostly by freelancers without direct observation of TM use. Based on the technique of Contextual Inquiry, our study provides more detailed examples on how translators in the LSP combine the TM with other tools and sources of information to coordinate their work.

Examples of teamwork (section 4.3) are of particular interest given that these occur when a job is split between several translators in order to be completed within tight deadlines. Previous explorations of the relation of MT or TM use with translation efficiency and quality do not appear to discuss teamwork in significant detail. We intend to report more extensively on the collaboration, coordination and knowledge sharing among team members at the LSP in a subsequent publication.

# 6 Conclusion

We have exemplified how investigating the work practice of translators situated in an LSP using qualitative methods has helped us gain insights about the prospective use of MT in a localisation setting. Our analysis highlights several aspects of the work practice that are related to translation efficiency and quality assessment but have not yet been integrated into research and development efforts investigating the appropriateness of MT for localisation. The importance of teamwork in a collocated setting is also identified as an issue worth of further investigation.

## Acknowledgments

## References

Barrachina, S., O. Bender, F. Casacuberta, J. Civera, E. Cubel, S. Khadivi, A. Lagarda, H. Ney, J. Tomás, E. Vidal, et al. 2009. Statistical approaches to computer-assisted translation. *Computational Linguistics*, 35(1):3–28.

Beyer, Hugh and Karen Holtzblatt. 1998. *Contextual Design: Defining Customer-centered Systems*. Morgan Kaufmann.

Dillinger, Mike and Arle Lommel. 2004. *LISA Best Practice Guides: Implementing Machine Translation*. Localization Industry Standards Association.

Esselink, Bert. 2003. Localisation and translation. In Somers, Harold, editor, *Computers and Translation: A translator's guide*, chapter 5, pages 67–86. John Benjamins.

Glaser, Barney G. and Anselm L. Strauss. 1968. *The Discovery of Grounded Theory*. Weidenfeld and Nicolson.

Groves, Declan. 2008. Bringing humans into the loop: localization with machine translation at Traslan. In *AMTA 2008*, pages 11–22. Keynote presentation.

Koehn, Philipp and Barry Haddow. 2009. Interactive assistance to human translators using statistical machine translation methods. In *MT Summit 2009*, pages 73–80.

Lagoudaki, Elina. 2006. *Translation Memories Survey*. Imperial College London.

Lagoudaki, Elina. 2009. Translation editing environments. In *MT Summit 2009 Workshop: Beyond Translation Memories: New Tools for Translators*, pages 9–18.

Macklovitch, Elliott. 2006. Transtype2: the last word. In *LREC 2006*, pages 167–172.

Norman, Donald. 2009. Systems thinking: a product is more than the product. *Interactions*, 16(5):52–54.

O'Brien, Sharon. 2006. Pauses as indicators of cognitive effort in post-editing machine translation output. *Across Languages and Cultures*, 7(1):1–21.

O'Brien, Sharon. 2007. An empirical investigation of temporal and technical post-editing effort. *Translation And Interpreting Studies*, 2(1):83–136.

Offersgaard, L., C. Povlsen, L. Almsten, and B. Maegaard. 2008. Domain specific MT in use. In *EAMT 2008*, pages 150–159.

Randall, D., R. Harper, and M. Rouncefield. 2007. *Fieldwork for design: theory and practice*. Springer.

Roturier, Johann. 2009. Deploying novel MT technology to raise the bar for quality: a review of key advantages and challenges. In *MT Summit 2009*, pages 1–8.

Sharp, Helen, Yvonne Rogers, and Jenny Preece. 2007. *Interaction Design: Beyond Human Computer Interaction*. John Wiley & Sons.

Somers, Harold. 2003. Translation memory systems. In Somers, Harold, editor, *Computers and Translation: A translator's guide*, chapter 3, pages 31–48. John Benjamins.

Turian, Joseph, Luke Shen, and Dan Melamed. 2003. Evaluation of machine translation and its evaluation. In *MT Summit 2003*, pages 386–393.

Viller, Stephen and Ian Sommerville. 2000. Ethnographically informed analysis for software engineers. *International Journal of Human-Computer Studies*, 53(1):169 – 196.

Wittner, Janaina and Daniel Goldschmidt. 2007. Technical challenges and localisation tools. In *Localisation Guide 2007*, pages 10–14. Multilingual Computing Inc.

Yanishevsky, Alex. 2009. The emerging role of machine translation. In *Localisation Guide 2009*, pages 12–13. Multilingual Computing Inc.