Machine Translation manuscript No.
(will be inserted by the editor)

Nikiforos Karamanis · Saturnino Luz · Gavin Doherty

Translation Practice in the Workplace

Contextual Analysis and Implications for Machine Translation

Received: date / Accepted: date

Abstract This paper reports the results of a qualitative study which investigated localisation activities performed by translators working in two Language Service Providers. It argues that maintaining the appropriate quality level in this setting is a collaborative task which involves several translators. This perspective entails taking a broader view of the translation process than usually found in the Machine Translation (MT) literature and detailing the various knowledge sources which are deployed in this collaborative effort. The impact of collaboration on trust is examined, and a comparison is made between the relatively seamless flow of work between translators and the more strained relationships with remote contributors. In support of this view, the paper contrasts the flexibility of the analysed work practices with the rigid ways which tend to be followed when introducing MT into this setting. We identify the need to support collaboration and communication more actively as a broader issue in translation settings. While current strategies for introducing MT tend to further isolate translators from remote contributors, we propose that MT can serve as the catalyst for establishing a more dynamic and collaborative relationship between them.

Keywords translation practice · localisation · qualitative study · contextual analysis · translation quality · translation process · collaboration · trust · flexibility

N Karamanis · S Luz · G Doherty
Department of Computer Science
Trinity College Dublin, Ireland
Tel.: +353-1-896 - {1885, 3858, 3686}
E-mail: {karamann,luzs,gdoherty}@cs.tcd.ie

Karamanis, N., Luz, S., and Doherty, G. Translation practice in the workplace: contextual analysis and implications for machine translation. *Machine Translation 25*, 1 (2011), 35–52.

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 uses of MT in localisation settings. However, progress has been slow and other reports by localisation experts still refer to the role of MT as "emerging" (Esselink 2003; Wittner and Goldschmidt 2007; Yanishevsky 2009).

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. To date, the most accurate MT evaluations rely on standardised judgements provided by human raters (Jurafsky and Martin 2008, pp.930). Automatic metrics are tested in terms of their correlation with such judgements (Przybocki et al. 2009).

Similar practices are followed in industry settings. Roturier (2009, p.2), for example, reports that linguists who work for the organisation that deploys the MT engine judge whether a machine translated sentence can be understood without reading the source text. When the MT engine achieves such quality levels, its output is passed on to professional translators to be post-edited.

Assessing MT quality has also motivated experimental attempts to measure post-editing effort. These studies deploy various methods such as measuring time-on-task (Offersgaard et al. 2008; Guerberof 2009), often combined with keyboard logging (Jensen and Jakobsen 2000; Macklovitch 2006; O'Brien 2007; Koehn 2009; Plitt and Masselot 2010) and, more recently, eye-tracking (Carl 2009; Doherty et al. 2010). Most of these studies report performance gains when MT output is post-edited compared to translating "from scratch". Some are also motivated by an interest in gaining insight on the translation process and in informing the design of interfaces tailored to post-editing (Carl 2009; Koehn 2009).

Performance studies require imposing controls which deviate from the daily work practice of professional translators. Macklovitch (2006), one of the few studies performed *in situ*, is a characteristic case. Despite reported gains in post-editing efficiency, translators stated "in no uncertain terms" (ibidem, p.171) that they did not intend to use the tool tested in the study unless it incorporated functionalities similar to that of a Translation Memory (TM). 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 (ibidem, p.21).

According to Groves (2008, p.11) "historically, translators are not known for their open acceptance and use of MT technology". A need emerges to "educate translators to only correct actual translation errors [...] rather than give into their subjective opinions or preferred translation style. Carrying out such unnecessary post-editing effort reduces the benefits and poten-

Karamanis, N., Luz, S., and Doherty, G. Translation practice in the workplace: contextual analysis and implications for machine translation. *Machine Translation 25*, 1 (2011), 35–52.

tial cost savings of the MT system" (ibidem, p.16). Closely monitoring the post-editing effort (e.g. by counting keystrokes or the time spent between segments) is often seen as a way of maximising the benefits of MT deployment. Guidelines are sometimes provided in an attempt to standardise the task.

Although three decades have passed since Martin Kay outlined a vision of MT technology as a tool for human translators, highlighting the importance of good user interface design (Kay 1998), the work practice of professional translators still does not seem to have been considered in significant detail by MT research. In order to partially fill this knowledge gap, Désilets et al. (2009) deployed ethnographic methods to study the use of linguistic resources and tools by professional translators. They report that translators use a wide variety of resources and tools. These are not trusted blindly but translators exercise much critical judgment in deciding which solutions to choose or reject.

In this paper, we report results of another qualitative study which investigated activities of translators working in two Language Service Providers (LSPs). Our aim is not to educate translators but to learn more about their actual work practices. In particular, we report *how translators themselves ensure quality* as opposed to collecting standardised judgements or assessing their post-editing performance and compliance with guidelines.

The paper extends our previous work (Karamanis et al. 2010) and that of Désilets et al. by providing more detailed accounts of translation practice and recommendations for the successful integration of MT in the investigated setting. While Désilets et al. report on the use of tools in various settings, our study focuses on localisation activities within the LSP. In this setting, projects are often split between several translators. Indeed, it is the large volume of content involved in these projects that often motivates the use of MT for "bulk localisation" (van Genabith 2009, p.9).

Our study is novel in several ways. First, we argue that maintaining the appropriate quality level in the observed setting is a collaborative task which involves several translators. To our knowledge, maintaining quality has been treated as a task accomplished by a single individual in previous studies. Our view of the translation process is broader than in previous work and entails detailing the various knowledge sources which are deployed in this collaborative effort. In addition to the various linguistic resources and tools (some of which were also identified by Désilets et al.), our analysis accounts for the communication flows between the various people involved in the observed activities. We also provide a more detailed account of how translators combine these knowledge sources to structure their work.

Moreover, for the first time we explicate how collaboration impacts on trust and compare the relatively seamless flow of work between translators with the more strained relationships with remote contributors. The need to support collaboration and communication more actively is identified as a wider issue in the setting. In addition to documenting the translators' daily practice (which does not usually involve MT), we discuss data obtained from the first ethnographic investigation of the emerging post-editing task. We comment on the flexibility of the habitual translation practices and

Karamanis, N., Luz, S., and Doherty, G. Translation practice in the workplace: contextual analysis and implications for machine translation. *Machine Translation 25*, 1 (2011), 35–52.

contrast it with the strict enforcement of post-editing guidelines and performance measures. While current strategies for introducing MT tend to further isolate translators from remote contributors, we propose that MT can serve as the catalyst for establishing a more dynamic and collaborative relationship between them.

2 Background

The need to investigate 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, there has been a shift towards research which investigates 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.

Our investigation includes site visits and engagement with various stakeholders including employees in the localisation division of a large software company and in two LSPs. The first LSP employs over 150 in-house staff and acts as a multi-language provider localising products in several languages. The software company is one of its significant clients. Translators working in the LSP are divided into language departments. The most senior translator in each department acts as the team leader.

Overall, we carried out over 55 hours of fieldwork in three sites. Initially, we visited the localisation division of the software company and the LSP. We carried out 14 semi-structured interviews at these sites, typically lasting between 45 minutes to an hour each. This provided us with an understanding of the main activities within and across sites and of the roles of people involved in them including that of the translator, the team leader, the project manager, the terminologist working for the client, etc.

To investigate how translators work in more depth, we carried out additional visits to the first LSP and to a smaller LSP which acts as a single-language provider and employs only four in-house staff. 13 additional interviews were conducted at this stage. Our methods for collecting and analysing these data are presented in the following section.

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) and design (Randall et al. 2007). Contextual Inquiry (Beyer and Holtzblatt

Karamanis, N., Luz, S., and Doherty, G. Translation practice in the workplace: contextual analysis and implications for machine translation. *Machine Translation 25*, 1 (2011), 35–52.

1998) has been employed for product development and has become a standard reference when discussing techniques for establishing requirements in HCI (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 five to ten 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).

We carried out 13 interviews with translators localising content from English into six different languages (French, Italian, German, Spanish, Greek and Hungarian). The participants have been working in the LSPs for at least one year and had at least three years of experience (including working as freelancers) before that. They all held professional qualifications on translation.

The interviews lasted between one and 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 reference material, stage of completion, etc). The participants were observed working on several jobs and performing various localisation tasks.[1] The sessions were audio recorded. The researcher also kept notes and asked the participant to provide screenshots as a visual reference for the investigated activities.

The translators' work practice does not involve MT on a daily basis as yet. This means that we had the opportunity to investigate the emerging post-editing task only in a few occasions. This is not a major limitation since our focus was on capturing how translators maintain quality irrespective of the technologies involved. In fact, as we will show in section 4, investigating activities that do not involve MT helped us gain insights on habitual translation practices and identify wider issues in the setting. These insights are relevant to the bulk localisation projects that MT is usually considered appropriate for and serve as the backdrop for discussing the data obtained from the post-editing interviews.

3.2 Data analysis and validation

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

[1] Section 4 provides relevant details.

[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.

A narrative was then composed to consolidate the analysed data into a coherent account. The narrative exemplifies patterns of work in significant detail without overwhelming the reader with everything that occurred in the field. It focuses on work which spans across the larger LSP and its clients, supplemented with data collected in the other sites.

The narrative was discussed with the interviewed translators during subsequent visits in the LSP. The researcher asked them 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.

A presentation was then prepared and discussed during a session hosted in the LSP and attended by other translators and company executives. The translators recognised the reported work practices as "familiar" and elaborated with their own examples.

In the following section, we present the part of the analysis which explicates the collaborative effort expanded by translators to maintain the appropriate quality level and demonstrates how collaboration impacts on trust. We contrast the flexibility of the analysed work practices to the rigid ways which tend to be followed when MT is introduced to this setting. The implications of this analysis for the integration of MT are then discussed and vision for better supporting those working in this setting is outlined.

4 Collaborative translation practices and trust

Our analysis suggests that producing consistent translations and maintaining the appropriate quality level in bulk localisation projects is a collaborative task which involves several translators. The following examples demonstrate the collaborative effort expanded by the translation team.

4.1 Collaboration via the Translation Memory

One of the main tools used by translators in both LSPs is the software which displays matches from the Translation Memory (TM) and their translations. The TM's back end is essentially a database of previous translations. As a translation job progresses, new translations are added. The user interface to the TM displays the sentence to be translated and a proposed translation. The proposed translation is the translation of a sentence in the TM which matches the sentence to be translated exactly or approximately. The interface colour codes the proposed translation according to the type of match. Most such software has a Concordance function which can be used to search the TM for text entered by the translator.

Translators were observed searching the Concordance frequently, mostly for subparts of the segment they were working on such as a single word or a short phrase (i.e. a term).

Searching the concordance is not always equivalent to looking up an unknown word. In the following incident, the translator was working on a

large project which was split between her,[4] another translator in the LSP and freelancers in order to be completed on time. A third translator from the LSP acted as a reviewer.

The participant searched the Concordance for a term and explained the motivation behind this action:

> 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 [team leader] thinks or what my team agrees with, rather than starting a debate with a freelancer whom I have never worked with

Searching the Concordance returns a list of segments from the TM which contain the searched term and their translations. The Concordance interface shows who translated each entry and when the translation was done. The same properties are displayed when the translation of a matching segment is retrieved from the TM. These properties make the translator aware of what his team agrees with:

> it's very important to look at properties [...] 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 interfaces to the TM also display whether a translation has been revised:

> I did this segment on [date] 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

In this example, the translator willingly accepts the change done by the reviewer (one of her team members). The use of the TM to support collaboration between team members is summarised in the following quote:

> 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

4.2 Creating resources collaboratively

Another large project was split between several translators. They noticed that the TM was "inconsistent", containing terms that can be translated in "a few ways". The team collected these terms, resolved them collaboratively and recorded their decisions in a spreadsheet called "terminology issues":

[4] We use "she" and "her" to refer both to men and women.

> 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

Again, the translator trusts the decisions by her team members:

> because this is a team tool I am not really interested in why exactly they found this and they decided that, I start after they have already worked on it for a while so I trust them

This "team tool" is updated as the work progresses and a need to communicate changes emerges. An email is sent to make team members aware of such developments:

> [Translator reads email] "For "quiet mode" we are going to use this, I added it to the list", she [team member] is warning me and her [team member] that she decided, probably together with him [team leader] and her [team member], that this is the decision for this mode, so guys let's make it consistent

Similarly to the use of the TM, maintaining quality in that way is a collaborative task involving several team members.

4.3 Reconciling differences collaboratively

As shown previously (section 4.1), the first point of contact between the team members is the TM. When differences arise, they are resolved through talk:

> sometimes you find different terms [in the TM], if you disagree you talk, this is the importance of talking and having four people here

Disagreements about "terminology issues" (section 4.2) are addressed in the same way:

> if she [team member] and he [team leader] now agree to translate "device" in a certain way and I jump into the project because they are good translators I build on what they have already decided, if I really don't agree we can just talk

These examples further demonstrate the collaborative effort expanded on maintaining quality. Face-to-face communication is facilitated by the physical arrangement: Several translators share a desk and face each other. Email and instant messaging (e.g. when a team member is not immediately available or when the translator wants to send them a problematic segment) are blended with verbal exchanges. Such informal communications occurred frequently in both LSPs when translators needed to verify decisions made by their team members.

As shown so far, mutual trust built on collaboration and communication plays an important role on how translators address problems. Conversely, relationships with remote contributors are more strained as we exemplify in the following sections.

4.4 Freelancers are trusted less

While work is shared between team members in the LSP relatively seamlessly, freelancers are trusted less. This is not an arbitrary bias against other professionals. Freelancers are trusted less because they are not full participants to the way the work is done within the LSP. These comments were made during another Concordance search:

> I'd 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, [...] when they send us the TM of their job that they did at home we import it to our TM but we also put this attribute "Freelance of [Product]", even if I did not remember that she is a freelancer I would know from here

Mis-trust towards freelancers has implications for several stages of the process. The first project for a new client is always done internally. There is a preference for not splitting projects between team members and freelancers although for large projects this cannot always be avoided. In that case, they try to give easier parts to freelancers. A freelancer's work tends to be reviewed "more thoroughly", particularly when the freelancer is used by the LSP for the first time. While email and instant messaging seem to work well between team members (section 4.3), communication with freelancers is more problematic:

> which [asking] I cannot do at home if I worked as freelancer, it's this concept of having all these people in the same room, you go and talk to people or send them an email or whatever, it's not difficult, someone comes here and says "how did you solve that?"

4.5 Results of online search are scrutinised

Translators are usually provided with a glossary and other references. Sometimes, however, references are not available or are not helpful enough. In those cases, the translator would search online.

Because "the internet is a dangerous place", as one translator said, the results of online search are subject to scrutiny. For example, term definitions found in an initial search are confirmed by performing additional searches. In a characteristic case, a translator was observed searching an online dictionary which included "forum discussions" for a term. After looking at the meanings of the dictionary entry, she checked the links in the forum. A hypothesis was formed based on one of these links and was verified by searching Google "to check if it means what I assume that it means". The destination websites (certain sites are trusted more than others) and the number of hitcounts were used as cues. Such verification is usually not necessary for decisions made by team members:

> when I do research online 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 [team members] because they have all the information to decide

This comment was made in relation to the "team tool" collaboratively made by trusted team members (section 4.2). On the other hand, if her online search is not conclusive, the translator will consult a team member (also see section 4.3):

> I'd say "can you come over when you have a minute" and I'd show him the sentence, that's also the point of working in a team because we can discuss things [...] I'd say "I've been looking here and there I found this and that but I'm not sure"

Because translators do not always start working on the same project at the same time, questions are often directed to a team member who has "already worked on it for a while" (cf. section 4.2). Less experienced team members (junior translators) seek advice from more experienced ones. Contributions by the team leader are particularly valuable: "They've been working in translation for 30 years so sometimes it's good to check with them".

The analysis suggests that translators do not consult resources arbitrarily. Trusted resources are given priority over less trusted ones. Indeed, on many occasions the translators were observed doing a Concordance search as their first step for resolving a problem. Moreover, they would normally consult the TM and their references (provided that these were available) first and then search online. Trusted team members were asked to verify an inconclusive search and their own decisions.

The translation process is structured. It consists of strategies which are applied *flexibly* to support collaborative work. For instance, the translators search the Concordance at their own discretion and are not prevented from doing so if they consider this action appropriate: "Anytime I can highlight a portion of the sentence that I am translating and do a concordance search". The same is true for the collaborative creation of resources (section 4.2), the way in which differences are reconciled (section 4.3) or the scrutiny that online search is subject to.

4.6 Constraints in communication with the client

Translators reconcile differences collaboratively using a range of informal and flexible strategies (section 4.3). Communication with the client, on the other hand, is more constrained. Translators cannot contact the client directly. If they want to raise an issue, they have to to query the client via the project manager. However, this may cause bottlenecks and some queries may not be answered on time or even acknowledged at all.

Translators stated that they use the option of querying the client with great discretion and were observed refraining from doing so on several occasions. Communication with the client is viewed as a last resort, contrasting with the more collaborative practices observed among translators despite the differences in experience and rank between them (see section 4.5). Consequently, even though the client may receive a few queries, the full scope of problems faced by the translators is not revealed.

More generally, the work of the translator is not particularly visible to remote contributors. Conversely, the translator does not have a clear view

of the effort expanded by a freelancer or by a terminologist working at the client's site. While reconciliation and verification are accomplished via informal communication within the confines of the translation team, there is little provision for such practices in cross-organisational interactions. In other words, the translator is relatively *isolated* from remote contributors.

4.7 Machine Translation and unmet collaboration needs

Because introducing MT is still at a "very early stage", translators need to invent solutions to emerging issues, often by adjusting their work practice. As one translator said: "we are still working on it and trying to find out if we can use it [MT] in a good way". This process mainly takes place within the LSP while the constraints in collaboration and communication with the client cause further strains.

The previously quoted translator described the background of the project that she was observed working on as follows:

> they [client] decided to go through MT this year because they wanted to improve productivity, that was the main reason, there was a pilot back in September, from the data they collected they noticed that productivity could be improved so they decided to implement this new method this year, this can be true but sometimes it's not

A number of issues arose after the pilot study. One of the main issues was that the translator could no longer rely on the Concordance:

> normally when you don't use MT the concordance is very useful because you get to see translations in strings included in this bundle so you are sure that what you are translating now is consistent with previous translations but with MT this is not reliable because it will show you strings translated by the machine so you cannot rely on this, there is no way to filter it unfortunately, this is one of the big issues

Unlike the examples discussed in sections 4.1 and 4.4, the client requested translators to work with a tool that did not display who translated each of the retrieved segments when the Concordance is searched. To work around this issue, the translator had to copy the translated files to a separate folder and use the "search and replace" function of the Windows file system instead of the Concordance:

> in this folder I put all the exported versions of the files that I previously post-edited or translated myself, so I know that everything that is in here is reliable because I did it, in the concordance I am never sure because this could have been translated by the machine, this is a very big issue, you waste a lot of time in doing this and timing is very important for us, we have to find a solution for this

Although the pilot study did not account for the additional effort involved in this workaround, the translator feels that it causes her to "waste a lot

of time". Because such solutions are sought primarily within the LSP, the client does not have a clear view of the additional effort involved.

Another issue was that some strings were not translated by the MT engine although they were expected to:

> 99% is the percentage of translated strings, we had problems with these bundles [...] in the previous bundle I noticed that there were some strings that were not pre-translated, I emailed the project manager to double check if this was right and she said that it wasn't so I had to download the bundle again but now I can see that there is still some untranslated text so I will need to check this again [...] there should be translated text here just like in here but something went wrong with MT, I don't know what

Although only a small percentage of strings are affected by this issue, the translator ends up expanding significant effort on it (by contacting the project manager, re-downloading the bundle, etc). This extra effort is not particularly visible to the client either.

The translator speculated on what might have gone wrong with MT for certain segments (e.g. "the first string is not translated, maybe the MT has problems with slashes"). However, she is not actively supported in verifying her hypotheses. Consequently, she views MT as a black box.

The black box perception is reinforced from differences within and across projects. For instance, the translator noticed that MT was more successful in the translation of documentation than of software strings in this project. Documentation contains longer sentences, which was proposed as a possible explanation. However, other projects provided contradictory evidence:

> for some reason it works best with long sentences which is very unusual because for other clients we know that it is the opposite

Again, the translator is not actively supported in finding out more about the reason why MT works better in some texts than in others and about the perceived differences between projects.

The constraints in communication and collaboration cause strains. The productivity gain anticipated by the client in this project was questioned:

> we want to keep track of the changes that we make to see if productivity can be improved as they suggest or if there are issues that suggest that MT is not good for software which has been the case so far

To keep track of changes, translators are asked to "make as few changes as possible". This deviation from the normal work practice is perceived as particularly demanding:

> reviewing what the machine translated is very different because you have to be very careful all the time that you are not letting something slip [...] it is also very risky because the engine picks translations from memories of different products so sometimes when you are tired you may think that the translation is correct when it is not

It was felt by both management and translators that it is very hard to be consistent over a whole shift when working with translations by MT. Regular discussions are held within the LSP to help translators working in a project maintain consistency (cf. section 4.3).

However, such interactions rarely cross organisational boundaries. In some projects, linguists working for the client review a sample of the MT output before passing it on to the LSP. Translators working with such material felt that the client "did not care about quality". Providing guidelines also tends to have an alienating effect since translators are typically not involved in their compilation and find it hard to generalise from a few examples. Constraints in communication prevent them from discussing specific cases in the same way as within the LSP.

Unlike the flexible work practices that the translators normally deploy (section 4.5), guidelines can be too rigid. Coercing such guidelines impacts on translators' morale. One translator was asked to correct only grammatical errors but not word order. The translator felt that she was making allowances for MT: "This is not German". Several translators expressed concerns about having to "comply with whatever the machine translates".

Projects that involve MT can become so unpleasant that the work may be split to more translators than usual. With more translators involved, the need to support collaboration becomes more pressing. This is particularly true with respect to cross-organisational interactions. Current practices for introducing MT into the setting are not sensitive enough to the effort that translators expand on emerging issues and tend to further isolate them from contributors at the client's site. This causes additional strains as exemplified in this section.

5 Discussion

We have presented above a detailed account of how participants approach their work in this setting, and have exposed the rich, collaborative and flexible practices used to achieve it. In this section, we consider the main findings of the study (Table 1), and discuss their relevance to the deployment of MT, drawing on related work on the integration of MT. We also outline possibilities for better supporting the work based on our analysis.

The proposed vision is based on the main insight that we gained from our analysis, namely the need to support collaboration, communication and flexibility more actively. We compare our findings with related studies on the integration of MT into the translation process and claim that addressing this need is not only relevant to the adoptation of MT but will be beneficial for localisation in general.

5.1 Support collaboration between translators and across sites

Extending previous studies of translation in context (Désilets et al. 2009; Karamanis et al. 2010), we have identified and described a range of collaborative activities involved in producing consistent translations and maintain-

Translation and Collaboration	Maintaining quality in localisation projects is a collaborative task involving several translators. Translators are rarely "completely stuck" but need to decide between alternative translations. Various resources are used for this purpose. Colocated translators (team members) are one such valuable resource.
Communication	Communication between team members plays an important role in solving problems. Communication with freelancers is constrained. Communication with the client is a last resort. Communication around MT issues is not supported adequately.
Trust and Acceptance	Freelancers are trusted less than team members. Online sources are also trusted less. Trust impacts on how resources are used. Mixing trusted translations with untrusted MT output can increase translation effort. Translators speculate on what might have gone wrong with MT but are not supported in verifying their hypotheses. While habitual translation practices are flexible, MT tends to be introduced rigidly.

Table 1 Main findings of contextual study on translation practice

ing an appropriate quality level. Translators collaborate tacitly via the TM (section 4.1) as well as explicitly to create "team tools" (section 4.2). Informal communication (section 4.3) supports such collaborative practices by addressing emerging reconciliation and verification needs. This concerted effort contrasts with the suggestion that translators tend to give into their subjective opinions.

Compared to the relatively seamless flow of work within the translation team, relationships with remote contributors are more strained. Freelancers are trusted less than colocated team members (section 4.4). Translations found online are scrutinised (section 4.5). Consulting the client is seen as a last resort solution and may be subject to bottlenecks (section 4.6). Current practices for introducing MT into this setting tend to further alienate translators, causing more strains (section 4.7). In all these cases, collaboration is constrained, communication is rigid and the need for verification is not supported adequately.

Previous work seems to have regarded translation as an individual activity. The TM, for example, is primarily viewed in terms of its use to recycle translations (Somers 2003; Lagoudaki 2006); its role in supporting collaboration and in establishing trust has not been documented so far. Likewise, the typical view of the localisation process (Esselink 2003; Wittner and Goldschmidt 2007) does not consider such issues. Communication patterns in this domain have not been studied in significant detail either. The metrics used to assess MT quality in competitive evaluations (Przy-

Karamanis, N., Luz, S., and Doherty, G. Translation practice in the workplace: contextual analysis and implications for machine translation. *Machine Translation 25*, 1 (2011), 35–52.

bocki et al. 2009) and the industry (Roturier 2009) also appear to overlook the collaborative nature of the task.

Performance studies focus on the interaction of the individual translator with a particular interface in order to gain insight about underlying cognitive processes (Carl 2009; Koehn 2009). The aim there is to design systems that combine TM and MT components and improve user performance by anticipating their preferences in terms of post-editing effort (He et al. 2010).

Our view of the cognitive processes underlying the translation activity in the context of an LSP is somewhat broader. Solving translation problems in this context is an activity that relies not only on an individual translator's memory and cognitive abilities, but on interactions between teams members and with resources (such as the TM, references, etc) in the LSP setting. This broader view of the translation process is characteristic of a theory of *distributed cognition* (Hollan et al. 2000) which has been influential in the field of HCI.

MT is considered to be particularly relevant to very large jobs in which the volume of content to be translated cannot be dealt with by a single person. Such jobs are usually split between several translators, within and outside an LSP. In fact, the advent of MT may result in even more translators being involved in a project than usual (section 4.7). Hence, there is a need to integrate MT in a way that preserves the affordances of existing tools (such as the TM and Concordance) for supporting collaboration. The way in which terminology issues are resolved within the LSP (often with the expenditure of significant effort) should be particularly examined in this regard.

5.2 Support flexibility at work

We have argued that the translation process is structured, consisting of collaborative strategies which are applied flexibly (section 4.5). These flexible strategies contrast with the rigid practices that tend to accompany the deployment of MT into this setting (section 4.7).

Translators appreciate being in control when using the TM. Our analysis suggests that flexibility is not restricted to the use of the TM but extends to the broader translation process. Supporting flexibility has been recognised as an important requirement in HCI, e.g. in Cooper (2004, pp.168). A technology that is introduced too rigidly ends up disrupting the work that it intends to support (Bowers. et al. 1995). Coupling the deployment of MT with the inexorable application of performance measures and the coercion of guidelines has a similarly detrimental effect.

Further motivating the need for flexibility, different projects will have different quality requirements. This has an impact on the ways that we might support people in making quality decisions on MT, and support collaboration around quality decisions with MT output. Likewise the performance of the MT will also vary from job to job (depending on the nature of the content and the success of MT tuning/preparation work), which will impact on the way it should be dealt with in the LSP; this should be recognised within guidelines for deployment, as well as in the design of tools to

support the work. Involving translators in the development of such guidelines would be one route towards making them more suited to the varying demands of the real working environment.

5.3 Support communication

Translators stated that they are often more interested in "why a term is chosen" (section 4.5) than "what a term means" (section 4.1). We detailed how informal communication is used within the LSP to address this need (section 4.3) and the tensions caused when the communication protocol becomes too rigid (sections 4.4 and 4.6).

Translations need to be verified. This is particularly true for less trusted sources (section 4.5) but also applies to differences between team members (section 4.3) Supporting the translators' quest for verification can be seen as related to the more general principle of designing for error (Norman 2002, pp.131). Currently, although the researchers who developed an MT engine may be able to tell why a certain string has been translated in a particular way, for translators the technology remains a black box (section 4.7).

This need can be addressed by supporting collaborative practices within and, particularly, across sites more actively. This will require deeper organisational changes than simply (re-)educating translators so that they comply with the MT engine. We will discuss this option using the interdependencies between translators and terminologists as an example, but similar observations appear to apply to the relationship between translators and other remote contributors such as freelancers and authors.

In the typical localisation process, decisions about terminology are viewed as being made by a terminologist at the client's site and recorded in a glossary (Esselink 2003; Wittner and Goldschmidt 2007). However, as our examples show (see e.g. sections 4.1, 4.2, 4.3 and 4.5), such decisions are also made collaboratively by translators in the LSP. The terminologist may receive queries but the full scope of problems is not revealed since translators view contacting the client as a last resort (section 4.6). While email and instant messaging work well in the colocated setting (section 4.3), communication with remote contributors is more constrained. Although means for cross-organisational communication exist, the formal protocol for dealing with problems and the lack of visibility of each other's effort discourages collaboration across sites.

Supporting the cross-organisational creation and maintenance of shared resources about "terminology issues" (an informal practice which currently takes place within each translation team but does not cross organisational boundaries: see section 4.2) more explicitly would be one way of achieving a more dynamic and collaborative relationship between translators and terminologists. This could be augmented by the informal communication channels mentioned above. Such practices may, in the long run, help the distributed groups develop levels of trust similar to those of colocated workers as attested e.g. in Wilson et al. (2006).

The introduction of new technology in a setting can often become the catalyst for change. At the moment, the practice for introducing MT tends

to further isolate translators from remote contributors. However, MT can become an opportunity to question established practices and involve translators on a more equal footing. Translators (including freelancers) can be encouraged to provide more fine-grained feedback to linguists and others working at the client's site and contribute directly to the compilation of guidelines. For this to work however, such feedback would need to be extremely low overhead, and integrated with existing tools. Crucially, information on the effort expended in resolving certain types of problem could be communicated both to Linguists and MT developers. Viewing translation quality as the result of collaborative and flexible work practices that will span organisational boundaries supported by the appropriate communication channels is the first step in realising such a vision. Tools which provide more explicit support for collaborative work on terminology and quality judgements could support better collaboration with freelancers and improve communication not just with linguists but also with those responsible for the MT.

6 Conclusion

We have demonstrated how a contextual analysis can inform a vision for better supporting translators in a localisation setting. Our data suggest that despite the collaborative nature of the translation process as witnessed in the LSP, translators are relatively isolated from remote contributors. Current practices for introducing MT to this setting appear to overlook extant interdependencies of work and tend to reinforce this isolation causing additional tensions.

We argued that this situation could be addressed by a more flexible and collaborative approach which will actively support communication across sites. Re-educating translators to work with MT implies organisational change. Such change can take place at a deeper level than currently envisaged in order to deal with wider issues in the setting such as the requirement for more dynamic collaboration between translators and remote contributors.

Supporting cross-organisational collaboration is not an easy task. Conducting a contextual analysis and outlining a vision are the initial steps in this process. User-centered design methods such as storyboarding and low fidelity prototyping (Buxton 2007; Sharp et al. 2007) can then used to specify the details of the interaction. We intend to report on those efforts in subsequent publications.

Acknowledgements This research is supported by the Science Foundation Ireland (Grant 07/CE/I1142) as part of the Centre for Next Generation Localisation (www.cngl.ie). We are grateful to Fred Hollowood and Fernando Blasi for their help in arranging the study and to our participants for their time. Many thanks to Stephen Curran, John Moran, Alfredo Maldonado Guerra, Ielka van der Sluis, Stephan Schlogl, Anne Schneider, Ilana Rozanes and Cecily Morrison for their comments.

References

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

Bowers., J., G. Button, and W. Sharrock: 1995, 'Workflow From Within and Without: Technology and Cooperative Work on the Print Industry Shopfloor'. In: *Proceedings of the Fourth European Conference on Computer Supported Cooperative Work, ECSCW '95*. Stockholm, Sweden, pp. 51–66, Kluwer Academic Publishers.

Buxton, B.: 2007, *Sketching user experiences: getting the design right and the right design*. Morgan Kaufmann.

Carl, M.: 2009, 'Grounding translation tools in translator's activity data'. In: *MT Summit 2009 Workshop: Beyond Translation Memories: New Tools for Translators*. Ottawa, Canada. Available online at http://www.mt-archive.info/MTS-2009-TOC.htm.

Cooper, A.: 2004, *The inmates are running the asylum: Why high tech products drive us crazy and how to restore the sanity*. Toronto: SAMS Publishing.

Désilets, A., C. Melancon, G. Patenaude, and L. Brunette: 2009, 'How Translators Use Tools and Resources to Resolve Translation Problems: an Ethnographic Study'. In: *MT Summit 2009 Workshop: Beyond Translation Memories: New Tools for Translators*. Available online at http://www.mt-archive.info/MTS-2009-TOC.htm.

Dillinger, M. and A. Lommel: 2004, 'LISA Best Practice Guides: Implementing Machine Translation'. Technical Report 2, Localization Industry Standards Association (LISA), Domaine en Prael, Switzerland. Available online at http://www.lisa.org/products/bestPractice/.

Doherty, S., S. O'Brien, and M. Carl: 2010, 'Eye tracking as an MT evaluation technique'. *Machine Translation* **24**, 1–13.

Esselink, B.: 2003, 'Localisation and translation'. In: H. Somers (ed.): *Computers and Translation: A translator's guide*. John Benjamins, Chapt. 5, pp. 67–86.

Glaser, B. G. and A. L. Strauss: 1968, *The Discovery of Grounded Theory*. Weidenfeld and Nicolson.

Groves, D.: 2008, 'Bringing humans into the loop: localization with machine translation at Traslan'. In: *Proceedings of the Conference of the American Machine Translation Association (AMTA)*. Waikiki, Hawaii, pp. 11–22.

Guerberof, A.: 2009, 'Productivity and quality in MT post-editing'. In: *MT Summit 2009 Workshop: Beyond Translation Memories: New Tools for Translators*.

He, Y., Y. Ma, J. Roturier, A. Way, and J. van Genabith: 2010, 'Improving the Post-Editing Experience using Translation Recommendation: A User Study'. In: *Proceedings of the 9th Conference of the Association for Machine Translation in the Americas (AMTA 2010)*. forthcoming.

Hollan, J., E. Hutchins, and D. Kirsh: 2000, 'Distributed cognition: toward a new foundation for human-computer interaction research'. *ACM Transactions on Computer-Human Interaction* **7**(2), 174–196.

Jensen, A. and A. Jakobsen: 2000, 'Translating under time pressure: an empirical investigation of problem-solving activity and translation strategies by non-professional and professional translators'. In: A. Chesterman, N. G. S. Salvador, and Y. Gambier (eds.): *Translation in context*. Amsterdam & Philadelphia: John Benjamins, pp. 105–116.

Jurafsky, D. and J. H. Martin: 2008, *Speech and Language Processing*. Prentice Hall.

Karamanis, N., S. Luz, and G. Doherty: 2010, 'Translation practice in the workplace and Machine Translation'. In: *Proceedings of the 14th Annual Conference of the European Association for Machine Translation (EAMT)*. Saint-Raphaël, France, pp. 245–252.

Kay, M.: 1998, 'The Proper Place of Men and Machines inLanguage Translation'. *Machine Translation* **12**, 3–23.

Koehn, P.: 2009, 'A process study of computer-aided translation'. *Machine Translation* **23**, 241–263.

Lagoudaki, E.: 2006, 'Translation Memories Survey 2006: Users' perceptions around TM use'. In: *Proceedings of the ASLIB International Conference Translating & the Computer*, Vol. 28. pp. 1–29.
Macklovitch, E.: 2006, 'TransType2: The last word'. In: *Proceedings of the 5th International Conference on Languages Resources and Evaluation (LREC 06)*. Genoa, Italy, pp. 167–172.
Norman, D. A.: 2002, *The design of everyday things*. Basic Books.
O'Brien, S.: 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: *Proceedings of the 12th Annual Conference of the European Association for Machine Translation (EAMT)*. Hamburg, Germany, pp. 150–159.
Plitt, M. and F. Masselot: 2010, 'A Productivity Test of Statistical Machine Translation Post-Editing in a Typical Localisation Context'. *The Prague Bulletin of Mathematical Linguistics* **93**, 7–16.
Przybocki, M., K. Peterson, S. Bronsart, and G. Sanders: 2009, 'The NIST 2008 Metrics for machine translation challengeoverview, methodology, metrics, and results'. *Machine Translation* **23**, 71–103.
Randall, D., R. Harper, and M. Rouncefield: 2007, *Fieldwork for design: theory and practice*. New York: Springer.
Roturier, J.: 2009, 'Deploying novel MT technology to raise the bar for quality: a review of key advantages and challenges'. In: *MT Summit 2009*. pp. 1–8.
Sharp, H., Y. Rogers, and J. Preece: 2007, *Interaction Design: Beyond Human Computer Interaction*. Chichester: John Wiley & Sons, 2nd edition.
Somers, H.: 2003, 'Translation Memory systems'. In: H. Somers (ed.): *Computers and Translation: A translator's guide*. Amsterdam & Philadelphia: John Benjamins, Chapt. 3, pp. 31–48.
van Genabith, J.: 2009, 'Next Generation Localisation'. *Localisation Focus* **8**, 4–10.
Viller, S. and I. Sommerville: 2000, 'Ethnographically informed analysis for software engineers'. *International Journal of Human-Computer Studies* **53**(1), 169 – 196.
Wilson, J., S. Straus, and B. McEvily: 2006, 'All in due time: The development of trust in computer-mediated and face-to-face teams'. *Organizational Behavior and Human Decision Processes* **99**(1), 16–33.
Wittner, J. and D. Goldschmidt: 2007, 'Technical Challenges and Localisation Tools'. In: *Localisation Guide 2007*. Multilingual Computing Inc, pp. 10–14.
Yanishevsky, A.: 2009, 'The Emerging Role of Machine Translation'. In: *Localisation Guide 2009*. Multilingual Computing Inc, pp. 12–13.