

COACH: Designing a new CAT Tool with Translator Interaction

Laura Bota*

Timisoara, Romania*

Christoph Schneider+

Lingo24

Auckland, New Zealand+

Andy Way°

Greenfield, UK°

{laura.bota, christoph.schneider, andy.way}@lingo24.com

Abstract

In this paper we describe COACH, a new Computer-Assisted Translation (CAT) tool that we have developed in close conjunction with two focus groups of experienced translators. We explain why we thought it was essential for end-users to be involved from the start, and the process by which their feedback was gathered and acted upon for the ultimate benefit of Lingo24, the translators and our clients. We list a number of key changes that were made as a direct consequence of involving end-users from the get-go. As a result, we believe that COACH, which is already deployed in a live environment, has the potential to be a game-changer in our industry.

1 Introduction

Given the challenges they face in their day-to-day work, most human translators today would acknowledge the critical role of technology in their workflow. However, it is fair to say that some of this technology is more highly regarded than others. For example, most translators are happy to use Translation Memory (TM) tools (Heyn, 1998), while Machine Translation (MT) has met with much less widespread acceptance to date.

One of the main problems raised by translators is that they feel that tools are being imposed on them in a prescriptive manner. To give an example, in Sect. 2 we describe how the preference of English-to-German translators to translate multiple English sentences into one German output sentence was thwarted for the sole reason that this was not encouraged by CAT tools.

With respect to MT, some practitioners (e.g. Way & Hearne, 2011) have advocated that in order for statistical MT (in particular) to advance beyond the current ceiling which appears to have been reached, MT developers will have to liaise closely with linguists if further gains are to be made. Indeed, Way & Hearne note that it is peculiar that some prominent academics in the field of MT have cast doubt on the role that translators play in the building

of today's state-of-the-art statistical engines, especially when you consider where the training material – a prerequisite for any corpus-based system – has come from.

We have recently developed a new CAT tool – COACH – that we believe will prove very popular with users. From the very beginning, COACH was designed with translators in mind. The two focus groups (firstly for European languages, and then for Asian and right-to-left languages) comprised experienced translators with whom we had a long-standing relationship, whose opinions we respected. We describe why it was important to involve translators in the design and testing of COACH, how they were involved in the process, and some of the improvements made to COACH as a direct consequence of their involvement.

This paper describes the design issues underpinning COACH from cradle to grave, with particular focus on how we engaged closely with translators to ensure that they were happy with what we were building. In encouraging them to provide us with critical feedback, they were developing a degree of ownership in COACH. As the eventual users of the tool, creating a sense of stakeholdership from the very beginning was critical in gaining user acceptance of the finished product. As the evidence provided in this paper attests, this is vital if a tool is to be seen as the technology of choice by translators; as well as providing feedback, it is essential for translators to see their ideas being acted upon in subsequent iterations of a tool. Indeed, even though COACH has already been launched, we continue to encourage feedback from our users so that it can be continually improved, to the benefit of users, our clients, and of course ourselves.

The remainder of the paper is structured as follows. In Sect. 2, we provide some discussion of translators' perceptions towards technology, and the extent to which we believe this to be justified. In Sect. 3, we rationalise the decisions behind the critical components of our tool, especially in the light of existing technology. In Sect. 4 and 5, we hone in on how we elicited feedback from our focus groups, and their reactions to different aspects of the technology. Based on this feedback, we describe in Sect. 6 some of the changes made, and how these

were received by the users. In Sect. 7, we demonstrate typical productivity gains in COACH, and present our conclusions in Sect. 8.

2 Translators and Current CAT Tools

Traditionally, there has been a fair degree of translator resistance to the uptake of new technology. To some extent, this has been justified, given (i) their lack of consultation in tool design, and (ii) the fact that the use of such tools has caused them to alter the way they work.¹

When CAT tools were first introduced, the advent of computer-based file formats meant a paradigm shift for professional translators; simply put, paper was replaced by an array of electronic file formats, with electric typewriters replaced by word processors. Translation needs have changed so much that working with traditional paper documents comprises just a fraction of today's overall market.

In light of this shift, there were two main features that made the use of CAT tools almost inevitable, namely (i) the ability to deal with file formats, and (ii) the functionality to allow the insertion of in-line formatting at the appropriate points in the target language.

For many translators, the kind of 'help' built into these early CAT tools turned more into a hindrance. Take, for example, English sentences, and compare them to their German equivalents, which are typically longer, given that phrases and clauses are often connected together using conjunctions. A translator who may previously have had a preference to combine two English sentences into one German string will have found that typical CAT tools did not encourage this; additionally, of course, translators lose leverage in their TMs, too.

This sort of thing had a real effect on the way translators operate, and on the decisions they made in carrying out that work. It also has an effect on the target cultures; one example is that bold font faces were not encouraged as emphasis in CJK languages given their negative impact on readability, but the use of CAT tools nevertheless spread this sub-optimal practice over the Internet.

For many translators, then, having to change the way they worked proved to be a restriction. Other issues included the cost of the tools (mainly the need to continuously update the packages), the requirement for new hardware, etc. Nonetheless,

¹ For reasons of brevity if nothing else, we omit here an in-depth discussion of translators' reaction to the use of MT. For a critique of how and why translators react negatively to MT, see Way (2013). In contrast, see Brasler and Zetzsche (2013) as a peerless example of how a positive view of MT can really help translators.

these issues were probably less of an inconvenience given the increasing acceptance among the community that the job of a freelance translator was a profession that necessitated the use of professional tools, which develop and evolve over time.²

Clearly then, it makes sense to involve translators – those people who are actually going to use the tool – in the development of new technology. This is the approach we took in the design of COACH. In Sect. 3, we focus on the main issues we were confronted with in building the tool, with later sections detailing the reaction to subsequent versions.

3 Issues in CAT Tool Design

The main focal points that we wanted to concentrate on were technical as well as customer- and user-oriented, including:

- bringing high volumes to the translator,
- allowing clients to be more flexible with their quality expectations,
- having more content translated at predictable quality, and
- creating a post-edited MT (PEMT) interface that allows more and differently educated users to approach this task, which recognises and utilises a user's strengths and weaknesses and supports their intuition.

3.1 What is wrong with other CAT tools?

Some of the main problems we identified with other CAT tools included:

- *Difficulties in instant access to QA checks:* our CAT tool has, we believe, a very appealing way of filtering Quality Assurance (QA) check results and working through them, allowing the supplier to provide qualitative feedback about the QA findings,
- *No live spellchecking,*
- *Little flexibility in terms of how the content could be translated:* in COACH, the translatable content can be filtered out based on various criteria that allow translators to only see repetitions, non-translatables, matches, etc.

² Of course, the definition of what it means for a translator to be 'technically advanced' has changed dramatically over time. For some recent discussion, see: http://www.linkedin.com/groupItem?view=&gid=148593&item=208824910&type=member&commentID=120201238&trk=hb_ntf_COMMENTED_ON_GROUP_DISCUSSION_YOU_COMMENTED_ON#commentID_12020123

- *No fully online tool permitting shared TMs:* admittedly, there are several other browser-based tools, but we argue that COACH is fully integrated into a cloud-based Translation Management System that connects clients with huge translation requirements with the ideal suppliers, and
- *Limited QA checks for terminology:* the terminology capabilities in COACH are very important in providing clients with mature and fully fledged translation technology, while at the same time providing the suppliers with the key information they need in order to be efficient, consistent and reliable.

3.2 Why seek translators' input?

The main reason why we needed to build COACH was because we wanted to connect the translators with high volumes of our clients' content in an efficient way. However, without our translators, we would not be able to deliver on any of our jobs, so we wanted to provide as much content-driven coaching to specialists as possible. While our clients and our suppliers are extremely important, all along we tried to keep the bigger picture in mind; to provide an analogy, we were thinking in a similar way to a manager of a football team, who can coach a group of professional players using their individual skills while adding his own personal vision in order to make the club more successful.

Usability plays an important part in each stage of the design process. Consequently, obtaining direct feedback from end-users has become the main means by which five quality components are measured, namely:³

1. *Learnability:* How easy is it for users to accomplish basic tasks the first time they encounter the tool?
2. *Efficiency:* Once users have learned how to use the tool, how quickly can they perform tasks?
3. *Memorability:* When users return to the tool after a period of not using it, how easily can they re-establish proficiency?
4. *Errors:* How many errors do users make, how severe are these errors, and how easily can they recover from them?
5. *Satisfaction:* How pleasant is it to use the tool?

Given the above, it was important for translators to be involved in COACH from the earliest stages of development (when both the original design and the functionality were admittedly rather basic) so as to be able to develop a tool that would meet their requirements and expectations.

4 Translators' Feedback Loop: Round 1

As mentioned, the translators' feedback has been one of the driving forces in monitoring the tool's usability in the design process. This consists of three main stages, namely:

- General feedback on usability and utility,
- Project-based feedback, and
- Ongoing feedback on new functionality.

From Lingo24's point of view, there was a genuine interest in the translators' feedback on the usability and utility of COACH, which was the main reason why they were involved from a very early stage of development, as soon as basic versions of files could be processed straightforwardly in our new tool. The feedback loop followed multiple distinct steps, which we outline in the following sections.

4.1 Allocation of internal resources

From a logistical point of view, internal resources were allocated to coordinate communication and gathering of feedback. A Linguist Communication Manager was assigned to directly handle the relationship with the testers.

4.2 Identification of focus group

The initial focus group was carefully selected, in close consultation with the COACH Product Manager, from a wide range of translators, based on a diverse set of criteria, including:

- *Age and cultural background:* It is relatively uncontroversial to state that younger people typically have a more open-minded view towards change, in particular to new technologies. However, COACH has ultimately been designed to be used by a wide variety of translators. Accordingly, one of the objectives of the test was to obtain feedback from end-users of different ages, countries and cultural backgrounds, in order to better understand their expectations.
- *Technical background – relevant experience in working with CAT tools:* COACH was meant to be easy-to-use by any translator, irrespective of their experience in working with other CAT tools. Nonetheless, having

feedback from translators with different technical backgrounds was intended to cover a wider range of features than a translator would be used to seeing in existing CAT tools.

- *Previous working relationship with the translators:* Quality and speed are two of the main areas that COACH was intended to improve, so the focus group comprised translators who had previously done an extensive amount of work for our company with consistently high-quality results.
- *PEMT Experience:* COACH was designed to be Lingo24's main environment for PEMT. All the features developed were grouped around quality, speed and usability. Accordingly, gathering feedback from translators with extensive experience in PEMT was crucial so as to better assess those features that particularly distinguish PEMT from standard human translation projects.
- *Language Combinations:* COACH was initially designed to provide support mainly for European languages, so the first focus group included translators who covered these language combinations. A wider range of languages was covered during a later feedback stage.

The more distinct the members of the group were, the more relevant and varied the feedback was expected to be. The purpose of such a diverse group was to obtain a broader, clearer image of what translators (of different ages and backgrounds) would expect to see in a professional translation environment.

Although diverse, the original testing group was a very compact team of 15 translators. A larger group at such an early stage might potentially have led to less focused and ultimately unactionable feedback.

4.3 Topics of focus

Apart from obvious factors such as translation skills and the translator's expertise, there is a much wider combination of quite distinct aspects that may influence a translator's performance and impact implicitly on the quality of the output, including:

- Technical skills (translators' experience with technical translation tools),
- Translation environment (quick, easy and flexible access to translatable content and additional information, quick and reliable decision-making process, etc.), and

- Human factors, such as fatigue, state of mind, eye-strain, etc.

Questions related to more obvious factors include:

- *Speed:*
 - What features would be needed to increase efficiency during the translation process to achieve faster results?
 - What would help speed up the decision-making process?
- *Quality:*
 - What features would improve the quality output, with limited manual checks?
 - What types of automated quality checks would help the translation process and the quality output?
- *Usability and Utility:*
 - How intuitive is the tool?
 - How can the user's experience be improved?
 - Does the tool cater to the user's requirements?

4.4 Feedback prioritisation and implementation

All feedback was gathered and centralised in an online document available to all translators surveyed, both for full transparency and efficiency reasons. Each comment was grouped around the focus topics above (speed, quality, usability and utility), in various categories (bug, enhancement and new feature), as in Table 1.

The feedback implementation was prioritised in two different stages with direct involvement from the focus group:

- Prioritisation of each requirement by the translators as follows: *urgent* ("cannot launch CAT tool without it") and *not urgent* ("would like to have at a later stage").
- Individual assessment of feedback by the Product Manager, based on:
 - the type of request – which bugs were most urgent,
 - the priority previously proposed by the translators (urgent, not urgent),
 - the number of repeated feedback items, and
 - the focus topic of each comment (quality of translation, speed, usability).

Name	Interface	Focus topic	Suggestion Description	Category			Priority	
	(Translation editor/Translator's job list)	(speed/quality/usability)		Bug	Enhancement	Feature	Urgent (cannot launch CAT tool without)	Not urgent (would like to have)

Table 1: Gathering Feedback from our Focus Groups of Translators

5 Translators' Feedback Loop: Round 2

Once the suggestions had been developed and implemented, another round of feedback was carried out using the same focus group. The purpose of the follow-up test was to measure how efficiently the suggestions had been implemented, and also to assess the extent to which further urgent changes were needed, prior to the release of the tool. We followed the same steps as in the previous section for the sake of consistency.

5.1 Project-based feedback

More translators were subsequently involved in project-based feedback sessions to assist with the development of particular features.

Once COACH was stable and able to offer proper support for European languages, the next development stage was focused on expanding support for Asian and right-to-left languages.

Given the different linguistic characteristics of these languages, a different range of translators was involved to test and provide feedback on language-specific aspects. This second focus group was selected based on the same criteria applied during the initial testing phase.

The purpose of this feedback loop was mainly to have native speakers assess the quality of the language-specific segmentation rules implemented, as well as the reliability of QA checks.

5.2 Ongoing feedback

The ongoing monitoring of the translators' feedback on our tool has become a main priority for Lingo24 after the tool's release. Accordingly, a dedicated blog has been created for translators to encourage open discussions and sharing of feedback between the end-users and the management team.

5.3 Translators' reaction

Being directly involved in the design process of our new CAT tool has given translators a strong sense of

participation in and appreciation of its capabilities. Moreover, COACH is not only beneficial for Lingo24, but also for our end-users, so having them contribute to the building of a translator-friendly working environment has proven crucial.

The translators' response has been very enthusiastic, leading ultimately to a 100% acceptance rate among the translators surveyed. There was also a mutual feeling of appreciation and teamwork among the translators when asked to take part in the testing stages, as some of the following testimonies demonstrate:

Very excited about the new COACH tool you have mentioned! I've always said that LINGO24 deserves its very own translation tool :) Sure, I will be honored to take part in the test process and will do my best to improve its Version 1! (Andrey P, Russian translator)

I am really pleased and honoured with your offer and would love to participate. (Rosa-Elena A, Spanish translator)

This all sounds really exciting and I'd love to take part in this phase! (Marion T, German translator)

Firstly, thank you very much for your e-mail and trust with this project. It is a privilege to work on something that would benefit both LINGO24 and translator colleagues. I would be glad to help you as much as I can to produce a reliable CAT tool to make things go faster with consistency and the highest quality possible. (Heba S, Arabic translator)

6 Changes to COACH

The feedback loop resulted in important changes that significantly improved the tool. In what follows we describe some of the changes that were proposed and/or refined after the initial testing stage, grouped around the focus areas.

6.1 Speed

- Automatic prioritisation of translators' tasks based on the type of task (translation, review etc.) and the number of words to be translated given the deadline,
- Measuring real versus expected progress, and overall time spent (cf. Fig 1),

Progress	Task/Batch Deadline	Overall time	Editing distance
<div style="width: 100%; height: 10px; background-color: green;"></div> 100%	2013-04-22 11:00:00	00:30:00	11.4%
<div style="width: 100%; height: 10px; background-color: green;"></div> 100%	2013-04-19 17:00:00	00:11:38	0.27%

230 words translated, 230 expected by now

Fig 1: Real vs. Expected Progress

- Automatic propagation of repetitions across files, upon approval of first occurrence,
- Selective propagation of changes made to identical segments across files (cf. Fig 2),



Fig 2: Repetition Propagation

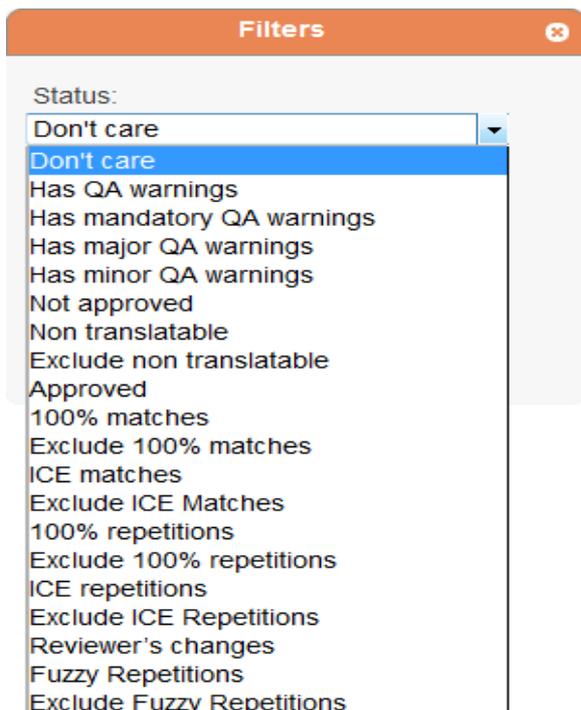


Fig 3: Available Filters in Coach

- Filter translatable content based on type of segments (cf. Fig 3),
- Multiple segment approval feature, etc.
- Enhanced internal matching: in-context repetitions & fuzzy repetitions (inc. display of TM matches for fuzzy repetitions),
- Filter on selection (Under Find & Replace): option to filter segments based on various key words (cf. Fig 4),

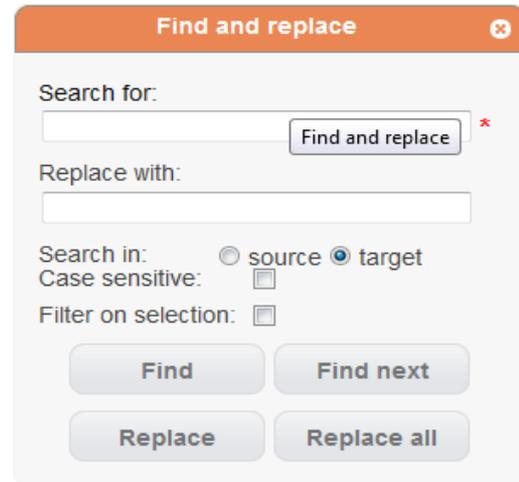


Fig 4: Filter Segments on Selection

- Track changes: allows editors to quickly see, approve or reject the changes made during the proofreading stage,
- My revisions and workflow revisions: allows reversion to older versions of a translation..

6.2 Quality

- Live highlighting of spelling mistakes and multiple spaces within the target segment,
- Terminology check to flag up inconsistent translations of the same term, forbid the use of 'rejected' terms,
- Customisable quality checks (optional – cf. Fig 5 – versus mandatory QA checks),
- Inconsistent translations QA check with possibility to view identical source segments in context,
- Ignore multiple false QA warnings simultaneously (exclude QA checks from current file),
- Add words to dictionary,
- Language-specific QA checks: Number formatting & Punctuation,

- Possibility of inserting TM match in target segment with or without tags.

Which QA Checks do you want to enable ?

<input type="checkbox"/>	Major QA Checks
<input checked="" type="checkbox"/>	Empty Translation
<input checked="" type="checkbox"/>	Terminology Check
<input checked="" type="checkbox"/>	Inconsistent TM Match
<input checked="" type="checkbox"/>	Spell Check
<input checked="" type="checkbox"/>	Inconsistent repetitions
<input type="checkbox"/>	Minor QA Checks
<input checked="" type="checkbox"/>	Consecutive Spaces
<input checked="" type="checkbox"/>	Consecutive Punctuation Marks
<input checked="" type="checkbox"/>	Different Brackets
<input checked="" type="checkbox"/>	End Punctuation
<input checked="" type="checkbox"/>	Entire Capitalisation
<input checked="" type="checkbox"/>	Inconsistent Spaces
<input checked="" type="checkbox"/>	Initial Capitalisation
<input checked="" type="checkbox"/>	Number Values
<input checked="" type="checkbox"/>	Non-matching Brackets
<input checked="" type="checkbox"/>	Non-matching Question Marks

✕ Close

Fig 5: Optional Quality Checks

6.3 Usability and Utility

- Ctrl+S shortcut for saving segments,
- Customisable font types and sizes,
- Customisable shortcuts (cf. Fig 6),

Select previous segment shortcut:	<input type="text" value="ctrl+up"/>	default: ctrl+up
Select next segment shortcut:	<input type="text" value="ctrl+down"/>	default: ctrl+down
Approve segment shortcut:	<input type="text" value="ctrl+s"/>	default: ctrl+s
Copy source shortcut:	<input type="text" value="shift+ctrl+c"/>	default: shift+ctrl+c
Insert previous tag into translation:	<input type="text" value="alt+p"/>	default: alt+p
Insert next tag into translation:	<input type="text" value="alt+n"/>	default: alt+n
Insert TM match in target:	<input type="text" value="shift+ctrl"/>	default: shift+ctrl+(1-5)

Fig 6: Customisable Shortcuts

- Detailed tool tips for quality warnings, etc.
- Statistics to reflect translators' overall performance (number of words/hour – cf.

Fig. 7 – edit-distance, time spent) and customised data based on the time period, type of task and workflow.



Fig. 7: Post-editor's Words per Hour

These were all signed off by the focus groups following further rounds of testing. It goes without saying that they were very happy with the finished tool, especially seeing as their feedback had been acted upon to their satisfaction.

6.4 Discussion

It is worth discussing how COACH differs from other tools which may have much of the same functionality. We contend that while other large translation enterprises may indeed have centralised their workflows, it very much remains the case that they continue to have to rely on desktop satellite software to enable suppliers to work inside their CAT tools.

By contrast, the key differentiator for COACH is that we make those important functions available in a centralised tool. For example, the QA checks are of the standard of third-party QA tools, made available in a cloud environment.

Indeed, the mere fact that such third-party QA tools exist demonstrates that QA is a weakness in many specialised tools. The core function for us is that the translator has access to a QA report that they can work through and that this information is also shared, depending on the workflow.

Furthermore, we bring together full workflows beyond the capacity of single users of desktop tools. Once again we deliver the core functions of common CAT tools in a collaborative environment.

Probably more important still is the fact that we actually connect content and supplier. Our goal is to unlock content that would not be translated unless COACH was available. The key functionality lies behind the scenes, and has to do with content splitting, more granular workflows and fine-tuned tasks, which means that COACH (with all its CAT functions) can be geared specifically to the task at hand. This would not normally happen in a project,

but we can facilitate this owing to our ability to assure the end-client of scalable workflows and processes.

In this context, obtaining positive feedback on COACH and the functions it offers is a fundamental cornerstone of the philosophy underpinning our service connecting the client and the supplier. Ultimately, we enable all parties – client, LSP and supplier – to get the most of content that would otherwise not be unlocked/translated.

7 Testing Productivity in COACH

As COACH has been designed to be able to support both standard projects and (in particular) PEMT jobs, we ran a series of tests to objectively measure the results, so as to be able to assess the productivity and quality gains brought about by the tool. Apart from the regular jobs that have been completed in COACH, we also organised a set of tests that were carried out in real-life conditions, with a range of different parameters, including:

- Different language combinations: EN-to-FR, DE-to-ES, EN-to-DE;
- Different domains: engineering, IT, marketing, business & finance;
- In-house customised MT engines;
- Translators with/ without PEMT experience.

When it came to measuring the results, we carried out both human and automatic evaluations. We used high-quality reference translations against which results were compared. We involved reliable translators to assess the output quality and provide detailed feedback based on an error typology model. The proofreaders assessed both the PEMT output and the reference translation, unaware of how the translations had been produced.

The automatic evaluation utilised well-known MT evaluation metrics such as BLEU, as well as using edit-distance versus overall time spent, all of which are readily available inside COACH.

Quality-wise, the output was similar to a human translation, with little variations, whereas in terms of efficiency, the increase in productivity ranged from 15.8—117%, assuming a baseline human translation rate of 2000 words/day, and a working day comprising 8 hours.

The number of words covered per hour ranged from 301 to 544, whereas the edit-distance was between 63.09% and 12.08%.

In sum, it is apparent that significant gains can be gained by using COACH and its integrated productivity tools.

8 Concluding Remarks

In this paper, we have described the building of a new CAT tool – COACH – which we believe has the capability of being a game-changer in the localisation and translation industries. It has the capacity to deliver cost-savings internally in our company, as well as making translators themselves more efficient, as demonstrated here. This benefits everyone: Lingo24, our translators, and ultimately our clients, who will benefit from quicker turnaround times as well as cheaper translation jobs; having complete control over both our CAT tool and our own MT technology affords us the possibility of coming up with creative pricing models as yet unseen in the industry.

We described why we considered a new CAT tool to be necessary. We discussed the reactions of typical translators to the introduction to new technology, and made the point that to obtain buy-in from our end-users it was essential to involve them from the outset in the planning and testing of the tool.

We outlined the two main phases of this testing regime, and provided testimonials from translators who had been involved in our two user groups. Finally, we listed some of the many changes that were brought about by involving translators at all stages in the design and testing phases.

As a result, we are convinced that the CAT tool that has resulted from this cooperation is far superior to what we might have achieved had we not involved our translators from the outset. Accordingly, they feel more favourably disposed both towards the company and COACH than might otherwise have been the case had they been excluded, and the tool imposed on them, as has been typical in the past.

References

- Brasler, C. and J. Zetzsche. 2013. Machine Translation: 50 Shades of Grey. *The ATA Chronicle*. http://www.atanet.org/chronicle/feature_article_february2013.php
- Heyn, M. 1998. Translation Memories – Insights & Prospects. In L. Bowker, M. Cronin, D. Kenny and J. Pearson (eds) *Unity in Diversity? Current Trends in Translation Studies*, Manchester: St. Jerome, pp. 123—136.
- Way, A. 2013. Who says Machine Translation will replace translators? <http://www.lingo24.com/blogs/company/who-says-machine-translation-will-replace-translators.html>
- Way, A. and M. Hearne. 2011. On the Role of Translations in State-of-the-Art Statistical Machine Translation. *Language and Linguistics Compass* 5:227—248.