

Juho Suokas & Kaisa Pukarinen & Stuart von Wolff & Kaisa Koskinen

Testing Testing: Putting Translation Usability to the Test

Abstract

User-centred translation (UCT) aims at providing translators with tools that can be used to enhance translations with the aid of usability research methods. This article reports an experimental case of UCT. The aim was to test the feasibility of usability testing for a team translation project involving web-based course material. Two different tests were designed: one followed a task-based model, where the participants were asked to perform a clearly defined task, while the other was an attempt to create an empirical test measuring the usability of a text intended to be read and understood by its users. The test participants were recruited from among potential representative users, who resemble the intended target audience. The results from the two test sessions indicate that the usability tests were indeed meaningful for the translation team, and while the methods themselves require refinement, UCT offers promising new avenues for research and practical applications.

1 Introduction

Translators have always been willing to adapt their translations to their readers' needs and to suit their predispositions. How this is accomplished in practice, however, has remained rather implicit, abstract, and left up to the translators' own understanding of what those needs and predispositions might be. By contrast, user-centred translation (UCT) aims at providing translators with tools that can be employed to enhance the usability of translation in more explicit ways (Suojanen/Koskinen/Tuominen 2015). At the core of this endeavour is an attempt to tap the resources of actual readers, involving them in the translation process.¹ In order to achieve this, various methods developed in user-centred design, usability testing and user experience design can be adapted for the translation process. UCT also implicates translation quality. In the translation industry, quality assessment efforts are typically concentrated at the end of the translation process, when potential shortcomings and mismatches are costly to fix and time is limited. In UCT, these end-of-process quality assessment methods are replaced by usability methods that are employed throughout all stages of the

¹ UCT is not to be confused with user-generated translation (UGT). In UGT, users themselves also translate (which could be described as an extreme case of UCT); in UCT, the professional translator is in charge of the translation process, but users are invited in as testers, commentators and contributors.

translation process, from planning and design through to the end product, incorporating client feedback. This iterative approach to quality assessment is also in line with more recent developments in the translation industry where increased attention is paid to controlling and improving process quality as well, not merely the quality of the end product (Drugan 2013). Quality is notoriously difficult to define and to measure. Changing the focus from end-product quality to usability allows for better alignment of expectations among stakeholders and provides more measurable targets to be achieved.

Suojanen, Koskinen and Tuominen (2015) promote user-centred translation from an academic position, and while its ideology has been well received among practising translators, questions have been raised about how well it truly functions in real life. Some elements included in UCT, such as mental models, designing personas, and heuristic expert analysis are fairly easy to introduce into a traditional translation project, but usability tests with actual users, however, may prove more difficult to inculcate. Nevertheless, usability testing, with its focus on interaction with actual users and its ethos of measurable effects, is somewhat alien to traditional translation quality work, and applying its methods to real translation cases is thus a valuable test case for the validity of UCT. In this article we report on one such experiment, which applied usability tests in the context of an actual translation project. The case is a team translation project commissioned by the University of Eastern Finland in the spring of 2013. In this case, two sets of usability tests were conducted during the translation process as a means of evaluating and improving the translation in progress.

2 Usability

Usability, as seen here, has its roots in Human-Computer Interaction (HCI) studies (Suojanen/Koskinen/Tuominen 2015: 13). Jacob Nielsen – a well-known HCI usability expert – defines usability as “a quality attribute that assesses how easy user interfaces are to use” and further specifies that “[t]he word ‘usability’ also refers to methods for improving ease-of-use during the design process” (Nielsen 2012: ¶3).

Thus, according to Nielsen, usability is not restricted to the assessment of certain qualities, rather it also includes the aspect of improving these qualities as well. The incorporation of usability into the translation process then could clearly benefit the overall quality of the final translation.

Nielsen presents five quality components that define usability: learnability (ease of use when the product is first encountered), efficiency (the users’ performance speed with the product), memorability (ease of use when returning to the product after a period of time), errors (the number, severity and ease of recovery from errors users make with the product) and satisfaction (how pleasant is the product for the user) (Nielsen 2012). Nielsen focuses mainly on internet and intranet user interface designs, but these can also easily be extended to texts, including translated texts.

Recently, interest in usability issues has emerged increasingly in Translation Studies. Although the term 'user' has not traditionally been in common currency in Translation Studies, the notion of end users surfaces with increased frequency in contemporary literature. For example, Drugan (2013) employs the terms 'user' and 'end user' throughout the text of her book on quality assurance in the translation industry. Similarly, 'user' and 'usability' are also integral to the works of Jody Byrne (2006, 2012), a pioneer of usability discussions in Translation Studies. Byrne's focus is technical translation and how usability strategies can be applied to improve their quality. For instance, Byrne defines usability of texts as follows:

When applied to texts usability measures the extent to which readers can read a text, understand its content and perform whatever task is required by the text quickly and accurately and the extent to which they find the experience difficult or easy.

(Byrne 2012: 201)

We wish to draw the reader's attention to three aspects of Byrne's definition of usability: (1) the readers/users are the focus of the text; (2) the readers use the text to perform a task; and (3) the experience is defined by the users themselves.

While Byrne's focus may well be technical translation, usability should be understood broadly as applicable to other forms of translation, as well. This broader application is examined by Suojanen, Koskinen and Tuominen (2015). They consider usability in translation on a larger scale. The authors adopt a functionalist perspective, emphasising that translation is instrumental, that it is always produced for a particular purpose. In addition to Byrne's aforementioned definition of usability, Suojanen, Koskinen and Tuominen (2015) see usability as user-specific and context-specific. They emphasise both social aspects, such as accessibility and social acceptability, and user experience aspects, such as personal intuition and affective factors.

3 Usability Research and Usability Testing

The concept of usability research must be clearly distinguished from that of usability testing. To elucidate this point, the following section highlights the differences between the two concepts – usability research in general and usability testing in particular.

3.1 Usability Research

The concept of usability may be subsumed under the broader concept of user experience, which includes everything the user undergoes when dealing with a company and its products and services (Nielsen 2012). HCI usability research is based on studying computer user interfaces; in HCI, usability is seen as a quality attribute of the interface. Usability research in its current form has evolved from the 1980s onwards, concurrent with the rise of personal computers. Since that time, usability has branched out from computers' graphical user interfaces to other forms of interfaces, including texts. Byrne (2006) discusses user interface and suggests that software user guides can be seen as an interface, if the interface is defined focusing on the use of the

product (or interface), and if the definition is kept flexible in terms of concrete physical characteristics. It is fairly straight-forward to link user guides and usability; as instructive texts they are typically used in clearly-defined contexts for clearly defined aims (Byrne 2006: 151-152). However, as mentioned above, Suojanen, Koskinen and Tuominen (2015) take this notion further and apply the notion of usability to other genres of translation as well. As text types and contexts of use become more varied, the questions of how to define usability and how to measure its success become more complex.

In order to create a product with high usability, the user should be integral to the design and development process. This approach to process, in which the user is taken into account at every stage, is known as user-centred design (UCD), and user-centred translation, as suggested by Suojanen, Koskinen and Tuominen (2015: 3-6), is based on this idea. In practice, user-centredness is achieved by employing different usability engineering methods throughout the translation process, beginning from design and modelling, continuing iteratively throughout the production and revision of the translation, right up to feedback and evaluation of the final product. In this study, we focus on evaluating methods that are used iteratively during the translation process.

Byrne (2006) divides empirical usability evaluation into two categories: methods which do include user testing and those which do not (Byrne 2006: 180). This roughly corresponds to Suojanen, Koskinen and Tuominen's (2015) distinction between heuristic models on the one hand, and empirical research on actual users on the other hand (Suojanen/Koskinen/Tuominen 2015: 93-94). Nielsen (2012) presents user testing as the most basic and useful method of studying usability, and similarly Byrne (2006: 180) suggests that those evaluation methods which involve actual users yield more relevant information. These user-based methods include various testing possibilities, including methods already in use in Translation Studies (even if less commonly used in the translation industry), such as eye-tracking, thinking aloud, and the use of interviews and questionnaires.

As an alternative to – or preferably in addition to – actual user testing, usability can be evaluated through heuristic evaluation, or expert evaluation, whereby the material is evaluated according to a list of heuristic principles. The evaluators can span the range from usability experts to novices, or experts with knowledge of both usability and the evaluated product (Nielsen 1997; Suojanen/Koskinen/Tuominen 2015: 80). Since this approach is closer to traditional revision and quality assessment procedures familiar to translators, it is thus easier to promote to the translation industry than usability tests involving actual users. While it is true that heuristic evaluation may be carried out summatively, in order to evaluate a finished product, this approach is not ideal, even if it is indeed how evaluations are often performed in the translation industry. Evaluation should more ideally be performed formatively, or iteratively, throughout the (design and translation) process, right from the beginning, since doing so would be more in keeping with UCD and UCT.

3.2 Usability Testing

The main aim of usability testing is to gather information on how actual users behave, and Byrne (2006: 181) suggests that user observation is one of the best ways to gather this data. Observation can be carried out, either directly or indirectly, in a specifically designed setting (laboratory) or in the users' natural environment (field study). Direct observation requires the users to perform the task while being scrutinised by one or more observers, who gather data from the test. However, the presence of one or more observers may also affect the users' performance and the data gathered in this method relies on the observer's unwavering attention. Indirect observation methods, on the other hand, do not involve an observer's presence while the users perform their task(s), rather their actions are recorded via different means, for instance video cameras, software logging or eye-tracking. (Byrne 2006: 181-184) Usability tests aim at providing realistic information on how actual users perform while completing actual tasks. One should keep in mind, however, that user tests are always created, artificial situations, and as such cannot be completely relied upon to expose all usability problems (Suojanen/Koskinen/Tuominen 2015: 97).

Usability tests are often complemented by survey methods, which include questionnaires, focus groups and different interviews (Nielsen 1997; Suojanen/Koskinen/Tuominen 2015: 102). While observation methods are extremely useful when evaluating Nielsen's usability quality components, (such as learnability, efficiency and errors), survey methods address primarily Nielsen's fifth component, i.e. satisfaction. Survey methods in particular can provide information concerning users expect of the product, which would not otherwise be evident in user testing (Nielsen 1997). Byrne (2006) suggests that objective information gathered by observational methods is not sufficient, since users' subjective opinions are a very important part of usability and can highlight "problems which may not have been anticipated by the designers or evaluators" (Byrne 2006: 187).

Prior to conducting a usability test, though, careful planning is required. Rubin and Chisnell (2008: 67) enumerate the following components, which are most commonly included in all user-based usability test plans:

- Purpose, goals and objectives of the test
- Research questions
- Participant characteristics
- Method (test design)
- Task list
- Test environment, equipment and logistics
- Test moderator role
- Data to be collected and evaluation measures
- Report contents and presentation

As is readily apparent from this list, the researcher's first step is to justify the usability test, or to justify usability testing in this instance, to decide whether or not it suits its purpose. The second component, research questions, is according to Rubin and Chisnell (2008: 69) the most important one, since research questions dictate how testing will be conducted, by defining the questions the test wishes to answer. Rubin and Chisnell maintain that this is equally important in experimental, less structured tests, since the test conductors need to be aware of what they wish to learn from the test (Rubin/Chisnell 2008: 69).

The third component, participant characteristics, defines the test group. The test group should be representative of the actual users of the product being tested. This, thus, requires either previous familiarity with the product's users or conducting specific user profiling in order to be able to select suitable test participants (Byrne 2006: 194-195). The number of participants is important, too, since too few participants can lead to non-representative, inaccurate test results. Rubin and Chisnell (2008: 72) suggest selecting 10–12 participants per situation when conducting a formal usability test. Less formal usability testing, however, can be conducted using only 3–5 representative participants. This is so, because research has found that even such a smaller group can reveal approximately 80 per cent of the test product's usability problems (Nielsen 2000; Rubin/Chisnell 2008: 72; Suojanen/Koskinen/Tuominen 2015: 95). None the less, Rubin and Chisnell (2008) do point out that the remaining 20 per cent could uncover important problems. Hence, a larger group of participants is suggested if the moderator of the test does not have sufficient experience in conducting usability testing, in order to minimise the risk of overlooking important problems. Moreover, a larger group provides lesser-experienced moderators ample opportunities to refine their skills (Rubin/Chisnell 2008: 72-73).

The fourth component, the method of the test, describes how the test will be conducted, and what to expect from the moment the participants arrive to the moment they leave. The task list describes what will occur during the test, and it should include tasks that correspond with the actual use of the product/text being tested. For the test to succeed, it needs to have clearly set targets for observation and measurement. Byrne (2006) discusses a number of general task lists, and also emphasises the need to select the most relevant of these for each case, as trying to deal with too many criteria at once may be overwhelming for the team (2–3 criteria are suggested as an appropriate number). Byrne lists learnability, retention of information over time, comprehensibility, accessibility of information, and speed of processing as the five observation targets most relevant for translation (Byrne 2006: 198-199). Byrne does not, however, supply a task list for measuring these in practice; this same oversight applies to Suojanen, Koskinen and Tuominen (2015). Although there is a general discussion of how to set up such tests, they provide no target focus guidelines.

Suojanen, Koskinen and Tuominen (2015: 96) point out that the language used to set these tasks should be unambiguous, direct, yet natural, nor should it manipulate (or prime) the user towards certain outcomes. Rubin and Chisnell (2008: 80) also

recommend defining beforehand what counts as successful completion of a task, since there could be differing views on this matter. The test environment should resemble or simulate the actual user environment; the equipment described here includes only that used by the test group, not that used by the researchers.

The seventh item on the list, the role of the moderator, must also be defined beforehand, since the moderators are the only ones who should intervene in a test situation (Rubin/Chisnell 2008: 87-88). The data being collected should be based on the second stage of planning, the research questions, noting that it is also constrained by the data gathering equipment. The data can include measured variables such as error rates, tracked eye movements and time taken to complete tasks, but may also include non-quantified factors such as data gathered by questionnaires or interviews. The final component, report contents and presentation, includes a summary of the test report and how the research results will be disseminated (Rubin/Chisnell 2008: 67-91; Suojanen/Koskinen/Tuominen 2015: 97-98)

4 The case: Usability of Translations

In this section, we describe in detail our case study conducting usability testing for a Finnish-to-English translation project. We utilise Rubin and Chisnell's (2008: 67) structure, as explained above in section 3.1. The practical logistics of the tests are outlined in the sections below.

4.1 Background

In autumn 2013, the University of Eastern Finland (UEF) inaugurated a new international Master's Degree Program in Linguistic Sciences, which includes a sub-programme in Translation Studies and Translation Technology. The basics of translation technology were already being taught to UEF students in Finnish through an online course. This same course now needed to be translated into English for the international programme. To achieve this goal, a working team of translators was recruited the preceding spring, its members consisting entirely of advanced translation students, all of whose A-language was Finnish and B-language English. The translation was requested to meet the needs of the target audience: international students, who would be mainly non-native speakers of English, who would be using English as a lingua franca (ELF). It was deemed highly likely that the true target audience students would speak a wealth of different first languages, thus unnecessary culture specificity was considered important. In other words, the focus of the translation project was on internationalising the locally produced course materials. The client also requested the use of UCT, but the final decision on how exactly to operationalise this was left up to the translation team.

The translation's language was to be kept in line with the target audience, and this is clearly in alignment with user centred translation, since the target audience was the

essential factor regarding the translation strategies employed. The team decided to avoid potentially problematic language features, such as overly complex sentence structures and difficult or vague wording; the aim was to keep the text relatively simple and easy to read, yet maintaining the informational value of the course material.

In co-operation with their mentors from the Department of Foreign Languages and Translation Studies, the team produced the internationalised version of the course material during spring 2013. The project was conducted in a course setting, but the aim was to simulate a real commission project, as well as to produce actual translations that would be used in the international MA programme. The project was entirely student-led.

At this juncture, it is worth mentioning that initially this undertaking did not start out as a research-oriented project, but rather its focus was practical applicability of usability methods; the client had commissioned the translation team to familiarise itself with, and utilise UCT methods, with which the team were not previously familiar. The methods were applied so as to enhance the translation to better suit the users; the question of testing the methods themselves as a research project – which this paper is evidence of – developed later on around the translation project. The potential added benefits of having a student-led UCT project were considered particularly heightened in light of Stelmach and von Wolff's (2010: 62) observations concerning the millennial generation who have led more "digitally networked lives", and are thus better suited to embrace and conduct UCT project research.

4.2 Purpose, Goals and the Research Question

One of the essential features of this translation project was to incorporate UCT methods into the process. This included organising and conducting formative and iterative usability tests on the internationalised text. The main goal of the study was to acquire information about how to improve the translation, so that it would better meet the needs of an international target audience. This information included identifying problems in the first draft of the translation, such as potential inconsistencies, translation errors, register changes, inappropriate word choices, complicated structures perceived as too difficult, as well as any other issues that the test group might identify. This information was used to improve the overall quality and comprehensibility of the translation by improving its usability.

The secondary goal of the project was to study UCT methods in practice, in an actual translation project. This involved the project group familiarising itself with usability and practising how to conduct a usability test, since nobody in the translation team had any previous experience with usability. While the student team did have some experience with quality-control principles employed during translation processes, usability and UCT were up until this point unfamiliar.

The usability tests were conducted approximately one month before the translation project specific deadline. This made for a tight schedule but provided the team with

ample time to revise the translation based on the usability test results (and user evaluation).

4.3 Participant Characteristics

The fourteen test participants were recruited from the department's *Culture Colloquium* course, concurrent with the translation project. The language of instruction for the *Culture Colloquium* was English and its course participants consisted mainly of international students. The participants were all volunteers and represented several different nationalities and languages. It was deemed purposeful to gather a test group of international students in particular, since the anticipated true target audience of the translation were international MA students from various language backgrounds. Since the MA course was not yet operational at the time of testing, the team had to rely on the anticipated backgrounds of the programme applicants, i.e. the true target audience of the translation. In this sense, the test group corresponded well to expectations, since the test group were mainly international students. Testing was conducted in two phases, both sessions taking place directly after the *Culture Colloquium* course meetings. Four participants (two male, two female) who participated in the first session did not participate in the second session.

Of the fourteen participants, eight were female and six were male. Their ages ranged from 20 to 34 and they represented 10 different first languages, including one self-identifying bilingual participant. All the participants were also asked about their knowledge of additional languages, however, English was not part of the demographic questionnaire, since English was a pre-requisite for all *Cultural Colloquium* participants. Table 1 presents the participants' first languages and knowledge of languages in addition to English, in the order in which this was reported.

Two points concerning the test group are noteworthy. Firstly, amongst the group were two students, whose first language was Finnish, and whose major area of study is English Language and Translation. Secondly, bearing in mind the internationalisation focus of the project, it is important to remark that not one of the participants was a native speaker of English, i.e. as mentioned earlier (4.1), the participants were using ELF to a large extent. As Rubin and Chisnell (2008: 115) suggest, an ideal group of participants would reflect the actual users. With regard to language, the test group closely represented the assumed target audience, since the target audience was perceived as international students whose first language would likely not be English. We also wish to acknowledge that although the assumed target audience of the translation are international MA students, not all of the test group participants were necessarily students at the MA level. However, the *Culture Colloquium* is classified as Two an advanced BA-level course, thus participation requires comparable knowledge of the kind expected of an MA student. This test group's international and multilingual background took precedence whilst considering test group participant suitability for usability testing. In addition, participation in usability testing was voluntary, so we felt we could not exclude participants who were not yet MA-level students.

First language(s)	Additional languages
Chinese (Cantonese + Mandarin)	French
Czech	French
Finnish	German, Japanese, Swedish
Finnish	German, Swedish
French	Spanish, Finnish
French	Spanish
French	Spanish, Catalan
German	Finnish
Greek	French, Finnish, Russian
Italian	French, Russian
Romanian	Spanish, French
Russian	French, German, Latvian
Spanish	French, Finnish, German, Italian, Chinese
Spanish	Finnish, German, Italian, Japanese

Table 1: Participant language information

Due to logistical concerns, we decided to conduct the tests using the group of participants as a whole, as opposed to conducting individual tests with 3–5 participants at a time, as suggested by Nielsen (2000). This had an effect on the methods chosen. For instance, going through individually recorded computer interactions of 14 participants was deemed an unnecessarily large amount of work. Instead, the methods were selected based on the criterion of how well they suited the large group setting, and in order to provide suitable data for translation improvement. During the tests, there were four moderators present to facilitate the large participant group. Since all moderators were present at the same time, their observations supported and complemented each other. Our large group selection is also consistent with Rubin and Chisnell’s (2008: 72) suggestion that, testing with more than 10 participants is advisable, if the moderators are inexperienced in usability testing – as was the case here.

4.4 Methods and Task Lists

The main goal of our usability test was to acquire information on how to improve the translation to better suit its target audience. Several potentially interesting methods for usability testing were considered by the translation team members, including ideas such as testing different language pairs, as well as recording the test subject’s on-screen actions through software. However, given the organisational and time constraints, the team decided it could not pursue these ideas at this stage.

The team decided to focus on usability testing in particular. Direct observation and focus group interviews were selected as the most appropriate information gathering methods for our circumstances. While admittedly not ideal in terms of collecting data

from individual participants, the direct observation method was, however, deemed the most appropriate solution to facilitate testing organisation more easily and in a less time-consuming manner, considering the large test group. In addition, the focus group also provided the opportunity to gather subjective, spontaneous and perhaps even unexpected data that could be used to improve the translation.

Usability testing was conducted upon two excerpts of the course material: a section introducing translation memory (TM) software at a general level, and an instructive text on how to begin a new project with the TM programme used on the course. The tests were conducted in two phases with a week's break between each session. Of the total fourteen participants, four students (two male and two female) who were present at the first session were absent from the second one.

The sessions began with the facilitator providing instructions and gathering consent forms from the participants before beginning the task. The rest of the moderator group worked as note-takers and observers. After performing each task, the participants were offered refreshments (juice and coffee) and small snacks during the focus group interview. The interviews were semi-structured, only the general outlines of topics and questions were planned ahead.

The first usability test combined observation and survey methods, aimed at gathering the users' subjective comments on the material and discovering usability issues on a textual level. As test material, a general text regarding translation memory software was selected. The chosen test material included the first six pages of the first draught of the English translation. The participants were asked to go through the text while writing down comments, and possibly noting any parts of the text that somehow caught their attention. These written comments were collected from them after the focus group interview. The participants were encouraged to consider the text from their own perspective as a student, on a course where the test material might be actually used. They had 22 minutes to go through as much as the text as they could while writing down their comments.

After going through the text, the facilitator led a focus group interview while three other members of the translation team served as observers and note-takers. The pre-determined discussion questions were:

- Are there linguistic features that make understanding the text difficult? (long sentences, difficult grammar, etc.)
- Is there unfamiliar vocabulary or unclear terminology?
- Are there strange or unfamiliar expressions?
- Are there spelling mistakes or other minor problems?

In addition to these, the interviewer followed up on various participant comments and in a few cases asked whether others in the focus group agreed or disagreed with individual comments. Nine out of the fourteen participants said they were already familiar with the subject of translation memory software. While not everyone in the focus group took part in the discussion – approximately one third of the group actively

contributed to the discussion – the individual comments and notes gathered from the participants provided much supportive data in addition to what was learnt and observed during the interview.

The second test session was held a week later in a classroom where the participants had access to translation memory software. The aim of the second test was to test the usability of the user instructions that the software had translated. The text included screenshots as well as written instructions. The students were asked to create a new project with the software using the translated instructions. They had half an hour to complete as much of the task as possible. No detailed task list was used. Based on what was available from existing literature (see above), task completion was taken to be the most significant indicator of usability of the instructions.

Participants were also surveyed regarding their IT backgrounds. All participants were first asked to rate their overall IT skills on a scale from 1 to 5 where 1 represented below average or insufficient skills and 5 represented very good skills. The scale of answers ranged from 2 to 4, so it can be concluded that the participants did not consider their IT proficiency entirely insufficient, nor did they consider themselves particularly expert, either. They were also asked whether or not they had previous experience using translation memory software. Only three participants reported such previous experience, which is interesting compared to the previous test where nine participants stated they were at least somewhat familiar with the subject. In addition, all three participants who reported having previous experience with translation software rated their IT skills differently.

The participants had half an hour to complete the task. Similarly to the first test, the facilitator led a focus group discussion after the task. This was meant to elicit participants' thoughts regarding TM software usage, and how helpful they deemed the translation for the process of working with the software.

4.5 Test Moderator Roles

Regarding the first test session, the methods included presenting the material, followed by a focus group interview. Of the four moderators present, one served as the facilitator and the other three as observers and note-takers. The facilitator led the focus group interview, while the observers took notes on the feedback, the test groups' reactions, etc. The interview was planned as semi-structured in order to give the facilitator the possibility to follow up on interesting or unexpected topics.

In the second session, the four moderator roles were assigned as in the first session: one served as facilitator, who instructed the test participants regarding the task ("using the translated instructions, create a translation memory project"), while the others served as observers. The observers' responsibility was to ensure that the test was performed according to the instructions, and also to observe if the participants seemed to have any technical or other difficulties (i.e. computer or software failures). It must be stressed that in the second session, the participants were to rely on the

translated instructions only, and the facilitator or observers were not to answer any questions from the students regarding the content of the instructions or how to operate the software.

5 Results

5.1 First Session

In the first session, the test group had 22 minutes to go through the text excerpt. Reading the whole text was not expected – only four participants reported having gone through the entire text. After going through the text, or reading as much as they could, the participants were offered refreshments from the nearby cafeteria. The facilitator then led a focus group interview session to evaluate the translation, based on the participants' comments. These comments were analysed and used to improve the tested translation and other translations made during the course. The major issues with the tested translation, as reported by the participants, are discussed below.

On the whole, the translation was considered genuinely understandable. However, the participants did identify several issues. First of all, most of the constructive criticism concerned sentence structure and/or wording. Occasionally the participants deemed the text difficult to read. This was, of course, at odds with the original goal, which was to formulate a clear, easily understandable and simply structured internationalised English version of the original. Issues in sentence structure included long sentences and the draft quality of the English text at some points. Shorter sentences were suggested, especially since the material was to be used in an online course environment.

One specific aspect that the participants criticized was the use of “he or she”, and addressing the reader directly as “you” when talking about a theoretical, generic person. Some participants suggested using only the generic “he”; others suggested using “they”, “one” or “society”. Since the translation team did not want to use the “generic he” pronoun repetitively, it was agreed that the team would rather use perceived gender-neutral expressions like “they” or “one”, or avoid using pronouns altogether, where possible. A significant factor in this particular instance is certainly the fact that Finnish, a non-Indo-European language and the language of the source text, does not make any gender distinction in its third-person pronoun “hän”. Hence, translating such texts from Finnish will often set this particular obstacle in front of the translator, depending of course on the target language. Regarding the comments on direct address, the team decided to use this only when the source text was addressing the reader in a specifically instructive manner.

It was also noted that certain words and expressions occurred too frequently in the text. These included such expressions as “etc.”, “advantage/disadvantage”, “this results” and “at the moment”, the last of which was considered to have been used especially at the beginning of sentences. In addition, unfamiliar terminology was

identified at some points, such as referring to “software programs” instead of just “programs” or “software”.

In addition, some comments were made on the formality and register of the text. The participants thought that certain word choices in the text conveyed either a positive or negative additional value, and seemed objectionable or judgemental. Below are a few such examples:

“The translation may *easily* follow the conventions of the source text.”
“[...] which differed *remarkably* from the earlier version”

The italicized words were considered to have positive connotations, which was deemed unfitting for learning material; the expectation seems to be that the language of learning material should remain neutral, which in the example is not the case. Objectionable word choices such as these were removed from the final translation.

The participants noted that occasionally the text’s register varied from formal to colloquial, which was considered undesirable, and even perhaps unacceptable in academic learning material. In this case, the expectations of the test audience seemed to be that the language in academic learning material should be formal or highly formal in register. Here are some examples the test group thought were too colloquial:

“Naturally, when choosing a suitable program, translators should also take into account the programs that their commissioners use, provided *that the translator knows what they are.*”
“*If it feels difficult to use a new program*, it could be useful to attend training sessions or webinars...”

In this case the changes in register may be a negative transfer from the source language. In Finnish, the conventions of learning material – including at the academic level – allow some colloquial expressions to be used to a certain degree. In this sense, Finnish academic language conventions might be somewhat less formal when compared with the academic conventions in the English language.

5.2 Second Session

For the second test session a computer lab was required, since the test users’ task was to create a translation project using TM software. As stated earlier, two male and two female participants were not present for this second session. In addition, one participant arrived late. Of the ten participants, eight managed to create a translation memory project successfully within the half-hour set time limit. Of the two who did not succeed in the task, one had arrived late, and thus did not have as much time as the other participants, while the other encountered various difficulties (including computer failure), but did none the less manage to proceed with the task, even if completion was not possible, due to running out of time.

All in all, the group’s performance was very good and the majority of the participants indeed succeeded in their task within the given time frame. Since none of the participants rated their IT skills as very poor or very good, it can be assumed that the participants see themselves as “average” or satisfactory IT users. Since the

majority of the participants managed to complete their task using the instructions within the allotted time, it can be assumed that the instructions themselves are indeed usable for a user with sufficient enough IT skills. Given that they have enough time, a user with sufficient IT skills should be able to operate the TM software using the instructions.

This time round there were significantly fewer comments from the participants. This was taken to be a sign that the participants most likely did not encounter many difficulties in interpreting the instructions and/or producing the required translation memory project. Participants did offer some comments regarding the text not making much sense without using the programme itself, but then again, user instructions are not intended to be read or used separately from within their bound context and topic. The screenshots in the text were also considered useful.

During this second usability test, the participants did not comment on any linguistic problems in the translation, with the exception of a few missing articles. Obviously, the participants' attention had not purposefully been drawn to textual and linguistic matters as it was during the first session. Nevertheless, one can also infer that the results of the initial test session had indeed helped to improve the overall translation, so that the text was fit for purpose. The participants also seemed to agree that the text seemed to fulfil its purpose. This is also supported by the fact that the majority of the test participants managed to complete the task relying on the translated instructions only.

6 Discussion

It would certainly appear that usability testing was justified and beneficial for the translation project. Usability research, being focused on tools, appliances, devices and programmes, offers a wealth of previous studies on how to use texts that are not instructive by nature. This being the case, the first test was an experiment in creating a usability test that would directly address textual and linguistic issues. Its aim was to help the translation team mainly in gathering information in order to improve the usability of the translations. The test participants identified several problems with the draught version, which in turn helped the translation team revise and improve the text. In other words, the test fulfilled its goal and may be considered a success.

According to Rubin and Chisnell (2008: 72), a suitable group size for formal usability testing is 10 to 12 participants. The test group in our case had a total of 14 participants. This large group size was evident in the somewhat time-consuming analysis stage of our research project, which entailed documenting all the received feedback and comments. In addition, as stated earlier, the methods selected had to be suitable for a group of this size. This led to ruling out certain possible methods, for example recording each participant's individual on-screen actions. Then again, a smaller group might have worked as well, since many participants represented similar backgrounds and might have thus provided similar feedback. Examples of common feedback from participants with similar (linguistic) backgrounds include comments proffered regarding the usage of the "generic he" pronoun and the academic register of the text.

A number of participants pointed out that the register used in the translation was occasionally perhaps too casual or colloquial for their taste. We reflected on this observation, seeing that in Finland, the student-teacher relationship is often quite informal. Teachers and students talk to each other casually, using informal speech and using colloquial language, and first names as default, something which often confuses foreign students and academics, who might be used to more formal and perhaps hierarchical higher education learning systems and environments. This casual approach is apparently present in the Finnish source material, where it was consequently transferred to the internationalised version of the text, as well, since several participants noticed the less formal register employed throughout much of the first draught. This was certainly an aspect of the translation the translation team students had not taken into consideration when producing their first draught. One remarkable consideration might be whether preservation of less formal language and register in the translation of learning materials could actually assist international students in acclimatising to Finland and the Finnish custom of fairly casual student-teacher relationships. Still, language deemed too informal by such users could be seen as a sign of translator incompetence – or it could be considered impolite, unfitting or otherwise unsuitable for academic learning materials, and even a hindrance to learning. Consequently, the team decided in the end to employ a more formal register throughout the translation to better suit the customs and likely expectations of the intended international audience.

The focus group interviews worked well, especially during the first phase of usability testing. The participants provided abundant feedback, which was at times quite blunt. Thus, these sessions' goals were met, at least in the sense of obtaining constructive criticism of the translation. In this case, the participants were all language students and as such are often used to verbalising their thoughts on language-related issues. When considering other target audiences, using a similar test might not prove to be as useful. However, since the target audience and participant group both consist of language students, the methods were appropriately selected and well suited for their purpose.

In hindsight, Nielsen's (2000) suggestion of using various smaller test groups (3–5 people) would have provided more opportunities to apply observation methods. Now, one might ask whether the direct observation of the large group was the most suitable approach to adopt. This is an interesting result, as it indicates that traditionally applied usability metrics, such as time for task completion, may not yield significant input for the purposes of translation. However, it may also be the case that the translators conducting the usability tests were more attuned to the linguistic issues that came up during the first test, but were less able to observe and measure relevant aspects during the second test. Attentive observation is a skill that one must develop, and there are as of yet no methodologically tested and validated task lists to be employed in these contexts, and this combination of lack of observation experience and lack of reliable guidelines both contributed to lesser relative success in obtaining usable information from the second test.

Another important, complex point of consideration on a more abstract level has to do with testing the application of an instructional text by allowing informants to carry out an assigned task, as in our second usability test: How can any researcher ever know the degree to which the software is “user-friendly” or even intuitively understandable, or whether at issue is the comprehensibility of the text? Paradoxically, this is an inherent problem with usability testing involving an object and a document. It is, thus, worth considering whether the users would have been able to operate the TM programme without the help of its documentation. It is impossible to know with absolute certainty in this type of evaluation whether the documentation’s usability is the reason behind the users’ performance level, or whether performance could be attributed to the software programme itself, if it can be seen as easily understandable. Similar observations and examinations of usability testing involving software user documentation have been made by others, significantly by Salmi (2003). In her dissertation, Salmi applied usability testing to studying problems users reported when employing user documentation for word-processing software in different languages. While her study would suggest that many of the problems were actually related to the users’ IT skills and background knowledge concerning related software, a similar problem arose with how users actually read the documentation whilst using any given software, or whether they in fact read it at all (Salmi 2003: 208). In the Salmi test, users were not explicitly told to read the instructions whilst operating the software, unlike our test, at the start of which users were instructed to use the software based on the given material. None the less, considering the fact that the evaluated text was meant to be used as material in an independent, online learning situation, one could rightly assume that most of the actual test users did in fact read the instructions – at least to some extent – whilst using the software.

Overall, usability testing was considered an efficient approach to gaining information about the participants’ reactions to the translation and its use. The test participants were all international students and thus corresponded well linguistically to the anticipated true target audience of the internationalised English-language learning material. Thus, the tests provided the team with valuable information, especially with regard to text’s comprehensibility in terms of style, structure and terminology, which in turn allowed the translators to enhance the final translations. Also, the participants performed well on the second phase of the testing (creating a new translation memory project), which suggests that the translation was understandable and usable. In addition, the subjective feedback received during the first phase had an improving effect on the second test translation, since hardly any issues at all came up during the second phase.

7 Conclusions

In this article we report on an experimental case of applying the principles of user-centred translation to a real-life translation project. Its aim was to test the usefulness of

usability testing in a translation project, and for this purpose we designed two different tests: one followed a task-based model familiar from usability research of user instructions (the participants were asked to perform a clearly defined task) and the other was an attempt at creating an empirical test measuring the usability of a text that was meant to be read and understood by its users. The results from the two test sessions indicate that the usability tests were indeed meaningful. The first, textual test provided a wealth of feedback, and the second test allows us to infer that the overall quality of the translation had indeed improved due to the feedback received at the first test session, since there were significantly fewer participant comments (generally, as well as concerning the language of the translation) at the second test session. We can thus conclude that in the translation project in question, the use of usability testing was justified and provided positive results, and the case study thus supports further research on usability methods and, in particular, it lends support to more experimentation with various methods involving actual users in test and authentic settings.

As will be the case with such experimental case research, there were also some lessons learnt, as the test team students were inexperienced and the methods untested. First of all, as previously acknowledged, the test group's size was larger than the often suggested 3–5 participants, which complicated the results analysis stage concerning both usability test sessions, and also led to the exclusion of some potential observation methods (such as eye-tracking or Jääskeläinen and Tirkkonen-Condit's (1991) think aloud protocols). In the future, a smaller group would be advisable for more time-efficient data analysis, which would also allow for the possibility to utilise different and perhaps more precise observation methods. In addition, observing a smaller group could expose more detailed insight into an individual test user's actions and perhaps encourage less forthcoming participants to join possible focus group discussion. Then again, when contrasted with traditional translation quality assessment (mainly carried out by the translator and/or the editing staff and/or commissioner), where the expectations are projected by the translator and/or the publication staff and/or other parties, usability testing shines the spotlight on the intended target audience during the actual translation process. Given that the test group corresponds to the intended target audience to a notable degree, usability testing allows for translation staff to test the expectations of their target audience in place of (or in addition to) projecting presumed expectations onto that group. Applying this type of usability testing, the translator can save time and effort, since the translator can edit the translation to suit the target audience's expectations during the actual translation process, rather than having to incorporate a separate phase at the end of the process. However, since usability testing can be a time-consuming process, it would likely be most beneficial if employed whilst translating large entities aimed at a particular target audience, such as user manuals, educational materials, instructions, and the like.

Since usability testing here was student-led, conducting the testing required the translation team to expand their roles and moderate discussions (as either a facilitator

or an observer) concerning the translation at hand. This left the translators in a position where their translation was being criticised for its problems only. This may not be the most encouraging approach, and for an inexperienced moderator in particular, it may feel awkward or annoying. The role of moderator required detaching oneself from the translation for the time being, since the aim of testing was to gain information about how to improve the translation's usability for that particular target audience, and not to discuss potential disputable translation solutions or errors. Having conceded this, testing as a whole was deemed useful, although for an inexperienced moderator, conducting testing might feel somewhat uncomfortable, since the test group's feedback may be very blunt, even to the point of tactless. As stated earlier, some of this feedback was delivered in a very blunt manner, which might have caused discomfort for some moderators.

One distinct topic for further research lies in the fact that the usability testing was conducted in English, for material that was also translated into English, and the moderators themselves were also the producers of that material. As Suojanen, Koskinen and Tuominen (2015) point out, the language used to conduct usability testing should not manipulate the participant(s) towards particular outcomes, and this may obviously be an issue in tests that focus on the usability of the linguistic material itself (Suojanen/Koskinen/Tuominen: 22). Moreover, it is important to acknowledge the ELF variable which was present during the testing and focus group discussions. Whether the use of the same language in the test situation as that of the material to be tested has an effect on the outcomes, is also an important point upon which to reflect. In this case, any effects will have been unintentional, and went unnoticed, and the test sessions were conducted with language students directly after a course conducted in English. Still, more in-depth analysis about using the language to be tested in conducting the testing is warranted.

Similarly, further research into the testing methods themselves is a topic that could yield significant benefits, bearing in mind that some of the methods employed here proved to either suit the purpose well or be somewhat problematic. Usability testing methods could be applied to various translation projects of different sizes and topics in order to refine suitability and methods for practical application.

Acknowledgements

The authors would like to thank Jenni Kelin, Tuomas Kuosmanen, Veera Saarelainen, Paula Sivonen and Virpi Vuorijärvi for their participation in the translating and user testing of the course materials.

References

- Byrne, Jody (2006): *Technical Translation: Usability Strategies for Translating Technical Documentation*. Dordrecht: Springer
- Byrne, Jody (2012): *Scientific and Technical Translation Explained: A Nuts and Bolts Guide for Beginners*. Manchester: St Jerome
- Drugan, Joanna (2013): *Quality in Professional Translation. Assessment and Improvement*. London: Bloomsbury
- Jääskeläinen, Riitta; Sonja Tirkkonen-Condit (1991): "Automatised Processes in Professional vs Non-professional Translation: A Think-aloud Protocol Study." Sonja Tirkkonen-Condit (ed.): *Empirical Research in Translation and Intercultural Studies*. Tübingen: Narr
- Nielsen, Jakob (1997): "The Use and Misuse of Focus Groups." – <http://www.nngroup.com/articles/focus-groups/> (28.10.2015)
- Nielsen, Jakob (2000): "Why You only Need to Test with 5 Users." – <http://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/> (10.02.2015)
- Nielsen, Jakob (2012): "Usability 101: Introduction to Usability." – <http://www.nngroup.com/articles/usability-101-introduction-to-usability/> (09.02.2015)
- Rubin, Jeffrey; Dana Chisnell (2008): *Handbook of Usability Testing. How to Plan, Design, and Conduct Effective Tests*. Indianapolis: Wiley Publishing
- Salmi, Leena (2003): *Documents multilingues pour logiciels et utilisabilité*. (Annales Universitatis Turkuensis B 269.) Doctoral Thesis Turku: University of Turku. Department of French Studies
- Stelmach, Bonnie; Stuart von Wolff (2010): "The Construction of Parental Roles at Eight Western Canadian Universities." *Journal of Educational Administration and Foundations* 21 [1]: 60-68
- Suojanen, Tytti; Kaisa Koskinen, Tiina Tuominen (2015): *User-centered Translation*. London: Routledge

trans-kom

ISSN 1867-4844

trans-kom ist eine wissenschaftliche Zeitschrift für Translation und Fachkommunikation.

trans-kom veröffentlicht Forschungsergebnisse und wissenschaftliche Diskussionsbeiträge zu Themen des Übersetzens und Dolmetschens, der Fachkommunikation, der Technikkommunikation, der Fachsprachen, der Terminologie und verwandter Gebiete.

Beiträge können in deutscher, englischer, französischer oder spanischer Sprache eingereicht werden. Sie müssen nach den Publikationsrichtlinien der Zeitschrift gestaltet sein. Diese Richtlinien können von der **trans-kom**-Website heruntergeladen werden. Alle Beiträge werden vor der Veröffentlichung anonym begutachtet.

trans-kom wird ausschließlich im Internet publiziert: <http://www.trans-kom.eu>

Redaktion

Leona Van Vaerenbergh
University of Antwerp
Arts and Philosophy
Applied Linguistics / Translation and Interpreting
Schilderstraat 41
B-2000 Antwerpen
Belgien
Leona.VanVaerenbergh@uantwerpen.be

Klaus Schubert
Universität Hildesheim
Institut für Übersetzungswissenschaft
und Fachkommunikation
Universitätsplatz 1
D-31141 Hildesheim
Deutschland
klaus.schubert@uni-hildesheim.de

Authors

Juho Suokas is a PhD student and part-time teacher at the University of Eastern Finland. His research is focused on practical application of usability methods in translation. While not busy playing guitar, his other interests include translation technology and translation of music-related topics.

E-mail: juho.a.suokas@gmail.com

Kaisa Pukarinen is a Master of Arts and an alumna of the University of Eastern Finland. The topic of her MA thesis (2014) is the translation of foul language and Anglicisms in a Finnish dubbed version of a children's animated cartoon. She is a practicing interpreter and translator in English and Finnish.

E-mail: kaisa.pukarinen@gmail.com

Stuart von Wolff, Junior Lecturer, University of Eastern Finland, has taught translation and interpreting at a number of universities, and is fluent in English, Finnish, French and German. His research interests are interdisciplinary and in addition to interpreting and translation studies, they embrace issues in pedagogy in higher education and philosophy.

E-mail: stuart.vonwolff@uef.fi

Kaisa Koskinen is professor of Translation Studies at the University of Eastern Finland in Joensuu. She is the author of *Translating Institutions. An Ethnographic Study of EU Translation* (St. Jerome 2008), co-author of *User-Centered Translation* (Routledge 2015) and co-editor of *Translators' Agency* (TUP 2010). Her current research interests include translation and affect, literary retranslations in Finland, multilingualism and translatorial action, and the city of Tampere as a translational space.

E-mail: kaisa.koskinen@uef.fi

Neu bei Frank & Timme

TRANSÜD. Arbeiten zur Theorie und Praxis des Übersetzens und Dolmetschens

Herausgegeben von
Prof. Dr. Klaus-Dieter Baumann,
Prof. Dr. Dr. h.c. Hartwig Kalverkämper,
Prof. Dr. Klaus Schubert

Susanne Hagemann: **Einführung in das trans-
lationswissenschaftliche Arbeiten.** Ein Lehr-
und Übungsbuch. ISBN 978-3-7329-0125-8.

Xenia Wenzel: **Die Übersetzbarkeit philoso-
phischer Diskurse.** Eine Übersetzungskritik an
den beiden englischen Übersetzungen von Hei-
deggers *Sein und Zeit*. ISBN 978-3-7329-0199-9.

Ralph Krüger: **The Interface between Scientific
and Technical Translation Studies and Cogni-
tive Linguistics.** ISBN 978-3-7329-0136-4.

FFF: Forum für Fachsprachen-Forschung

Herausgegeben von
Prof. Dr. Dr. h.c. Hartwig Kalverkämper

Raimund Drommel: **Sprachprofilung – Grund-
lagen und Fallanalysen zur Forensischen
Linguistik.** ISBN 978-3-7329-0158-6.

Peter Kastberg: **Technik der Kondensation
und der Expansion in der Technik.**
ISBN 978-3-7329-0221-7

Ost-West-Express. Kultur und Übersetzung

Herausgegeben von
Prof. Dr. Jekatherina Lebedewa und
Prof. Dr. Gabriela Lehmann-Carli

Jekatherina Lebedewa (Hg.), unter Mitarbeit
von Anja Holderbaum: **Tabu und Übersetzung.**
ISBN 978-3-7329-0034-3

TTT: Transkulturalität – Translation – Transfer

Herausgegeben von
Prof. Dr. Dörte Andres, Dr. Martina Behr,
Prof. Dr. Larisa Schippel,
Dr. Cornelia Zwischenberger

Cornelia Zwischenberger/Martina Behr (eds.):
**Interpreting Quality: A Look Around and
Ahead.** ISBN 978-3-7329-0191-3.

Vorankündigung:
Mehmet Tahir Öncü: **Basiswissen für Dol-
metscher – Deutschland und die Türkei.**
978-3-7329-0154-8.

