Is the concordance feature the new black? A workplace study of translators’ interaction with translation resources while post-editing TM and MT matches

Bundgaard, Kristine; Christensen, Tina Paulsen

Published in:
Journal of Specialised Translation

Publication date:
2019

Document Version
Publisher’s PDF, also known as Version of record

Link to publication from Aalborg University

Citation for published version (APA):
Is the concordance feature the new black? A workplace study of translators’ interaction with translation resources while post-editing TM and MT matches
Kristine Bundgaard, Aalborg University
Tina Paulsen Christensen, Aarhus University

ABSTRACT

This paper reports on a workplace study of industrial practices at a large Danish Language Service Provider and is concerned with how seven translators interact with a translation tool when post-editing translation proposals in an environment that combines translation memory (TM) and machine translation (MT). Recently conducted studies indicate that translation resources play a more important role in the translation process than has previously been acknowledged in process research, and those translators who apply computer-aided translation (CAT) tools find the concordance feature particularly useful. This paper investigates which types of translation resources the translators use when translating a technical and a marketing text from English into Danish at their usual workplace, and whether the concordance feature is the translators’ first choice of resource. Moreover, the paper analyses how the translators explain their interaction with the concordance feature retrospectively. Based on the results, the paper discusses whether the translators experience this type of interaction as an instance of cognitive friction. The study adopts an embedded mixed method design that combines an experimental and an ethnographically inspired approach. The results show that the concordance search is the translators’ preferred resource. Furthermore, the study indicates that, in some cases, concordance searches are instances of cognitive friction in the sense that they disrupt translators’ technology-aided cognitive processes. At the same time, the study shows that it is difficult to determine when they are disrupted.

KEYWORDS

Translator-computer interaction, post-editing, MT-assisted TM, translation resources, concordance feature, cognitive friction.

1. Introduction

Translation technology is now being used on a regular basis in the language industry. This caused O’Brien (2012) to characterize professional translation as a form of human-computer interaction (HCI). Unfortunately, the increasing professional use of translation technology has not been mirrored within translation studies (TS) research (Doherty 2016: 952; Munday 2009: 15; O’Hagan 2013). Although TS is beginning to focus on the use of translation tools (Christensen et al. 2017), “there is not an enormous amount of empirical data to speak of” (Pym 2011: 2). In particular, we need further research on how professional translators interact with translation tools, i.e. on translator-computer interaction (TCI), as such knowledge is needed if we are to prepare students for a future in which interaction with translation tools is an integral part of translation. Following the paradigm of situated, embodied cognition (Risku 2010), we should no longer see translation tools (for an overview of translation tools, see Alcina 2008) as isolated auxiliary artefacts but rather as part of a complex network. Along
the same lines, Ehrensberger-Dow (2014: 358) argues that we need knowledge about the activity of translation in situ, i.e. that we need to investigate this at the translators’ workplaces in order to ensure ecologically valid data. Several scholars have advocated combining the methods traditionally used in laboratory translation process research (TPR) with more qualitative, ethnographically-inspired methods in the workplace setting (Hubscher-Davidson 2011; Olohan 2011; Risku 2010; Risku et al. 2013; Ehrensberger-Dow and Massey 2014). According to Christensen (2011: 156), such workplace studies may investigate authentic translation assignments or be carried out as so-called “experimental field studies which adopt a combined approach to internal and external processes.” However, to our knowledge, the number of studies taking such a combined approach to the study of post-editing in practice is still very small. From a company’s perspective, a possible explanation is that such studies might negatively influence the company’s revenue. From a research perspective, it might be due to the methodological and ethical challenges related to the recording of real data from commercial companies and guaranteeing confidentiality and anonymity of research participants (Ehrensberger-Dow 2014: 378).

The study presented here, however, hopes to contribute to filling this gap by exploring the post-editing processes of seven professional translators in their usual workplaces at a large Danish Language Service Provider (LSP) when translating two authentic texts from English into Danish. At the time of data collection, the company was implementing machine translation (MT) into their workflows and the translation tool used by the LSP was a TM that incorporated MT. In this paper, we shall refer to this type of translation tool as MT-assisted TM translation. In the language industry, the use of MT-assisted TM tools has changed the translation process and industrial workflows as translators are no longer supposed to translate segments from scratch. Instead, their job is to post-edit pretranslated text that consists of 100% and fuzzy matches provided by the TM and machine translation output for all no matches (for an overview of match types in TM systems, see Bundgaard 2017a: 12-15). In this paper, we refer to matches retrieved from a TM as TM matches, whereas we refer to matches translated by the MT system as MT matches. In an MT-assisted TM environment, translators are assumed to switch between correcting fuzzy matches from the TM and correcting machine translation output (O’Brien and Moorkens 2014: 132). Silva (2014) refers to the former act as human post-editing, and the latter as post-editing MT. Traditionally, post-editing is defined as the task of editing, modifying and/or correcting pre-translated texts that have been processed by an MT system from a source language into a target language (Flanagan and Christensen 2014: 257) and the task of correcting TM matches has been referred to as e.g. revision or editing or simply translation. For simplicity reasons and since the boundary between TM and MT has become blurred with MT-assisted TM translation, we refer to the task of correcting both TM and MT matches as post-editing (O’Brien and Moorkens 2014: 131; Teixeira 2014: 184-185; O’Brien 2016; Christensen et al. 2017). This blurring between TM and MT is reinforced when MT
matches post-edited by translators are included in the TM and can then be retrieved as TM matches in new translation assignments, and also when TM data are used to train MT engines (Bundgaard 2017a: 15).

The present workplace study explores seven professional translators’ unrestricted use of resources in their usual workplace while post-editing TM and MT matches. Inspired by Raído’s (2011: 66) framework for the analysis of students’ web search behaviours when facing information needs relating to translation problem solving, our study focuses on the translation resources used during the search process to address a single or multiple information need(s). In this study, we define translation resources as all types of sources that translators consult to solve a translation problem. This covers functionalities integrated in the CAT tool, such as the use of the concordance feature, the termbase, and all other digital and printed resources such as Web searches, printed dictionaries or reference material.

As stated by Bundgaard (2017a), very few studies have addressed translators’ interaction with translation resources when using CAT tools, e.g. the concordance feature and web searches. Along the same lines, Valli (2014: 53) states that the “actual use translators make of external resources, i.e., what they systematically need to look up while translating, remains an area open for investigation.” The benefits of learning more about translators’ use of resources are illustrated by a recent experimental study (Hvelplund 2017) which found that professional translators spend around 25 per cent of their overall translation time on resource consultation when translating language-for-special-purposes texts without the use of a CAT tool.

Studies investigating which resources translators use in authentic CAT settings indicate that translators typically consult corpus-based resources, e.g. searchable TMs, websites, dictionaries, and termbases when they encounter a translation problem (LeBlanc 2013, 2017). Désilets et al. (2009) found that terminology problems are the most common type of translation problem and that translators use more bilingual resources than unilingual ones. Interestingly, they also found that in 35% of cases, translators conduct further searching after they have already found relevant information in one resource. The studies by Karamanis et al. (2010, 2011) found that the main tool used by the participating translators was the TM system, and that translators typically activate a concordance search when they encounter a translation problem or want to check for solutions. When they do not find a suitable solution, they typically turn to available reference materials, and if this is not fruitful, go online to search for possible solutions. Teixeira and O’Brien (2017: 88) found that the main task performed by professional translators in addition to translation ‘proper’ is the consultation of information resources. Their workplace study demonstrated that the only resource consulted by all translators participating in the experiment was the concordance feature. Ehrensberger-Dow (2014) and Ehrensberger-Dow and Massey (2014) found that professional translators typically switch between
the different available resources. The study by Bundgaard (2017a) indicates that when interacting with an MT-assisted TM tool, translators typically do not conduct concordance searches if they decide to accept a proposed match. When they decide to revise a match, translators typically use no resources in their post-editing of TM matches with high match values, whereas they do consult resources (in particular the concordance feature) when post-editing TM matches with low match values and MT matches. Hence, all these studies indicate that the concordance feature plays a central role in the TCI process.

1.1. The concordance feature

The concordance feature included in TM tools allows translators to search for and retrieve text fragments below sentence level (called sub-segment matching) from the TM database (O’Brien et al. 2010: 1). Translators activate the concordance feature using a shortcut and can then type or paste a word or a string of words in the appearing search window; they are then presented with a list of translation units from the TM (source and target segments in which the word or string of words searched for are highlighted). According to Valli (2014: 59), translators might conduct a spot search (one-time search event) or a search session (a repeated search for the same or changed text strings). If a suggestion seems useful to the translator, the translator can copy-paste it into the relevant target segment. Thus, the use of the concordance search reflects a “pull approach” by which the translator decides if and when to access the information (Warburton 2015: 656). As regards the potential cognitive impact of concordance searches, O’Brien et al. (2010) investigated how much time six professional translators spent on carrying out concordance searches, and how often they copy-pasted from the concordance window or reproduced the suggested content from the concordance window by typing it. The findings of the experimental study suggest that translators use the information provided by a concordance search and even give the sub-segment matches priority over the longer matches presented in the TM window. Interestingly, the use of the concordance feature influenced translators’ productivity negatively, but had a positive impact on translation quality.

1.2. Research questions

To learn more about translators’ interaction with translation resources in general and the concordance feature in particular, this paper explores the following research questions:

RQ1: Which types of resources do the professional translators use when post-editing TM and MT matches?
RQ2: How often is the concordance feature the translators’ first choice of resource?
RQ3: How do the translators explain their interaction with the concordance feature retrospectively?
Based on the results, the study discusses whether translators’ use of the concordance feature constitutes an instance of cognitive friction, drawing on the conceptual framework presented below.

2. Cognitive friction

Cognitive load theory is typically used as a line of inquiry in instructional psychology, but has recently been used in TPR, in particular in relation to the post-editing of MT (e.g. O’Brien 2006, 2008; Carl et al. 2015).

Cooper (2004: 19) defines cognitive friction as “the resistance encountered by a human intellect when it engages with a complex system of rules that change as the problem changes.” O’Brien et al. (2017) and Teixeira and O’Brien (2017) link cognitive friction to the concept of ‘flow’, which is used in psychology to refer to a state of being fully immersed in a task, causing this immersion to have an energising effect (Nakamura and Csikszentmihalyi 2002). O’Brien et al. (2017) assume that this flow is interrupted if humans encounter resistance when carrying out a task. Typically, what interrupts translators’ flow is some kind of encountered problem that they have to solve. The scholars distinguish between intrinsic and extraneous loads. The former is caused by the translation task itself, whereas the extraneous load is related to task-external aspects such as tool crashes, or results from the redundancy of information, e.g. incorrect metadata or too many or unusable translation proposals being presented on the screen. Based on this, they define cognitive friction as instances when, due to unhelpful or distracting CAT tool features, extraneous load is added to the intrinsic load of the translation task itself.

Applying Cooper’s (2004) concept of resistance, Olohan (2011) views CAT as interaction between a human agent (translator) and a non-human agent (the technology). She operationalises this type of interaction as a process in which translators need to carry out certain accommodating actions to enable the ongoing interaction between the tool and the translator in order to overcome potential resistance posed by the tool. Adopting this framework, Bundgaard et al. (2016) described how a professional translator repeatedly felt a need to accommodate the resistance posed by an MT-assisted TM tool. As regards the concordance feature, Valli considers concordance searches as manifestations of translation problems and argues that they interrupt “the translation workflow, in order to satisfy an information need” (Valli 2014: 55). The studies by Ehrensberger-Dow (2014) and Ehrensberger-Dow and Massey (2014) which investigated how translators and computers can reasonably be considered to impact on and adjust to one another in order to respond to disturbances and meet new demands adopted a similar approach. They found that translation tools might constrain the translation process and limit the translators’ autonomy since translators check even simple decisions against the contents in, for example, TMIs and style guides.
From the literature mentioned above, it is not clear whether the use of the concordance feature is necessarily to be considered as an instance of cognitive friction by means of extraneous load, even though Valli (2014) seems to believe so. For the purpose of this study, however, we assume that concordance searching is indicative of a point in the translation process where there might be an interruption or break in cognitive flow, where flow refers to a state of mind where the translator is being cognitively fully immersed in the translation process (Nakamura and Csikszentmihalyi 2002). Thus, we consider concordance searching as an instance of cognitive friction only if the translator’s cognitive processes are disrupted by the translation suggestions provided. Here, we operationalise disruption as instances where the translators want to deviate from or express doubt, mistrust or uncertainty in relation to the provided translation suggestion and due to this initiate a concordance search. Hence, concordance searching is not necessarily disruptive and may in fact be facilitative of translators’ cognitive processes.

3. Research design

The data used in this study were collected as part of a PhD dissertation (Bundgaard 2017a) which explored professional translators’ interaction with an MT-assisted TM translation tool. The study employed an embedded mixed methods research design consisting of a workplace study at a large Danish LSP, TextMinded Danmark A/S. The researcher stayed at the company for a month in order to learn about the company’s work practices and workflows and in order to conduct an experiment in which eight translators post-edited two authentic translation tasks.

In the experiment, the eight translators, who were all in-house translators and had between 6.5 and 23 years of experience in professional translation, each post-edited two texts from the Danish company Bang & Olufsen, a regular client of TextMinded. The source texts were 1) a Frequently Asked Questions (FAQ) text that related to a surround-sound speaker system (625 words in 76 segments), and 2) a newsletter about a music system (368 words in 25 segments). Both texts were translated from English into Danish, and both were authentic translation assignments. For the FAQ, the translators received a reference text with the fully formatted source text.

The translators (referred to as translators A-H) post-edited the texts at different times during a week. They carried out the tasks at their usual computers at their usual desks, and had unrestricted access to translation resources. The CAT tool used in the experiment was SDL Trados Studio 2011. Both source texts had been pretranslated using TM matches from TextMinded’s TM for Bang & Olufsen down to a match value of 70%, and segments with match values below 70% were translated using the MT engine SDL BeGlobal Enterprise, which had been trained with the TM data and a client-specific termbase. The termbase was also available to
translators during translation. In both texts, approximately half of the words were translated using the TM and the other half using the MT engine (in the FAQ, 47.4% of the words were translated by means of the TM and 52.6% by means of the MT engine. In the Newsletter, these numbers were 49.5% and 50.5%, respectively). During the experiment, data were collected using screen capture (BBFlashBack Express), keystroke logging (Inputlog) (Leijten and Van Waes 2013) and observation. The latter resulted in an observational protocol in which the researcher would note if the translator said anything to a colleague or used printed dictionaries, for instance. After the individual translation tasks had been completed, the researcher compared the translations to the pretranslated versions using software which highlighted the changes made by the translator. Next, each of the translators participated in a retrospective interview about their translation processes while watching the screen capture recording. Finally, they filled in a questionnaire about their background, experience and perceptions of the experiment (cf. Bundgaard et al. 2016, Bundgaard 2017a, Bundgaard 2017b). Accidentally, Translator F deleted his screen capture file after completing the translation tasks. Therefore, this paper draws on data from the remaining seven translators.

3.1. Methodology

This study investigates the seven translators’ use of resources, based on an analysis of their interaction with the CAT tool in the segments in the FAQ and the Newsletter which they individually chose to revise (46% and 90% of the included matches in the FAQ and the Newsletter, respectively).

In order to investigate RQ1 (cf. Section 1.1.), based on the screen capture recording, the process when post-editing each match in both texts was described qualitatively by the researcher in a step-by-step approach for each translator. The keystroke logging files were used for crosschecking purposes. Based on the observational protocols, the translators’ use of printed resources was analysed. From this process analysis we identified the types of resources used by the translators in a bottom-up fashion. The identified translation resources are shown in Figure 1.

<table>
<thead>
<tr>
<th>Concordance:</th>
<th>Search in the TM for one or more words using the concordance feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Termbase:</td>
<td>Search in the client-specific termbase provided to the translators in the MT-assisted TM tool</td>
</tr>
<tr>
<td>Google:</td>
<td>Search in a web browser using Google</td>
</tr>
<tr>
<td>Web page:</td>
<td>Visit to a web page appearing as the result of a Google search</td>
</tr>
<tr>
<td>Online dictionary:</td>
<td>Search in a dictionary on the Internet found by typing (a part of) the dictionary’s web address in the URL bar of the browser</td>
</tr>
</tbody>
</table>
To examine RQ2, for all 14 translation processes we identified the segments in which translators had consulted resources and the resource of their first choice.

We explored RQ3 by means of a descriptive analysis of translators’ verbalisations of their interaction with the concordance feature. First, we identified the segments in which the translators used the concordance feature. Next, if a translator commented on such a segment during the retrospective interview, this was analysed further in order to investigate how translators explain their interaction with the concordance feature. In the results section, we provide selected examples that illustrate this type of TCI and translators’ explanations. In doing so, we apply Valli’s (2014) distinction between spot searches and search sessions.

4. Results

This section presents and discusses the results of the analyses pertaining to the three research questions.

4.1. Types of translation resources

The results presented in this section address RQ1:

Which types of resources do the professional translators use when post-editing TM and MT matches?

<table>
<thead>
<tr>
<th>Resource type</th>
<th>Translator</th>
<th>Concordance search</th>
<th>Termbase search</th>
<th>Google search</th>
<th>Online dictionary</th>
<th>Local dictionary</th>
<th>Reference text</th>
<th>Resource consultations in total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>51</td>
<td>82.3%</td>
<td>0%</td>
<td>4%</td>
<td>6.5%</td>
<td>1.6%</td>
<td>2%</td>
<td>3.2%</td>
</tr>
<tr>
<td>B</td>
<td>8</td>
<td>72.7%</td>
<td>9.1%</td>
<td>9.1%</td>
<td>0%</td>
<td>0.0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>C</td>
<td>10</td>
<td>83.3%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>D</td>
<td>6</td>
<td>100%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>E</td>
<td>22</td>
<td>66.7%</td>
<td>0%</td>
<td>11%</td>
<td>33.3%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>G</td>
<td>7</td>
<td>63.6%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>4%</td>
</tr>
<tr>
<td>H</td>
<td>26</td>
<td>63.4%</td>
<td>0%</td>
<td>6%</td>
<td>14.6%</td>
<td>0%</td>
<td>4%</td>
<td>9.8%</td>
</tr>
<tr>
<td>Total/%</td>
<td>130</td>
<td>73.9%</td>
<td>1%</td>
<td>22%</td>
<td>12.5%</td>
<td>1%</td>
<td>6%</td>
<td>3.4%</td>
</tr>
</tbody>
</table>

Table 1. Resource consultation – FAQ.

For the FAQ and the Newsletter, respectively, Tables 1 and 2 show how often each translator consulted the different resource types as well as the share of the individual translator’s total number of resource consultations.
The last row shows the total number of consultations for each resource type together with the share of the total number of resource consultations.

<table>
<thead>
<tr>
<th>Resource type</th>
<th>Concordance search</th>
<th>Termbase search</th>
<th>Google search</th>
<th>Online dictionary</th>
<th>Local dictionary</th>
<th>Web page</th>
<th>Resource consultations in total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>35</td>
<td>92.1%</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>38</td>
</tr>
<tr>
<td>B</td>
<td>6</td>
<td>66.7%</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>C</td>
<td>12</td>
<td>100%</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td>D</td>
<td>3</td>
<td>37.5%</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>E</td>
<td>12</td>
<td>63.2%</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>19</td>
</tr>
<tr>
<td>F</td>
<td>3</td>
<td>60.0%</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>G</td>
<td>13</td>
<td>59.1%</td>
<td>1</td>
<td>4.5%</td>
<td>7</td>
<td>1</td>
<td>22</td>
</tr>
<tr>
<td><strong>Total/%</strong></td>
<td><strong>84</strong></td>
<td><strong>74.3%</strong></td>
<td><strong>1</strong></td>
<td><strong>0.9%</strong></td>
<td><strong>21</strong></td>
<td><strong>3</strong></td>
<td><strong>113</strong></td>
</tr>
</tbody>
</table>

Table 2. Resource consultation – Newsletter.

In the FAQ, the concordance search was the preferred resource for all translators. In fact, concordance searches accounted for between 63.4% and 100% of their individual resource consultations. Out of the total number of resource consultations, the concordance search accounted for 73.9%. The termbase was only consulted once by one translator. A plausible explanation for the very limited use of the client-specific termbase might be that terms contained in the termbase were highlighted in the source segment and the term entries were automatically displayed to the translators (by means of active terminology recognition). This means that the translators could see relevant termbase entries without actively searching in the termbase and, conversely, they knew that if a term was not highlighted, it was not included in the termbase (cf. Bundgaard 2017a: 157-158). Four translators conducted Google searches, thus making Google searches the second-most consulted resource (12.5%). Only one translator consulted an online dictionary (0.6%) and two translators consulted local dictionaries two and four times (3.4%), respectively. Five translators consulted the reference text (9.1%).

In the Newsletter, for all translators apart from translator D who conducted Google searches as often as concordance searches, the concordance feature was the preferred resource. Out of the total number of resource consultations, the concordance search accounted for 74.3%. For the remaining six translators, concordance searches constituted between 59.1% and 100% of their total individual resource consultations. The termbase was only consulted once by one translator (0.9%). Five translators conducted Google searches (18.6%), making Google searches the second-most consulted resource. Two of the translators visited Web pages appearing as the result of such Google searches (2.7%). Only one translator consulted an online dictionary (2.7%) and did so three times during the translation process, and one translator consulted a local dictionary once (0.9%). As translators were not provided with a reference text for this assignment, there are no data for consultation of this resource.
In other words, in both texts, concordance searches were the preferred resource, and Google searches were the second-most used resource. It is worth noting that most of these searches are what Hvelplund (2017: 81) refers to as shallow queries, i.e. informational queries during which the translator remains on the results page of a Google search and does not seek further information in the websites appearing as the result of a Google search (deep queries). The verbal reports indicate that translators sometimes conduct Google searches to check the number of hits on certain expressions. Dictionaries were used very seldom; printed dictionaries were not used at all, whereas online and local dictionaries were used a few times. Furthermore, our study – like the study of Teixeira and O’Brien (2017) - indicates that the consultation of resources might be a matter of individual style, as some translators consulted resources very frequently (e.g. Translator A carried out 62 resource consultations in the FAQ and 38 in the Newsletter) and some very infrequently (Translator D, for instance, carried out 6 and 8 resource consultations in the FAQ and Newsletter, respectively). However, all translators seem to consider the concordance feature the most useful resource - except for Translator D, who activated Google searches as often as concordance searches in the Newsletter. Due to space constraints, we will not elaborate any further on the individual resource consultation patterns.

In Table 3, for each type of resource consultation and for all translators, we show how many consultations were undertaken in TM and MT matches, respectively. We see that in both texts, in total, by far the most resource consultations are undertaken in MT matches (78.4% in the FAQ and 86.7% in the Newsletter). In comparison, in TM matches the translators used resources quite seldomly (21.6% in the FAQ and 13.3% in the Newsletter).

<table>
<thead>
<tr>
<th>FAQ</th>
<th>Concordance search</th>
<th>Term-base search</th>
<th>Google search</th>
<th>Online dictionary</th>
<th>Local dictionary</th>
<th>Reference text</th>
<th>Resource consultations in total</th>
</tr>
</thead>
<tbody>
<tr>
<td>TM</td>
<td>20</td>
<td>15.4%</td>
<td>1</td>
<td>100%</td>
<td>8</td>
<td>36.4%</td>
<td>0</td>
</tr>
<tr>
<td>MT</td>
<td>110</td>
<td>84.6%</td>
<td>0</td>
<td>0%</td>
<td>14</td>
<td>63.6%</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Newsletter</th>
<th>Concordance search</th>
<th>Term-base search</th>
<th>Google search</th>
<th>Online dictionary</th>
<th>Local dictionary</th>
<th>Reference text</th>
<th>Resource consultations in total</th>
</tr>
</thead>
<tbody>
<tr>
<td>TM</td>
<td>10</td>
<td>11.9%</td>
<td>0</td>
<td>0%</td>
<td>3</td>
<td>14.3%</td>
<td>1</td>
</tr>
<tr>
<td>MT</td>
<td>74</td>
<td>88.1%</td>
<td>1</td>
<td>100%</td>
<td>18</td>
<td>85.7%</td>
<td>2</td>
</tr>
</tbody>
</table>

**Table 3. Resource consultation in TM and MT matches.**

Furthermore, it is evident that concordance searches, Google searches and searches in online dictionaries were more frequent in post-editing MT matches than in TM matches in both texts. The reference text, which was only available for the FAQ, and Web pages were also primarily consulted in MT matches. By contrast, translators predominantly used local dictionaries when post-editing TM matches. The only resource that translators seem to use to the same extent in both match types was the termbase; however, here the amount of data is very small. It might not be surprising that
translators used the concordance feature more often when post-editing MT matches than when post-editing TM matches since the TM matches had been retrieved from the very memory on which the concordance feature draws. Based on the above, our data indicate that resources play a particularly vital role when post-editing MT matches.

4.2. The concordance search as the potential first choice of resource

The analysis presented in this section explored RQ2:

How often is the concordance feature the translators’ first choice of resource?

Table 4 shows the number of segments in the FAQ and Newsletter in which the individual translators consulted at least one resource and the number of segments in which translators decided to conduct a concordance search as their first choice of resource.

<table>
<thead>
<tr>
<th>FAQ</th>
<th>Translators</th>
<th>Segments with use of resources</th>
<th>Segments with concordance search as first choice of resource</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>28</td>
<td>24</td>
<td>85.7%</td>
</tr>
<tr>
<td>B</td>
<td>9</td>
<td>7</td>
<td>77.8%</td>
</tr>
<tr>
<td>C</td>
<td>10</td>
<td>9</td>
<td>90.0%</td>
</tr>
<tr>
<td>D</td>
<td>5</td>
<td>5</td>
<td>100.0%</td>
</tr>
<tr>
<td>E</td>
<td>18</td>
<td>16</td>
<td>88.9%</td>
</tr>
<tr>
<td>G</td>
<td>11</td>
<td>7</td>
<td>63.6%</td>
</tr>
<tr>
<td>H</td>
<td>19</td>
<td>14</td>
<td>73.7%</td>
</tr>
<tr>
<td>Total/%</td>
<td>100</td>
<td>82</td>
<td>82.0%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Newsletter</th>
<th>Translators</th>
<th>Segments with use of resources</th>
<th>Segments with concordance search as first choice of resource</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>14</td>
<td>14</td>
<td>100.0%</td>
</tr>
<tr>
<td>B</td>
<td>6</td>
<td>6</td>
<td>100.0%</td>
</tr>
<tr>
<td>C</td>
<td>7</td>
<td>7</td>
<td>100.0%</td>
</tr>
<tr>
<td>D</td>
<td>4</td>
<td>2</td>
<td>50.0%</td>
</tr>
<tr>
<td>E</td>
<td>10</td>
<td>10</td>
<td>100.0%</td>
</tr>
<tr>
<td>G</td>
<td>4</td>
<td>3</td>
<td>75.0%</td>
</tr>
<tr>
<td>H</td>
<td>9</td>
<td>5</td>
<td>55.6%</td>
</tr>
<tr>
<td>Total/%</td>
<td>54</td>
<td>47</td>
<td>87.0%</td>
</tr>
</tbody>
</table>

Table 4. Segments with concordance search as the first choice of resource.

In the segments in the FAQ where translators consulted resources, they generally chose the concordance feature first (in between 63.6% and 100% of the segments). In total, this was the case in 82% of the segments. In the Newsletter, the translators consulted the concordance feature first in between 50% and 100% of the segments, and in total this was the case in 87% of the segments. Thus, the analysis shows that when translators decided to consult a resource, they typically started with a concordance search.

It is also worth noting that on 15 out of the 18 occasions when the concordance search was not the first choice of resource in the FAQ, the translators first consulted the reference text (provided for the FAQ with the
fully formatted source text). In the remaining three cases, they conducted a Google search. In the Newsletter, the Google search was the first choice of resource in five out of the seven cases in which the concordance search was not used first, and the remaining two resource consultations comprised a termbase search and a local dictionary consultation.

4.3. Translators’ retrospective verbalisations and explanations

The analysis presented in this section examined RQ3:

*How do the translators retrospectively explain their interaction with the concordance feature?*

In 86 segments in the FAQ and 48 segments in the Newsletter, the translators conducted 130 and 84 concordance searches, respectively. In this section, we provide two examples from both texts that illustrate how translators interact with the concordance feature, giving a step-by-step description of their use of resources, and how they explain this interaction. We selected the examples from the 34 and 19 segments, respectively, which translators addressed in the retrospective interviews, i.e. when they explicitly or implicitly reflected on their use of the concordance feature. The examples illustrate spot searches and search sessions conducted in MT matches, since concordance searches were particularly frequent in MT matches. For each example, tables 5-8 show the segment number, match type, source text segment, provided match and resources used listed in chronological order as well as the translators’ verbalisations of their interaction with the CAT tool when post-editing the particular segment (our translation from Danish).

<table>
<thead>
<tr>
<th>Segment</th>
<th>Match type</th>
<th>Source text</th>
<th>Provided match</th>
<th>Chronological resource consultations</th>
<th>Verbalisation</th>
</tr>
</thead>
<tbody>
<tr>
<td>41</td>
<td>MT</td>
<td>To avoid noise on the LINE or AMP signal make sure that you have connected all sockets on BeoLab 14 to the corresponding sockets on the connected product.</td>
<td>For at undgå støj på Line eller AMP signal m ake sikker på, at du har tilsluttet alle stik på BeoLab 14 til de tilsvarende stik på det tilsluttede produkt.</td>
<td>Concordance: [make sure] [connected all sockets] [corresponding sockets]</td>
<td>&quot;Well, I don’t know, again I am a bit uncertain whether I should at all trust what that MT is writing. And I am also uncertain whether &quot;sockets” is just &quot;stik” [sockets] so therefore I looked that up too. So I probably have problems trusting it when I see that it's not coming from the TM. Well, apart from it phrasing it weirdly, you can live with that, but it's difficult for me to think &quot;well, that’s probably the right terminology&quot;. Then it just has to be double-checked.”</td>
</tr>
</tbody>
</table>

*Table 5. Translator A – FAQ– spot searches (example 1).*
In segment 41 (cf. Table 5), the only resource used by Translator A was the concordance feature. She searched for three strings of words, i.e. “make sure”, “connected all sockets” and “corresponding sockets”. We define these as three spot searches as they are three one-time search events on three different phrases from the source segment. In the interview, she commented specifically on the translation of the term “socket” which she was uncertain about and which was contained in the second and third searches. Also, she expressed a general uncertainty about whether to trust the output of the MT engine, stating that she felt a need to double-check it, implicitly saying that she would have trusted it had it come from the TM. Thus, the translator explained her interaction with the concordance feature in terms of accommodating a lack of trust in the MT output.

<table>
<thead>
<tr>
<th>Segment</th>
<th>Match type</th>
<th>Source text</th>
<th>Provided match</th>
<th>Chronological resource consultations</th>
<th>Verbalization</th>
</tr>
</thead>
<tbody>
<tr>
<td>23</td>
<td>MT</td>
<td>How should I set the bass position knob (FREE, WALL, CORNER) on BeoLab 14?</td>
<td>Hvordan skal jeg indstille basspositionsknappen (FRI, WALL, HJØRNE) på BeoLab 14?</td>
<td>Reference text: Concordance: [bass position knob] [position knob] [knob]</td>
<td>“that is where I encounter this “bass position knob” for the first time. And this I cannot find in the concordance and then I conduct some Google searches on what it might be called. And I look it up in dictionaries as well”</td>
</tr>
</tbody>
</table>

Table 6. Translator H – FAQ – search session (example 2).

In segment 23 (cf. Table 6), Translator H consulted the reference text and the concordance feature. First, she consulted the reference text in order to see the source segment in context. Afterwards, she activated the concordance feature to search for two strings of words (first “bass position knob” and then “position knob”) and one word (“knob”). We define this as a search session since the second and third searches are reduced versions of the prior searches. In both cases, the search strings are reduced by means of a so-called *left trim*, in which the left-most part is removed (Valli 2014). This indicates that when a concordance search does not provide a useful result, the translator searched for an increasingly more general term. In the interview, she stated that she could not find “bass position knob” in the concordance and therefore consulted Google and dictionaries. Interestingly, from the screen capture recording, it was evident that after the three concordance searches in segment 23, she entered segment 25, which also contained the term, and here she conducted another concordance search on “position knob” although she had already searched for this string of words in segment 23. As the concordance search again did not provide her with relevant results, she consulted two different bilingual dictionaries (in both cases, first searching on “position knob” and then on “knob”) and then conducted six Google searches on possible translations of the term:
“basknap” [bass knob]
“basknappen” [the bass knob]
“baspositionsknappen” [the bass position knob]
“beolab 14 basknap” [beolab 14 bass knob]
“beolab 14 baspositionsknapp” [beolab 14 bass position knob]
“beolab baspositionsknap” [beolab bass position knob]

Figure 2. Google searches relating to the term “bass position knob.”

She then returned to segment 23 and implemented her decision “basknappen”. The translator’s comment suggests that it was a matter of course for her to consult the concordance feature first, and only after recognising that this did not provide her with useful results, did she continue her search in other resources.

<table>
<thead>
<tr>
<th>Segment</th>
<th>Match type</th>
<th>Source text</th>
<th>Provided match</th>
<th>Chronological resource consultations</th>
<th>Verbalisation</th>
</tr>
</thead>
</table>
| 11      | MT         | Experiencing the brand’s acoustic innovations first-hand has never been more accessible or compelling. | Opleve den brand akustiske nyskabelser det har aldrig været nemmere lydfronten. | Concordance: [compelling]  
Google search: [compelling]  
Web page: [Compelling I Define Compelling at Dictionary.com]  
Concordance: [the brand] [innovations] | “there was that “compelling” which I was just looking up in the concordance to see what they had called it earlier, and then suddenly I became uncertain of what the word actually means in English. And then I just had to search on the Web, and then I found a definition and chose to use one of terms with which it had previously been translated, which I think fitted okay in the context if it had to be a bit interesting to read.”  
(…)  
“I remember I was thinking whether “brand” should be translated into “brand” or something else, and then I chose to write “Bang & Olufsens” instead. I just think that it sounded better and that is what they mean.”  
(…) |
"I was checking to see whether they had previously called it “nyskabelser” or if they had called it something else. I did find some examples with “nyskabelser.”

Table 7. Translator E – Newsletter – spot searches (example 3).

In segment 11 (cf. Table 7), the resources used by Translator E were the concordance feature, Google search and Web page. First, the translator conducted a concordance search on “compelling”. She then went on to Google and searched on “compelling”. One of the results of this search was a link to a definition of “compelling” at dictionary.com, which the translator clicked on.

This step can be seen as an example of a so-called deep query (Hvelplund 2017). Translator E then returned to the CAT tool and consulted the results of the previous concordance search. Next, she searched the concordance for “the brand” and finally, she ran a concordance search on "innovations". We define these concordance searches as three spot searches since they are searches on three different words and phrases from the source segment.

In the interview, the translator explained that she conducted a concordance search on “compelling” because she wanted to see which term it had previously been translated into. The word “compelling” had not been translated by the MT engine, and the translator’s explanation suggests that the concordance feature was a natural place to start when searching for a possible translation. As regards the search on “the brand”, she explained that she was uncertain as to whether to use the proposal “brand” (which can be used in Danish too) in the target text as well. Although one of the results of the concordance search was the word “mærket” [the mark], which, in our opinion, seems to be an acceptable translation of “the brand”, she chose to write the brand name instead (Bang & Olufsen) because “it sounded better”. Her explanation might indicate that since the suggested translation (“brand”) given in the match was not confirmed in the concordance search results, she chose a strategy of explicitation. With regard to the search on “innovations”, the translator explained that she wanted to check whether it had previously been translated into “nyskabelser” (which was the suggested translation given in the match) or whether a different term had been used. Her comment seems to indicate that she found “nyskabelser” to be an acceptable translation, but instead of relying on her own judgment, she wanted to check the suggested translation against the contents in the TM.
<table>
<thead>
<tr>
<th>Segment</th>
<th>Match type</th>
<th>Source text</th>
<th>Provided match</th>
<th>Chronological resource consultations</th>
<th>Verbalisation</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>MT</td>
<td>The all-steel industrial design is quite unique and the TV combines a lot of advanced technologies in a package that is more like a piece of furniture than consumer electronics.</td>
<td>Det unikke industrielle ståldesign og tv kombinerer en lang række avancerede teknologier i en samlet pakke, der er mere som et møbel end forbrugerel ektronik.</td>
<td>Concordance: [all-steel industrial design] [all-steel] [steel industrial] [consumer electronics]</td>
<td>“I am thinking about this “all steel design”. My reviewer might say that “it’s not necessary to write “ståldesign” [steel design] because you need to make clear that it’s all in steel” and I am thinking a bit about that, how I can include it and if it’s necessary” Researcher: “if it might just be “design” and then something afterwards with “stål” [steel] or?” “yes, “kun i stål” [only in steel] or “i 100 procent stål” [in 100 per cent steel] or something” (...) “and I am trying to make it a bit more attractive and then “consumer electronics” is translated into “forbrugerelektronik” and that is actually okay, but it’s just not brilliant in this sentence, so I just write “minder mere om et møbel end et fjernsyn” [resembles a piece of furniture more than a television]” Researcher: “because it’s a bit more elegant perhaps or?” “yes, that’s what it is and “forbrugerelektronik” might just as well be an immersion blender”</td>
</tr>
</tbody>
</table>

Table 8. Translator H – Newsletter – search session and spot search (example 4).

In segment 24 (cf. Table 8), Translator H conducted four concordance searches on “all-steel industrial design”, “all-steel”, “steel industrial” and “consumer electronics”. We define the first three searches as a search session since the first search is reduced in the second search (by means of a right trim), and the second search is both reduced and expanded in the third search (by means of a left trim and an addition at the end of the search string), whereas the fourth search is a spot search. In the interview, the
translator explained that she was wondering whether to include “stål” [steel] in a compound noun with “design” or whether to include it in a prepositional phrase after “design”. Although she did not express it explicitly, it seems that this is the reason for the search session on the different word strings that all included “steel”. In terms of the fourth concordance search on “consumer electronics”, the translator explained that she tried to make the text more “attractive” and therefore deviated from the translation given in the match, which was confirmed in the concordance search. Thus, her explanation suggests that, although she wanted to deviate from the translation given in the match, she felt a need to check the contents in the TM before doing so.

5. Limitations

This study was carried out as a relatively uncontrolled experiment at a Danish LSP, as we wanted to capture the translation processes under naturalistic conditions. Since the study was thus more exploratory than controlled, it is not possible to identify any general trends. The analyses only draw on data from 14 translation processes, i.e. two from each of seven translators. Also, the study exclusively investigates the translators’ use of resources in segments which they chose to revise and not the segments which they chose to accept or reject (cf. Bundgaard 2017a). We judged the revised segments to be the most relevant for our purposes; however, it would also be interesting to explore the translators’ use of resources in relation to the other choices. Moreover, as regards RQ2, we only explored the translators’ resource consultation in relation to the first translation problem in each segment, although translators might have faced more than one translation problem in a particular segment. Also, as regards RQ3, we focused on the problem-solving process and not on the quality of the translators’ final solutions.

6. Concluding discussion

In line with other studies (Karamanis et al. 2010, 2011; LeBlanc 2013, 2017; Teixeira and O’Brien 2017), the present study suggests that the concordance feature is the preferred resource for translators. As in Teixeira and O’Brien (2017: 88), all professional translators in our study used the concordance feature, and concordance searches accounted for the main part of translators’ resource consultations (approx. 74%). Furthermore, our study shows that the concordance feature is typically the translators’ first choice of resource. Typically, it was not until the translators recognised that the concordance feature was not helpful that they continued their searches using other resources. Interestingly, when seen in the context of Hvelplund’s (2017) study, our findings suggest that translators change their research pattern when a CAT tool (here MT-assisted TM) forms part of the complex translation process. Hvelplund found that in non-aided translation, bilingual dictionaries are translators’ preferred resource, representing
75.2% of all digital resource consultations, whereas in our study, only around 4% were dictionary consultations.

Our analysis of the translators’ verbalisation of their translation processes indicates that translators often consult the concordance feature in order to overcome uncertainty about a suggested translation, which may be caused by a lack of trust in MT output. The translators also wish to double-check the suggested translation and to check this against previous translations, even in cases when the translators want to deviate from the translation suggested in the match. Also, even when translators found a suggested translation to be acceptable, instead of relying on their own judgment, they sometimes checked it against the TM.

The analysis of the translators’ verbalisations provides very illustrative examples of how extraneous cognitive load caused by problematic MT output is sometimes added to the intrinsic load related to the pure transfer of the source message to the target text. For example, the extraneous load seemed to be present when translators were faced with MT output which they, for some reason, wanted to deviate from, and when translators experienced doubt, mistrust or uncertainty in relation to the MT output. Hence, in these cases, we consider the technology-aided cognitive processes of the translators investigated in this study as instances of cognitive friction, because the translators’ cognitive flow was disrupted by the concordance searches.

An interesting finding that emerges from our analysis is that translators seem to find the tool-internal concordance feature very helpful and facilitative of their cognitive processes, even in cases when it is assumed to interrupt their cognitive flow as defined by us. Actually, our data suggest that the translators view the concordance feature as an integrated part of the technology-aided translation process; in example 1, for instance, the translator seemed to use the concordance feature as a matter of course when she experienced uncertainty and felt a need to check a suggested MT match. This might indicate that she feels obliged to reuse matches stored in the TM, thus ascribing it a binding force. Furthermore, in examples 1 and 4, the concordance feature was the only resource used by the translators, which means that the translators never left the CAT tool. Interestingly, in examples 2 and 3, the translators did leave the CAT tool, but only when the concordance feature did not provide them with useful proposals or any proposals at all. When this happened, translators turned to resources such as the reference text, Web pages and Google searches. The fact that translators prefer the concordance search to other information resources might be a part of an accommodation strategy, where translators consciously or subconsciously tend to use resources that break the flow of the translation task the least. Another possible explanation could be that translators avoid switching attention between different windows on the screen (see also Teixeira and O’Brien 2017: 97) because this additional cognitive load would probably affect their productivity, which is important
in a business context. Also, an obvious explanation is that TMs contain translations produced by the translator him- or herself and/or other translators and thus in some sense “validated” contents.

The analysed verbalisations also point to the possibility that conducting a concordance search first may be an internalised action by translators when using a CAT tool, even when concordance searching seems to be unnecessary. By internalised we mean that the action is habitual by translators during post-editing, or as Risku (2014) puts it, constitutes an iterative operation pattern or a cognitive routine. For instance, example 3, in which the translator said that she used the concordance feature even though she found the suggested translation of a term to be acceptable, and example 4, in which the translator consulted the concordance feature although she wanted to deviate from the suggested translation, indicate that the translators seem to perform concordance searches as a default action, although the small number of examples do not allow us to make any definite claims about this. The examples are, however, interesting, since they suggest that the translators want to verify the suggested MT output against the TM, even when there is no perceived translation problem. Arguably, these examples are not necessarily instances of cognitive friction in the sense that translators are not experiencing a translation problem, but are simply indulging in a habit that is unnecessary. If translators are, however, subconsciously having doubts about the usefulness of the MT output, the concordance searches might be instances of cognitive friction. This illustrates that it is difficult to determine whether the concordance searches are instances of cognitive friction or not.

However, translators participating in the experiment seemed to consider the use of the concordance feature as facilitative of their cognitive processes. The assumption that, from a translation theoretical point of view, concordance searches initiated by translators looking for a solution to a translation problem, are said to represent an instance of cognitive friction, and at the same time are viewed as facilitating translators’ cognitive processes by the translators themselves, suggests that research should take a closer look at the context in which the concordance feature is consulted. This would allow us to learn more about cognitive friction as a theoretical concept as well as a practical phenomenon.

In conclusion, the study indicates that the concordance feature is the new black in CAT, as the study found that the concordance feature is the translators’ preferred and first choice of resource when post-editing TM and MT matches. This finding implies that when professional translators feel a need to consult information resources when working with an MT-assisted TM tool, they only very rarely leave the CAT tool. They only seem to do so when the tool-internal resources do not provide helpful suggestions. Thus, the tool-internal concordance feature seems to have substituted bilingual dictionaries as professional translators’ “go-to guy”. Further, the study has shown that concordance searches might be instances of cognitive friction.
Despite the important role that the concordance feature plays in professional translation workflows and translators’ technology-aided cognitive translation processes, much more research remains to be carried out to get a better understanding of professional translators’ search behaviour when repairing TM and MT matches in general and of their interaction with the concordance feature in particular. So far, the concordance feature has gone rather unnoticed in TPR. This study, however, has demonstrated that the feature deserves further attention to see how or whether it can be integrated in an even more helpful way into CAT tools, especially for MT post-editing, as MT-assisted TM tools, which are being used by LSPs more and more frequently (Christensen and Schjoldager 2016), should be designed in a way so that they do not become a potential source of cognitive friction for translators. It is our hope that this exploratory study will pave the way for future, more extensive research on translators’ use of the concordance feature.

Acknowledgement

The authors would like to thank Dr Sharon O’Brien for her invaluable input to this article.

References


• **Cooper, Alan** (2004). *The Inmates are Running the Asylum: Why Hi-Tech Products Drive us Crazy and How to Restore the Sanity*. Indianapolis: Sams Publishing.


• **Hubscher-Davidson, Severine** (2011). “A discussion of ethnographic research methods and their relevance for translation process research.” *Across Languages and Cultures* 12(1), 1–18.


**Biographies**

**Kristine Bundgaard**, PhD, is Assistant Professor at the Department of Culture and Global Studies at Aalborg University, Denmark, where she is a member of the research groups Language and Linguistics and Communication and Culture in Professional Contexts. She teaches translation and text production, and her primary research interests include translation technology, translation processes and business communication.

E-mail: kbundgaard@cgs.aau.dk

**Tina Paulsen Christensen**, PhD, is an Associate Professor at the Department of Communication and Culture, Aarhus University, Denmark, where she teaches translation tools, professional translation and
interpreting. With a background as state-authorised translator between Danish and German, her research interests cover computer-assisted translation, translation process research, legal translation and court interpreting.

E-mail: tpc@cc.au.dk