Raw machine translation use by patent professionals

A case of distributed cognition

Mary Nurminen
Tampere University

This article examines the use of raw, unedited machine-translated texts by patent professionals using the framework of distributed cognition. The goals of the study were to evaluate whether the concept of distributed cognition is a useful theoretical lens for examining and explaining raw MT reception, and to contribute to our knowledge of raw MT use through an analysis of a real-life use case. The study revealed that patent professionals often rely on a large network of artifacts and people to help them in the task of understanding raw MT, and therefore the concept of distributed cognition was applicable and useful. The study also contributed new knowledge to our overall understanding of the use of raw MT.

Keywords: raw machine translation; machine translation gisting; patent machine translation; machine translation users; distributed cognition; machine translation for assimilation

1. Introduction

The use of raw, unedited machine translation (MT) to acquire a basic understanding of a text is today a widespread phenomenon, with the number of users of Google Translate alone estimated in 2016 to be 500 million (Turovsky 2016). Raw MT is also being used in professional contexts, not only in professional translation but also in fields such as customer support and academia. As reported by Nurminen (2019), raw MT has been used on a broad scale in the field of intellectual property rights (IPR) for approximately a decade. Patent professionals, including patent counsels, attorneys, examiners and other professionals, use raw MT to read patent documents that are in languages they know only a little or not at all.

The case of raw MT use by patent professionals has gone largely unnoticed by researchers. In fact, our knowledge of the reception of raw MT in general is still lacking, and very few hypotheses or theories have been developed on the cognitive aspects of raw MT reception. There is a considerable body of research on cognition and translation (e.g., Shreve & Angelone 2010; Schwieter & Ferreira 2017), much of which might be applicable to the context of raw MT use, but few studies explore the application of that knowledge to raw MT cognition.

This article analyzes one case of MT gisting—the use of raw MT by patent professionals (definition in next paragraph)—through the lens of the theoretical framework of distributed cognition. My first goal in this article is to apply the principles of distributed cognition to the case study, evaluating whether it is a useful means for explaining raw MT reception. My second goal is to build on our understanding of the nature of MT gisting through this analysis of a real-life use case, so that we can begin to understand some of the conditions that are present in a functioning use case of MT gisting. In this inductive, exploratory study, the intent is not, and cannot be, to establish causality. Rather, the study aims at an initial identification of factors and conditions that appear to influence the use of raw MT, either positively or negatively. Those factors can then be further explored and tested in future research.

A few notes on terminology are in order. The first concerns languages. *L1* refers to the language a person has the best command of, which is often the language they use most often in their everyday lives. *L2* indicates a language a person has a very good command of and uses on a regular basis. *L3* applies to other languages, which might be languages a person has some competence in or languages they do not understand at all. The Scandinavian informants in this article represented three different L1s and all of them have English as an L2. They use it daily in their work and the raw MT texts they work with are translated from other languages into English.

The second note concerns the phenomenon in which a person reads raw, unedited MT output to gain at least a basic level of understanding of it. In this article, that phenomenon is referred to as *MT gisting*, while the text that is being read in MT gisting is termed either *raw* or *unedited* MT and the person reading raw MT for gisting purposes is referred to as a *raw MT user*. Finally, in this study, the term *patent professionals* refers to professionals who use their expertise in IPR processes to assist internal or external clients with their Intellectual Property Rights needs.

The next section reviews literature related to this study. Section 3 describes the methods used in the case study on patent professionals. Readers are not necessarily familiar with IPR work, so Section 4 introduces the environment in which patent professionals work. Section 5 discusses the results of the analysis, and conclusions are offered in Section 6.

2. Related work

Very little research has been done in the cognitive specifics of MT gisting. O'Brien (2017) devotes a chapter to MT and cognition in *The Handbook of Translation and Cognition*, but she comments frankly in the beginning that, since research on users of raw MT is in its early stages, it would receive little focus and the section would be devoted mostly to post-editing and evaluation processes. The one mention of raw MT users is in relation to evaluation methods, and the paragraph warns that the methods covered in the section:

[...] are taken, to some extent at least, as acting as a proxy for measuring the usability and acceptability of MT output by those for whom it is essentially produced: end users. However, these are not direct measures of interaction with MT and tell us very little about the cognitive processing effort that might be involved in reading, comprehending, and acting on a set of machine-translated task instructions.
(O'Brien 2017, 327–328)

However, a few studies on cognition and reception of various types of texts have included raw MT as one text type. O'Brien (2010) tested the effect of controlled language on readability for readers of texts in the original source language as well as in raw MT form in different languages. Other studies used eye tracking combined with other methods, such as post-task questionnaires, to investigate end users' evaluations of factors such as usability and acceptability in texts (Castilho et al 2014; Doherty & O'Brien 2009, 2014; Castilho & O'Brien 2017). These studies again tested different types of texts, including original texts, human translations, post-edited MT and unedited MT, and much of the focus of the studies was on comparing the results of those different types of texts, rather than on specifically exploring the cognitive aspects of raw MT reception.

2.1. Distributed cognition and translation

An important area of background research I relied on when writing this article dealt with situated and distributed cognition in the processes of translators. One might assume that it would make sense to use studies on areas closer to raw MT reception; for example, studies on cognition and reception of human translation. And, as O'Brien (2017) noted in the quote above, the cognitive situations of a translator or post-editor at work and a user of raw MT are different. However, it is in the area of situation cognition and translation that the greatest theoretical development has taken place and for that reason, material from that area was found to be the most helpful.

Muñoz (2010, 172) states that "cognition takes place in the context of inputs and outputs relevant to the task at hand." He applies this to the work of translators and the way they "act together to accomplish the complex task of creating a single, communicative product" (ibid.) Risku (2014, 335) discusses "situated, embodied and extended cognition" and how "cognitive processes are context-dependent, i.e., they are dependent on and partly constituted by the social

and physical environment in which they are carried out.” Muñoz (2017, 564) also illustrates how situation can affect meaning: “Meaning is encyclopedic, and it is a process, not a thing [...] Understanding is an activity that crucially depends on the environment—and also on experience—because environmental affordances foster and constrain meaning construal.” Although all of these descriptions are in articles about translators’ cognition while translating, they are fitting descriptions of the work in this use case of MT gisting.

The work in situated cognition in translation studies has also led to a new emphasis on qualitative research methods. Risku (2014, 335) argues that, while the controlled experiment research that is dominant in translation process research is necessary, “we also need other methodological paths of inquiry to model the cognitive processes in translation and to establish a deeper understanding of how translations are produced.” Ehrensberger-Dow (2019, 41) also emphasizes that study in this area relies on both experiments in laboratory settings and workplace observations.

2.2. Raw MT users

The number of studies published on raw MT users remains relatively small. One area of research consists of surveys and market studies, some of which focus on the use and users of a specific tool (Smith 2003; Yang & Lange 2003; Nuutila 2005; Burgett 2015; Nurminen & Papula 2018). A few analyze specific types of users, such as Gaspari’s (2007) survey of students and Henisz-Dostert’s (1979) survey of scientists. The latter of these differs from the others in that, although it was a survey, it was a qualitative one, conducted as a series of structured interviews. It also included a few questions on cognitive processes, such as mental corrections of awkward translations.

Another group of studies involve laboratory experiments in which volunteer participants, who are often not actual users of raw MT, are asked to evaluate specific aspects of raw MT use. Some of these were mentioned in the previous section. Others include experiments on aspects such as comprehensibility, confidence, and acceptability (Fuji et al 2001; Gaspari 2006; Bowker & Eghoetz 2007, to name a few).

A third area of research addresses MT for communication, which refers to raw MT used to enable communication between people who do not share a language. Many of these studies are experimental, but they are carried out in extended virtual environments that simulate real-life environments. Several studies involve working groups based on different sites that have different L1s. The groups are assigned a task to accomplish together, with all communication surrounding the task occurring in chat applications backed by MT. The communication transcripts from the chats are then analyzed. Often these experiments test the influence of individual factors on communication. For example, Gao et al (2013) tested the highlighting of keywords. Xu et al (2014) and Gao et al (2015) examined scenarios in which users were presented with two different MT outputs. Yasouka & Björn (2011) and Yamashita et al (2009) examined strategies to establish common ground among the participants.

Only a few use cases of professional groups using MT gisting have been studied. Although scientists were the first group studied, the research by Henisz-Dostert (1979) remains the only one of its kind. Some scholars have examined the use of raw MT in customer support (e.g., Stewart et al 2010; Burgett 2015). There is a growing body of research on the use of MT by academics (for example, Bowker & Eghoetz 2007; Jolley & Maimone 2015; Bowker & Buitrago 2019). A few projects have also studied the use of raw MT in health care (review in Liu & Watts 2019).

A large number of the aforementioned studies on MT gisting are quantitative and/or experimental. Risku’s (2014) call for more diverse methodologies in research could likewise be applied to raw MT reception, where our challenge is to model the cognitive processes so that we can acquire a deeper understanding of how raw MT users understand raw MT and what strategies they adopt when faced with challenges in understanding.

3. Methods

This article is based on a case study project on patent professionals' use of raw MT. The aim of the project (Nurminen 2019) was to explore the context in which raw MT is used by the group. During the data analysis phase, it became evident that distributed cognition might be an effective theoretical framework for explaining the phenomenon and that realization led this article. The study was conducted by adopting inductive, qualitative methods. Especially in such an under-researched area, a more nuanced analysis was needed that would not only contribute to an overall understanding of the phenomenon but would also help us identify the factors that could be tested in future controlled experiments.

The data for the study was gathered by conducting a series of semi-structured interviews between April 2018 and February 2019, either at informants' workplaces or through Skype audio calls, which were easier to arrange than video calls. Interviewing was chosen primarily because the time commitment it required from the highly busy informants was only 1.5 hours on average. Several sources helped me prepare interview questions. The patenting guidelines from the Finnish Patent and Registration Office (PRH 2018) and the European Patent Office (EPO 2018), as well as general books on patenting such as Oesch (2014) and Alberts et al. (2017) were instrumental in helping me understand the practicalities of patent professionals' work. Joho et al.'s (2010) survey of patent professionals was also useful. I considered gathering further data through other qualitative methods such as diaries, but was assured by patent professionals that people in their profession would not be willing to devote the time required for such activities.

The informants of the study were nine Scandinavian patent professionals from three types of organizations: five worked in companies that actively file and prosecute patents; two were from Intellectual Property Rights (IPR) service providers; and two worked for a governmental patent office. The patent-professional informants of the study were well educated: all nine had Master's degrees and four also had PhD degrees. Their experience in the IPR field ranged from 9 to 30 years, and their reported estimates on MT experience ranged from 4 to 15 years.

A total of 12 hours of interviews were recorded, transcribed, and analyzed using the thematic analysis method defined by Braun & Clarke (2006; 2013). This included the use of a member check after the initial analysis. Preliminary findings were compiled and then reviewed by three volunteer informants to validate that the findings accurately reflected their experience. Comments and clarifications were then considered in producing the final analysis. For this article, the themes and data were analyzed through the lens of distributed cognition.

4. Introduction to the work of patent professionals

This section briefly describes the aspects of the work conducted by patent professionals that relate to their use of raw MT, beginning with the cognitive activity under investigation. In analyses of translation and cognition, the activity that translators are involved in is the act of translating. In studies on reception and cognition, the activity readers are engaged in is reading a human translation. The activity under investigation in the case of patent professionals and cognition is one of reading and trying to understand a raw MT text.

4.1. *Texts and processes*

When inventors work on an invention that they are considering patenting, they must ensure that it is actually a new idea that does not exist, either in granted patents or in patent applications. Consequently, a large part of the work of patent professionals involves searching for and reviewing existing patent documents of inventions closely related to the one in question, with the intent of ensuring nothing in its claims has already been claimed in another patent. The goal is to identify a small set of patent documents that are the most relevant to the case at hand, and then to shape their own patent's claims so as to prevent infringement on any of them. The decision to include or exclude something into that set of relevant patent documents is one of the primary decisions patent professionals make (Nurminen 2019, 34–35).

Patent professionals and inventors are responsible for reviewing all existing patent documentation on inventions that closely resemble their own, including documents filed in other languages and in other countries. In 2012, an estimated 30% of a typical patent search resulted in patent documents in languages other than English (Tinsley et al 2012, 1); it is reasonable to assume that the percentage is now higher. As they may be searching for and reviewing a significant number of documents, sometimes “hundreds and hundreds at a time” (PP3), it would be impossible for patent professionals to acquire human translations of all of them, and they therefore rely heavily on MT in their review.¹

However, patent professionals and inventors do have the option of turning to human translation at any point if it becomes necessary. As described in Nurminen (2019), the decision to use human translation involves an assessment of risk and benefits. If the patent document that is in another language is needed for an especially risky or important IPR process, or if the patent professional cannot understand the MT, human translation might be introduced early in the process. On the other hand, if the IPR process is less risky or if the understanding they have of the raw MT is good enough, they will proceed with their case using the relevant documents in raw MT form. The alternative of using human translation is, however, always available. A similar access to human translation when needed has been reported in other studies of raw MT use, for example, Ishida (2016) and Pituxcoosuvann et al (2018).

4.2. MT for patents

The patent genre is challenging for MT systems. The genre is highly formulaic—patent applications and patents have a standard structure and a writing style that is occasionally (more or less jokingly) referred to as *Patentese*’ as described by Tinsley et al (2012, 69): “[it] typically comprises a mixture of highly specific technical terminology and legal jargon and is often written with the express purpose of obfuscating the intended meaning.” The style also commonly involves difficult syntactic structures. Considering the evidence that the use of controlled language can improve the quality of MT output (cf. O’Brien 2010; Marzouk & Hansen-Schirra 2018), Rossi & Wiggins (2013, 117) first reviewed the eight main components of controlled language: sentences are short, grammatically complete, use the active form, have simple syntax, and express only one idea; spelling and punctuation are correct; and texts use a limited lexicon. They then analyzed the typical features of patent language against these components and reported that patent language was incompatible with all eight.

Despite these challenges, MT systems designed specifically to handle patent documents have been developed for more than 20 years, beginning with the first implementations in the 1990s in Japan (Cavalier 2001) and the eventual launch of a free MT service in Japan. This was followed by MT services offered by the State Intellectual Property Office of China in 2008 (Wang 2009), the World Intellectual Property Organization in 2011 (Pouliquen 2016), and the European Patent Office in 2012 (Battistelli 2012). Commercial patent databases started to include MT in their solutions in the early-to-middle 2000s.

Patent MT tools handle a large variety of languages. The most frequently translated languages are Chinese, Japanese and Korean, primarily because China, Japan and Korea produce a significant number of patents and patent applications every year. For example, the top four countries filing patent applications in 2017 were China, with approximately 1.3 million; the USA, with 600,000; Japan, with 300,000; and the Republic of Korea, with 200,000 (WIPO 2018, 40). All informants in this study mentioned the predominance of these languages and the effects of working with them.

4.3. Tool environment

One of the most important tools patent professionals use in their work are patent databases, which aggregate patent documents from a number of different national patent offices and offer them through one user interface. The databases provide patent professionals with a broad

¹ In this and other quotes from informants, spoken data has been edited so that it conforms to written language conventions. The identifier *PP* refers to patent professional.

coverage of patents, advanced search tools, and other features that facilitate the complex task of searching. MT is one of those features. Sometimes the patent database manufacturer produces the MT in a bulk process, so that patent documents are already in machine-translated English when they are loaded into the database. Others provide tools to perform on-demand translations. MT is so commonly included in these systems that informant PP2 described the feature as “very routine, normal. You don’t even ask for it separately”. This ubiquity makes it possible to integrate MT directly into IPR processes, and the reports in this study illustrated a tight integration of MT into IPR tools and workflows.

An important element of the technical environment is that it provides functionality for saving machine-translated documents. Way (2013, 2018), Moorkens (2017), and Nitzke et al (2019) suggest that good use cases for raw MT most often involve, and should involve, highly perishable texts. However, the patent use case involves texts which are very long-lived. In addition to that, the machine translations of those texts can be long-lived, with translations being saved, stored, and shared among multiple people. Our informants described different ways they saved and shared machine translated texts. Some mentioned copy/pasting machine-translated excerpts of patent documents into Word documents or Excel sheets to share with technical teams. Two informants also mentioned that they also post-edited the excerpts a bit before sharing them. Informants also reported that they routinely save entire machine-translated documents in Word or PDF format to share with their teams or clients. Patent examiners store machine-translated documents in the systems that they use to communicate with patent applicants, allowing the applicants to access them directly from the system.

The technical environment also uses indicators to ensure that readers are aware that what they are reading is machine-translated. Informants described indicators such as colored frames, stamps labeled [*machine translation*], and documents with both original texts and translations shown in consecutive paragraphs or in side-by-side columns. These indicators often include the name of the system that produced the MT as well as a time stamp, which is either provided automatically by MT systems or added by patent professionals. These extra pieces of information can be important because the use of an inferior MT system or a translation considered to be dated may trigger the patent professional to perform a new translation. The informants of this study displayed an awareness of the advancement of MT technology and the improvements it brings, and mentioned that they sometimes re-translated documents because a newer translation might bring better results.

5. Distributed cognition: understanding through interaction with network

As mentioned, patent professionals are engaged in the cognitive activity of understanding, which in some cases might even be described as *deciphering*, a machine-translated text. In the middle of this process is one person reading and attempting to understand one machine-translated text. If the MT output is of low quality, a person unfamiliar with the use of raw MT might easily consider the task to be impossible. However, what if we envision the same activity as a socio-cognitive exercise in the environment in which it actually occurs, with the patent professional having access to “the roles and cognitive contributions of all co-workers,” as occurs for translation (Muñoz 2010, 179)? In that light the task—and the chance for success in achieving the task—appear to be very different.

Clark & Chalmers (1998) present the idea of cognition coming through a *coupled system*, or a combination of a human mind and an external entity. They argue (p. 8) that this combination can be considered a cognitive system in its own right in that:

All the components in the system play an active causal role, and they jointly govern behaviour in the same sort of way that cognition usually does. If we remove the external component the system’s behavioural competence will drop, just as it would if we removed part of its brain.

At the core of the task performed by the patent professionals in our case study—understanding a MT of a patent document—is the coupled system of one patent professional, with the knowledge and competences they bring into the situation, and at least one external entity, an MT engine.

Without the MT engine, the patent professional could make little or no sense of a patent document in a different language. Thus, as reasoned by Clark & Chalmers, the external MT engine is an “active” external component. If it is removed, the behavior of the human patent professional changes drastically.

In the case of patent professionals, the distributed nature of cognition encompasses far more than this core coupled system. Similarly to the translation project environment reported on in Risku (2014), patent professionals also work within a large network of artifacts and people, and they often achieve an understanding of patent documents in other languages through interacting with the elements of this network. In fact, the task of achieving understanding can be considered a process of trying different alternatives, using different elements of the network, until a sufficient level of understanding is achieved (Nurminen 2019, 37). The workings of such a network are perhaps best illustrated by Muñoz (2017, 564): “Cognition is also often *distributed*, in that several cognizing and not cognizing agents conjointly perform complex tasks, such as translating and producing large digital texts.”

This section examines the distributed network of artifacts and humans that patent professionals access for assistance in understanding machine-translated patent documents. Table 1 introduces the elements and it is followed by a discussion of each.

Element	Description
Original source document	The patent document in its original language, complete with multimodal elements such as drawings
Inventors or technical experts	Closest human element
Alternative machine translations	Machine translations retrieved through a second MT engine
Larger network of stakeholders	Other people who have an interest in the IPR case in which the machine-translated patent document is being used

Table 1. Elements in patent professionals’ network that contribute to understanding machine-translated documents

5.1. *Original source document*

The source-language patent document is one of the first elements in the network that patent professionals use to augment their understanding of the MT of a patent document. These documents frequently contain non-textual material, such as drawings or chemical formulas, that are not language dependent. They are “universally understood and should not be overlooked as important tools for reviewing [patent] documents in any language” (List 2012, 195) The informants of this study described how they used those parts of the original document as follows:

- (1) PP9: *Sometimes you just have to use the figures of the application and then you have to compare and see what it might be.*
 PP7: *It’s a combination of the original, if I see the chemical formulas or whatever they are using, because those abbreviations, they are not translated, like in carbohydrate variations, they are not translating those [...] and then I have the original and I have the translation. So then I combine them.*

Depending on the language of the original document, patent professionals also reported relying on their own language skills to use the source document. They all used English in their daily work and as a target language for MT, although none spoke English as their L1. In addition, all reported having some competence in German and most also had some command of other languages, such as French and Swedish. They reported reading patents in the original language when they felt their L3 command was good enough, but they also described situations in which they combined what they apprehended in the original with their grasp of the MT output to arrive at an understanding.

- (2) PP2: *Well there's the thing that when you understand German, you check the original as well...a little bit. So that kind of helps you understand because you're not just solely reading the machine translation [...] it of course helps because you can understand the structure quite well. You basically just need support with a few words you're unfamiliar with.*

Other studies have claimed a similar beneficial effect of language competence for raw MT reception. They have focused on the influence of source language command (Henisz-Dostert 1979; Morland 2002; Gaspari 2004; Nurminen 2016; Nurminen & Papula 2018), on target language skills (Henisz-Dostert 1979; Morland 2002; Smith 2003), or on proficiency in a pivot language (Ogura et al 2004).

The information sources listed thus far—the background knowledge of the patent professionals, the machine-translated text, and the original patent document—were very often enough to enable informants to acquire a satisfactory understanding of a document. One of the questions I asked in all interviews was, “Of all the times you use machine translation, how often is it successful, meaning you get enough information from the translation that you are immediately able to act on it?” A large majority reported that the success rate was very high. In six interviews, informants estimated their success rate at 90% or above.² One gave a somewhat lower rate, 75%:

- (3) PP6: *It's a combination of understanding the figures and understanding the machine translated text, so I would say that maybe 75%? In those cases I'm pretty convinced if this document is relevant or not.*

Another gave an even more conservative estimate, stating that every other case required more checking. Despite these high success rates, the informants reported an occasional need to rely on other elements in the network to acquire understanding.

5.2. *Inventors and technical experts*

The humans in the patent professional's network that they work with most closely are the inventors or technical experts who have created the invention that requires a patent or is already protected by a patent. Patent professionals often work in an advisory role to the technical experts, assisting them through patenting and other IPR processes. In some scenarios, these two professional groups work in the same company, and the *client* of the patent professional is *internal*. Other patent professionals, such as those who work in patent service providers and governmental patent offices, work for external clients.

In some cases, the technical experts perform the initial task of searching through patents and selecting the most relevant ones; in others, it is the patent professional who does it. In either case, both sides share the responsibility for understanding L3 patent documents, and this is often achieved through cooperative discussions. Sometimes, those are informal conversations between two colleagues, but informants also described reviews of machine-translated texts in meetings:

- (4) PP1: *But many times we are in a meeting. I have the people there. I have [the tool] open and we are reading from it [...] and then when there's a German or French or Chinese text, we realized that oh, that's what we're dealing with, and the button is there and ready, Translate, and then it does it.*

When the patent professionals were working with external clients, these dialogues were described as more formal, but the process of achieving understanding through dialogue between various participants was similar. Even in the formal situation of the governmental patent office examining patent applications, understanding appeared to be an exercise in distributed cognition:

² One of the informants reported that the initial 90% success rate had dropped after the technical field they worked in changed, resulting in an increased need to review documents from China and Japan.

- (5) PP6: *So if my interpretation of some kind of document based on the machine translation is wrong, I can change my mind later [...] It takes usually over two years to get a patent. So we get the answer from the applicant and we probably write another office action and then the applicant replies again. So it's a conversation.*

One factor that appeared to contribute to this patent professional-inventor pair's ability to understand MT documents was the combination of their different areas of contextual expertise, with patent professionals knowing the genre of patents and technical experts knowing about the subject matter or technical area of the patent in question. The informants in this study were experts in the patent genre due to their average of 17 years' experience in the IPR industry. They reported this to be helpful in understanding MT:

- (6) PP2: *And when you understand the... if we're talking about patent publications, there's a certain structure and there's a certain format that they're in. Then it's in a way easier to follow.*

While they had the best understanding of the genre, the informants indicated that the technical experts were the "best expert[s] in the technical field" (PP6). Many also pointed out that it was the combination of their own expertise in the patent genre and the inventors' competence in the subject matter that was important:

- (7) PP1: *So the system goes that way that, when we got the search results I send them out to the technical experts. They read them first by themselves. And they pick out those that they are worried about or where they want to have more information, where they're not sure what it really means. So then we have a meeting and we go through them together. So I can tell them what it really means and how to read it.*

Other studies have similarly noted the influence of contextual knowledge on the use of unedited MT. In Henisz-Dostert's (1979) study of scientists, informants were asked what the understandability of MT primarily depends on. They were given the choices of: (a) familiarity with subject matter; (b) translation of words; (c) sentence structure; (d) format; and (e) general style. The most cited option, by an overwhelming majority, was *familiarity with subject matter* (Henisz-Dostert 1979, 189). The influence of contextual knowledge on raw MT reception has also been discussed by Smith (2003), Yamashita et al (2009), Yasouka & Björn (2011) and Bowker & Buitrago (2019).

5.3. *Alternative machine translation*

The next element patent professionals use to gain understanding of machine-translated texts is output from an alternative MT engine. The patent professionals either compare two different outputs and select the better one, or they combine what they understood of both to construct an overall general understanding of the text. This tendency to compare different translations to arrive at an understanding has likewise been reported earlier. For example, Tinsley et al (2012) and Gao et al (2015) discuss how users of unedited MT compare different MT outputs, while Pym & Matsushita (2018) report on users comparing human translations. In the patent case, some tool manufacturers facilitate access to alternative MT engines by embedding links to them directly in the tools. The most common alternative tools that informants reported using were the MT engines embedded in the patent databases of individual country patent offices, such as J-PlatPat in Japan and the Patent Search and Analysis tool in China, but Google Translate was also used occasionally.³

³ J-PlatPat <https://www.j-platpat.inpit.go.jp/> Patent Search and Analysis <http://pss-system.cnipa.gov.cn/sipopublicsearch/inportal/i18n.shtml?params=902F004CA61084A284089435EAAEA94F59CC921A916ADCEB>

5.4. *Larger network of stakeholders*

Beyond the dyad with technical experts, the network that patent professionals rely on includes many other human stakeholders, all of whom have some level of interest in the IPR case in which a machine-translated patent document is being used. In the patenting process, one or more IPR service providers may be employed during the initial drafting of the application, and at least one governmental patent office will examine and prosecute the patent application. Furthermore, the IPR world is highly international and it is rare that a patent would be applied for in only one country. For this reason, further IPR service providers in other countries are often employed to interact with further governmental patent offices, frequently with one of the service providers coordinating the activities. The result is a large and complex network of people and long chains of communication. However, it also means that patent professionals have people they can rely on for help in their task of understanding an L3 patent document. If a patent search results in a document with only one detail that is not understood, patent professionals often tap into this network of people for help instead of ordering a human translation for that part of the document. One of the service providers from the same country the relevant document originated from can be asked to explain the part that was not understood.

Due to the nature of IPR work, even competitors can assume a role in helping to achieve an understanding of a patent document. The patenting process includes a nine-month phase for third-party observation and opposition. According to the Finnish Patent and Registration Office (PRH 2018, 19), the amount of material to review in a patent examination is so large that it would be impossible to cover it all. The purpose of this nine-month phase is therefore to allow third parties to identify relevant cases and information that might have been overlooked during the examination. During this phase, competitors can also challenge a patent and its claims. This often leads to discussions about differing interpretations of claims by different parties. If the case involves relevant patent documents that were machine-translated, those are discussed along with all other relevant documents.

Challenges to the interpretation of MT output might also arise, such as in a case brought up by one of the study informants in which they challenged the MT output used by a national patent office with a translation they took from the European Patent Office's MT engine, which was commonly considered to be a better one. Another informant mentioned a situation in which they challenged the scope of a patent document by offering more information than was contained in the (required) English abstract with a machine translation of a more specific part of the full document. In both cases, the discussions led to a better understanding of the raw MT output, so that they are both good examples of how meaning-making in this MT use case often comes through negotiation.

5.5. *Meaning-making through negotiation on a higher level*

An important point about meaning-making through distributed cognition described in this section is that it is embedded in an environment where meaning-making through negotiation is a common practice. Even when all information is in English, IPR texts tend to be dynamic. Applications change during the long patenting process and, more importantly, the meaning of the texts is interpreted, debated, and challenged by different parties. This offers two important supports for the use of raw MT. First, the space allocated in IPR processes for discussions on interpretation and meaning also gives space for machine-translated texts to be examined and errors to be spotted. Second, the people working in this environment can be assumed to be accustomed to the idea of meaning coming through negotiation. The assurance that MT output can be challenged and corrected, and the familiarity with the idea of meaning-making through negotiation, may both contribute to making it easier for patent professionals to accept the practice of relying on raw MT in important processes.

6. Conclusions and future research

The first goal of this case study on the use of raw MT by patent professionals was to analyze whether distributed cognition is a useful theoretical lens through which to examine patent professionals' use of raw MT. The second goal was to contribute to our understanding of the nature of MT gisting through an examination of the environment surrounding a functioning use case.

The study revealed that patent professionals perform their work as part of a large network of artifacts and people, and that the task of understanding machine-translated texts is often achieved through distributed cognition. At the core, the patent professional and MT engine work together as a *coupled system*, where each component in the system plays an active role in cognition. This is augmented through artifacts such as the original source document and alternative MTs, and through interaction with people, including the technical experts, a large network of other stakeholders, and even competitors. The concept of distributed cognition proved to be a useful framework to analyze and explain the understanding of raw MT in this case, and it might be applicable to other cases of MT gisting as well.

The article fulfilled the second goal by providing a nuanced account of the use of raw MT in a real-life scenario. Analyzing patent professionals' MT gisting as a case of distributed cognition expanded our overall understanding of raw MT use and contributed to our knowledge of the conditions that are present in the environment of a functioning case of MT gisting.

There were, however, certain limitations to this case study. First, the study used only one data gathering method, interviewing. This was caused by feedback from informants and potential informants that they could not afford to participate in more time-consuming processes. As discussed by Risku (2010) and others, studies in this area can include a variety of approaches and methods. A more comprehensive picture of distributed cognition could be gained through further studies which employ different methods, for example, surveys would be a way to gather information from more participants and it would not require much time from them.

A second issue is with interviewing itself, which results in a description of informants' ideas of what they are doing, but is not an actual account of their true actions. However, participants' viewpoints on what they are doing is a good starting point, and further studies can build on and test the ideas that come from such a study. A final limitation was that, although one of the main motivations for the study was to examine the role of environment, it was not planned specifically to be an application of a distributed cognition perspective. Perhaps if it had been initially designed as such, it might have revealed even further aspects and nuances regarding the distributed nature of patent MT cognition.

Further study of this use case and this professional group is essential. Their use of raw MT is a widespread and long-term practice, and there is much we can learn about the ways people have found to use raw MT, their strategies to ensure understanding, and their attitudes towards the practice. The knowledge gained from the use case may help us understand current and future uses of MT and other forms of artificial intelligence. Future studies could include a larger number as well as a wider diversity of patent professional users. It would also be beneficial to conduct studies employing varying methods so as to develop a more comprehensive understanding of the phenomenon.

Also, this article's focus on distributed cognition was not an exhaustive account of the different aspects of cognition in relation to this use case; future studies could explore aspects such as embodied or embedded cognition, which would contribute new viewpoints to the phenomenon. More studies would be welcome that focus on other professional environments where raw MT is being used, for example, in academia or customer support scenarios. That might provide new insights into the idea of raw MT as distributed cognition while also building on our knowledge of the environments and conditions in which MT gisting occurs and the factors that contribute to successful use of raw MT. Most of all, it is evident that raw MT use is a largely under-researched area and is in need of additional studies of all kinds.

References

- Alberts, Doreen, Cynthia Barcelon Yang, Ken Koubek, Suzanne Robins, Matthew Rodgers, Edlyn Simmons, and Dominic DeMarco. 2017. "Introduction to Patent Searching." In *Current Challenges in Patent Information Retrieval*. 2nd ed. Edited by M. Lupu, K. Mayer, N. Kando and A. J. Trippe, 3–45. Berlin: Springer.
- Battistelli, Benoît. "Patent Translate Service Goes Live!" <https://blog.epo.org/the-epo/patent-translate-service-goes-live/> Accessed 5 July 2019.
- Bowker, Lynne, and Jairo Buitrago Ciro. 2019. *Machine Translation and Global Research: Towards Improved Machine Translation Literacy in the Scholarly Community*. Bingley, UK: Emerald Publishing.
- Bowker, Lynne, and Melissa Ehgoetz. 2007. "Exploring User Acceptance of Machine Translation Output: A recipient evaluation." In *Across Boundaries: International perspectives on Translation Studies*. Edited by D. Kenny and K. Ryou, 209–224. Newcastle, UK: Cambridge Scholars.
- Braun, Virginia, and Victoria Clarke. 2006. "Using Thematic Analysis in Psychology." *Qualitative Research in Psychology* 3 (2): 77–101.
- Braun, Virginia, and Victoria Clarke. 2013. *Successful Qualitative Research: A practical guide for beginners*. Los Angeles: SAGE.
- Burgett, Will. 2015. "Unmoderated Remote Usability Testing of Machine Translation Content." *TAUS Review of Language Business and Technology* IV: 30–37.
- Castilho, Sheila, Fabio Alves, Sharon O'Brien, and Morgan O'Brien. 2014. "Does Post-Editing Increase Usability? A study with Brazilian Portuguese as target language." In *Proceedings of the 17th Annual Conference of the European Association for Machine Translation*, 183–190. <http://www.mt-archive.info/10/EAMT-2014-Castilho.pdf>
- Castilho, Sheila, and Sharon O'Brien. 2017. "Acceptability of Machine-Translated Content: A multi-language evaluation by translators and end-users." *Linguistica Antverpiensia* 16: 120–136.
- Cavalier, Tim. 2001. "Perspectives on Machine Translation of Patent Information." *World Patent Information* 23 (4).
- Clark, Andy, and David Chalmers. 1998. "The Extended Mind." *Analysis* 4 (1): 7–19.
- Doherty, Stephen, and Sharon O'Brien. 2009. "Can MT Output be Evaluated through Eye Tracking?" In *Proceedings of the 12th Machine Translation Summit*, 1–8. <http://www.mt-archive.info/MTS-2009-Doherty.pdf>
- Doherty, Stephen, and Sharon O'Brien. 2014. "Assessing the Usability of Raw Machine Translated Output: A user-centered study using eye tracking." *International Journal of Human-Computer Interaction* 30: 40–51.
- Ehrensberger-Dow, Maureen. 2019. "Ergonomics and the Translation Process." *Slovo.Ru: Baltijskij Accent* 10 (1): 37–51. <https://doi.10.5922/2225-5346-2019-1-3>.
- EPO: European Patent Office. 2018. *How to Get a European Patent: Guide for applicants*. Vienna: European Patent Office Vienna.
- Fuji, M., N. Hatanaka, E. Ito, S. Kamei, H. Kumai, T. Sukehiro, T. Yoshimi, and H. Isahara. 2001. "Evaluation Method for Determining Groups of Users Who Find MT "Useful". In *Proceedings of the 8th Machine Translation Summit*, 11–21. <http://www.eamt.org/events/summitVIII/papers.html>
- Gao, Ge, Hao-Chuan Wang, Dan Cosley, and Susan R. Fussell. 2013. "Same Translation but Different Experience: The effects of highlighting on machine-translated conversations." 27 April–2 May 2013.
- Gao, Ge, Bin Xu, David Hau, Zheng Yao, Dan Cosley, and Susan R. Fussell. 2015. "Two is Better than One: Improving multilingual collaboration by giving two machine translation outputs." 14–18 March 2015.
- Gaspari, Federico. 2004. "Online MT Services and Real Users' Needs: An empirical usability evaluation." In *Machine Translation: From Real Users to Research 6th Conference of the Association for Machine Translation in the Americas, AMTA 2014*, 74–85. <https://link.springer.com/book/10.1007/b100780>
- Gaspari, Federico. 2006. "The Added Value of Free Online MT Services: Confidence boosters for linguistically-challenged Internet users, a case study for the language pair Italian-English." In *Proceedings of the 7th Conference of the Association for Machine Translation of the Americas*, 46–55. <http://www.mt-archive.info/05/AMTA-2006-TOC.htm>
- Gaspari, Federico. 2007. "The Role of Online MT in Webpage Translation." PhD dissertation, University of Manchester.

- Henisz-Dostert, Bozena. 1979. "Users' Evaluation of Machine Translation." In *Machine Translation*. Edited by W. Winter, 149–244. The Hague: Mouton.
- Ishida, Toru. 2016. "Intercultural Collaboration and Support Systems: A brief history." In *PRIMA 2016: Principles and Practice of Multi-Agent Systems, 19th Annual Conference*, 3–19. <https://link.springer.com/book/10.1007/978-3-319-44832-9>
- Joho, Hideo, Leif Azzopardi, and Wim Vandervauwhed. 2010. "A Survey of Patent Users: An Analysis of Tasks, Behavior, Search Functionality and System Requirements." In *Proceedings of the Third Symposium on Information Interaction in Context: IliX '10*, 13–24. <https://dl.acm.org/citation.cfm?id=1840784.1840789>
- Jolley, Jason R., and Luciane Maimone. 2015. "Free Online Machine Translation: Use and perceptions by Spanish students and instructors." In *Learn Languages, Explore Cultures, Transform Lives*. Edited by A. J. Moeller, 181–200. Minneapolis, MN: Central States Conference on the Teaching of Foreign Languages.
- List, Jane. 2012. "Review of Machine Translation in Patents - Implications for search." *World Patent Information* 34 (3): 193–195.
- Liu, Nancy Xiuzhi, and Matthew Watts. 2019. "Mobile Translation Experience: Current state and future directions." In *Impacts of Mobile use and Experience on Contemporary Society*. Edited by X. Xu, 193–212. Hershey, PA: IGI Global.
- Marzouk, Shaimaa, and Silvia Hansen-Schirra. 2019. "Evaluation of the Impact of Controlled Language on Neural Machine Translation Compared to Other MT Architectures." *Machine Translation* 33 (1-2): 179–203. <https://doi.org/10.1007/s10590-019-09233-w>
- Moorkens, Joss. 2017. "Under Pressure: Translation in times of austerity." *Perspectives: Studies in Translation Theory and Practice* 25 (3): 464–477.
- Morland, D. Verne. 2002. *Nutzlos, Bien Pratique, Or Muy Util? Business users speak out on the value of pure machine translation*. Dayton, OH: NCR Corporation.
- Muñoz Martin, Ricardo. 2010. "Leave no Stone Unturned: On the development of Cognitive Translatology." *Translation and Interpreting Studies* 5 (2): 145–162.
- Muñoz Martin, Ricardo. 2017. "Looking Toward the Future of Cognitive Translation Studies." In *The Handbook of Translation and Cognition*. Edited by J. W. Schwieter and A. Ferreira, 555–572. London: Wiley-Blackwell.
- Nitzke, Jean, Silvia Hansen-Schirra, and Carmen Canfora. 2019. "Risk Management and Post-Editing Competence." *The Journal of Specialised Translation* (31): 239–259.
- Nurminen, Mary. 2016. "Machine Translation-Mediated Interviewing in Qualitative Research: A Pilot Project." *New Horizons in Translation Research and Education* 4: 66–84.
- Nurminen, Mary, and Niko Papula. 2018. "Gist MT Users: A Snapshot of the Use and Users of One Online MT Tool." *Proceedings of the 21st Annual Conference of the European Association for Machine Translation*, 199–208. <http://rua.ua.es/dspace/handle/10045/76049>
- Nurminen, Mary. 2019. "Decision-making, Risk, and Gist Machine Translation in the Work of Patent Professionals." *Proceedings of the 8th Workshop on Patent and Scientific Literature Translation*, 32–42. <https://www.aclweb.org/anthology/W19-7204/>
- Nuutila, Pertti. 2005. "Rough Machine Translation in the Communication Process." Licentiate thesis, University of Tampere.
- O'Brien, Sharon. 2010. "Controlled Language and Readability." In *Translation and Cognition*. Edited by G. M. Shreve and E. Angelone, 143–165. Amsterdam: John Benjamins.
- O'Brien, Sharon. 2017. "Machine Translation and Cognition." In *The Handbook of Translation and Cognition*. Edited by J. W. Schwieter and A. Ferreira, 313–331. London: Wiley-Blackwell.
- Oesch, Rainer, Heli Pihlajamaa, and Sami Sunila. 2014. *Patenttiolakeus* [Patent Law]. 3rd edition. Helsinki: Talentum.
- Ogura, Kentaro, Yoshihiko Hayashi, Saeko Nomura, and Toru Ishida. 2004. "User Adaptation in MT-Mediated Communication." *The First International Joint Conference on Natural Language Processing (IJCNLP-04)*, 596–601.
- Pituxcoosuvarn, Mondheera, Toru Ishida, Naomi Yamashita, Toshiyuki Takasaki, and Yumiko Mori. 2018. "Machine Translation Usage in a Children's Workshop." *CollabTech 2018: Collaboration Technologies and Social Computing: 10th International Conference, Proceedings*, 59–73. Edited by Takashi Yoshino et al, 59–73. Saskatoon, SK, Canada: Springer.
- Pouliquen, Bruno. 2016. *Practical use of Machine Translation in International Organizations*. Varanasi, India: World Intellectual Property Organization (WIPO).
- PRH: Patentti- ja rekisterihallitus [Finnish Patent and Registration Office]. 2018. Patenttiopas [Patent Guide]. Helsinki: Patentti- ja rekisterihallitus (PRH).

- Pym, Anthony, and Kayo Matsushita. 2018. "Risk Mitigation in Translator Decisions." *Across Languages and Cultures* 19 (1): 1–18.
- Risku, Hanna. 2010. "A Cognitive Scientific View on Technical Communication and Translation: Do embodiment and situatedness really make a difference?" *Target* 22 (1): 94–111.
- Risku, Hanna. 2014. "Translation Process Research as Interaction Research: From mental to socio-cognitive processes." In *Minding Translation*. Edited by R. Muñoz, 331–353. *MonTI*, special issue 1. Alicante: Universidad de Alicante.
- Rossi, Laura, and Dion Wiggins. 2013. "Applicability and Application of Machine Translation Quality Metrics in the Patent Field." *World Patent Information* 35 (2): 115–125.
<http://doi.org/10.1016/j.wpi.2012.12.001>.
- Schwietzer, John W., and Aline Ferreira, eds. 2017. *The Handbook of Translation and Cognition*. Hoboken, NJ: John Wiley & Sons.
- Shreve, Gregory M., and Erik Angelone, eds. 2010. *Translation and Cognition*. Amsterdam: John Benjamins.
- Smith, Ross. 2003. "Overview of PwC/Sytranet On-Line MT Facility." In *Proceedings of the Twenty-fifth International Conference on Translating and the Computer*. <http://mt-archive.info/Aslib-2003-TOC.htm>
- Stewart, Osamuyimen, David Lubensky, Scott Macdonald, and Julie Marcotte. 2010. "Using Machine Translation for the Localization of Electronic Support Content: Evaluating end-user satisfaction." In *Proceedings of the 9th Conference of the Association for Machine Translation in the Americas*. <http://www.mt-archive.info/10/AMTA-2010-TOC.htm>
- Tinsley, John, Alexandru Ceausu, Jian Zhang, Heidi Depraetere, and Joeri Van de Walle. 2012. "IPTranslator: Facilitating patent search with machine translation." 28 October–1 November, 2012.
- Turovsky, Barak. 2016. *Google Translate Blog: Ten years of Google Translate*.
<https://Blog.Google/Products/Translate/Ten-Years-of-Google-Translate/>. Accessed 30 June, 2016.
- Wang, Dan. 2009. "Chinese to English Automatic Patent Machine Translation at SIPO." *World Patent Information* 31 (2): 137–139.
- Way, Andy. 2013. "Traditional and Emerging Use-Cases for Machine Translation." Paper presented at Translating and the Computer 34.
https://www.computing.dcu.ie/~away/PUBS/2013/Way_AS LIB_2013.pdf
- Way, Andy. 2018. "Quality Expectations of Machine Translation." In *Translation Quality Assessment: From principles to practice*. Edited by J. Moorkens, S. Castilho, F. Gaspari and S. Doherty, 159–178. Cham, Switzerland: Springer.
- WIPO: World Intellectual Property Organization. 2018. *World Intellectual Property Indicators 2018*. Geneva: World Intellectual Property Organization.
- Xu, Bin, Ge Gao, Susan R. Fussell, and Dan Cosley. 2014. "Improving Machine Translation by Showing Two Outputs." In *CHI '14: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 3743–3746. <https://dl.acm.org/citation.cfm?id=2556288>
- Yamashita, Naomi, Rieko Inaba, Hideaki Kuzuoka, and Toru Ishida. 2009. "Difficulties in Establishing Common Ground in Multiparty Groups using Machine Translation." In *CHI '09: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 679–688.
<https://dl.acm.org/citation.cfm?id=1518701>
- Yang, Jin, and Elke Lange. 2003. "Going Live on the Internet." In *Computers and Translation: A Translator's Guide*. Edited by H. Somers, 191–210. Amsterdam: John Benjamins.
- Yasouka, Mika, and Pernille Bjorn. 2011. "Machine Translation Effects on Communication: What makes it difficult to communicate through machine translation?" In *Proceedings of the Second International Conference on Culture and Computing*, 110–115.

Author's address

Mary Nurminen

Faculty of Information Technology and Communication Sciences

Tampere University

Kalevantie 4

FI-33100 Tampere

Finland

mary.nurminen@tuni.fi

Biographical note

Mary Nurminen is currently completing her PhD on non-translator users of machine translation. She is a teacher of Translation Studies at Tampere University. Before joining academia in 2013, Mary spent 13 years at Nokia and 5 at Lionbridge, working in the areas of technical communication and localization as (in this order) practitioner, manager, and solution architect.