Decision-making, Risk, and Gist Machine Translation in the Work of Patent Professionals

Mary Nurminen
Tampere University
Kalevantie 4
FI-33100 Tampere, Finland
mary.nurminen@tuni.fi

Abstract

This is the first study on how patent professionals use gist machine translation (MT) in their work. Inductive, qualitative research methods were adopted to explore the role of gist MT specifically in decision-making. Results show that certain decisions by patent professionals rely on gist MT, that the decision to involve human translation is often based on a risk assessment, and that certain factors in the patent environment give affordances for the use of gist MT. The study contributes to the body of knowledge on patent MT users and on gist MT users in general.

1 Introduction

Machine translation (MT) for patents has been developed for a few decades and a broad body of research is devoted to the technologies and techniques for producing patent MT. The professionals who work with patents – patent attorneys, counsels, examiners, etc. – use this MT in its raw, unedited form to obtain a basic understanding, or gist, of patent documents that they need but that are in languages they do not understand. Although their use of this raw MT (termed *gist MT* in this article) has been widespread for approximately a decade, very little research has been conducted on these MT users. In fact, while the number of studies on one group of professionals who use MT in their work, translators, has increased in recent years, research on other professional groups who use the technology remains scarce.

The main objective of this article is to provide the first study focused specifically on the users of patent MT. The article presents the results of a qualitative, exploratory study based on interviews with a small group of patent professionals who use MT in their daily work. Three themes were investigated for the article: the types of decisions patent professionals make based on machine-translated information, the risk assessment they use when deciding between relying on gist MT or opting for human translation, and finally, the environmental factors that appear to give affordances for the use of gist MT in this context.

Two important aspects of patent MT are not in the scope of this study. First, the article does not focus on the issue of quality of MT output, as that has already been studied in numerous other articles. Instead, I wanted to concentrate on exploring other factors that influence gist MT use. Second, another key application of patent MT is its use by professional post-editors to enhance their translation process. These users are not included in the scope of this study.

The article will help to inform research and solution development in the patent MT field. It will also contribute to studies of different professions' use of gist MT and to a general understanding of gist MT users. Better knowledge of how MT is used in different contexts and what contributes to successful use will help us to define what makes a potential use case good, or conversely poor, for gist MT use. In addition, research on experienced users of this form of artificial intelligence can give us insights into the needs of users of other AI technologies.

The structure of the article is as follows: the next section contains a review of related work. This is followed by a description of the data and methods used in this study. Section 4 discusses the types of decisions that patent professionals report making based on gist MT. Section 5 describes the risk assessments that informants appeared to undertake when deciding on ordering human translation. Section 6 focuses on the factors in the work environment that appear to support the use of gist MT. Final conclusions and suggestions for further studies are presented in Section 7.

2   Related work

To the best of my knowledge, thus far no studies have been conducted on how patent professionals use gist MT in their everyday work. A few experimental studies have been done. Larroyed (2018) and Tinsley et al. (2012) describe evaluation experiments in which one evaluation is performed by real patent professionals. A number of studies describe technical solutions for patent MT, and some of those include discussions of some aspects of MT in patent professionals' work, for example, Tinsley (2017), Rossi and Wiggins (2013), and List (2012). In addition to these, a few studies that focus on patent searchers also allude briefly to MT in patent search, including Joho et al. (2010) and McDonald-Maier (2009).

To date there is only a small body of research on professional areas where gist MT is used. Professional translation has been studied to some extent, though in that industry MT is predominantly used for dissemination and not for gisting. Industries with reported use of gist MT include customer support, academia, medicine and the legal field. Customer support groups began to offer multilingual access to knowledge base articles through gist MT in the early 2000s. However, although several articles describe these solutions (e.g. Stewart et al., 2010; Dillinger and Gerber, 2009), very little user experience research has been undertaken, as stated in one of the few studies on actual users (Burgett, 2015: 30). A growing body of research focuses on the use of MT in academia. Much of this focuses on the effects of MT on education and students, but some of the studies also cover educators' viewpoints, such as Bowker and Eghoetz's (2007) study on the acceptability of MT in a university setting and Bowker and Buitrago's (2019) book on using MT in research. Health care is another field where gist MT is beginning to be researched. Liu and Watts (2019) give a good overview of current studies on mobile MT use in health care. Most recently, John Tinsley describes the emergence of new use cases for gist MT in two different industries: legal and life sciences (Beninatto and Stevens, 2019). Both cases are similar to the patent case in that MT is mainly used to sift through large numbers of documents to categorize and then locate the ones that need further scrutiny and possibly human translation.

Work in the area of risk and translation has examined risk assessment and management strategies either as part of the individual translator's work (Pym, 2015; Pym and Matsushita, 2018), or from the perspective of the translation process and service provider (Canfora and Ottmann, 2018). Canfora and Ottmann (2016) present a model for risk management for internal translation processes, including a risk matrix combining the probability of risk and the potential consequences. A recent paper by Nitzke, Hansen-Schirra and Canfora (2019) introduces a model for assessing the risk associated with using post-edited or gist MT. Nonetheless, the focus of that study is primarily the post-editing context, while scenarios involving unedited MT remain mostly unexplored.

3   Methods

The main data for this study was gathered through interviews with nine patent professionals working in Scandinavia. The term *patent professional* in this study refers to professionals in the intellectual property rights (IPR) field who use their expertise in patents to assist and guide others (internal or external groups) in their IPR processes. These professionals hold a variety of titles, such as Patent Counsel, Patent Attorney, and Patent Examiner. The informants for the study are presented in Table 1.

Type of informant	N
Patent professionals working in companies that are active in filing and prosecuting patent applications	5
Patent professionals working in an IPR service provider	2
Patent professionals working in a government patent office	2
Total	**9**

Table 1. Informants interviewed for the study

I included informants from the key areas where patent professionals work: private companies, IPR service providers, and governmental organizations. The largest group consisted of professionals working in companies that file patent applications. This is somewhat reflective of

the 2010 survey by Joho et al., in which 88% of respondents reported working predominately with internal clients (Joho et al., 2010: 16), which indicates that they worked in patent-filing and prosecuting companies. In addition to the interviews, I gathered background information through talking to people involved in creating and maintaining patent MT solutions.

The average age of the informants was 47 and the average length of experience in the IP field was considerable, 17 years. The group was highly educated; all had at least a master's level education and four of the nine held a PhD degree. This is similar to the educational levels reported in Joho et al. (2010).

The interviews were all semi-structured discussions that occurred either at informants' workplaces or through Skype audio calls. They were conducted in the time frame of April 2018 to February 2019. The two themes of context and transparency were explored in the interviews. I used a variety of sources in the development of the questions. ISO standard 9241-210:2010 (ISO, 2010) defines *context of use* through the broad categories of users, goals and tasks of users, and environment, and this was a good starting point. I relied on descriptions of patent processes in official documents (PRH, 2018; EPO, 2018) and other sources (Alberts et al., 2017; Oesch et al., 2014; Joho et al., 2010) to identify the touchpoints users might have with MT and to develop questions around those touchpoints. The questions also developed somewhat over the course of the data-gathering phase.

Most of the interviews were recorded, transcribed with the aid of automatic transcription tools, and then post-edited. One interview was not recorded due to technical difficulties, so the data from that interview consisted of my notes taken during the interview. A total of 12 hours of interviewing was conducted, and 229 pages of transcription and note data compiled for analysis.

The data was analyzed by closely following the thematic analysis method outlined by Braun and Clarke (2013, 2006, n.d.) with additional guidance from Merriam and Tisdell (2016). The data was approached from a semantic perspective, wherein "coding and theme development reflect the explicit content of the data" (Braun and Clarke, n.d.) rather than searching for underlying meanings in the data. One reason for this was that the topic of the use of technology at work was fairly straightforward. Also, the focus was on the context, as described by informants, instead of each informant's personal experiences.

At a point later in the analysis process, a summary of findings was compiled and a member check performed by three of the informants. They were asked to compare the results against their own experiences and to comment on any incongruences they may detect. These comments were then reviewed and incorporated into the analysis.

A qualitative method was chosen for this study for specific reasons. First, it was necessary because this is the first study on how this group uses gist MT, and research at such an early stage often requires inductive, exploratory methods. When designing the study, there simply was not enough knowledge on these users to allow for the crafting of a quantitative study such as a survey. A second reason was that the small body of research on gist MT users in general tends to rely on surveys and laboratory experiments. I believed there was a need for in-depth studies that would give us a more nuanced view of the use of gist MT. I selected interviewing for data-gathering because it proved difficult to persuade exceedingly busy patent professionals to participate in a study using more time-consuming qualitative methods such as diaries. The interviews required a commitment of only 1.5 hours, which seemed to be more tenable.

4 Decisions that rely on gist MT

Rossi and Wiggins (2013: 116) argue that "In the patent field, MT is used as a support tool for performing novelty, validity, infringement or state-of-the-art searches, and to provide a first understanding of the content of retrieved publications." However, is gaining a "first understanding" really the only way patent professionals use MT, or do they actually make decisions based on gist MT? For example, Henisz-Dostert's study of scientists' use of MT to understand Russian scientific articles reported that, contrary to predictions by early scholars that MT would be used only for scanning, scientists used MT "more as a tool of information than as a tool for the selection of information." (Henisz-Dostert, 1979: 180). One goal of this study was to explore the ways in which patent professionals use gist MT and the decisions they make with its help.

4.1 Relevance

One of the primary uses for patent MT, as defined by Tinsley (2017: 411) is "[to] provide an on-demand 'gist' translation of foreign patents for information purposes to determine relevance." The primacy of using gist MT for this purpose was

also found in this study. Informants described how they used gist MT when searching for 'prior art' (patent documents that show that an invention is not new and therefore present an obstacle to patenting it). For each patent document (either a patent or a patent application) found in their search, they need to decide whether it is relevant to the IPR process they are working on or not. Informants discussed four main IPR processes in which they use machine-translated information to determine relevance: (1) the patenting process (Does this invention show enough novelty that it could be patented?), (2) freedom to operate (Can we launch our products in this market or are there patents that we would be infringing on if we launched?), (3) monitoring (Is this patent application sufficiently relevant to our work that I should monitor its progress?), and (4) infringement (Is this patent infringing on one of our patents or are we in danger of infringing on someone else's patent?).

The results of this study reveal that the decision on relevance is very often made without the help of human translation. Therefore, the first decision made is not whether or not a patent document should be sent for human translation, but whether or not it appears to be relevant, and that is determined largely on the basis of gist MT:

I would say it's [the use of MT] *successful in 90 percent of the time because the conclusion is, this is not relevant…So rejecting things from further analysis I think is done 9 out of 10 reviews of the machine translated documents.* (PP4)[1]

It is important to note that the decision on relevance is not as minor a decision as it may seem. The consequences of mistakenly discarding a patent document that seems irrelevant can be considerable, as was reported by informants:

…for example I can decide about a patent that it is not in any way relevant for us, which is a pretty strong decision, because then we shut it out completely, the whole followup of the patent, and we just think that that won't be harmful, but then it could be that if there's a mistake in the translation then it turns out that it really is harmful. But those are the kinds of decisions I make. (PP1)[2]

*At work we talk about how most mistakes take place because someone **overlooks** a relevant patent…when a mistake happens, it is most likely to be caused by that. But mistakes can come from*

other reasons than the machine translation. There are just so many patents to go through. But putting it into the 'not interesting' pile is a risk. (PP3)

4.2 Monitoring

A second type of decision that is very often made based solely on gist MT is the decision to tag a patent application for monitoring. If an application is deemed relevant, patent professionals may decide to follow its progress throughout its prosecution. Besides being used to determine enough relevance for monitoring, Gist MT is also used to understand communications on the application's prosecution or to review changes in the application. Tagging an application for monitoring also often means that the decision on human translation is postponed, because the application will most likely change before it is granted:

…if it's about pending patents then the claim scope is changing all the time, so therefore even if you would translate it and get it kind of right in the beginning, it's something different when it's granted…so therefore there's no point maybe to get it human translated at the early stage (PP2)

4.3 Patenting and opposition

A third area in which informants reported using gist MT in decision-making was during the patenting process. Within the European context, the role of MT in the examination process is explained in official guidelines:

In order to overcome the language barrier constituted by a document in an unfamiliar non-official language, it might be appropriate for the examiner to rely on a machine translation of said document…A translation has to serve the purpose of rendering the meaning of the text in a familiar language…Therefore mere grammatical or syntactical errors which have no impact on the possibility of understanding the content do not hinder its qualification as a translation. (EPO, 2018, Part G, Chap. IV-4)

Patent examiners typically share the results of their patentability search with patent applicants, and any relevant patent document that is in another language is provided in machine-translated form. Unless the applicant decides it is so important that they will provide a human

---

[1] Here and elsewhere: PP = Patent Professional. Also, quotes have been edited for fluency.

[2] Some quotes and passages translated by author.

translation, prosecution proceeds based on the machine translation. MT is occasionally used in opposition proceedings as well:

PP9: *I mean normally in opposition cases at the EPO, European Patent Office, you can use machine translations.*

Interviewer: *OK. And in the Finnish patent office as well?*

PP9: *Yes you can do that. I have never been asked to provide a human translation about any of these.*

5 Deciding on human translation: an exercise in risk assessment

As far as translation is concerned, the most important decision patent professionals or patent applicants make is whether to rely on gist MT for understanding a relevant document or to have it translated by a human. Nitzke et al. (2019) proposes that this type of decision involves risk and that an assessment of those risks should be part of the process of decision-making. Evidence of such a risk assessment emerged in this study, with patent professionals weighing various factors before deciding on gist MT or human translation. The factors that supported human translation of a patent document included the riskiness of the IPR process in which the document would be used, the assumed relevance and importance of the other-language document, and the potential consequences if a misunderstanding would occur due to an error in the gist MT. The factors supporting the use of gist MT were lower costs, quicker access to information, and trust that the patent document is adequately understood. This assessment was summarized by some informants:

...the more important decision, the less you do the decision based only on the machine translation. (PP8)

...if the context is clear then it's OK as I see it, I trust the machine translations enough, but sometimes when we are in borderline decisions it's required to have a proper human translation...So it's more a question of the uncertainty margin of the translation with respect to what we are deciding. (PP4)

Each side of this assessment is examined more closely below.

5.1 Arguments for human translation

One of the top considerations for triggering human translation was the IPR process the other-language relevant document would be used in,

with some processes being seen as more high-risk than others. Cases that involved infringement or freedom to operate might involve considerable costs and legal involvement, and these were consequently cited as cases in which human translation is often needed:

It depends on the costs involved in the case...if you are in a patent battle, if there is an infringement case...there's a lot of money involved. If you want to be absolutely sure then you have to have a human translator. (PP9)

If, based on the gist MT, a patent document appears to be highly relevant and important to a case, that would also serve as a strong argument for triggering human translation:

...probably that also depends a little bit on the case. If it is highly important then I would choose immediately to get it translated, or claims or parts of it, translated with human translation. (PP2)

Informants also mentioned potential consequences as a factor in the decision on human translation:

If we make the wrong decision and allow a product to the market which does not have freedom to operate, there is a risk of using time and money and goodwill in a court case and potentially being responsible to cover the damages of a client. (PP4)

5.2 Arguments for relying on gist MT

The main arguments for using gist MT are clearly that translation is very quick and does not generate extra costs. MT is provided at no cost by various national and international patent offices such as the Japan Patent Office and the EPO, and it is commonly included by default in commercial patent search tools. Its use is also made easy through tight integration to patent search tools and processes.

A complete understanding of the arguments for relying on MT in the risk assessment, however, requires consideration of another important element: how strong is the patent professional's trust that they have a sufficient understanding of a patent document? Much of this depends on the quality of the machine translation, of course. However, past studies have shown that other factors can enhance users' abilities to understand MT, and those were reported as helpful in this study as well. Two factors appeared to contribute to trust in understanding in this case study: the fact that patent professionals rely on other resources than the gist

MT, and the background knowledge that patent professionals possess. These are discussed further in the following subsections.

Understanding does not depend on MT alone

The understanding of the machine translation of a patent document can be seen as a process of trying different alternatives until a sufficient level of understanding is achieved. The first alternative is to combine the gist MT with other resources, such as drawings and chemical formulas in the original-language patent documents, to enhance understanding. This combining of MT and auxiliary, often multimodal, information to obtain an understanding of other-language texts has been reported in other studies on MT users (Nitzke et al., 2019; Pituxcoosuvarn et al., 2018; Suzuki and Hishiyama, 2016; Way, 2013; Gaspari, 2004; Henisz-Dostert, 1979). Auxiliary information had a clear role in patent professionals' reports of their work in this study as well:

> When it's good enough that I can see that it's relevant? It's a combination of understanding the figures and understanding the machine translated text. (PP6)

> Oh yeah then you have to look at the original because it doesn't translate any of those chemical formulas…And then if they are totally different then it could be that I don't even make any translation because then I know that, well, they are talking about totally different things. (PP7)

A second alternative patent professionals resort to are alternative machine translations from other MT tools, a practice that has been noted in earlier studies (Gao et al., 2015; Tinsley et al., 2012). At least one commercial patent search tool offers users access to both their own MT solution and the alternative of Google Translate in the same window. Although a few informants mentioned using a general tool such as Google Translate for alternative translations, a more common method was to try the MT tools provided by specific governmental patent offices:

> I do the EPO machine translation first and if that's not more understandable then I go directly to the patent office that the publication came from, so Chinese or Japanese. (PP5)

> …for instance if it's a Chinese document I go to Chinese Patent Office website and try to find the same application there…and usually it's a different machine translation and that actually helps sometimes, when you have two machine translations you can read them at the same time

> and maybe it gives you a better impression. (PP6)

The next alternative professionals can turn to are the other patent professionals they collaborate with. Instead of ordering a human translation of a text that is not sufficiently understood, they can ask the patent professionals who work more closely with the inventors (for example, the patent professionals in the country which the patent originates from) to clarify unclear passages for them.

Background knowledge aids understanding

As mentioned previously, the informants of this study were both highly experienced in the IPR field and well educated. Their contextual knowledge and competences in languages appeared to be important factors in helping them understand and use machine-translated information effectively.

The importance of MT users' knowledge of context in helping them understand machine-translated texts has been reported in a number of studies. Henisz-Dostert (1979) found that a user's familiarity with the subject matter was seen as the main factor in determining the understandability of machine-translated texts. Other studies that have highlighted the importance of contextual knowledge include Bowker and Buitrago (2019), Yasouka and Björn (2011), Yamashita et al. (2009) and Smith (2003).

In the patent context, contextual knowledge is often divided between the patent professionals, who know the patent genre, and the inventors or researchers behind the patents, who know the subject matter better. These competences, their role in helping to understand machine-translated patents, and the division of expertise between patent professionals and inventors were a common theme in the interviews.

> And when you understand…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 easy easier to follow. (PP2)

Several previous studies have examined the role of users' language competence in gist MT scenarios (Nurminen and Papula, 2018; Nurminen, 2016; Henisz-Dostert, 1979). In the current study, this background competence also appeared to be a factor in successful use of MT. Although none of the informants spoke English as their native language, all used English daily in their work. Their MT use was mainly from other

languages into English, not into their native languages. Besides English, all informants had varying levels of competence in other languages, with German being the most often mentioned, followed by French, Spanish, Swedish, Italian, Dutch, and Japanese. Several informants indicated that competence in the source language helped them to understand texts that were machine-translated from those languages:

And quite often I actually combine a machine translation and original reading…the complementarity of understanding the structure of the language better than the machine, and the machine understanding more words than I do, is a good complementarity. (PP4)

However, the reality is that the major languages patent documents are translated from are Chinese, Japanese, and Korean because these countries are significant producers of patents. China became the world's largest patent producer in 2011. By 2017, China had filed 1.3 million patent applications, more than double the number filed by the second country, the U.S. (WIPO, 2018: 40). The predominance of China was mentioned in all interviews. We can assume from this that competence in the three major patent languages of Chinese, Japanese, and Korean would be particularly useful for patent professionals.

6 Affordances in the patent context

Thus far this article has presented a scenario in which patent professionals can and do use gist MT to make certain decisions. The article has also discussed the factors involved in their decisions to rely on gist MT or to order human translation. However, in the analysis of this study's data, certain contextual factors emerged which appeared to make affordances for the use of gist MT. These affordances must be considered when discussing this specific case because they appear to play an important role in making the use of gist MT tenable, and an understanding of this ecosystem's use of gist MT is incomplete without them. The following sections explore two factors of affordance, risk tolerance and legitimacy.

6.1 Risk-tolerant environment

In the book titled *Translation Quality Assessment*, Andy Way states that MT systems need to be evaluated with the knowledge of what the system would be used for. Way also notes that "[o]f course, some objectives could be more tolerant of MT errors than others" (Way, 2018: 170). Certain features of the patent environment, while perhaps

not fully error-tolerant, appeared to make affordances for the risks and potential errors tied to the use of gist MT.

The patenting process is long and iterative, with multiple parties often reviewing the same or similar texts. Different stakeholders may have different interpretations of a patent application's scope and claims. To address these issues, the process contains space for discussion and mechanisms for stakeholders to examine and challenge each other's work. One of these is explained in the Finnish Patent and Registration's Patent Guide:

Even though inventions must show absolute novelty, it is not possible for patent authorities to clarify all public information when examining an application. For this reason, the examination process is augmented by the third-party observation and opposition processes, in which third parties, for example competitors, can bring to the attention of the authorities issues that did not emerge during the examination of the patent application. (PRH, 2018: 19)

The nature of this process means that there are also multiple stages where errors in the understanding of gist MT can be detected and corrected. This was described by one informant:

Well of course you can get the wrong impression of the subject matter in the document, but I don't see that it's a really big risk because the patent application process is a long process, so if my interpretation of some kind of document based on the machine translation is wrong, I can change my mind later, if I see it. 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. So during the process there's many times when these things can be dealt with. (PP6)

Another informant described a case when parties examined and challenged each other's MT work:

We've had these cases where the examiner used Google Translate and we translate it using EPO's official site and then we can explain to them that 'now we would like to kindly point out that the translation used by the examiner contained a mistake in this spot, and that we have this in that same spot, and our version uses the terms in this way.' And we rely on the [machine] translation completely…the examiner doesn't understand Japanese and we don't understand Japanese. We are both relying on machine translation and there is nothing else. (PP1)

Besides the risk tolerance present in processes, the informants in this study displayed a tendency to accept the risk involved with using MT and making decisions based on it. One reason for this might be that the informants were vastly experienced. Another reason might be that the IPR field contains other risks besides the use of gist MT, so the organizations they work in might have a higher willingness to take risks, or "risk appetite" as defined by Nitzke et al. (2019). Finally, the acceptance of risk might be an acknowledgement that the risk is simply necessary due to the impossibility of relying on human translation for the large volumes of documentation they regularly encounter, as voiced by one informant:

Yes, there is always risk involved. But we have so many patents to go through. Hundreds and hundreds at a time. It would be impossible if all of those had to be translated by a human. Always a risk though. (PP3)

6.2   Legitimacy of MT

One aspect of the use of MT in the patent environment that I did not expect when I began my research was the legitimacy that it enjoys. One of Merriam-Webster's definitions for *legitimate* best reflects what it means in this context: "conforming to *recognized* principles or *accepted* [emphasis by author] rules and standards."[3] Three different themes in this study illustrated this legitimacy: MT use was transparent, the boundaries of its legitimacy were documented and generally agreed upon by users, and its users had a relatively high level of 'MT literacy.'

Transparency

Transparency in gist MT use has been addressed in a few reports, most recently in a 2019 Globally Speaking Radio podcast in which John Tinsley reported that the legal profession is beginning to use MT for e-discovery, and that its use is fully transparent in that context: "So you go into the court and say to the judge, 'We are taking this position on the basis of a machine translation of this document into English,' and that's legally defensible" (Beninatto and Stevens, 2019).

At least in the European context of this study, the first sign of transparency was the inclusion of MT in EPO guidelines. Second, descriptions by study informants depicted an environment in which the role of MT is transparent to all. They also reported that MT is transparent to secondary users of patent MT, the internal and external clients the patent professionals work with. The results of searches these clients receive from patent professionals often include documents that are machine-translated. These are clearly marked as machine translations and they often also include the date and MT tool that produced the text. Patent professionals discuss MT with the secondary users, as in this example:

I would point out that this is a machine translation and, depending on if the client is a knowing patent engineer, then I would maybe give my opinion if we need a proper translation or not, but then ask what they think. (PP8)

Boundaries of legitimacy

An important aspect of legitimacy is that it is bound to a specific scope. The 'recognized principles or accepted rules and standards' referred to in the definition provided earlier are agreed upon by a certain group for a certain purpose, and the boundaries of applicability are recognized by the participants. In this study, the boundaries of legitimacy for MT were sometimes mentioned during answers to other questions: "For information purposes, it's fine. For use as a legal text, no." (PP3) But in the interviews I also asked directly, "In what situations is it not OK to use machine translation?" The responses indicated some agreement on the areas in which MT should not be used, such as when filing patent applications:

…when you're translating your application to other languages – like we have seen some kind of, I think they are usually applicants from Asia, that file an application here and usually they apply in English, but you can really see that their application is machine translated from the Chinese version – not OK. (PP6)

There was very clear agreement that MT should not be used in legal settings, as in this example in which an informant described a process involving another company's potential infringement of their patent:

We would start with searching prior art and use MT. Our aim is to see if there's overlap with our patent or not. If we find something that looks in-

---

[3] https://www.merriam-webster.com/dictionary/legitimate

teresting, then we would order human translation of it. We would not just go ahead on that with machine translated information. (PP3)

MT literacy

In 1993 Church and Hovy defined six "desiderata" for a good use case for MT. Among the six were: "it should set reasonable expectations" and "it should be clear to the users what the system can and cannot do" (Church and Hovy 1993: 257). Bowker and Buitrago (2019) expanded this idea by coining the term *MT literacy,* and then applying it to the case of MT use in academic work. On the basis of their definition, I described MT literacy for the context of this study as a patent professional's ability to: (1) comprehend the basics of how machine translation systems process texts, (2) understand how machine translation systems are or can be used (by oneself or others working with patents) to find and read patent documents within the context of IPR processes, and (3) appreciate the wider implications associated with the use of machine translation. Based on this definition, the informants in this study displayed a generally high level of MT literacy. They appeared to understand the basics of MT technologies, knew how to access different MT tools, and were aware of the possibility and consequences of translation errors. They also had experience with different types of MT tools and noticed improvements in quality over time:

> *They all* [languages] *have become better, and especially nowadays if you make a machine translation for some 'normal' language, for German or French, they are really good.* (PP9)

Perhaps one of the clearest signs of the high level of general MT literacy was an observation I made throughout the study: the hype issues currently visible in other spheres (for one example, see Hassan et al., 2017 followed by Toral et al., 2018) do not seem to be occurring in patent MT. In the present study, MT was considered to be one tool among others and people were aware of its uses and limitations. I heard no reports of overreaching claims on MT capabilities.

7   Conclusions

The main objective of this study was to explore the types of decisions patent professionals make based on machine-translated information, the risk assessment they use when deciding between relying on gist MT or using human translation, and the environmental factors that appear to support the use of gist MT in this context. The results revealed that patent professionals routinely make decisions on relevance and monitoring based on gist MT, and that the patenting process also relies on it. In the key decision of initiating human translation, patent professionals tend to weigh the riskiness of the IPR process in which the translated patent document would be used, the assumed relevance and importance of the document, and the potential consequences of misunderstanding against the lower costs, quicker access to information, and trust in a good enough understanding of the patent document. That understanding is often based not only the gist MT, but also other factors, such as auxiliary information sources and patent professionals' contextual and linguistic knowledge. The environmental factors of risk tolerance and legitimacy for gist MT also support the use of MT.

The study contributes to our knowledge of how people, and specifically professional groups, use gist MT. It explores factors that can enhance the use of gist MT, and this understanding will help us to define the characteristics of good, as well as poor, contexts for gist MT use. In addition, this analysis contributes to the growing body of research on users of various types of artificial intelligence.

This study had certain limitations. The group of informants was small and somewhat homogeneous, and this influenced the results. Data was gathered through only one method, interviewing. The results also focused on patent work in one geographical area and one specific point in time, and the results cannot be considered representative of the larger population of patent professionals. Nevertheless, as the first exploratory study of this very experienced group of MT users, it fulfilled one of the main purposes of inductive research in that it uncovered new themes and hypotheses on how a specific group uses gist MT and on the contextual factors that contribute to their use of it.

Further studies on this gist MT user group could target an expanded group of informants, including more diverse participants, other patent MT user groups, and less experienced patent MT users. Studies incorporating other methods such as contextual inquiry, diaries, or quantitative methods such as surveys could verify some of the findings of this study and might reveal further insights on this user group. In addition, it is hoped that we will see a growth in the research, and number of researchers, devoted to studying all types of users of gist MT.

References

Alberts, Doreen, Cynthia B. Yang, Ken Koubek, Suzanne Robins, Matthew Rodgers, Edlyn Simmons, and Dominic DeMarco. 2017. Introduction to Patent Searching. In Lupu, Mihai, Katja Mayer, Noriko Kando and Anthony J. Trippe, editors, *Current Challenges in Patent Information Retrieval*, Second ed. Vol. 29, Springer, Berlin Heidelberg.

Beninatto, Renato and Michael Stevens. 2019. What's the Latest with Neural MT? *Globally Speaking Radio* (podcast), episode 75. RWS Moravia.

Bowker, Lynne and Jairo Buitrago Ciro. 2019. *Machine Translation and Global Research: Towards Improved Machine Translation Literacy in the Scholarly Community,* Emerald Publishing, Bingley, UK.

Bowker, Lynne and Melissa Ehgoetz. 2007. Exploring User Acceptance of Machine Translation Output: A Recipient Evaluation. In Kenny, Dorothy and Kyongjoo Ryou, editors, *Across Boundaries: International Perspectives on Translation Studies*, Cambridge Scholars Publishing, Newcastle, UK.

Braun, Virginia and Victoria Clarke. 2013. *Successful Qualitative Research: A Practical Guide for Beginners.* SAGE, Los Angeles, California, USA.

———n.d. Thematic analysis: reflexive approach, accessed March-June, 2019. https://www.psych.auckland.ac.nz/en/about/our-research/research-groups/thematic-analysis.html

———2006. Using thematic analysis in psychology. *Qualitative Research in Psychology,* 3(2):77–101.

Burgett, Will. 2015. Unmoderated Remote Usability Testing of Machine Translation Content. *TAUS Review of Language Business and Technology,* IV:30–37.

Canfora, Carmen and Angelika Ottmann. 2018. Of ostriches, pyramids, and Swiss cheese: Risks in safety-critical translations. *Translation Spaces,* 7(2):167–201.

———2016. Who's afraid of translation risks? Presentation at 8th EST Congress, Aarhus, Denmark.

Church, Kenneth W. and Eduard H. Hovy. 1993. Good Applications for Crummy Machine Translation. *Machine Translation,* 8:239–258.

Dillinger, Mike and Laurie Gerber. 2009. Success with Machine Translation: Automating Knowledge-Base Translation, Part 1. *ClientSide News*, 10.

EPO: European Patent Office. 2018. Guidelines for Examination in the European Patent Office. European Patent Office, Munich, Germany.

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. *18th ACM Conference on Computer Supported Cooperative Work & Social Computing,* Vancouver, BC, Canada, 852–863.

Gaspari, Federico. 2004. Online MT Services and Real Users' Needs: An Empirical Usability Evaluation. *6th Conference of the Association for Machine Translation in the Americas, AMTA 2014,* Washington DC, USA 74–85.

Hassan, Hany, Anthony Aue, Chang Chen, Vishal Chowdhary, Jonathan Clark, Christian Federmann, Xuedong Huang, Marcin Junczys-Dowmunt, Will Lewis, Mu Li, Shujie Liu, Tie-Yan Liu, Renqian Luo, Arul Menezes, Tao Qin, Frank Seide, Xu Tan, Fei Tian, Lijun Wu, Shuangzhi Wu, Yingce Xia, Dongdong Zhang, Zhirui Zhang, and Ming Zhou. 2018. Achieving Human Parity on Automatic Chinese to English News Translation. https://arxiv.org/abs/1803.05567

Henisz-Dostert, Bozena. 1979. Users' evaluation of machine translation. In Winter, Werner, editor, *Machine Translation*. Mouton Publishers, The Hague.

ISO. 2010. Ergonomics of Human-System Interaction. Part 210, Human-Centred Design for Interactive Systems (ISO 9241-210:2010*).* Suomen standardisoimisliitto SFS, Helsinki, Finland.

Joho, Hideo, Leif Azzopardi, and Wim Vandervauwheded. 2010. A Survey of Patent Users: An Analysis of Tasks, Behavior, Search Functionality and System Requirements. *3rd Information Interaction in Context Symposium*, New Brunswick, NJ, USA 13–24.

Larroyed, Aline A. 2018. Machine Translation and Disclosure of Patent Information. *IIC - International Review of Intellectual Property and Competition Law,* 49(7):763–786.

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 Xu, Xiaoge, editor, *Impacts of Mobile use and Experience on Contemporary Society,* IGI Global, Hershey, PA, USA.

McDonald-Maier, Lisa. 2009. esp@cenet: Survey reveals new information about users. *World Patent Information,* 31(2):142–143.

Merriam, Sharan B. and Elizabeth J. Tisdell. 2016. *Qualitative Research: A Guide to Design and Implementation.* John Wiley & Sons, San Francisco, CA, USA.

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. *21st Annual Conference of the European Association for Machine Translation,* Alicante, Spain 199–208.

Oesch, Rainer, Heli Pihlajamaa, and Sami Sunila. 2014. *Patenttioikeus* [Patent Law]. 3rd ed. Talentum, Helsinki, Finland.

Pituxcoosuuvarn, Mondheera, Toru Ishida, Naomi Yamashita, Toshiyuki Takasaki, and Yumiko Mori. 2018. Machine Translation Usage in a Children's Workshop. *10th International Conference on Collaboration Technologies,* Costa de Caparica, Portugal, 57–73.

PRH: Patentti- ja rekisterihallitus [Finnish Patent and Registration Office]. 2018. *Patenttiopas* [Patent Guide].Patentti- ja rekisterihallitus, Helsinki, Finland.

Pym, Anthony. 2015. Translating as risk management. *Journal of Pragmatics,* 85(Aug):67–80.

Pym, Anthony and Kayo Matsushita. 2018. Risk Mitigation in Translator Decisions. *Across Languages and Cultures,* 19(1):1–18.

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.

Smith, Ross. 2003. Overview of PwC/Sytranet on-line MT Facility. *Twenty-Fifth International Conference on Translating and the Computer,* London, UK.

Stewart, Osamuyimen, David Lubensky, Scott Macdonald, and Julie Marcotte. 2010. Using Machine Translation for the Localization of Electronic Support Content: Evaluating End-User Satisfaction. *The 9th Conference of the Association for Machine Translation in the Americas,* Denver, Colorado, USA.

Suzuki, Hiroshi and Reiko Hishiyama. 2016. An Analysis of Expert Knowledge Transmission using Machine Translation Services. *Seventh Symposium on Information and Communication Technology,* 352–359.

Tinsley, John. 2017. Machine Translation and the Challenge of Patents. In Lupu, Mihai, Katja Mayer, Noriko Kando and Anthony J. Trippe, editors, *Current Challenges in Patent Information Retrieval,* Second ed. Vol. 29, Springer, Berlin Heidelberg.

Tinsley, John, Alexandru Ceausu, Jian Zhang, Heidi Depraetere, and Joeri Van de Walle. 2012. IP-Translator: Facilitating Patent Search with Machine Translation. *AMTA-2012: Tenth Biennial Conference of the Association for Machine Translation in the Americas,* San Diego, CA, USA.

Toral, Antonio, Sheila Castilho, Ke Hu, and Andy Way. 2018. Attaining the Unattainable? Reassessing Claims of Human Parity in Neural Machine Translation. *Third Conference on Machine Translation,* Brussels, Belgium 113–123.

Way, Andy. 2018. Quality Expectations of Machine Translation. In Moorkens, Joss, Sheila Castilho, Federico Gaspari & Stephen Doherty, editors, *Translation Quality Assessment: From Principles to Practice,* Springer, Cham, Switzerland.

Way, Andy. 2013. Traditional and Emerging use-Cases for Machine Translation. *Translating and the Computer,* London, UK.

WIPO: World Intellectual Property Organization. 2018. World Intellectual Property Indicators 2018. World Intellectual Property Organization, Geneva, Switzerland.

Yamashita, Naomi, Rieko Inaba, Hideaki Kuzuoka, and Toru Ishida. 2009. Difficulties in Establishing Common Ground in Multiparty Groups using Machine Translation. *CHI 2009, the SIGCHI Conference on Human Factors in Computing Systems,* Boston, MA, USA 679–688.

Yasouka, Mika and Pernille Bjorn. 2011. Machine Translation Effects on Communication: What Makes it Difficult to Communicate through Machine Translation? *IEEE,* Kyoto, Japan 110-115.