PRE-PRINT - AUTHORS' FINAL VERSION
To appear in: James Hadley, Kristiina Taivalkoski-Shilov, Carlos S. C. Teixeira and Antonio Toral (eds) *Using Technologies for Creative-Text Translation* by Routledge, 2022.

Translating with technology: How digitalisation affects authorship and copyright of literary texts

Maarit Koponen, University of Eastern Finland, ORCID: 0000-0002-6123-5386

Sanna Nyqvist, University of Helsinki, ORCID: 0000-0002-6046-2851

Kristiina Taivalkoski-Shilov, University of Turku, ORCID: 0000-0002-3299-0817

Abstract

This chapter examines questions of copyright, authorship and textual ownership in the context of literary translation, and particularly the effects of increasing use of translation technology. Overall, translation is an anomaly in terms of copyright, which seeks to protect the expressions of original works, and not the ideas contained in them. Translations, however, are derivative works where the ideas of the source text are expressed in new form. Since the 19th century, copyright regimes have struggled to accommodate the moral and economic rights of the translator and the author of the original text. More recently, new complications have been introduced by technologies like translation memories and machine translation, which are increasingly used in various domains. Such technological tools enable the reuse of translations as data, and machine translation in particular has the potential to extend this reuse to new contexts, including literary translation where translation technology has so far had less impact. The question of textual ownership becomes ambiguous in a technology-assisted translation process where a translator employs suggestions based on previous works. We examine the notions of textual ownership and authorship from legal, technical and ethical perspectives, and suggest possible solutions to the open questions.

1 Introduction

In copyright terms, translation is an anomaly. Copyright laws protect the expressions (not ideas) of original works, but translators create derivative works that use new expressions to convey the ideas and forms of originals. The copyright regime has struggled to accommodate the rights of translators to those of the original authors and to maintain the priority of the latter despite the inevitable transformation of expressive originality in translation (see Nyqvist 2018; Venuti 1995). A new complexity has arisen with digital resources, such as translation memories (TM) and machine translation (MT), that automatise the production of translated texts and multiply the collaborative aspect of translation, thus heightening the already problematic status of translation as intellectual property. The use of TMs, which enable the reuse of translations, raises questions of authorship and copyright, which are made even more complex when translated texts are used as training data for developing MT systems. The training corpora contain millions of translated segments originating from different sources, from TMs made available by international organisations to public domain literary works and webcrawled texts. A translation produced by an MT system presents an amalgamation of the corpora rather than a copy of any given translation.

To date, TM and MT have been used mainly in domains like localisation and administrative texts, but recently interest in translation technology also for literary translation has grown (e.g. Toral and Way 2015; Moorkens et al. 2018; Kenny and Winters 2020). In this chapter, we examine authorship and copyright issues and the potential effects of digitalisation and translation technology in literary translation. Similarly to the approach taken by Hadley (2017, 183), this chapter can be called meta-analytical in its nature: we review scholarship that

addresses questions of copyright and authorship involved in translation as a process and translations as data. Rather than focusing on specific case studies, this approach allows us to take a broad view of issues arising in the context of translation in general, and more particularly in the intersection of translation technology and literary translation.

In what follows, we discuss copyright issues and the notions of textual ownership vs authorship first from a legal and technical perspective, then moving on to ethical considerations. We start with a historical overview and analysis of the legal status of translation in the context of international copyright in Section 2. Section 3 examines translation technology in more detail, with a particular focus on the reuse of translations in TMs and MT training. Section 4 discusses the notion of authorship in literary translation. We focus on the ownership of the translation product, achieved with the help of translation technology and ask whose text is a translation rendered fully or partly with the help of translation technology. Finally, in Section 5 we reflect on the implications raised by the earlier sections, discussing open questions and possible solutions.

2 Translation and copyright

Copyright laws developed across Europe from the 18th century onwards in response to the need to control the production and distribution of books and the ideas contained in them. Early copyright laws aimed to regulate the book market by distributing rights between authors, printers and booksellers, to control the flow of information in the interest of censorship, and eventually to protect the rights of individual authors in order to encourage the creation of new works (Borghi 2017; Saunders 1992; Woodmansee 1984). Following the expansion of the international book trade in the 19th century, the ideas of copyright and *droit d'auteur* were extended to translations of literary works (Hemmungs Wirtén 2009). One aim of the first multilateral international copyright agreement, the Berne Convention of 1886, was to regulate the division of rights between the original author and the translator in order to determine the status of translations as intellectual property (Ricketson 1986; Ricketson and Ginsburg 2006).

2.1 The legal status of translation in the Berne Convention

The Berne Convention and its subsequent revisions harmonised key elements of the national copyright laws of its signatory states (176 states according to WIPO 2020). Influenced by the continental European legal tradition which emphasises the *droit d'auteur*, the treaty acknowledges both the moral and economic rights of authors. Moral rights include the right of attribution (the author's right to claim the authorship of his or her own work) and the right of integrity (the author's right to preclude modifications to his or her own work). These rights remain with the author even when the author surrenders the economic rights. Economic rights include the rights to reproduce and distribute works and the right to authorise derivative works, such as adaptations.

In the 19th century, it was unclear how translations should be classified, whether to regard them as reproductions or adaptations, where a permission from the original author was necessary, or independent creations outside the control of the original author (Hemmungs Wirtén 2009; Ricketson and Ginsburg 2006, 35). A main concern in the discussion on translations was also the benefit of the reading audiences: it was suggested (and indeed implemented in the 1886 Berne Convention) that the original author's right to authorise translations should be limited to a period of ten years from the publication of the original, allowing for free translation rights thereafter (Ricketson and Ginsburg 2006, 66-67).

The sections concerning translation and translation rights were repeatedly revised during the first decades of the Berne Convention, resulting in the still operative categorisation of translations as derivative works (Art 8), whereby original authors were granted the exclusive right of making or authorising translations during the period of protection (the minimum length currently prescribed by the Berne Convention is the author's lifetime plus 50 years) (Art 7 § 1). However, the Convention also acknowledges the creative input of translators in Art 2 § 3, which stipulates that translations and other kinds of adaptations "shall be protected as original works without prejudice to the copyright in the original work."[1]

Copyright in translation is thus based on the division of rights between the author of the translated work and the translator. The translator's copyright in his or her work does not limit the moral and economic rights of the translated author in the original work, and the publication and reproduction of the translation requires permission from the author. Conversely, the translator's copyright limits the original author's rights to the translation: for instance, the author may not reproduce or modify the translation without the translator's consent (Cabanellas 2014, 53).

2.1.1 Challenges of translation in relation to current copyright systems

The legal status of translation as codified in the Berne Convention and applied to the national laws of its signatories provided mechanisms for managing the problems of rampant piracy and unfair treatment of both authors and translators in international contexts, but it also manifests the difficulty of assimilating translation as creative activity in the traditional copyright framework. The reiteration of the key term "originality" in Art 2 § 3 cited above suggests dual standards of originality. By classifying translations as secondary, derivative works, the law ascribes originality in translation to the primary text and its author (Nyqvist 2018; see also European Commission[2]* 2014, 99).

The idea of transcendent originality, however, contradicts another fundamental principle of copyright protection, namely the division between ideas and expression. Copyright law grants protection only to manifestations of ideas, not the ideas themselves. The originality and basis for protection of literary works lies in their expression. Yet a translation is original in its expression, not the content, and the novelty of its expression could form an argument for refuting the original author's right over the translation. A commentary to the Berne Convention, the 1978 *WIPO Guide to the Berne Convention* defined translations as works that express an author's thoughts in a different language – an interpretation which recategorises the original work as belonging to the realm of ideas rather than expression (WIPO 1978, 19; see also Venuti 1998). Original and derivative at the same time, translation undermines the concept of original authorship as the foundation of copyright protection. Yet, as Cabanellas (2014, 54) notes, the practical reason for distinguishing original and derivative originals lies in the economics of translation: the value of the original surpasses that of a translation and hence "[t]he extension of the copyright in the original work to translations permits a market remuneration for translators and the international

[1] The current version of the Berne Convention for the Protection of Literary and Artistic Works (as amended on Sept. 27, 1979) can be accessed in https://wipolex.wipo.int/en/text/283698 (accessed 25 Jan 2021)

[2] The report "Translation and Intellectual Property rights", to which we refer as European Commission 2014, was commissioned by the European Commission to the law firm Bird & Bird LLP. The team of authors at Bird & Bird was led by Jean-Christophe Troussel and Julien Debussche. Hence some researchers refer to the text as Troussel & Debussche 2014. As it is a report issued by the European Commission, we prefer to attribute the text to the issuing organisation.

effectiveness of copyright, which is one of the underlying goals of the contemporary copyright regime."

In the 21st century, the rapid advance of MT technology, the reuse of translation data, and the increasing use of TMs as translation tools has further complicated the legal status of translation. Whereas copyright law is based on the notion of an original work attributable to an author, in the context of MT existing translations and their originals are transformed into bits of data in vast corpora used to automatically produce new translations. Machine-assisted and machine-generated translations thus challenge the human/machine division inherent in copyright laws, which grant protection only to original human achievements, not mechanical output.

As Section 3 illustrates in more detail, MT raises a wide array of copyright concerns. The production and harvesting of translation data, crediting and remuneration of authors and translators whose texts are utilised in the training of artificial intelligence, transformation of original works into bits of data, the creative input and authorship of the programmers involved in developing translation software (which may in itself merit copyright protection), eligibility of translation corpora to compilation protection (a type of copyright protection), and authorship and ownership of the translations produced by, or with the assistance of, MT are currently an object of speculation as hardly any case law exists and academic research is only now emerging. (It is telling that Cabanellas's thorough *The Legal Environment of Translation* from 2014 does not discuss MT.)

The prevailing uncertainty and lack of established codes of conduct cause difficulties to all stakeholders from original authors and translators to software developers and users of TM and MT tools. Strong enforcement of the existing, conflicting rights of different parties might result in the so-called anticommons problem (Moorkens and Lewis 2019, 476), where utilising valuable materials becomes impossible, which discourages the development of potentially beneficial technologies. Bringing MT from more information-oriented translation tasks to creative texts like fiction will undoubtedly electrify the discussion, as stakes are higher in the genres which highlight originality of expression and the role of the author. The increasing emphasis on Open Access publication both in public and private sectors as well as the possibility of more nuanced distribution of rights through mechanisms such as the Creative Commons licensing system[3] have opened new possibilities for reusing texts in TM and MT. Nevertheless, transitions between open and protected spheres require special attention to ensure fair treatment for both producers and users of translations.

3 Translation technology and its effects

Where the previous section provided an overview of copyright and translation, this section focuses on the effects of expanding digitalisation and technology use in translation workflows. Translation technology, which covers tools like TM and MT, appears to have had comparatively little impact so far in the literary translation context. Such tools are, however, commonplace particularly in technical translation and localisation, and increasingly used in many contexts. Translation technology depends on parallel corpora formed by alignments of source texts and their translations. Parallel texts have been collected in electronic format that enables the reuse of translations since approximately the 1980s (Simard 2019). In this section, we discuss how translations in the form of parallel texts are reused in TMs and MTs, and the implications for copyright and authorship.

[3] See https://creativecommons.org (accessed 20 Jan 2021).

3.1 Translation memories and machine translation systems

A TM is essentially a database consisting of parallel texts aligned as translation units (sentences or similar segments like titles and captions). When a translator uses the TM in a computer-assisted translation environment, the software segments the source text and compares the unit being translated against the source language units stored in the TM. If one or more similar segments are found, the software presents them to the translator as a suggestion to reuse the previous translation. The translator may accept a suggested segment, revise it as necessary or reject the suggestion completely and write a new one. The new or revised translation unit is then saved in the TM for potential reuse. As noted by Lewis et al. (2016, 1601), TMs often contain units from multiple source texts, multiple translators and even multiple clients because they are commonly reused between different projects by language service providers.[4] On cloud platforms, TMs may be shared with and utilised by multiple translators, particularly in crowdsourcing (Moorkens and Lewis 2019, 7), and translated target texts may in turn form source texts for indirect translation (Moorkens and Lewis 2020, 473).

Translation memory ownership and related questions of copyright are unclear. Even in the simplest case, where a TM contains only one source text translated by one translator, it involves elements with different copyright claims: the source text, the target text and the TM database itself (European Commission 2014, 128-129). The situation is further complicated by the inclusion of legacy data from previous projects. This introduces ownership claims involving each author of the source content and each translator of the target content already in the database, as well as potentially the language service provider or client through database rights covering the alignment of the translation units (Lewis et al. 2016, 1602-1603). During the translation project, new source and target content is added, with potentially new authors and translators. Producing the translation generally involves translation suggestions offered by the TM or a MT system (see discussion on MT below). Revising or post-editing these suggestions may give the translator a claim to copyright of the new translation, although this depends on the level of editing: the claim is weaker if very little editing is done (European Commission 2014, 102-103; Lewis et al. 2016, 1603). Although the translator is in most cases considered the author of the translation and explicit transfer of rights is therefore required if the translation is reused (European Commission 2014, 113-114), it is common practice in localisation, technical and administrative translation to deliver TMs to the language service provider or client as a "by-product of a translation effort" even if transfer of rights is not explicitly addressed in contracts (Moorkens and Lewis 2019, 7). In part, this is likely due to the fact that translators see value in benefiting from each other's work by sharing TMs, as Moorkens and Lewis (2019, 8) discuss. However, widespread freelancing and practices in the field may also disempower translators and limit their ability to assert intellectual rights to their work (Moorkens et al. 2016, 1).

While TMs reuse previous translations through suggesting similar segments, MT systems aim to translate new, previously "unseen" content. Since the early 2000s, MT has been dominated by data-driven approaches, first statistical MT and more recently neural MT. Both statistical and neural MT utilise machine learning, which involves compiling a corpus of data, generally some sort of pre-processing and annotation, and then training a learning algorithm based on the corpus (Eckart de Castilho et al 2019). MT relies on corpora of translated texts, although also monolingual data may be used. While a detailed description of MT technologies is not within the scope of this chapter (see e.g. Forcada 2017; Simard 2019), some notes about the processing of parallel texts are in order. In (phrase-based) statistical MT, the aligned source and target

[4] Some clients may, however, specifically forbid the inclusion of their texts into TMs used for other projects (European Commission 2014, 128).

segments in a corpus are chopped into sequences of consecutive words (termed phrases) and translation probabilities are calculated for pairs of source and target phrases. The translated segment is then generated by piecing together phrases to form the most likely candidate. Neural MT processes parallel texts as even smaller units – words or sub-word units (individual characters or sequences of characters) – and builds the translation by predicting the most likely next word taking into account the context of the full sentence (or sometimes multiple sentences). MT systems have mostly been trained for specific source and target language pairs, but more recent developments include multilingual systems (e.g. Tiedemann and Thottingal 2020). For these systems, parallel corpora from many different language pairs are used to train a single model to translate between any of the languages involved.

3.2 Translation corpora

Vast quantities of text are needed to train MT systems, and the most valuable for this purpose are aligned translations created by translators (Moorkens and Lewis 2020). Obtaining these resources, however, is complicated. When building a corpus of parallel texts, agreements are needed from the author and publisher of the original text, and the translator and publisher of the translation (De Clercq and Montero Perez 2010, 3384; Lewis et al. 2016). The complexity of the situation is observed, for example, by Bywood et al. (2013) who describe a corpus of subtitles collected as part of a research project. The project partners included subtitling companies that contributed data for MT development, but Bywood et al. (2013) discuss the difficulty of obtaining permissions, as copyrights were held by the client, not the subtitling company. Based on their experiences in compiling the multilingual Dutch Parallel Corpus intended for both research and commercial use, De Clercq and Montero Perez (2010, 3386) also state that most copyright holders were reluctant to donate texts and often stipulated that the texts should be unrecognisable and not downloadable in full, for example.

Some translation memories and parallel texts may be made available by public service institutions in the European Union, for example, with provision of reuse for any commercial or non-commercial purpose (Moorkens et al. 2016, 3). One example is the Europarl corpus of translated European parliament speeches (Koehn 2005). TMs may be shared also on specific vendor platforms like the TAUS Data Cloud, which has a common intellectual property agreement covering the corpora shared (Lewis et al. 2016, 1601). Corpora also include crowdsourced translations, such as the parallel corpus of TED talks and their translations created by volunteers and shared under a Creative Commons license (Cettolo, Girardi and Federico 2012), or the OpenSubtitles parallel corpus consisting of subtitle translations uploaded by the users of a website (Tiedemann 2012, 2215). Parallel corpora may also be built by "crawling" the internet. This involves automatically collecting texts from multilingual websites, detecting parallel documents (with the help of clues like URLs, html structure, content similarity and images), identifying their languages and aligning sentences in the documents (Toral et al. 2017, 1022-1023). Although such data can contain misidentified or misaligned translations, Toral et al. (2017, 1045-1046) suggest that it offers a solution for building parallel corpora particularly in under-resourced languages. For instance, the ParaCrawl corpus contains webcrawled parallel texts in 23 European languages, with numbers of aligned sentences ranging from 195,000 sentences for Maltese-English to over 30 million for French-English (Esplà-Gomis et al. 2019).

An example illustrating the scale of parallel corpora, the variety of sources and text types is provided by the OPUS site, which hosts freely accessible parallel corpora and tools intended to support research and development of MT (Tiedemann 2012). OPUS is the largest collection of openly available parallel corpora: in 2020, the site contained 57 parallel corpora covering more

than 700 languages and totalling over 9.2 million documents across all corpora and language pairs (Aulamo, Sulubacak and Virpioja 2020, 3782-3783). Types of texts in OPUS include administrative/legislative texts published by institutions like the European Commission Directorate General of Translation and United Nations, newspaper texts (for which translations are often compiled and released by research projects and evaluation campaigns like the annual WMT conferences[5]), religious texts, digitised books, open source localisation files, subtitles, and Wikipedia articles (Tiedemann 2012; Aulamo, Sulubacak and Virpioja 2020). The number of retrievable sentence pairs in each corpus generally ranges from tens of thousands up to tens of millions, although numbers may be smaller for specific corpora and language pairs (Aulamo, Sulubacak and Virpioja 2020, 3788).

The use of MT also involves unsolved questions of ownership and intellectual property rights. In general, the legal status and copyright of various corpora used for training MT and other natural language processing systems is unclear (Eckart de Castilho et al. 2019). The report on translation and intellectual property rights commissioned by the European Commission (2014, 115) states that the reuse of translations can constitute an infringement on the source author's or translator's rights. As noted above, the owner of the compiled TM may also have separate rights. Although extracting individual translation suggestions and using those to create a new translation would not be substantial enough to infringe on the rights of the database owner, the situation may be different for repeated and systematic extraction for example for system training (European Commission 2014, 124-125). Varying use scenarios – publishing translated content vs using TMs for new translation projects vs using aligned translations for MT training – may require different usage rights, which should be covered separately in contracts (Lewis et al. 2016, 1603; European Commission 2014).

3.3 Machine translation and literary texts

The textual similarities employed by TMs are generally useful only within narrow domains, and a TM is unlikely to provide for texts of a different genre, as noted by Moorkens et al. (2016). MT, however, extends the potential reuse of translations to other domains and genres, where strict text-based comparisons of segments might not provide usable matches. Due to the vast amounts of data required, MT system training generally combines parallel texts from different genres – for example, all the corpora available in OPUS for the relevant language pair. This may be followed by fine-tuning with a smaller amount of data in the relevant domain. Although literary texts may not be particularly useful in TMs, training corpora for MT systems used in various domains often contain public domain literary translations.

Recent work has also explored using MT for literary translation. Toral and Way (2015) experimented with a statistical MT system for translating a novel from Spanish into Catalan. Their MT model was first trained on approximately 630,000 sentences of parallel news texts, and then fine-tuned with around 22,000 sentences from two novels by the same author, with a Catalan language model trained on news, novels, and over 16 million sentences of web texts (Toral and Way 2015). In small-scale manual evaluation of 101 sentences, two readers ranked 60% of the MT sentences of equal quality or even higher than a human translated sentence (Toral and Way 2015, 129). Subsequent work reported by Toral, Wieling and Way (2018; see also Moorkens et al. 2018) addressed English-to-Catalan MT of a novel. This study compared both a statistical and a neural MT system trained on over a million parallel sentences from 133 translated novels, 400,000 sentences of parallel subtitles (statistical system only), and over 5 million sentences from more than 1,000 novels written in Catalan (Toral, Wieling and Way 2018,

[5] http://statmt.org/wmt20/

3). Translators participating in the study rated output of both systems as helpful, mostly fluent and adequate, with slight preference to the neural MT (Moorkens et al. 2018, 254-255). Although translated novels were used as training data, it is difficult to determine domain in the case of literary texts. As Toral, Wieling and Way (2018) state, their "in domain" material included novels from various different genres. Style of expression may, however, be very different in for example historical works compared to science fiction or contemporary romance novels, and in works by different authors.

4 The effect of technologisation on authorship

Even though translators' rights are partly protected by laws and conventions, their relation to the texts they create while translating can also be viewed from the perspective of general ethics of translation that investigates translators' rights, in addition to their duties (see e.g. Chesterman, 2001, 143). This section broadens the discussion to ethical questions concerning authorship and textual ownership in literary translation, which are actually age-old issues in translation theory. The line between creative writing and translating has never been clear-cut, since translators can take liberties when translating, original texts can be misleadingly presented as translations (pseudo-translations), and the author and translator can be the same person (self-translation). Additionally, translators' attitudes toward the author and the source text have varied from servility to hubris over the centuries (see e.g. Ballard 1995). This has affected their translation strategies and consequently the make-up and identity of the translated text.

4.1 Authorship vs ownership from an ethical perspective

Today's norms of translation and copyright laws give literary translators less room to manoeuvre compared to previous centuries, but even in copyright terms the status of the translated text is still ambiguous, as discussed above. The same applies to the translator's status, since no clear rules exist concerning what amount of visibility is suitable for translators vis-à-vis the source text author (see Flynn 2013). Even though some prominent translation scholars have encouraged translators to claim more authorship for their translations (see e.g. Venuti 1995), practicing translators tend to opt for less. This does not mean, however, that they do not claim ownership of their translation. Jansen (2019), who studied contemporary Scandinavian literary translators' attitudes toward the author and the source text, makes a useful distinction of two kinds of "mental" ownership that literary translation entails:

> **There might, in fact, be two kinds of belonging at play, namely ownership and authorship**. The translators [who responded to Jansen's questionnaire] do claim ownership, meaning thereby that it is 'their' text, because
>
> (a) they have **written or created** the translated text, for which they **take responsibility** […];
>
> (b) they hold the copyright to the translated text and **they want the final say in the editing process** – in other words, it's their text, not the publishers' (or as one respondents says, 'the translator is not the editor's servant'); and
>
> (c) the authors as a rule do not know the target language and are thus unable to value or control the translation.
>
> However, […] the large majority does not seem to claim ownership in the sense of authorship, that is of taking over the authorial role, replacing the original author, and rewriting the source text according to their own agenda. (Jansen 2019, 684; emphasis added)

Jansen's clarifications are pertinent not only for distinguishing between contemporary human authors and translators, but also for pondering whether a MT system used to translate a literary text without human involvement could be considered the owner of the target text similarly as a

human translator. According to European Commission (2014, 102-103), running a text through a MT system alone does not make a work creative and raw MT is not protected by copyright. Neither does the MT system meet the requirements for mental ownership of the translated text, as defined by Jansen (2019). Obviously, the system is unable to take responsibility for the target text or to have any say in the editing process. Nor does it need copyright or textual ownership: as we know, machines have no need to make a living or to gain symbolic capital. Copyright legislation and (inter)national copyright conventions have been created to protect the rights of flesh-and-blood authors and translators and to foster and reward their creativity (Nyqvist 2018, 10).

As to those individuals who create MT programs, the trend in this field appears to emphasise sharing the fruits of invention rather than claiming authorship or textual ownership. As discussed in Section 3.2, MT models are built using translation corpora and training algorithms. Toolkits, such as the MarianNMT framework[6] (developed and maintained by the Microsoft Translator team together with academic and commercial partners; see Junczys-Dowmunt et al. 2018), are commonly distributed as open source and can be used by any company, organisation or private user to train and deploy their own MT models. Trained MT models are also available as open-source software (see e.g. Tiedemann and Thottingal 2020). Business models for commercialising MT systems by language service providers or technology companies can take the form of selling clients the use of generic MT systems as a cloud-based service (often charged in terms of number of words translated), or services where a dedicated system is trained and maintained on behalf of the client. These practices would deserve an ethical investigation of their own.

The question of ownership of a literary translation is thornier in the case of machine-assisted human translation or human-assisted machine translation. If resources like TMs and MT (with their associated questions, see Section 3) were used collectively to create a new translation, determining the owner of the translated text becomes a complex issue with no clear, generally applicable solution. To date, such dilemmas are mainly theoretical owing to the practicalities of literary translation. Unlike other clients in the translation industry, publishing houses do not currently appear to impose the use of TMs in literary translation projects, not to mention reusing TMs between different projects and sharing them between translators. Professional literary translators tend to work alone or in pairs and if they resort to TMs, these are most probably in private use. A literary translator who utilises a legally purchased or free TM tool and creates a TM for a translation project does not transgress any ethical guidelines, especially if the database is not reused for translating other authors, which would increase the risk of homogenizing the voices of different authors to sound like a single author in the target language (see Taivalkoski-Shilov 2019, 697). As to post-editing machine-translated literary texts, copyright stipulates creative, human input, which raises potential questions when a translation is produced by recombining prior texts and involves post-editing by the translator (European Commission 2014, 102-103). The situation may be even more complicated if, instead of post-editing a static machine-generated text, the translator is working with a so-called interactive system, which adapts the suggestion shown according to edits made by the translator and may even learn from those changes in real time to generate translations adapted to the specific text or the specific user (see Peris and Casacuberta 2019). Selecting and post-editing MT suggestions "could give rise to copyright protection in case the translator would be able to imprint his [sic] personality and make such work original", as concluded in the report by European Commission (2014, 103). Consequently, even a translator – or other agent of literary

[6] Available at: https://marian-nmt.github.io/

translation, such as a line editor in charge of post editing – who post-edits could claim textual ownership for a literary translation, depending on how much the post-editor's "personality" has left a mark on the translated text.

4.2 Machine-assisted literary translation and multiple translatorship

In fact, machine-assisted literary translation presents a particular case of what Jansen and Wegener (2013) have termed *multiple translatorship*. This concept has been inspired by Stillinger's notion of *multiple authorship* that underlines the collaborative nature of artistic writing. Traditionally, literary works officially attributed to a single author or translator have been discussed as if no one other than the source text author or the translator had a share in the (re)creation of the text. Nevertheless, as Stillinger (1991, v-vi and *passim*) and Jansen and Wegener (2013) have convincingly argued, most (literary) texts, whether original or translations, tend to be the fruit of conscious joint, composite or collaborative production of some sort.

The notion of multiple translatorship is helpful in reminding that the technology-assisted translation process is not exceptional in introducing more voices to the translated text: textual fragments from previous human translations, machine-generated suggestions and interactively created ad hoc translation solutions. As Alvstad et al. (2017, 4) suggest, "translation is a matter of circulation of and confrontation between voices". From the perspective of textual ownership, what matters is the creative input, textual design as a whole, and moral responsibility for the text. Having said that, to counter translators being gradually pushed to ever more ancillary roles in technology-assisted literary translation processes, it is necessary to highlight their role in translation technology development and create new ways to foster and reward their creativity. Such practices would be in line with the "human-centric approach" to artificial intelligence adopted by the EU, and the ethical principles of this approach, one of which is fairness and "ensuring equal and just distribution of both benefits and costs" (European Commission 2019, 12).

5 Discussion and concluding remarks

The advent of translation technology in literary translation provides an opportunity to redefine the translator's relationship to the source text author. Separating authors and translators dichotomously has been commonplace even though, as Jansen (2019, 684) argues, the distinction between them "may occasionally be blurred and should perhaps be seen rather as a continuum" (see also Walkowitz 2015). Sometimes authors and translators even appear as antagonists, for instance when Venuti, criticizing current copyright laws as being unfair to translators, writes:

> From the viewpoint of translators and translation, these limitations [on the translator's control of the translated text] carry some troubling consequences, both economic and cultural. By subordinating the translator's rights to the author's, the law permits the author to shrink the translator's share in the profits of the translation. (Venuti 1998, 47)

And yet, the millennial relationship between authors and translators has covered a wide range of differing attitudes and forms of interaction. Jansen's (2019, 678) recent study indicates that some translators interact with their source text authors simply for the purpose of bonding. Several of Jansen's respondents wrote that they cherish friendships with their authors. Authors' attitudes to their translators have been less studied (for some exceptions, see Chesterman 2004/2017; Greenall 2019, 653-655). According to Washbourne (2017, 25), "[t]he writer may view translation as a threat to the original and to his or her identity, or, on the opposite pole, as a transformative, life-extending experience." Judging by Washbourne's (2017, 16 and passim)

article, the experience of *being translated* tends to be upsetting in one way or another. However, Washbourne's study did not include authors whose text has been machine translated or crowdsourced from amateur translators. Authors' attitudes toward these new forms of translation probably vary (see e.g. the case of Richard Powers in Besacier 2014). Still, it is likely that authors prefer professional human translators over these new options that have appeared with the technologisation of translation.

A hypothesis can be formed that, from the original authors' perspective, human translators stand for quality. The termination of the author's (and subsequent rightholders') control with the expiration of the copyright opens up possibilities for myriad translation practices, among them MT. In the context of literature, MT has been applied to texts in the public domain by some businesses to publish print-on-demand versions of classics of world literature into a variety of languages. Such machine-translated texts have undergone little, if any, post-editing resulting in barely legible "translations" (see Taivalkoski-Shilov 2019, 699n5). Copyright laws present no impediments for such predatory publishing, except in Denmark, Finland, Norway and Sweden, where copyright protection is extended to works that can be considered classics (Fredriksson 2019, 10-12). If they are transformed or republished publicly in a form that "violates cultural interests", to quote the Finnish Copyright Act (51 § 1), such practices can be prohibited by the authorities. In Finland, the landmark ruling on the protection of classics concerns precisely translations of classic literature which the court deemed to be of inferior quality (Korkein oikeus [The Supreme Court of Finland] 1967 KKO 1967 II 10). As the example regarding the protection of classics demonstrates, copyright laws sometimes provide alternative models for the author-centred and temporally limited basic form of copyright that can be problematic vis-à-vis translations, as we have argued. Yet the protection of classics and its possibilities for more extensive and ethically sensitive modes of copyright protection are limited to the Nordic countries and cannot therefore address the global trend of expanding MT use for publishing literary classics.

Indeed, rethinking the protection and fair treatment of all the stakeholders in translation in the digital age requires more than minor adjustment of existing national and international copyright laws. The internationalisation of book trade in the 19th century led to substantial copyright reforms and to international regulation that is still in force today (the Berne Convention and the inclusion of most of its statutes in the WTO-governed TRIPS treaty from 1995 onwards). The rapid digitalisation of literature and translation from the late 20th century onwards represents a similarly urgent challenge for the regulation of intellectual property. However, literature and its translation are marginal areas in the world-wide trade on intellectual properties and the copyright regimes have proved very resistant to attempts at fundamental revision. Therefore, it seems unlikely that legal copyright alone could provide a framework for the regulation of translation in the digital age.

One measure for addressing the situation is through contracts. Various guidelines for translation contracts emphasise that the translator retains copyright to the translation and that the translation can only be used for the purpose agreed upon in the contract. Moorkens and Lewis (2020, 472-473) discuss this perspective in recommendations by the Netherlands Association of Interpreters and Translators, for example. The principle underpins also the guidelines by the Finnish Translators' and Interpreters' Association (SKTL n.d.), and by the European Council of Literary Translators' Associations (CEATL 2018), among others. Although such guidelines recommend explicit written agreements covering the ownership of translation data and the transfer of rights between the translator and client, contracts tend to ignore copyright issues, as shown by the review of translation contracts from different countries reported in European Commission (2014, 133). Furthermore, even when contracts cover transfer of rights, the

translator may not realise the translation could be further shared as part of TMs or other translation data (Drugan and Babych 2010, 8). Contracts should therefore define which rights are transferred and how the "translator's work products" may be exploited, as well as the type and form of authorised exploitation, duration and compensation (European Commission 2014, 133). The guidelines addressing literary translation also stipulate that contracts "must not call for the summary transfer of all rights", rather, "each licensed right shall be mentioned in the contract", and the "right to exploit the work through technologies that do not yet exist" would not be transferred (CEATL 2018). Although the literary translation guidelines do not address uses such as MT evaluation and training, other recommendations have been made that these should be specified separately, including in cases where post-edited MT is used (Lewis et al. 2016, 1603).

In addition to transfer of rights in contracts, compensation can also be considered. In the literary field, it is considered fair practice that the translator is paid royalties, as well as "a share of the profits derived from secondary uses" (CEATL 2018). Although this refers to uses like e-books and audiobooks, a parallel could be drawn to a secondary use of the translation as data. Drawing on the 1976 UNESCO *Recommendation on translators' status and rights*, Drugan and Babych (2010, 8) suggest that translation contracts should include a provision for supplementary payment if the translation is to be used in a way not originally specified in the contract. It is important to note, however, that these guidelines focus on contracts between the translator and client. They do not address situations where translations are exploited by a third party. In particular, tracking an individual translator's contribution is often impossible, because identifying metadata is typically removed when TMs and similar resources are shared (Moorkens et al. 2016, 3). This situation could perhaps be addressed by suggestions regarding digital knowledge commons and the possibility of non-exclusive data ownership right, which would enable translators to record their contribution to a dataset, thereby asserting partial data ownership, and provide for more sustainable control of translation resources and compensation (Moorkens and Lewis 2019, 11; 2020, 476).

The questions concerning copyright, authorship and textual ownership in the technologised literary translation process are very complex, with no clear answers. On one hand, it is crucial that the rights and needs of flesh-and-blood authors and professional translators continue to be recognised and protected. On the other hand, it is important to allow the development of new technologies and foster the creativity of the parties that contribute to them. Professional literary translators can also benefit from the advances of technology, and technology can widen our understanding of the processes of (literary) translation. For a more sustainable future, the rights and needs of different stakeholders of translation – from authors to translators, developers of technological solutions to the various users of translation products and technologies -– should be taken into account. Copyright laws are clearly insufficient as means of regulation, and in practice, copyright tends to protect those who have power while it is the weak that need protection and the financial security copyrights can offer. The current uncertain circumstances clearly manifest the necessity for a more comprehensive and circumspect approach to translation in the digital age.

References

Alvstad, Cecilia, Annjo K. Greenall, Hanne Jansen, and Kristiina Taivalkoski-Shilov. 2017. "Introduction: Textual and contextual voices of translation." In *Textual and contextual voices of translation*, edited by Cecilia Alvstad, Annjo K. Greenall, Hanne Jansen, and Kristiina Taivalkoski-Shilov, 3–17. Amsterdam: John Benjamins.

Aulamo, Mikko, Umut Sulubacak, and Sami Virpioja. 2020. "OpusTools and Parallel Corpus Diagnostics." In *Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020)*, 3782–89.

Ballard, Michel. 1995. *De Cicéron à Benjamin. Traducteurs, traductions, réflexions*. Lille: Presses Universitaires de Lille.

Besacier, Laurent. 2014. "Traduction automatisée d'une oeuvre littéraire: Une étude pilote." In *Proceedings of TALN 2014 (Volume 2: Short Papers)*, 389–94.

Borghi, Maurizio. 2017. "Copyright and the Commodification of Authorship in 18th- and 19th-Century Europe." In *Oxford Research Encyclopedia of Literature*. New York: Oxford University Press USA. https://doi.org/10.1093/acrefore/9780190201098.013.268

Bywood, Lindsay, Martin Volk, Mark Fishel, and Panayota Georgakopoulou. 2013. "Parallel Subtitle Corpora and their Applications in Machine Translation and Translatology." *Perspectives: Studies in Translatology* 21 (4): 595–610. https://doi.org/10.1080/0907676X.2013.831920.

Cabanellas, Guillermo. 2014. *The Legal Environment of Translation*. Abingdon: Routledge.

CEATL. 2018. "Guidelines for Fair Translation Contracts." https://www.ceatl.eu/wp-content/uploads/2018/05/Guidelines_for_fair_contracts.pdf.

Cettolo, Mauro, Christian Girardi, and Marcello Federico. 2012. "WIT3: Web Inventory of Transcribed and Translated Talks." In *Proceedings of the 16th Annual Conference of the European Association for Machine Translation, EAMT 2012*, 261–68.

Chesterman, Andrew. 2001. "Proposal for a Hieronymic Oath." *The Translator*, 7 (2): 139–154.

Chesterman, Andrew. 2004/2017. "Kundera's sentence." *Reflections on Translation Theory. Selected Papers 1993-2014*, 281–294. Amsterdam: John Benjamins.

Copyright Act of Finland. 1961/2015. English Translation. https://www.finlex.fi/fi/laki/kaannokset/1961/en19610404.pdf

De Clercq, Orphée, and Maribel Montero Perez. 2010. "Data Collection and IPR in Multilingual Parallel Corpora: Dutch Parallel Corpus." In *Proceedings of the 7th International Conference on Language Resources and Evaluation, LREC 2010*, 3383–88.

Drugan, Jo, and Bogdan Babych. 2010. "Shared Resources, Shared Values? Ethical Implications of Sharing Translation Resources." In *JEC 2010: Second Joint EM+/CNGL Workshop "Bringing MT to the User: Research on Integrating MT in the Translation Industry", AMTA 2010*, 3–9.

Eckart de Castilho, Richard, Giulia Dore, Thomas Margoni, Penny Labropoulou, and Iryna Gurevych. 2019. "A Legal Perspective on Training Models for Natural Language Processing." In *LREC 2018: 11th International Conference on Language Resources and Evaluation*, 1267–74.

Esplà-Gomis, Miquel, Mikel Forcada, Gema Ramírez-Sánchez, and Hieu Hoang. 2019. "ParaCrawl: Web-Scale Parallel Corpora for the Languages of the EU." In *Proceedings of Machine Translation Summit XVII Volume 2: Translator, Project and User Tracks*, 118–19.

European Commission. 2014. "Translation and Intellectual Property Rights: Final Report." Luxembourg: European Union. https://doi.org/10.2782/72107.

European Commission. 2019. "Ethics Guidelines for Trustworthy AI: Independent High-Level Expert Group on Artificial Intelligence." Brussels: European Commission.

Flynn, Peter. 2013. "Author and translator." In *Handbook of Translation Studies* online. Vol. 4. Amsterdam: Benjamins. https://doi.org/10.1075/hts.4.aut1

Forcada, Mikel L. 2017. "Making Sense of Neural Machine Translation." *Translation Spaces* 6 (2): 291–309. https://doi.org/10.1075/ts.6.2.06for.

Fredriksson, Martin. 2019. "Between Intellectual and Cultural Property: Myths of Authorship and Common Heritage in the Protection of Traditional Cultural Expressions." *Cultural Analysis* 17 (1): 1–24.

Greenall, Annjo K. 2019. "The discursive (re-)construction of translational ethics." *Perspectives*, 27 (5): 648–663.

Hadley, James. 2017. "Indirect translation and discursive identity: Proposing the concatenation effect hypothesis." *Translation Studies* 10 (2): 183–97.

Hemmungs Wirtén, Eva. 2009. *Cosmopolitan Copyright: Law and Language in the Translation Zone*. Uppsala: Uppsala Universitet. http://uu.diva-portal.org/smash/get/diva2:458015/FULLTEXT01.pdf

Jansen, Hanne. 2019. "I'm a translator and I'm proud: How literary translators view authors and authorship." *Perspectives* 27 (5): 675–88.

Jansen, Hanne, and Anna Wegener. 2013. "Multiple translatorship." In *Authorial and Editorial Voices In Translation*, edited by Hanne Jansen and Anna Wegener, 1–42. Vita Traductiva 2 & 3. Montreal: Éditions québécoises de l'oeuvre.

Junczys-Dowmunt, Marcin, Roman Grundkiewicz, Tomasz Dwojak, Hieu Hoang, Kenneth Heafield, Tom Neckermann, Frank Seide, Ulrich Germann, Alham Fikri Aji, Nikolay Bogoychev, André F.T. Martins, and Alexandra Birch. 2018. "Marian: Fast Neural Machine Translation in C++." In *Proceedings of ACL 2018, System Demonstrations*, 116–21.

Kenny, Dorothy, and Marion Winters. 2020. "Machine translation, ethics and the literary translator's voice." *Translation Spaces* 9 (1): 123–49.

Koehn, Philipp. 2005. "Europarl: A Parallel Corpus for Statistical Machine Translation." In *Proceedings of the 10th Machine Translation Summit (MT Summit X)*, 79–86.

Lewis, David, Kaniz Fatema, Alfredo Maldonado, Brian Walshe, and Arturo Calvo. 2016. "Open Data Vocabularies for Assigning Usage Rights to Translation Memories." In *Proceedings of the 10th International Conference on Language Resources and Evaluation, LREC 2016*, 1601–9.

Moorkens, Joss, and Dave Lewis. 2019. "Research Questions and a Proposal for the Future Governance of Translation Data." *Journal of Specialised Translation* 32: 2–25.

Moorkens, Joss, and Dave Lewis. 2020. "Copyright and the Re-Use of Translation as Data." In *The Routledge Handbook of Translation and Technology*, edited by Minako O'Hagan, 469–81. London: Routledge.

Moorkens, Joss, Antonio Toral, Sheila Castilho, and Andy Way. 2018. "Translators' Perceptions of Literary Post-Editing Using Statistical and Neural Machine Translation." *Translation Spaces* 7 (2): 240–62. https://doi.org/10.1075/ts.18014.moo.

Moorkens, Joss, David Lewis, Wessel Reijers, Eva Vanmassenhove, and Andy Way. 2016. "Translation Resources and Translator Disempowerment." In *Proceedings of ETHI-CA^2 2016: ETHics In Corpus Collection, Annotation & Application*, 49–53.

Nyqvist, Sanna. 2018. "Literature and International Copyright After the Berne Convention." In *Oxford Research Encyclopedia of Literature*. New York: Oxford University Press USA. DOI: 10.1093/acrefore/9780190201098.013.267

Peris, Álvaro, and Francisco Casacuberta. 2019. "Online Learning for Effort Reduction in Interactive Neural Machine Translation." *Computer Speech & Language* 58: 98–126.

Ricketson, Sam. 1986. "The Birth of the Berne Union." *Columbia-VLA Journal of Law and the Arts* 11 (9): 9–32.

Ricketson, Sam, and Jane C. Ginsburg. 2006. *International Copyright and Neighbouring Rights: The Berne Convention and Beyond.* Vol. 2nd ed. Oxford: Oxford University Press.

Saunders, David. 1992. *Authorship and Copyright*. London: Routledge.

Simard, Michel. 2019. "Building and Using Parallel Text for Translation." In *The Routledge Handbook of Translation and Technology*, edited by Minako O'Hagan, 78–90. London: Routledge. https://doi.org/10.4324/9781315311258-5.

SKTL. n.d. "General Terms of Contract for Translations." Accessed 29 January 2021. https://www.sktl.fi/@Bin/81708/General+Terms+of+Contract+for+Translations.pdf

Stillinger, Jack. 1991. *Multiple Authorship and the Myth of Solitary Genius*. New York: Oxford University Press.

Taivalkoski-Shilov, Kristiina. 2019. "Ethical issues regarding machine(-assisted) translation of literary texts." *Perspectives* 27 (5): 689–703.

Tiedemann, Jörg. 2012. "Parallel Data, Tools and Interfaces in OPUS." In *Proceedings of the 8th International Conference on Language Resources and Evaluation, LREC 2012*, 2214–18.

Tiedemann, Jörg, and Santhosh Thottingal. 2020. "OPUS-MT-Building Open Translation Services for the World." In *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation*, 479–80.

Toral, Antonio, Miquel Esplá-Gomis, Filip Klubička, Nikola Ljubešić, Vassilis Papavassiliou, Prokopis Prokopidis, Raphael Rubino, and Andy Way. 2017. "Crawl and Crowd to Bring Machine Translation to Under-Resourced Languages." *Language Resources and Evaluation* 51 (4): 1019–51. https://doi.org/10.1007/s10579-016-9363-6.

Toral, Antonio, and Andy Way. 2015. "Translating Literary Text between Related Languages Using SMT." In *Proceedings of NAACL-HLT Fourth Workshop on Computational Linguistics for Literature*, 123–32. https://doi.org/10.3115/v1/w15-0714.

Toral, Antonio, Martijn Wieling, and Andy Way. 2018. "Post-Editing Effort of a Novel With Statistical and Neural Machine Translation." *Frontiers in Digital Humanities* 5: 2–12. https://doi.org/10.3389/fdigh.2018.00009.

Venuti, Lawrence. 1995. "Translation, Authorship, Copyright." *The Translator* 1 (1): 1–2.

Venuti, Lawrence. 1998. *The Scandals of Translation: Towards an Ethics of Difference*. London: Routledge.

Walkowitz, Rebecca. 2015. *Born Translated: The Contemporary Novel in an Age of World Literature*. New York: Columbia University Press.

Washbourne, Kelly. 2017. "'Am I Still There?: On the Author's Sense of Becoming in a New Language." *The AALITRA Review: A Journal of Literary Translation* 12: 16–30.

WIPO. 1978. *Guide to the Berne Convention for the Protection of Literary and Artistic Works*. Geneva: World Intellectual Property Organisation.

WIPO. 2020. "Contracting Parties > Berne Convention." https://wipolex.wipo.int/en/treaties/ShowResults?start_year=ANY&end_year=ANY&search_what=C&code=ALL&treaty_id=15

Woodmansee, Martha. 1984. "The Genius and the Copyright: Economic and Legal Condition of the Emergence of the 'Author.'" *Eighteenth-Century Studies* 17 (4): 425–48.