

Author-accepted manuscript. To cite, please use the following reference:

Kenny, Dorothy and Marion Winters (2024) 'Customization, Personalization and Style in Literary Machine Translation' in Marion Winters, Sharon Deane-Cox and Ursula Böser (eds) *Technological Change in Translation and Interpreting*. Bloomsbury Advances in Translation, London: Bloomsbury, 59-79.

Customization, Personalization and Style in Literary Machine Translation

Dorothy Kenny and Marion Winters

Introduction

Literary machine translation has come in from the cold. Once the preserve of the avant-garde (see Lennon 2014; Kenny 2020), it is now a focus of attention in mainstream machine translation (MT) and translation studies circles. As interest in the field has grown, so too has the variety of questions addressed and the range of methodologies adopted. Some studies follow well-trodden paths in MT research,¹ focusing on the automatic and human evaluation of MT outputs (e.g. Toral and Way 2018). Others also draw on familiar traditions in translation studies, homing in, for example, on putative universal features of post-edited machine translated texts (Toral 2019; Castilho and Resende 2022). But only a small body of research has so far integrated the figure of the literary translator, to consider how they might experience literary MT, how the use of MT might affect their style, and what a shift to MT might mean for their oeuvre. In a previous study (Kenny and Winters 2020), we began this line of enquiry. In this chapter we continue in this vein by laying the groundwork for further investigations of translator/post-editor style in scenarios where MT is used. In our ongoing research (Kenny and Winters 2020; Winters and Kenny forthcoming), we are particularly interested in how the style of one prominent literary translator, Hans-Christian Oeser, is affected when he works in post-editing mode. Our broad investigation obviously relies on empirical observation of texts post-edited by Oeser, but it also requires us to have a clear understanding of what we mean by “style” to begin with. At the same time, we are keenly aware that since we began our studies of translator style (see, for example, Kenny 2001, Winters 2007), the world has changed: MT has improved to the extent that its use in literary translation is no longer principally associated with the production of disfluencies and comic effects (see Kenny 2020); MT researchers have become more and more interested in the translation of literature (e.g. Voigt and Jurafsky 2012; Toral and Way 2018); and there is increasing focus in general on the customization of MT engines to make them more suitable for specific tasks, such as the translation of specific literary genres (Hansen 2021) or even the translation of works by a specific author, based on previous translations by a specific translator (Hansen 2022). As the horizons of MT researchers expand, so too does their interest in notions like style, and their understanding of this concept begins to feed into an already crowded space. Meanwhile, long-established understandings of style in the field that we call home, namely corpus-based translation studies, also need to be revisited given the more expansive treatment the concept has recently received in digitally-informed literary

¹ By ‘MT research’ we mean research conducted primarily by computer scientists specializing in MT.

studies (Herrmann et al. 2015). In short, before we proceed to further empirical investigation of translator style in post-edited texts we need to take stock.

We thus begin our exploration in the current chapter by surveying recent work in literary MT, drawing a tentative map of the various approaches taken in the field. We home in on studies that address the related concepts of customization, personalization and style in the contexts of machine and human translation, before making the case for the adoption of a literary studies approach to style in MT research and characterizing literary post-editing as a kind of downstream translator-specific personalization. We conclude, by way of illustration, by giving brief details of one study which has already taken this approach.

Literary Machine Translation

We understand “literary machine translation” as an emerging interdisciplinary field that embraces a range of phenomena related to the application of MT to the translation of literary texts. It touches upon or intersects with, among other areas: speculative and empirical enquiry into the features that make literary texts particularly amenable to or (more often) difficult for MT (e.g. Voigt and Jurafsky 2012; Taivalkoski-Shilov 2019a); computational techniques that might meet the attendant challenges (e.g. Van de Cruys 2020); the application of generic MT to literary texts (e.g. Fonteyne et al. 2020); the customization of MT systems for use with literary texts (e.g. Toral and Way 2018; Kuzman et al. 2019; Hansen 2021, 2022); the systematic evaluation of literary texts translated by MT (e.g. Toral and Way 2018; Moorkens et al. 2018; Matusov 2019); literary translators’ interactions with MT (Kenny and Winters 2020; Oeser 2020); reaction to the use of MT in the literary translation profession (Oeser 2020); readers’ reception of machine translated literature (Guerberof-Arenas and Toral 2020, 2022); style in human, raw and post-edited literary MT (Kenny and Winters 2020; Lee 2021); and studies of so-called “universals” in post-edited literary texts (e.g. Castilho and Resende 2022). It would clearly be impossible to do justice here to all these interweaving strands. In what follows we focus on studies that engage in particular with questions of customization, personalization and style in literary MT.

Customization, Personalization and Style in Literary MT

The current dynamism of the field shows that literary translation is now seen as a valid object of study in MT circles, having previously been considered beyond the reach of the technology. The increased attention given to literary MT coincided with the step change in quality that accompanied the transition first from rule-based to statistical MT and then, more rapidly, from statistical to neural MT in the mid-2000s and mid-2010s respectively. A detailed discussion of how and when these transitions came about is beyond the scope of this chapter. The reader is referred instead to treatments in Bentivogli et al. (2016), Kenny (2018) and Hansen (2021, 2022), among others. Suffice it to say here that statistical MT (SMT) and neural MT (NMT) not only offered superior quality to their respective predecessors, based on both human and automatic evaluations of their outputs, they also allowed *customization*. MT engines could now be trained specifically on *literary* texts from which they would

presumably learn mathematical models of *literary* translation. Some early research in the area (e.g. Toral and Way 2015a, 2015b) focused on adapting generic SMT engines by adding relatively modest quantities of literary text to their training data, yielding “literary-adapted MT” (Toral and Way 2015b). As efforts scaled up, and deep learning became embedded in MT, larger corpora of literary texts — usually novels — and their translations were used to train NMT engines that could now claim to be doing fully-fledged “literary MT”, even without the support of “out-of-domain” (i.e. non-literary) data.² At the time of writing, custom NMT engines designed for use with literary texts may be trained on both generic and literary text or on literary text alone. The NMT engines reported on in Guerberof and Toral (2022) serve as a case in point: the English-to-Catalan engine is trained on both out-of-domain and in-domain (literary) text, while the English-to-Dutch engine is trained exclusively on literary text.³

Customization can take even more specific forms: texts written by particular authors and translations by particular translators can be integrated into the training data for a given engine, steering literary MT towards increasing personalization. What we call “translator-specific personalization” is mooted in Oeser (2020: 23), while Kuzman et al. (2019) experiment with both author- and translator-specific training data in literary MT, thus yielding both author- and translator-specific personalization, albeit on a very small scale using a single translated text as parallel data, and with limited success. Likewise, Hansen (2022) reports on an MT engine trained on six novels produced by one particular author and translated by one particular translator working in the heroic fantasy genre. The genre appears to be especially problematic for MT, whether generic or custom (ibid.), but Hansen still concludes that early experiments point to the potential of translator-specific personalization in literary MT. Such personalization based on training data appears to us to be a promising way to give “textual voice” (Alvstad et al. 2017) to human translators in the context of literary MT, following an ethical imperative that we have explored in previous work (Kenny and Winters 2020).

Generic systems have, meanwhile, proved capable of competing with and sometimes outperforming customized literary systems – even (modestly) personalized ones – in research settings (Kuzman et al. 2019; Matusov 2019), no doubt because of the sheer size of the

² Toral and Way’s (2018) early experiments with NMT, for one, were based on training data composed entirely of novels translated from English into Catalan, and a monolingual corpus of novels in Catalan.

³ “Generic” engines are understood here as MT engines that are trained on sentences taken from texts covering “a wide range of topics, styles and genres, and not specialized in any particular domain” (Ramírez-Sánchez 2022). These are typically contrasted with “custom,” “domain-adapted” or “bespoke” engines, which are designed for use with specific genres in specific domains (e.g. legislation, patents, novels, etc.). The distinction between “systems” and “engines” meanwhile, roughly follows that made in Kenny (2022: 45): an MT system is understood as an MT product or service that is made available by a single supplier or developer. Familiar examples include Google Translate and Microsoft Translator. Many such systems often come in a free online version, yielding the acronym FOMT. An MT engine, on the other hand, is an MT “program (or even a “model”) that has been trained to deal with a particular language pair and, often, domain or genre” (Kenny ibid.). A single system provider may thus offer several different generic and/or custom engines. In cases where it is not necessary to make the distinction between system and engine or model, the term “system” serves as a useful superordinate.

training data they learn from. More tellingly perhaps, the improved performance that accompanied the rise of NMT has meant that even generic MT has reached a level of quality at which some literary translators can begin to countenance experimenting with it (see Kenny and Winters 2020: 133; Oeser 2020:20). This opens the door to an alternative approach to translator-specific personalization in our view: rather than changing the training data in an upstream effort to produce translations that reflect the preferences or style of a particular translator (although we do not deny the merit of this approach), it is also feasible to personalize downstream. In this case a literary translator would post-edit the output of a generic MT engine, so that it conforms to their expectations of what a good translation of the source should look like. In so doing, we hypothesize, they will put their own “stamp” on the final translation. What we are suggesting is that a literary translation created by a machine and post-edited by an experienced human literary translator will contain elements that reflect the style of that translator. If we have independent evidence as to what constitutes the translator’s style, then we should be able to show empirically whether their style as a post-editor coincides with their style as a translator. We characterize our approach as involving “downstream translator-specific personalization.”

Before describing precisely how we proceed, we revisit some key concepts in literary MT, addressing the motivations behind “literary-adapted” MT, and especially translator-specific personalization based on training data, problematizing in particular the understanding of style in MT research.

The Interaction between Corpus, Domain and Style in Machine Translation Research

The basic idea behind any kind of customization or “domain adaptation” in MT is that the best translation engine will be the one that has been trained (or fine-tuned) on “relevant” data (Koehn 2020: 239).⁴ Relevance here suggests that the texts in the corpus used to train the system share characteristics with the texts the system will subsequently be used to translate.⁵ The term “domain” is used very broadly in MT research, however: texts from the same domain might be expected to share “similar topic[s], style, level of formality, etc.” according to Koehn (2020: 239), but in practice it is typically the *provenance* of a corpus that marks it as belonging to a given domain, or in Koehn’s (ibid.) words, “domain...typically means a corpus that comes from a specific source.” The assumption in much of the work in MT seems to be that similar “topic, style, level of formality, etc.” (ibid.) can be inferred from provenance.

In this light, the treatment of “literary texts” as a domain in literary MT seems problematic: there is no single source (e.g. author, publisher or even country) from which the 133 English-language novels in the parallel corpus used in Toral and Way (2018) might be said to hail, for example. The same can be said of the 1,000 books in the Catalan and the 1,600 books in the

⁴ A discussion of how domain adaptation is achieved in practice (e.g. through fine-tuning, data interpolation, etc.) is beyond the scope of this chapter. The reader is referred instead to Koehn (2020: 239-261).

⁵ An engine trained on a parallel corpus of legislation, for example, should be well adapted to the task of translating previously unseen legislative texts in the language pair in question.

English monolingual corpora used in their study. It also seems unlikely that the texts in the corpora in question can be united by topic or level of formality, as per Koehn’s description. We are left with the assumption that the texts in each corpus perhaps share a certain literary *style*, although this is not explicitly pursued in Toral and Way (2018) or related studies.⁶

The picture gets fuzzier still when one considers how the term “style” itself is understood in MT research. It seems that the general vagueness around the term “domain” is mirrored in the treatment of “style”. Korotkova et al. (2018: 2), for example, treat the term “text style” as “loosely as covering concepts like text domain, genre, formality and other text characteristics” (Korotkova et al. 2018: 2), making “domain” an element of “style”. The three “styles” included in their study are OpenSubtitles, Europarl, and JRC-Acquis, where “style” effectively becomes coterminous with “corpus,” in the same way as “domain” became coterminous with “corpus” in Koehn (2020).⁷ Korotkova et al. (2018) provide examples of how sentences in French and German are automatically translated into English in the three different “styles” under consideration. The French string *on se lance?*, for example, is translated as “let’s go” in the OpenSubtitles “style”, “are we getting started?” in the Europarl “style”, and “are we going?” in the JRC-Acquis “style”. Conceptual vagueness, it turns out, does not prevent the machine learning techniques used in NMT from learning (at least some of) those features of each corpus that mark it as stylistically different from the other two.⁸

Michel and Neubig (2018) also attempt to create NMT outputs in varying styles, but this time the focus is on mimicking the style of individual source text speakers in a corpus of translated TED talks. Wang et al. (2021) pursue the related aim of mimicking the styles of individual translators in a similar corpus of translated TED talks. In both cases, the authors treat each individual (speaker or translator) as a “domain”, and add discrete speaker or translator tokens to the training data to create personalized NMT engines.⁹ Interestingly, Michel and Neubig (2018) do not refer to “style” as such. Rather they use the phrases “speaker traits” and “speaker-related variations” to capture what becomes labelled as “speaker and author style” in Wang et al. (2021). Wang et al. (2021) commit to the term “style”, however, and claim that their “style-augmented translation models are able to capture the style variations of translators and to generate translations with different styles on new data” (ibid.: 1). They also claim that “translator information has more impact on NMT than the speaker information”, although one possible explanation for this is that there are fewer translators than speakers in

⁶ The impressive corpora used in this study have also been used in subsequent studies, including Moorkens et al. (2018) and Guerberof Arenas and Toral (2020, 2022), none of which, however, problematize the notions of domain or style.

⁷ Stylistic differences are, in turn, assumed to exist between these corpora (Korotkova et al. 2018: 3), and despite acknowledged intra-corpus heterogeneity in OpenSubtitles, many “stylistic traits” are said to be “similar” even within this corpus, “which means that these common traits can be learned as a single style” (Korotkova et al. 2018: 3).

⁸ Note here also that, in line with other natural language processing studies in “style transfer”, Korotkova et al. (2018) distinguish between “semantic content” and “style”, while acknowledging that in practice the two concepts are sometimes dealt with together.

⁹ For example, the speaker token is added at the start of the target sentence for each speaker in Michel and Neubig (2018).

their training data, as some translators translate speeches by several different speakers. Hence there is more training data per individual translator than per individual speaker, making the translator signal stronger (ibid.:5). They note that style-augmented translation outperforms the baseline in both human and automatic evaluations, and that “human evaluation confirms that observed differences are all about style and not translation quality” (ibid.: 1-2). Wang et al. (ibid.), like some of the above-mentioned researchers, thus suggest that there is merit in translator-specific personalization in MT research, and that this merit resides in the ability of personalized MT output to capture translators’ styles. They do not define “style” formally however, choosing instead to give examples of what has been considered to constitute “style” in the MT literature, mentioning “verbosity”, “formality”, “politeness”, “demography” and “personal traits” (ibid.:2). In their own study they exemplify “style” using three criteria: verbosity (how many words a translator uses), and lexical and grammatical choices. For a study that purports to be interested in translator style, however, references to the translation studies literature on the subject are conspicuous by their absence.

Overall, the MT literature is often vague or non-committal when it comes to defining style. This vagueness means that it is difficult to distinguish between style and related concepts in MT research, and that it would be difficult to rely on definitions from the area in any attempt to operationalize style in investigations such as ours. Rather than rely on the under-conceptualized notion of style encountered in much MT research, we thus turn for guidance to a field that has been concerned with literary style for centuries (Herrmann et al. 2015:25) and that has influenced corpus-based translation studies, namely literary studies.

Style in Literary Studies

There is a long history of scholarship on style in literary studies. Herrmann et al. (2015) provide a comprehensive overview of contributions made since 1945, taking in hermeneutic, structuralist, pragmatic and digital humanities approaches, among others. The definitions of style encountered in these traditions, according to Herrmann et al. (ibid:26) cover:

style as revealing a higher-order aesthetic value, as the holistic ›gestalt‹ of single texts, as an expression of the individuality of an author, as an artefact presupposing choice among alternatives, as a deviation from a norm or reference, or as any formal property of a text.

In the interests of establishing “a common ground for literary scholars and stylisticians when talking about style” (ibid.: 28), Herrmann et al. go on to propose “an operational definition of style that incorporates a minimal common ground for interdisciplinary empirical research and the application of new, digital methods” (ibid.). Their subsequent definition of style is one that resonates with us, given its commitment to empiricism and the ease with which it can accommodate digital methods. It is also broad and non-normative, in the sense that it does not promote coherent style as a sign of aesthetic quality. Nor does it suggest that style is unique to literary texts. It presents style as an empirical textual category, that is, as something that can be observed in texts, where observation can be primarily based on quantitative or qualitative methods, or a mixture of both. Herrmann et al.’s definition of style is as follows:

Style is a property of texts constituted by an ensemble of formal features which can be observed quantitatively or qualitatively. (Herrmann et al. 2015: 44, italics in original)¹⁰

This definition is largely consistent with that adopted in digital corpus stylistics, one of the areas on which Herrmann et al. (ibid.) draw, and which has been influential in corpus-based research into style in translation studies (e.g. Baker 2000; Saldanha 2011). However, the definition does not *require* style to be identified contrastively, an approach often taken in corpus-based translation studies, but neither does it exclude such an approach. In quantitative stylistic analyses, in particular, inter-corpus investigations are common. In such cases:

Frequencies of selected elements can be quantitatively related *across* a collection of individual texts, for example by using the collection itself as the relative norm, or using a large(r) reference corpus as a backdrop. (Herrmann et al. 2015: 47)

The fact that style does not have to be identified contrastively *by definition* is another factor that draws us to Herrmann et al.'s definition. In the past we have used Saldanha's (2011) influential definition of "translator style" (Kenny and Winters 2020), which integrates contrast in the form of "deviance from a norm" as a necessary condition for the identification of an individual translator's style. More specifically, a translator's style is identified partly on the basis of how their work differs from that of other translators. Such contrast is not a sufficient condition for the recognition of style in Saldanha's definition, as translator style also has to meet other criteria. Saldanha's definition of translator style is as follows:

A 'way of translating' which:

- is felt to be recognizable across a range of translations by the same translator,
- distinguishes the translator's work from that of others,
- constitutes a coherent pattern of choice,
- is 'motivated', in the sense that it has a discernible function or functions, and
- cannot be explained purely with reference to the author or source-text style, or as the result of linguistic constraints.

(Saldanha 2011: 31)

The "contrast with other translators" condition seems to us now to be somewhat problematic in that it suggests that two translators cannot have the same style, or share elements of the same style, which in our view is an unnecessarily restrictive position to take. Investigations of style in literary translation do not have the same priorities as investigations of style in

¹⁰ Herrmann et al. (ibid.: 30) point out that their contribution "centers on style definitions that are valid at the textual level of usage" and not "those approaches that deal explicitly with style at the level of the reader's processing, such as psychological theories of style processing and cognitive-linguistic theories in their pure form." The "eminently important" analytical distinction between "the levels of reader response and textual structure" is however, acknowledged, as is the role of cognition in subjective judgements about deviation and norms, for example (ibid.).

forensic linguistics, for example, where aims like author (or translator) attribution mean that techniques that *distinguish* individuals from each other are paramount.¹¹

Herrmann et al.'s (2015) definition also differs from Saldanha's (2011) in dropping the requirement for "coherence", which they associate with "more normative views of style which see coherent style as a sign of aesthetic quality" (ibid.: 44). They fall back on the notion of an "ensemble" of formal features instead:

By »ensemble« we mean that style is constituted by the combination of many possible features and should be seen as a complex system, with features situated at different linguistic levels. Such an ensemble does not necessarily exhibit a coherent unity; rather, it can have various degrees of unity or harmony, or, on the contrary, contrasts or incoherence. (Herrmann et al. 2015: 44)

This approach seems to us to allow for very broad descriptive explorations of style that are not unnecessarily restricted from the outset, but also that allow for the fact that an individual translator's style might vary over a range of texts or over time, even if some features of their style remain constant. Having said that, there is some evidence that individual author styles can persist over decades, and despite changing editorial environments. This is the case, for example, with the judges of the US Supreme Court investigated in Jockers et al. (2020). Similarly, this persistence of individual style over time can also apply to translators, as we have demonstrated for Oeser (Kenny and Winters 2020).

Causality and Style in Literary Machine Translation

Herrmann et al.'s (2015) general approach also has the merit of allowing for *causality* to be integrated into discussions of style in many ways:

Even in the absence of conscious intentions, causal relationships may be hypothesized: genre can cause style (e.g. by means of conventions: form and themes), authors can cause style (e.g. by means of idiosyncrasies), theme and topic can cause style. The interpretability of style relative to categories such as authorship, literary genre, or literary period, is hence paramount. This means that any stylistic phenomenon can ultimately be considered the trace of or the index towards such categories (or, in other terms, may be ›characteristic‹ of them). (Herrmann et al. 2015: 46)

To Herrmann et al.'s list, and drawing on the literature in corpus-based translation studies and literary MT, we would add the following hypotheses, some of which have been tested in the field:

¹¹ Note that corpus linguists in general have also been criticized in the past for setting too much store by difference at the expense of recognizing similarity (see, e.g., Baker 2004; Taylor 2018). Note also that the dropping of a requirement of contrast with others from the definition of (translator) style does not mean that we do not use contrastive *techniques* in the exploration of translator style. On the contrary, Winters and Kenny (forthcoming) make heavy use of (and defend) one such technique, namely keywords analysis.

Translation can cause style.

The very fact of translation may influence the formal features of a target text, regardless of which translator performed the translation, what the source language was, etc. This is the basic idea behind research into so-called “universals” of translation (see Baker 1993).

Translators can cause style.

The formal features of a target text may be seen as a trace of a particular translator (Baker 2000; Kenny 2001; Saldanha 2005, 2011; Winters 2009, 2013).

Machine translation can cause style.

Even if we allow for the fact that the particular texts on which an engine is trained, the genres to which they belong, and the language pair involved can be reasonable causes of style, we still have to admit that the algorithms used in contemporary machine translation themselves can cause style, most notably by amplifying characteristic features of texts in the domain in question, leading to over-use of those features – for example, contractions like *I'll* in subtitle translation (Korotkova et al. 2018) – or to gender and other biases in their outputs (Hovy et al. 2020; Vanmassenhove et al. 2021).

Post-editing can cause style.

Previous research has shown that post-editing – or more accurately, and for obvious reasons, the combination of machine translation and post-editing, also known as “PEMT” – can lead to target texts that differ in systematic ways from human translations. Toral (2019), for example, using a number of different corpora involving several different language pairs, shows that there are statistically significant differences between post-edited texts and human-translated texts on certain easily computed scores (e.g. type-token ratio and lexical density). In general, however, it is difficult to factor out post-editing from machine translation itself as a cause in this kind of study. Čulo and Nitzke (2016), who focus on terminological variation, thus find evidence of machine translation “shining through” in PEMT texts, while Farrell (2018) similarly concludes that the PEMT texts in his study contain “machine translation markers.” The latter two studies serve as examples of work that combines both quantitative and qualitative perspectives, identifying as they do the particular terms or expressions that prove to be characteristic of PEMT texts.

Post-editors can cause style.

Although “post-editing” is considered the condition of interest in many hypothesis-testing studies, some studies depart from what we might call a “post-editing as monolith” position by pointing to differences between post-editors, among whom individual variation has been described as “inevitable” (O’Brien and Simard 2014: 160). This is despite the fact that, for years, post-editors have been instructed to change as little of the machine translation output as possible (O’Brien 2022) so one would expect their room for manoeuvre to be limited. To date, however, there have been few if any studies that set out to investigate how stylistic phenomena can be considered “the trace of or the index towards” (Herrmann et al. op. cit.) a particular post-editor. Kenny and Winters (2020) is one exception. It is described briefly in the next section.

The literary translator as post-editor: a case of downstream translator-specific personalization

As already indicated, we are interested in the potential of post-editing as a type of downstream translator-specific personalization in literary MT. Such personalization could be said to occur in cases where, by editing a machine translated text in a particular way, a literary translator is shown to assert their own style in the post-edited text. In Kenny and Winters (2020), we attempted to capture data on such personalization in an experiment in which Hans-Christian Oeser post-edited a short text translated by DeepL. The text in question was a chapter from F. Scott Fitzgerald’s *The Beautiful and Damned* (1922), which Oeser had himself translated (from scratch, that is, without the use of MT) some twenty years earlier (*Die Schönen und Verdammten*, 1998). His post-edited version was compared with his from-scratch translation in order to ascertain whether, as a post-editor, Oeser put his own stylistic stamp on the MT output. By tracking individual edits made on the short text, we found that Oeser’s style, as familiar from previous corpus-driven studies, was somewhat attenuated in his post-editing. The study had some shortcomings however. The translation commission was fictitious, and the use of MT was not the only variable that might have affected the outcome. Oeser himself commented on the passage of time as a factor affecting his performance as a translator/post-editor. In Winters and Kenny (forthcoming) we address both of these issues. In this case the commission is a real one and we make comparisons with a corpus of translations published by Oeser around the same time as he completed the post-editing task in question, as well as with a reference corpus of texts originally written in German at roughly the same time. The passage of time is thus not a confounding variable.¹² Space constraints prevent us from giving fuller details of this study. In what follows we confine ourselves to its general design and some principal findings.

¹² Note here that a further difference between Winters and Kenny (forthcoming) and Kenny and Winters (2020) is that the latter involved a three-way comparison (in the first instance) between a human translation, a machine translation, and a post-edited machine translation. Winters and Kenny (forthcoming), on the other hand, involves (in the first instance) a two-way comparison between a machine translation and a post-edited machine translation. Further comparisons with the larger translator-specific corpus and reference corpus of original German literary texts are then carried out.

In Winters and Kenny (forthcoming) we analyse a version of Christopher Isherwood's (1954) novel *The World in the Evening* that was post-edited by Hans-Christian Oeser on the basis of a machine translation into German produced by DeepL (free version used in 2019). Oeser's post-edited version was subsequently published as *Die Welt am Abend* (Isherwood 2019). As elaborated upon in Winters and Kenny (ibid.), Oeser's work on the DeepL translation does not conform to the image of "post-editing" promulgated in many language industry sources. In fact, Oeser himself does not describe the process as post-editing but rather as one of "painstaking retranslation" (Oeser 2020: 20), adding that "there was hardly a sentence that did not have to be thoroughly revised and rebuilt" (ibid.). We received from Oeser the full raw machine translated version of the novel and his own post-edited version, both in digital form, and can attest to the extent of the changes made to the MT output, which in turn suggested that there would be ample data to study both the style of the MT and the style of Oeser's subsequent post-edited version. What interests us most in our study is the extent to which, in editing the MT output, Oeser not only fixes errors, which is the main purpose in conventional post-editing (O'Brien 2022), but also makes changes that have the effect, whether intended or not, of imposing his style on the final version.

From previous quantitative and qualitative empirical studies (esp. Winters 2007, 2009) we have a comprehensive picture of Oeser's style as it then stood (see also Winters 2015). We have also created a reference corpus, *Oeser 12*, containing twelve of Oeser's recent translations (novels and novellas translated as sole translator between 2016 and 2021¹³) and a reference corpus of 57 works originally written in German over a similar period (2000 to 2012, the most recent time interval covered by). Our data sets allow us to study the style of the machine translated text and the post-edited text in their own rights, by observing "an ensemble of formal features" (Herrmann et al. 2015: 44, and see above) in each. Knowing how both texts came into being, we can then hypothesize MT and post-editing as "causes" of the various features in the latter text. This part of the analysis will certainly draw on contrastive analysis between the machine translated and the post-edited texts, but it can also allow us to see similarities between these two texts. Meanwhile, by comparing the style of his post-edited text with what we already know about Oeser's style as a translator and with what we can glean about this style from the contemporaneous reference corpus of his translations, we can work out how consistent his 'style when post-editing' is with his 'style when translating'. If there is a high level of consistency, then the argument that literary MT post-editing can be seen as a kind of downstream translator-specific personalization becomes more

¹³ As already indicated, by using a contemporaneous corpus to get a snapshot of Oeser's style at a particular point in time, we implicitly acknowledge that his style (or any translator's style) might change over time. (Such change over time would be best captured using a diachronic corpus design.) The 'synchronicity' of *Oeser 12* with his post-edited work could, however, be questioned given the longer timeframe over which the source texts behind *Oeser 12* were written: while the majority of those source texts were published between 2010 and 2020, some were originally published earlier, e.g., in the 1940s, 50s and 70s. We also acknowledge that both the time of publication of the source text and the time period of the fictional setting might influence style in translation, motivating, for example, the use of 'elevated' lexis. In the case of the word we focus on later in this chapter, however, we note that while there is much variation in the frequency of *weshalb* across the texts in *Oeser 12*, we have not found any evidence of a relation between its frequency and the publication date of the originals.

compelling. Comparisons with the reference corpus of original German literature, finally, allow us to investigate whether certain features of Oeser’s style as a translator and post-editor are particularly indicative of his work, or whether they are also typical of German letters in general.

Of course, in such a study, or rather series of studies, there are hundreds of formal features – involving varying levels of abstraction, quantification and qualitative analysis – that could be investigated (see Herrmann et al. 2015: 45). For reasons elaborated upon in Winters and Kenny (forthcoming) we start our investigative journey by tracing keywords in the various data sets. The analysis is thus, in the first instance, a lexical one, and focuses on those lexical items that distinguish Oeser’s post-edited text from the DeepL MT output. The findings reported in Winters and Kenny (forthcoming) suggest that there is hardly a keyword in Oeser’s post-edited version that cannot be explained with reference to his style. By way of illustration, and due to space constraints already alluded to, we give just a single example here.

The word form ranked fifteenth in the keywords list for Oeser’s post-edited text is *weshalb* (‘why’).¹⁴ It occurs 19 times in Oeser’s post-edited text and not at all in the raw MT output.¹⁵ Inspection of the two texts conducted using a parallel concordancer shows that in every case, Oeser inserts *weshalb* to replace *warum* (a more common word also meaning ‘why’) in the MT output.¹⁶ A comparison of the relative frequency of *weshalb* in Oeser’s post-edited text, the reference corpus of his translations (*Oeser 12*) and the *German Original Literature* reference corpus, as depicted in Table 1, suggests that *weshalb* is a word characteristically used by Oeser.

Table 1: Frequency per 100,000 tokens of *weshalb* in Oeser’s post-edited Isherwood translation, *Oeser 12* and *German Original Literature*

keyword	Oeser’s post-edited Isherwood translation	per 100,000	reference corpus: Oeser 12	per 100,000	reference corpus: German Original Literature	per 100,000
<i>weshalb</i>	19	18.27	168	23.97	257	7.15

It occurs nearly 24 times in 100,000 words (tokens) in *Oeser 12*. It is far less commonly used by writers of original prose in German, however, occurring slightly more than 7 times in 100,000 words (tokens) in the *German Original Literature* corpus. Interestingly the use of *weshalb* is mentioned by Oeser in his comments on the post-editing work we reported on in Kenny and Winters (2020). In that source, Oeser characterized its use as one of his ‘quirks’:

¹⁴ Keywords are computed using WordSmith Tools version 8.0 (Scott 2022) and using the raw MT output as the reference text. They are ranked using log likelihood scores in conjunction with Bayesian Information Criterion (BIC) scores, which measure effect size. Full details of the keyword analysis are given in Winters and Kenny (forthcoming).

¹⁵ The log likelihood score for *weshalb* in this text is 14.33 (BIC=26.58).

¹⁶ For ease of analysis the MT raw output and post-edited versions were aligned and searched using Tetrapla, developed by David Woolls (2008-2022).

Auch so eine Marotte von mir: lieber "weshalb" als "warum" zu verwenden.

(‘Another one of my quirks: I prefer using *weshalb* to *warum*.’)

(Oeser 2020 in Kenny and Winters 2020:143)

The presence of so many instances of *weshalb* in his post-edited text thus brings that text more into line with what is known about Oeser’s style, as observed in a corpus of his translation work, and as revealed in Oeser’s own commentary on his previous work.

Conclusions

In this chapter we have reflected on current strands in research into literary MT and have pointed up the relative lack of conceptualization regarding style in the field. We have revisited the concept of style in the light of recent thinking in literary studies and made the case for adopting Herrmann et al.’s (2015) broad and digitally accommodating definition in studies of translator style whether or not those translators are working with MT output. We have exemplified our approach with one simple example, using just one analytical technique, and focusing on just a single translator and a single post-edited text. In ongoing work (Winters and Kenny forthcoming; Kenny and Winters 2022) we account for far more of our data, and apply additional techniques. We hope here, nevertheless, to have offered “proof of concept” that a literary translator’s post-editing activity can be regarded as downstream translator-specific personalization of MT output.

References

- Alvstad, C., A. K. Greenall, H. Jansen, and K. Taivalkoski-Shilov (eds) (2017) *Textual and Contextual Voices of Translation*, Amsterdam/Philadelphia: John Benjamins.
<https://doi.org/10.1075/btl.137>
- Baker, M. (1993) ‘Corpus Linguistics and Translation Studies’, in Baker, M., Francis, G. and Tognini-Bonelli (eds) *Text and Technology*, Amsterdam: Benjamins.
- Baker, M. (2000) ‘Towards a Methodology for Investigating the Style of a Literary Translator’, *Target* 12(2): 241–266.
- Baker, P. (2004) ‘Querying keywords: Questions of difference, frequency and sense in keywords analysis’, *Journal of English Linguistics*, 32(4): 346–59.
- Bentivogli, L., A. Bisazza, M. Cettolo and M. Federico (2016) ‘Neural versus Phrase-Based Machine Translation Quality: A Case Study’, in EMNLP 2016. arXiv:1608.04631v1 [cs.CL] 16 Aug 2016.
- Castilho, S. and N. Resende (2022) ‘Post-Editese in Literary Translations’, *Information* 13(66). <https://doi.org/10.3390/info13020066>
- Čulo, O. and J. Nitzke (2016) ‘Patterns of Terminological Variation in Post-editing and of Cognate Use in Machine Translation in Contrast to Human Translation’, *Baltic Journal of Modern Computing*, 4(2): 106-114.

Farrell, M. (2018) 'Machine Translation Markers in Post-Edited Machine Translation Output', *Proceedings of the 40th Conference Translating and the Computer*, pages 50–59, London, UK, November 15-16, 2018. <https://www.transcreate.it/tempfolder/wp-content/uploads/2018/11/TC406.pdf>

Fitzgerald, F. Scott (1922) *The Beautiful and Damned*. New York: Scribner's.

Fitzgerald, F. Scott (1998) *Die Schönen und Verdammten*. Zürich: Diogenes. Translated by Hans-Christian Oeser.

Fonteyne, M., A. Tezcan and L. Macken (2020) 'Literary Machine Translation under the Magnifying Glass: Assessing the Quality of an NMT-Translated Detective Novel on Document Level', *Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020)*, 3790–3798. <https://aclanthology.org/2020.lrec-1.468.pdf>

Guerberof-Arenas, A. and A. Toral (2020), 'The Impact of Post-editing and Machine Translation on Creativity and Reading Experience', *Translation Spaces*, 9(2): 255-282

Guerberof-Arenas, A. and A. Toral (2022), 'Creativity in translation. Machine translation as a constraint for literary texts', *Translation Spaces*, online first, DOI: <https://doi.org/10.1075/ts.21025.gue>

Hansen, D. (2021) 'Les lettres et la machine : un état de l'art en traduction littéraire automatique', *Actes de la 28e Conférence sur le Traitement Automatique des Langues Naturelles Volume 2 : 23e REcontres jeunes Chercheurs en Informatique pour le TAL (RECITAL)*, Lille, France, 28 June to 2 July, 2021 pages 28–45.

Hansen, D. (2022) 'La traduction littéraire automatique : Adapter la machine à la traduction humaine individualisée', Manuscript available from <https://hal.archives-ouvertes.fr/hal-03583562/> (last accessed 28/11/2022).

Herrmann, J. B. , van Dalen-Oskam K. , Schöch C. (2015) 'Revisiting style, a key concept in literary studies', *Journal of Literary Theory*, 9 (1): 25 – 52.

Hovy, D., F. Bianchi, and T. Fornaciari (2020) 'You Sound Just Like Your Father' Commercial Machine Translation Systems Include Stylistic Biases', in Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 1686–1690, Online. Association for Computational Linguistics.

Isherwood, Christopher (1954) *The World in the Evening*. London: Methuen.

Isherwood, Christopher (2019) *Die Welt am Abend*. Translated by Hans-Christian Oeser. Hamburg: Hoffmann und Campe.

Jockers, M. L., F. Nascimento and G. H. Taylor (2020) 'Judging style: The case of Bush versus Gore', *Digital Scholarship in the Humanities* 35(2): 319-327.

Kenny, D. (2001) *Lexis and Creativity in Translation. A Corpus-based Study* Manchester: St. Jerome.

Kenny, D. (2018) 'Machine Translation' in Piers Rawling and Philip Wilson (eds) *The Routledge Handbook of Translation and Philosophy*, London and New York: Routledge, 428-445.

Kenny, D. (2020) 'Machine Translation', in Mona Baker and Gabriela Saldanha (eds) *The Routledge Encyclopedia of Translation Studies*, 3rd edition. London: Routledge, 305–310.

- Kenny, D. (2022) 'Human and machine translation', in Dorothy Kenny (ed.) *Machine translation for everyone: Empowering users in the age of artificial intelligence*, 23–49, Berlin: Language Science Press. DOI: 10.5281/zenodo.6759976
- Kenny, D. and M. Winters (2020) 'Machine Translation, ethics and literary translator's voice', *Fair MT: Towards ethical, sustainable Machine Translation*, Special Issue of *Translation Spaces* 9(1): 123-149.
- Kenny, D. and M. Winters (2022) 'A text of one's own: post-editing and the affirmation of style', Paper presented at Translation in Transition 6, Charles University, Prague, Czech Republic, 22-23 Sept. 2022.
- Koehn, P. (2020) *Neural Machine Translation*, Cambridge: Cambridge University Press.
- Korotkova, E., M. Del and M. Fishel (2018) 'Monolingual and cross-lingual zero-shot style transfer', arXiv preprint arXiv:1808.00179.
- Kuzman, T., Š. Vintar, and M. Arčan (2019) 'Neural Machine Translation of Literary Texts from English to Slovene', in *Proceedings of the Qualities of Literary Machine Translation*, pages 1–9, Dublin, Ireland. European Association for Machine Translation.
- Lee, C. (2021), 'How do machine translators measure up to human literary translators in stylometric tests?', *Digital Scholarship in the Humanities*. <https://doi.org/10.1093/lc/fqab091>
- Lennon, B. (2014) 'Machine translation: A tale of two cultures', in S. Bermann & C. Porter (Eds.) *A companion to translation studies*, Oxford: Wiley, 136–146.
- Michel, P. and G. Neubig (2018) 'Extreme adaptation for personalized neural machine translation', in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics* (Volume2: Short Papers), pages 312–318, Melbourne, Australia. Association for Computational Linguistics.
- Matusov, E. (2019) 'The Challenges of Using Neural Machine Translation for Literature', in J. Hadley, M. Popović, H. Afli, A. Way (eds) *Proceedings of the Qualities of Literary Machine Translation*, pages 10–19, Dublin, 2019. <https://www.aclweb.org/anthology/W19-7302.pdf>.
- Moorkens, J., A. Toral, S. Castilho and A. Way (2018) 'Translators' perceptions of literary post-editing using statistical and neural machine translation', *Translation Spaces* 7 (2): 240–262. <https://doi.org/10.1075/ts.18014.moo>
- O'Brien, S. and M. Simard (2014) 'Introduction to special issue on post-editing', *Machine Translation* 28, 159–164. <https://doi.org/10.1007/s10590-014-9166-8>
- O'Brien, S. (2022) 'How to deal with errors in machine translation: Postediting', in D. Kenny (ed.) *Machine translation for everyone: Empowering users in the age of artificial intelligence*, Berlin: Language Science Press, 105–120. DOI: 10.5281/zenodo.6759982
- Oeser, H.-C. (2020) 'Duel with DeepL. Literary translator Hans-Christian Oeser on machine translation and the translator's voice', *Counterpoint* 4: 19-123. https://www.ceatl.eu/wp-content/uploads/2020/12/Counterpoint_2020_04_article_05.pdf
- Ramírez-Sánchez, G. (2022) 'Custom machine translation', in D. Kenny (ed.) *Machine translation for everyone: Empowering users in the age of artificial intelligence*, Berlin: Language Science Press, 165–186. DOI:10.5281/zenodo.6760022

- Saldanha, G. (2005) *Style of translation: an exploration of stylistic patterns in the translations of Margaret Jull Costa and Peter Bush*, Unpublished PhD Thesis, Dublin City University.
- Saldanha, G. (2011) 'Translator Style: Methodological Considerations.' *The Translator* 17(1): 25–50. <https://doi.org/10.1080/13556509.2011.10799478>
- Scott, M. (2022) 'How Key Words are calculated', *WordSmith Tools Help*. Consulted 31 March 2022. https://lexically.net/downloads/version_64_8/HTML/keywords_calculate_info.html
- Taivalkoski-Shilov, K. (2019a) 'Free indirect discourse: an insurmountable challenge for literary MT systems?', in J. Hadley, M. Popović, H. Afli, A. Way (eds) *Proceedings of the Qualities of Literary Machine Translation*, Dublin, 2019, pages 35–39. <https://aclanthology.org/W19-7305.pdf>
- Taylor, C. (2018) 'Similarity', in C. Taylor and A. Marchi (eds) *Corpus Approaches to Discourse: A critical review*, London: Routledge, 19-37.
- Toral, A. and A. Way (2015a) 'Translating Literary Text between Related Languages using SMT', in *Proceedings of NAACL-HLT Fourth Workshop on Computational Linguistics for Literature*, 123–132, Denver, Colorado. <https://doi.org/10.3115/v1/W15-0714>
- Toral, A. and A. Way (2015b) 'Machine-assisted translation of literary text: A case study', *Translation Spaces* 4(2): 240-267. <https://doi.org/10.1075/ts.4.2.04tor>
- Toral, A. and A. Way (2018) 'What level of quality can Neural Machine Translation attain on literary text?', in J. Moorkens, S. Castilho, F. Gaspari, and S. Doherty (eds) *Translation Quality Assessment: From Principles to Practice*, Heidelberg: Springer International Publishing, 263–287. https://doi.org/10.1007/978-3-319-91241-7_12
- Toral, A. (2019) 'Post-editeese: an Exacerbated Translationese', in *Proceedings of Machine Translation Summit XVII: Research Track*, pages 273–281, Dublin, Ireland. European Association for Machine Translation. <https://aclanthology.org/W19-6627>
- Van de Cruys, T. (2020) 'Articuler ou trahir. La poésie des machines à traduire', Paper presented at *La traduction littéraire et SHS à la rencontre des technologies de la traduction : enjeux, pratiques et perspectives*. Translating Europe Workshop. Université Toulouse Jean Jaurès. 7 February 2020. http://timvandecruys.be/media/presentations/pres_tradlit_070220.pdf
- Vanmassenhove, E., D. Shterionov, and M. Gwilliam (2021) 'Machine Translationese: Effects of Algorithmic Bias on Linguistic Complexity in Machine Translation', in *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 2203–2213, Online. Association for Computational Linguistics. 10.18653/v1/2021.eacl-main.188
- Voigt, R. and D. Jurafsky (2012) 'Towards a literary machine translation: The role of referential cohesion', in *Proceedings of the NAACL-HLT 2012 Workshop on Computational Linguistics for Literature*, 18–25.
- Wang, Y., C. Hoang and M. Federico (2021) 'Towards Modeling the Style of Translators in Neural Machine Translation', in *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. <https://aclanthology.org/2021.naacl-main.94>

Winters, M. (2007) 'F. Scott Fitzgerald's *Die Schönen und Verdammten*: A Corpus-based Study of Speech-act Report Verbs as a Feature of Translators' Style', *Meta: journal des traducteurs/Translators' Journal* 52(3):412-425.

Winters, M. (2009) 'Modal particles explained. How modal particles creep into translations and reveal translators' styles', *Target: International Journal of Translation Studies* 21(1):74-97.

Winters, M. (2013) 'German modal particles – from lice in the fur of our language to manifestations of translators' styles', *Perspectives: Studies in Translatology* 21(3):427-445 (online first), <https://doi.org/10.1080/0907676X.2012.711842> [7 Aug 2012].

Winters, M. (2015) 'Hans-Christian Oeser: Translation strategies of a literary translator', in S. Egger (ed) *Cultural/Literary Translators: Selected Irish-German Biographies II* (Irish-German Studies; vol. 9), Trier: Wissenschaftlicher Verlag Trier, 210–224.

Winters, M. and D. Kenny (forthcoming) 'Mark my keywords: a translator-specific exploration of style in literary machine translation', in A. Rothwell, A. Way and R. Youdale (eds) *Computer-assisted literary translation: the state of the art*, London: Routledge.

Woolls, D. (2008-2022), Tetrapla, CFL Software Limited.