Chapter 3 in *Computer-Assisted Literary Translation.* Edited By Andrew Rothwell, Andy Way and Roy Youdale. New York: Routledge. © 2024
DOI: 10.4324/9781003357391-5

Mark My Keywords

A Translator-Specific Exploration of Style in Literary Machine Translation

Marion Winters and Dorothy Kenny

Introduction

Since the mid-2010s, growing attention has been paid to how machine translation (MT) might be applied in the translation of literary texts, to the extent that literary MT is now emerging as a field in its own right. The interests of researchers working in the area are many and varied (see Kenny and Winters, forthcoming; and the other chapters in the present volume), but one important strand focuses on the customisation of MT engines to improve their performance in translating literary texts. Such customisation can be individualised: engines can be (partly) trained on texts written by particular authors (see Chapters 1 and 6, this volume) and/or translated by particular translators, yielding what we call 'author-specific personalization' and 'translator-specific personalization', respectively (Kenny and Winters, op. cit.). This kind of personalisation has as its implicit or explicit aim the recreation of a certain 'style' in MT, namely that of the source-text author or the individual translator. It happens upstream, that is, at the training stage, before any new translation is attempted. It is also possible to imagine downstream personalisation, however, in the guise of literary post-editing. In Kenny and Winters (op. cit.) we thus characterise the post-editing of a machine-translated text by an experienced literary translator whose brief is to 'make the text his [i.e. Oeser's] own' as a kind of downstream translator-specific personalisation (TSP). Whether or not the TSP in question is successful would depend not on how the outputs were rated using standard MT evaluation metrics (e.g. Castilho et al., 2018; Way, 2018), but rather on whether the post-edited version could be shown to reflect the translator's known style. In this chapter, we continue on this track, presenting the results of an empirical investigation into literary translator Hans-Christian Oeser's style when he works in post-editing as opposed to 'conventional' translation mode.[1] In what follows, we first describe how we approach style in this study and share what we already know about Oeser's style as a translator. We then describe the specific method we use to analyse translator style in Oeser's post-edited text, which relies on keyword analysis, before outlining the rest of our research design. Finally, we present our results and discuss our findings in light of what they can tell us about downstream TSP.

Literary Style and Translation

Interest in literary style goes back centuries, but scholars have not always proceeded on the basis of a shared understanding of the term (Herrmann et al., 2015, p. 25). Even within the bounds of Translation Studies treatments of style vary (see, especially, Saldanha, 2014), but there is a general consensus that one can study the style of source texts, target texts or translators. Saldanha's (2011) definition of *translator style* has been particularly influential, both in our own work and elsewhere (e.g. Kenny and Winters, 2020; Youdale, 2020). For reasons outlined in Kenny and Winters (forthcoming), however, we prefer to fall back on the broader definition of style of Herrmann et al. (2015, p. 44) as:

> *a property of texts constituted by an ensemble of formal features which can be observed quantitatively or qualitatively.* (italics in original)

and then to hypothesise *causes* for the observed formal features of texts, in the form of particular translators, post-editors, and so on.

The particular formal features that can be studied and the ways in which they can be processed run into the hundreds (Herrmann et al., 2015, p. 45). Previous corpus-based and corpus-driven studies in translator style (e.g. Baker, 2000; Saldanha, 2011; Youdale, 2020) have started, for example, by focusing on basic statistics like lexical density, average sentence length, (standardised) type-token ratios and other measures of lexical variety. These studies usually branch into richer qualitative analyses on the basis of their initial quantitative findings. Similar metrics have been used in studies that

1 We are using 'conventional' here to designate translation completed without the use of MT for lack of a better term. Alternative terms used in the literature create as many problems as they solve: 'human translation', for example, is often used to contrast with 'machine translation' but post-editing is very much a human activity too, a fact that reduces the discriminating capacity of the epithet 'human' in this instance. Likewise, the term 'from scratch' seems inadequate as it appears to present translation completed without the use of MT as a kind of *ex nihilo* activity, which is clearly not the case. That said, 'from scratch' is so commonly used to designate translation completed without a machine-translated first draft that it is difficult to displace.

compare the style of machine-translated and human-translated texts (e.g. Lee, 2021) and in investigations of 'post-editese' (Toral, 2019), although neither Lee nor Toral focuses on any particular translator or post-editor. Somewhat more sophisticated quantitative analyses use Burrows' (2002) Delta to calculate distances between texts. This approach is favoured by scholars working in the wider Digital Humanities (e.g. Rybicki, 2012) where the predominant concern is with author/translator attribution. In related work that uses the most frequent 1-, 2- and 3-grams (1-, 2- and 3-word sequences, not necessarily phrases *per se*) as the features of interest, Lee (2021) applies standard machine learning and statistical techniques (support vector machines and principal component analysis respectively) to differentiate human and machine translations in what is essentially a text classification task; as already indicated, Lee's focus is not on the style of any particular translator, however. For him, 'human translators' remain an undifferentiated mass, whereas 'machine translators' (i.e. MT systems) are identified by name.

In the current study, we investigate Oeser's style using lexical features—in this case word forms—that stand out because of their unusually high frequency in a text he has post-edited compared to the MT output that he starts with. Their unusual frequency in Oeser's post-edited text makes them keywords (see below) in that text. We go on to compare Oeser's post-edited text with a corpus of his other recent translations, and with a corpus of original German literary prose, in a bid to see whether Oeser's post-edits are consistent with his wider work and with German fiction in general. We do this using a second, more focused, iteration of keyword analysis.

Oeser's Style as a Translator and Post-Editor

Hans-Christian Oeser is an internationally acclaimed literary translator with more than 40 years' experience and over 220 titles to his name as translator, editor or author.[2] In previous work, Winters (2007, 2009) used corpus-driven and comparative methods to study his style as a translator. Winters (2015) uses interview data to triangulate findings from these studies. We have thus built up a rich picture of Oeser's observed and self-reported style, some prominent aspects of which are that he:

- favours subject-verb inversion
- is committed to preserving lexical richness in translation
- very consciously uses higher register when appropriate
- attempts to replicate features of natural spoken language

2 See https://hanschristianoeser.wixsite.com/hcoeser for further details.

In Kenny and Winters (2020), we use an experimental design to see what happens to Oeser's style when he is called upon to post-edit rather than translate conventionally, finding that, overall, it is somewhat diminished. The current study improves on this previous work in particular by drawing on a real translation brief.

Keywords as a Method in Investigations of Style

Herrmann et al. (2015, p. 45) point to the vast number and heterogeneity of potential stylistic features and measures encountered in the literature. Nevertheless, they contend that 'most style markers have so far been relatively simple in nature'. They include:

> frequencies and frequency distributions of characters, words, lemmata, word classes or syntactical structures, taken by themselves or in sequences (n–grams); and the length, and distribution of lengths, of words, sentences, paragraphs or other units.
>
> (ibid.)

Higher-order stylistic features are derived from these basic style markers by relating selected markers to each other and/or through the application of various statistical techniques and tests. Among the 'well understood' (ibid.) methods used to generate such higher-order features, Herrmann et al. (ibid.) list *keyness* measures. In short, an item is considered 'key' in a given text or corpus if it occurs with unusual frequency compared with its frequency in another text or (often larger) corpus. The second (larger) corpus is usually called the *reference corpus*, although Scott (2022) uses the term *comparison corpus*, a usage we follow in this chapter. Although keyness can be attributed to any of the items listed above (lemmata, word classes, etc.) it is frequently word forms that are studied in corpus-based translation studies. In the paradigm case (see Rayson, 2019), a word form (or 'type') is said to be a *keyword* in a given text, if it occurs with greater than expected frequency in that text given its frequency in the comparison corpus. Such a keyword is described as a *positive keyword*. A *negative keyword*, in contrast, is one that occurs with less than expected frequency in the text in question.

The computation of keyness usually involves comparing the relative frequencies of word forms in the focus text and the comparison corpus, and then conducting a test to ascertain whether any difference between these relative frequencies is statistically significant. The score produced by this test can then be used to *rank* word forms in terms of their keyness in the text in question.

There has been much debate over which statistical significance test is most suitable for the identification and ranking of keywords, as well as general criticism over assumptions made by many of the tests in question (see Rayson, 2019). The *log likelihood* test has been favoured for some time (ibid.), although it is not without its detractors (e.g. Gabrielatos, 2018). Jeaco (2020, p. 149), however, holds that log likelihood-based keyword calculations ‘can be used effectively for a range of different kinds of research, but often work best with texts and moderately large collections of texts rather than with very large corpora at the entire corpus level’. He adds (ibid.) that log likelihood can be used effectively in combination with related measures such as Bayesian Information Criterion (BIC). The corpus-processing software WordSmith Tools now integrates BIC alongside a variety of keyness measures including log likelihood, the settings for which are often user-adjustable.[3]

Keyword analysis has also come under criticism for its prioritisation of difference over similarity in corpus studies (see e.g. Taylor, 2018). But even the techniques most associated with the privileging of difference can be turned to the analysis of similarity (ibid., p. 21) and it is possible, for example, to investigate whether keywords generated for a given text remain key in other texts by the same author or translator, given a different comparison corpus. This is the approach taken in this study.

All told, keyword analysis offers a number of advantages in corpus stylistic studies. Mastropierro (2018, p. 66), for example, argues that using keywords means that the analyst works with a controlled number of automatically generated items, whose frequency is statistically significant, which makes for ‘an efficient way to begin a study’ and helps to minimise researcher bias. Keyword lists typically contain content and function words, both of which can characterise an author’s, translator’s or character’s style (Culpeper, 2009) and thus are of interest to stylisticians. This fact can differentiate keyword analysis from ‘most frequent word’ analyses, which typically revolve around function words. Keywords also provide a useful exploratory bridge between more quantitative analyses of textual features and more qualitative, interpretative analyses, as is customary in corpus stylistics (Herrmann, 2017). And, as Mastropierro (2018, p. 67) reminds us, the generation of a keyword list is not an end in itself; rather, it can be the starting point for further quantitative and qualitative examination of what, upon careful analysis, is likely to turn out to be important meaningful features of the text in question.

3 www.lexically.net/wordsmith/.

Research Design

Data

In this study, we use as primary data two German versions of Christopher Isherwood's (1954) novel *The World in the Evening*. The first is a machine translation initiated by Hans-Christian Oeser using the free version of DeepL in 2019.[4] The second is Oeser's post-edited translation of the DeepL output, subsequently published as *Die Welt am Abend* (Isherwood, 2019). In what follows we label these texts *DeepL MT* and *Oeser PE,* respectively. We received both texts directly from Oeser.

If using DeepL was 'a cakewalk' (Oeser, 2020, p. 21), then post-editing its output to create a publishable translation that would meet Oeser's normal standards turned out to be less straightforward. Although he describes the experience as 'somewhat less time-consuming' than translating 'from scratch', the process entailed 'painstaking retranslation' given that 'there was hardly a sentence that did not have to be thoroughly revised and rebuilt' (ibid., p. 22). The post-editing was completed in Oeser's normal working environment using Microsoft Word. Given that the maintenance of ecological validity was particularly important to us and to Oeser, no observation techniques were used to track Oeser's progress, and no attempt was made to time Oeser or to gather any kind of user-activity data.

Table 3.1 gives a basic quantitative overview of the texts,[5] presenting token counts, (unlemmatised) type counts and type–token ratios—unstandardised (ttr) and standardised (sttr) using a base of 1,000—all as computed by WordSmith version 8.0 using the software's default settings. Table 3.1 shows that Oeser decreases the token count and simultaneously increases lexical variety, as evidenced by higher type-token ratios in *Oeser PE* than *DeepL MT*. This superficial analysis tells us nothing about the particular edits Oeser makes, however. Our keywords analysis below will shed some light on this issue.

Table 3.1 Basic for statistics for DeepL MT and Oeser PE.

	tokens	*types*	*ttr*	*sttr*
DeepL MT	104,892	10,603	10.11	45.84
Oeser PE	103,511	12,210	11.80	47.55

3 www.deepl.com/translator (last accessed 30 September 2022).

4 The corresponding statistics for the source text are: 104,096 tokens; 8,123 types; ttr 7.80; sttr 43.35.

We are interested not just in how Oeser's post-edited version differs from the MT version, but also in whether the interventions Oeser makes as a post-editor are consistent with his translatorial style. We thus compare *Oeser PE* with a purpose-built corpus, *Oeser 12*, containing 12 novels/novellas translated by Oeser as the sole translator over a roughly contemporaneous period (2016 to 2021).[6]

A third comparison, this time with a corpus called *Original German Literature*, sheds further light on Oeser's style, indicating whether it is distinctive from or consistent with that of other German-language authors. *Original German Literature* runs to 3,596,676 tokens. It contains 57 novels and novellas (excluding crime fiction) extracted from the 53 billion-word[7] German Reference Corpus of the *Institut für Deutsche Sprache* (IDS). These 57 works were published between 2000 and 2012, the most recent year covered by the IDS.[8]

Although our analysis is primarily target-oriented, our qualitative analysis sometimes requires us to look back at the source text to seek extra contextual information about what prompted particular translations. For this purpose, we use a small parallel corpus consisting of the source text (*Isherwood ST*) and two target texts (*DeepL MT* and *Oeser PE*) aligned at paragraph level and accessed using Tetrapla (Woolls, 2021).[9]

Procedure

We start our analysis by generating a keyword list for *Oeser PE* using the keywords function in WordSmith Tools 8.0 and taking *DeepL MT* as our comparison corpus.[10] In general, the closer the comparison corpus is to the text under study in terms of genre and other extratextual factors the better (Culpeper, 2009). In our case, the comparison corpus is extremely close to the study text, differing only in the translation condition (post-edited MT vs raw MT). This reduces the chances that the keyword analysis will highlight words that indicate what the study text is 'about', which might serve as a distractor in a study of style (Scott & Tribble, 2006). The output is a

5 *Oeser 12* contains 699,315 tokens and 52,427 types and has a standardized ttr of 50.47.

6 Correct as of 08 March 2022.

7 The IDS allows users to build a customised sub-corpus from the German Reference Corpus. However, there is no function to create frequency lists or other statistics. Counts for selected types (see Section 3.6.4) were thus done by looking them up individually.

8 We would like to thank David Woolls for developing *Tetrapla* and optimizing it for German.

9 The relevant settings are: maximum wanted = 500; minimum frequency of occurrence of candidate keywords in the study text = 3; minimum BIC score = 2.5; minimum log ratio = 0; p value = 0.1. The p value is set high as it can effectively be ignored if BIC values are used (Scott, 2022).

Table 3.2 Positive keywords in Oeser PE compared to DeepL MT.

N	*Key word*	*Freq. in Oeser PE*	*% in Oeser PE*	*Freq. in DeepL MT*	*% in DeepL MT*	*BIC*	*Log_L*
1	ACH	144	0.14	0	0.00	189.20	201.45
2	DOCH	275	0.27	62	0.06	135.88	148.13
3	GAR	81	0.08	4	0.00	74.30	86.55
4	S	70	0.07	4	0.00	60.05	72.30
5	SIE	2,370	2.29	1,875	1.79	52.01	64.26
6	ELIZABETHS	49	0.05	2	0.00	42.17	54.42
7	VERMUTLICH	36	0.03	0	0.00	38.11	50.36
8	BESTIMMT	31	0.03	0	0.00	31.12	43.37
9	SCHON	183	0.18	83	0.08	27.56	39.80
10	FALLS	37	0.04	2	0.00	26.48	38.73
11	EBEN	33	0.03	1	0.00	26.27	38.52
12	ELIZABETH	372	0.36	226	0.22	25.62	37.87
13	FURCHTBAR	36	0.03	3	0.00	21.08	33.33
14	GEWISS	20	0.02	0	0.00	15.73	27.98
15	WESHALB	19	0.02	0	0.00	14.33	26.58
16	HAB	24	0.02	1	0.00	14.30	26.55
17	O	30	0.03	3	0.00	13.74	25.98
18	MERKTE	18	0.02	0	0.00	12.93	25.18
19	JA	230	0.22	138	0.13	12.17	24.42
20	GERN	17	0.02	0	0.00	11.53	23.78
21	STETS	17	0.02	0	0.00	11.53	23.78
22	JENEM	17	0.02	0	0.00	11.53	23.78
23	SCHWARZENSEE	16	0.02	0	0.00	10.13	22.38

list of candidate keywords ranked in descending order of their log likelihood and BIC score (see Table 3.2).

Because we are also interested in whether Oeser's post-edits reaffirm his style as observed elsewhere, and whether these features of his style differentiate him from other writers in German, we go on to compare the relative frequencies of keywords identified in *Oeser PE* with their relative frequencies in our two reference corpora: *Oeser 12* and *Original German Literature*. We test whether any differences observed are statistically significant, again using log likelihood and BIC scores, this time calculated using Rayson's log-likelihood and effect size calculator.[11] This effectively means that we test the keyness in other data sets of a handful of positive keywords identified in *Oeser PE*.

10 Available from https://ucrel.lancs.ac.uk/llwizard.html.

Results and Discussion

Table 3.2 shows all 23 positive keywords in *Oeser PE* generated using *DeepL MT* as a comparison corpus.[12] Space constraints prevent us from discussing all of them, so in the following, we attempt to account primarily for groups of related keywords.[13]

Proper Nouns and Inconsistent MT

All keywords proposed in Table 3.2 tell us something interesting about changes that Oeser has made to the machine-translated text. Even proper nouns (ranked 6, 12 and 23) throw up surprises: although it is common to find proper nouns in keyword lists, this was not expected to happen in the current case, given the extremely close relationship between the study text and the comparison corpus and thus apparently complete overlap of their 'aboutness'. The explanation for this finding is trivial, however. DeepL is inconsistent in the spelling (*s* vs *z*) of its translation of *Elizabeth* and *Elizabeth's*. Such inconsistency in lexical translation is a well-known problem even in state-of-the-art neural MT systems. Oeser fixes the problem by standardising to *Elizabeth*/*Elizabeths*. What appears in Table 3.2 as a keyword in his post-edited text is actually the trace of error correction rather than a reflection of his style. The other proper noun (*Schwarzensee*, ranked 23) appears because Oeser changes the spelling for the lake known as the *Schwarzsee* in both the source text and *DeepL MT*.

Fictional Dialogue

More interesting are the seemingly innocuous words *s* (ranked 4; a contracted form of *es* 'it') and *hab* (ranked 16; a contracted form of [*ich*]

11 125 negative keywords were generated in *Oeser PE* using the same settings. Due to space constraints, we will not elaborate on them here except to the extent that they shed light on the positive keywords we discuss.

12 Log likelihood (LL) values can be interpreted as follows, according to Rayson (See fn. 11):
LL of 3.8 or higher is significant at the level of $p < 0.05$
LL of 6.6 or higher is significant at $p < 0.01$
95th percentile; 5% level; $p < 0.05$; critical value = 3.84
99th percentile; 1% level; $p < 0.01$; critical value = 6.63
99.9th percentile; 0.1% level; $p < 0.001$; critical value = 10.83
99.99th percentile; 0.01% level; $p < 0.0001$; critical value = 15.13
BIC scores can be interpreted as follows, according to Gabrielatos (2018):
below 0: not trustworthy (or evidence in favour of H0 according to Rayson (ibid.) 0–
2: not worth more than a bare mention
2–6: positive evidence against H0 6–
10: strong evidence against H0
>10: very strong evidence against H0

habe '[I] have'). They are used in depictions of direct speech by Oeser but not by DeepL. The elevated frequency of these forms in *Oeser PE* is consistent with what is already known about Oeser's skill in writing convincing dialogue.

Oeser's frequent use of interjections and (potential) modal particles such as *ach* (ranked 1), *doch* (ranked 2), *gar* (ranked 3), *schon* (ranked 9), *eben* (ranked 11) and *ja* (ranked 19) may also be interpretable in this light, although it should be noted that instances of *ach* and *o* (ranked 17) tend to be straightforward replacements for the untranslated particle *Oh* in *DeepL MT*, and that the other forms mentioned here are generally polysemous, so each instance needs to be inspected individually to come to sound conclusions. Even the overrepresentation of the polysemous pronoun *sie*/*Sie* (ranked 5), one use of which is as a formal translation of 'you', is at least partly attributable to its use in Oeser's fictional dialogue.

Nearly half of the keywords identified in Table 3.2 may thus be linked to Oeser's particular way of handling fictional dialogue. Space restrictions prevent us from analysing these data further here. For now, we simply note the importance of keyword analysis in leading us towards higher-order features of texts (such as the treatment of dialogue) that differentiate the post-editor's work from that of the machine.

Lexical Preferences and Marker Words in Oeser PE

Of the other keywords in Table 3.2, *vermutlich* (ranked 7), *bestimmt* (ranked 8) and *gewiss* (ranked 14) form a group: all three function primarily as epistemic modal adjuncts and thus indicate the speaker's assessment of the truth of a proposition, for example, whether it is possible, probable or certain. They translate roughly as 'probably' (in the case of *vermutlich*) and 'certainly' (in the case of *bestimmt* and *gewiss*). The combined frequency of these three forms in *Oeser PE* is 87, while there are no occurrences at all in *DeepL MT*. All three represent 'marker words' for Oeser, that is words that reflect particular likes or dislikes of an author, compared to a 'competing' author (Kenny, 1982, p. 8), in this case DeepL. We analyse each of these three keywords in more detail below.

Vermutlich

A parallel concordance for *vermutlich* in *Oeser PE* alongside the corresponding segments in *DeepL MT* and *Isherwood ST* reveals that DeepL had output *ich nehme an* for 'I suppose' in 17 instances and *wahrscheinlich* for 'probably' in 16 instances. In two cases where Oeser used *vermutlich*, DeepL had output *ich denke*. In one case *vermutlich* replaced *ich schätze* (see Table 3.3).

Table 3.3 Vermutlich in Oeser PE and corresponding items in DeepL MT and Isherwood ST.

Oeser PE	*DeepL MT*	*Isherwood ST*
Vermutlich (36)	Ich nehme an (17)	I suppose (17)
	Wahrscheinlich (16)	Probably (16)
	Ich denke (2)	I expect (1)
		I think (1)
	Ich schätze (1)	I guess (1)

In no case does Oeser's edit affect the meaning or epistemic stance of the speaker. The changes from the verbal *ich nehme an* to the adverbial *vermutlich* do, however, influence the sentence structure, which changes from a hypotactic structure to a simple main clause in 12 instances (Example 1) or from a main clause to an incomplete sentence in five instances (Example 2).[14] The former allows inversion, which Oeser is known to like (see Section 3.3). The latter, it could be argued, yields a better approximation of spoken language.

Example 1

(1a) Isherwood ST I **suppose** I still regarded marriage as a kind of game.
(1b) DeepL MT Ich **nehme an**, ich betrachtete die Ehe immer noch als eine Art Spiel.
(1c) Oeser PE **Vermutlich** betrachtete ich die Ehe noch immer als eine Art Spiel.

Example 2

(2a) Isherwood ST 'Yes. I **suppose** so.'
(2b) DeepL MT „Ja. Ich **nehme** es **an**."
(2c) Oeser PE „Ja. **Vermutlich**."

Likewise, Oeser's choice of *vermutlich* to replace the synonymous *wahrscheinlich* has little impact on meaning or register (Example 3).

13 Note that particular examples are chosen as they attest the use of the word form in question in a relatively short sentence with few 'distractors'. By their very nature they may underrepresent the amount of editing that Oeser does across the text as a whole.

Example 3

(3a) Isherwood ST Yes, the Jane-situation still existed, and would continue to exist, **probably**, for a long time.
(3b) DeepL MT Ja, die Jane-Situation existierte noch und würde **wahrscheinlich** noch lange bestehen bleiben.
(3c) Oeser PE Ja, die Jane-Situation existierte noch und würde **vermutlich** noch lange existieren.

Indeed, Oeser could have maintained all instances of *wahrscheinlich* in his post-edited version, but he appears to actively dislike the word.[15] Not only does he change it to *vermutlich* in the 16 instances already noted, but on no occasion does he introduce *wahrscheinlich* (e.g. as a possible replacement for *ich nehme an* in the DeepL output). Oeser thus reduces the overall frequency of *wahrscheinlich* from 61 in the MT output to 42 in his post-edited text. And when he does use *wahrscheinlich*, it is always already present in the MT output.

Bestimmt

Table 3.4 presents findings for *bestimmt*. Out of 31 instances, 28 have an epistemic function. The other three were used non-epistemically to mean 'firm' or 'firmly'.

Where Oeser uses *bestimmt* it is frequently to replace *sicher* ('sure') (12 instances) or *sicherlich* ('certainly') (seven instances). Edits from *sicher* to *bestimmt*, it could be argued, involve a slight decrease in assertiveness,

Table 3.4 Bestimmt in Oeser PE and corresponding items in DeepL MT and Isherwood ST.

Oeser PE	*DeepL MT*	*Isherwood ST*
bestimmt (28)	sicher (12)	sure (9), surely (2), certainly (1)
	sicherlich (7)	certainly (7)
	müssen (4)	must (3), bound to (1)
	definitiv (1)	definitely
	wahrscheinlich (1)	probably
	gar (1)	certainly
	nicht wahr (1)	didn't he
	bin ich gefesselt (1)	I'll be bound

14 This appears to be a conscious preference: when asked in an interview (conducted on 22 April 2022) whether he liked the word *wahrscheinlich*, Oeser immediately replied: 'Nein, ich sag meistens vermutlich'. ('No, I mostly say *vermutlich*'.)

and some edits (five instances) may have been made in an effort to create natural-sounding dialogue (Example 4).

Example 4

(4a) Isherwood ST 'I'm **sure** you look cute in it.'
(4b) DeepL MT „Ich bin **sicher**, du siehst darin süß aus."
(4c) Oeser PE „Du siehst **bestimmt** niedlich darin aus."

The instances where *sicherlich* is changed to *bestimmt* (with or without inversion), on the other hand, seem to be motivated exclusively by Oeser's preference (see Example 5, which also exemplifies inversion).

Example 5

(5a) Isherwood ST I would miss her, **certainly**.
(5b) DeepL MT Ich würde sie **sicherlich** vermissen.
(5c) Oeser PE **Bestimmt** würde ich sie vermissen.

3.6.3.4 Gewiss

Table 3.5 shows findings for *gewiss*.

Of the 20 instances of *gewiss* in *Oeser PE*, 14 replace *sicherlich* and five replace *sicher* (see Examples 6 and 7). These edits may result in a slight elevation of register (Oeser, personal communication 22/04/2022), although sources like the Duden dictionary suggest that *sicher*/*sicherlich* and *gewiss* are synonyms.

Example 6

(6a) Isherwood ST **Certainly** not because I imagine you'll disapprove of him.
(6b) DeepL MT **Sicherlich** nicht, weil ich mir vorstelle, dass du ihn missbilligen wirst.
(6c) Oeser PE **Gewiss** nicht, weil ich glaube, dass du ihn missbilligen wirst.

Table 3.5 Gewiss in Oeser PE and corresponding items in DeepL MT and Isherwood ST.

Oeser PE	*DeepL MT*	*Isherwood ST*
gewiss (20)	sicherlich (14)	certainly (13), surely (1)
	sicher (5)	certainly (2), sure (3)
	gar (1)	certainly (1)

Example 7

(7a) Isherwood ST '**Sure**, I understand all about that, Bob.
(7b) DeepL MT **Sicher**, ich verstehe das alles, Bob.
(7c) Oeser PE **Gewiss**, das alles verstehe ich, Bob.

The flip side of Oeser's preference for *gewiss* seems to be a dislike of the synonymous *sicher* and *sicherlich*. He reduces their frequency considerably, by a third in the case of *sicher* (from 168 to 106 instances) and by three-quarters in the case of *sicherlich* (from 46 to ten instances). Unsurprisingly, both *sicher* and *sicherlich* thus appear as negative keywords for *Oeser PE* when *DeepL MT* is the comparison corpus.[16]

3.6.3.5 Weshalb

While the above three keywords can be grouped on semantic and pragmatic grounds, *weshalb* stands alone as the only interrogative form in the keyword list in Table 3.2. It translates as 'why' and is used by Oeser on all occasions as an interrogative or relative adverb to replace the more common synonymous form *warum* in the DeepL output. It is of particular interest to us, as its use has previously been identified as one of Oeser's 'quirks' (Kenny & Winters, 2020, p. 143). Although Oeser regards *weshalb* as being of higher register than *warum* (personal communication, 22/04/2022), the decision to change *warum* to *weshalb* is based entirely on his personal preference.

Oeser PE *Keywords in the Reference Corpora*

The analysis so far has concentrated on what keywords in *Oeser PE* (using *DeepL MT* as a comparison corpus) tell us about the particular changes Oeser has made to the machine-translated text, and we have referred obliquely to how these changes are manifestations of Oeser's style as a translator. More compelling direct evidence comes from comparisons between the frequencies of these keywords in *Oeser PE* and *Oeser 12*, and *Oeser PE* and *German Original Literature* (see Section 3.5.1).

15 *Sicher*: rank = 132, frequency in Oeser PE = 106, % in Oeser PE = 0.10, frequency in DeepL MT = 168, % in DeepL MT = 0.16, BIC = 1.13, LL = -13.38, p = 0.0002541700. *Sicherlich*: rank = 142, frequency in Oeser PE=10, % in Oeser PE = 0.01, frequency in DeepL MT = 46, % in DeepL MT = 0.04, BIC=12.38, LL = –24.63, p = 0.0000006923.

Table 3.6 gives the absolute and relative frequencies for the four keywords addressed above in the texts/corpora in question, sorted in descending order of absolute frequency in *Oeser PE*.[17]

As the bolding in Table 3.6 highlights, *vermutlich*, *bestimmt* and *gewiss* are all relatively more frequent in *Oeser PE* than in his wider work (*Oeser 12*). Only *weshalb* occurs relatively less frequently in *Oeser PE* than it does in *Oeser 12.* All four forms are relatively more common in *Oeser PE* than in *German Original Literature*. The comparison between *Oeser 12* and *German Original Literature* suggests that Oeser generally uses *vermutlich* and *weshalb* more than other writers in the target language while he is less likely than others to use *bestimmt* and *gewiss*. In his post-editing work then, he appears to be asserting his attested lexical style in the use of *vermutlich* and *weshalb.* He uses *bestimmt* and *gewiss*, however, not because they are characteristic of his style, but rather to avoid using *sicher* and *sicherlich*, which he generally disprefers. That said, he 'overuses' them in his post-edited work most likely because of the influence of the MT (although the source text could also be exerting an influence here) (see Table 3.7).

As indicated above, differences in frequencies can be tested for statistical significance using log likelihood, which compares observed and expected

Table 3.6 Frequency comparison of keywords in Oeser PE.

keyword	*Oeser PE*	*per 100,000*	*DeepL MT*	*per 100,000*	*Oeser 12*	*per 100,000*	*German Original Lit.*	*per 100,000*
vermutlich	36	**34.62**	0	0.00	127	18.12	269	7.48
bestimmt	31	**29.82**	0	0.00	118	16.84	654	18.18
gewiss	20	**19.24**	0	0.00	34	4.85	281	7.81
weshalb	19	18.27	0	0.00	168	**23.97**	257	7.15

Table 3.7 Lexical items Oeser generally disprefers (sorted by freq. in Oeser PE).

keyword	*Oeser PE*	*per 100,000*	*DeepL MT*	*per 100,000*	*Oeser 12*	*per 100,000*	*German Original Lit.*	*per 100,000*
sicher	106	101.95	168	**160.05**	197	28.11	1,456	40.48
wahrscheinlich	42	40.39	61	**58.11**	64	9.13	706	19.63
sicherlich	10	9.62	46	**43.82**	6	0.86	133	3.70

16 As keyness scores—and hence ranking—depend crucially on the comparison corpus in use, it is less useful to compare ranked keywords given changing comparison corpora. For this reason, we present relative frequencies in the first instance.

Table 3.8 Keyness of selected items Oeser PE vs Oeser 12.

	observed frequencies		*expected frequencies*		*over/under -use*		
	Oeser PE	*Oeser 12*	*Oeser PE*	*Oeser 12*		*LogLikelihood*	*BIC*
1. vermutlich	36	127	21.02	141.98	+	10.42	–3.17
2. bestimmt	31	118	19.21	129.79	+	7.19	–6.40
3. gewiss	20	34	6.96	47.04	+	20.14	6.54
4. weshalb	19	168	24.11	162.89	–	1.33	–12.27

Table 3.9 Keyness of selected items Oeser 12 vs German Original Literature.

	observed frequencies		*expected frequencies*		*over/under -use*		
	Oeser 12	*German Original Literature*	*Oeser 12*	*German Original Literature*		*Log Likelihood*	*BIC*
1. vermutlich	127	269	64.46	331.54	+	59.78	44.51
2. bestimmt	118	654	125.67	646.33	–	0.57	–14.70
3. gewiss	34	281	51.28	263.72	–	7.72	–7.55
4. weshalb	168	257	69.18	355.82	+	130.88	115.60

frequencies. Table 3.8, based on Rayson's log-likelihood and effect size calculator, suggests that *vermutlich*, *bestimmt* and *gewiss* are all significantly overused in *Oeser PE* compared to his wider work in *Oeser 12*, although negative BIC scores for the former two suggest that this result is 'untrustworthy' (Gabrielatos, 2018). The evidence for the underuse of *weshalb* also appears to be untrustworthy.

Comparisons between Oeser's wider work and other texts in German might be more fruitful, however, as the influence of individual texts is mitigated. Table 3.9 thus indicates that there is strong evidence that *vermutlich* and *weshalb* are overused by Oeser in general, which supports the interpretation of their use in *Oeser PE* as a manifestation of Oeser's style. Meanwhile, evidence to claim that Oeser underuses *bestimmt* and *gewiss* is far weaker.

Conclusions

This chapter has explored the use of keywords as a way of eliciting data for the analysis of a post-editing translator's style. The generation of an unlemmatised keyword list proved to be an efficient way of eliciting unbiased data for further examination, leading us to at least one higher-order feature that merits further attention, namely Oeser's treatment of fictional monologue and dialogue. Our comparative approach also allowed us to see keywords in Oeser's post-edited text not just as evidence of systematic editing, but also as indices of his lexical style as attested in a corpus of his translation work. Two words, *vermutlich* and *weshalb*, turned out to be extremely strong marker words for Oeser, and other keywords remain to be investigated in full. Although not anticipated at the outset, negative keywords also turned out to be of interest, given their ability to indicate the post-editor's dislikes. That disliked items remained to the extent they did in the post-edited version of the novel serves as a reminder of the strong priming influence of the machine-translated text. Also interesting, from the point of view of studies of style at least, is the fact that the keywords technique tends to draw the analyst's attention not to instances of error correction, which is the focus of much post-editing research, but to the ways in which the translator/post-editor asserts his style in the target text by making what are known in the post-editing literature as 'preferential' changes (see O'Brien, 2022, p. 118). Finally, this research differs from many existing inquiries into post-editing in that it is not concerned with productivity, and it focuses on a single named translator/post-editor, thanks to the availability of—to our knowledge—a unique data set.

References

Baker, M. (2000). Towards a methodology for investigating the style of a literary translator. *Target 12*(2), 241–266. https://benjamins.com/online/target/articles/target.12.2.04bak

Burrows, J.F. (2002). Delta: A measure of stylistic difference and a guide to likely authorship. *Literary and Linguistic Computing 17*, 267–287. https://doi.org/10.1093/llc/17.3.267

Castilho, S., Doherty, S., Gaspari, F. & Moorkens, J. (2018). Approaches to human and machine translation quality assessment. In J. Moorkens, S. Castilho, F. Gaspari & S. Doherty (Eds), *Translation quality assessment: From principles to practice* (pp. 9–38). Springer. https://www.researchgate.net/publication/326386260_Approaches_to_Human_and_Machine_Translation_Quality_Assessment_From_Principles_to_Practice

Culpeper, J. (2009). Keyness: Words, parts-of-speech and semantic categories in the character-talk of Shakespeare's *Romeo and Juliet*. *International Journal of Corpus Linguistics 14*(1), 29–59. https://www.researchgate.net/publication/233493422_Keyness_Words_parts-of-speech_and_semantic_categories_in_the_character-talk_of_Shakespeare's_Romeo_and_Juliet

Gabrielatos, C. (2018). Keyness analysis: Nature, metrics and techniques. In C. Taylor & A. Marchi (Eds.), *Corpus approaches to discourse: A critical review* (pp. 225–258). Routledge. https://www.researchgate.net/publication/319208347_Keyness_analysis_Nature_metrics_and_techniques

Herrmann, J.B. (2017). In a test bed with Kafka: Introducing a mixed-method approach to digital stylistics. *Digital Humanities Quarterly 11*(4). http://www.digitalhumanities.org/dhq/vol/11/4/000341/000341.html#

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. http://www.jltonline.de/index.php/articles/article/view/757/1776

Isherwood, C. (1954). *The World in the Evening*. Vintage.

Isherwood, C. (2019). *Die Welt am Abend* (Trans. H.-C. Oeser). Hoffmann und Campe.

Jeaco, S. (2020). Key words when text forms the unit of study. Sizing up the effects of different measures. *International Journal of Corpus Linguistics 25*(2), 125–154. https://doi.org/10.1075/ijcl.18053.jea

Kenny, A. (1982). *The computation of style*. Pergamon Press. https://www.sciencedirect.com/book/9780080242811/the-computation-of-style

Kenny, D. & Winters, M. (2020). Machine translation, ethics and the literary translator's voice. *Translation Spaces 9*(1), 123–149. https://doi.org/10.1075/ts.00024.ken

Kenny, D. & Winters, M. (2024). Customization, personalization and style in literary machine translation. In M. Winters, S. Deane-Cox & U. Böser (Eds.), *Technological change in translation and interpreting*. Bloomsbury. (In press).

Lee, C. (2021). How do machine translators measure up to human literary translators in stylometric tests? *Digital Scholarship in the Humanities 37*(3), 813–829. https://doi.org/10.1093/llc/fqab091

Mastropierro, L. (2018). *Corpus stylistics in heart of darkness and its Italian translations*. Bloomsbury.

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* (pp. 105–120). Language Science Press. https://doi.org/10.1080/0907676X.2021.1905673

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

Rayson, P. (2019). Corpus analysis of key words. In C.A. Chapelle (Ed.), *The concise encyclopedia of applied linguistics* (pp. 320–326). John Wiley. https://eprints.lancs.ac.uk/id/eprint/140803 /1 /Rayson_2019 _CorpusAnalysisofKeyWords_Submitted.pdf

Rybicki, J. (2012). The great mystery of the (almost) invisible translator: Stylometry in translation. In M. Oakes & M. Ji (Eds.), *Quantitative methods in corpus-based translation studies* (pp. 231–248). John Benjamins. https://doi.org/10.1075/scl.51.09ryb

Saldanha, G. (2011). Translator style: Methodological considerations. *The Translator 17*(1), 25–50. https://doi.org/10.1080/13556509.2011.10799478

Saldanha, G. (2014). Style in, and of, translation. In S. Berman & C. Porter (Eds.), *A companion to translation studies* (pp. 95–106). John Wiley. https://www.researchgate.net/publication/301104802_Style_in_and_of_Translation

Scott, M. (2022). How key words are calculated. WordSmith tools help. https://lexically.net/downloads/version_64_8/HTML/keywords_calculate_info.html

Scott, M. & Tribble, C. (2006). *Textual patterns: Key words and corpus analysis in language education*. John Benjamins. https://www.researchgate.net/publication/238737537_Textual_Patterns_Key_Words_and_Corpus_Analysis_In_Language_Education

Taylor, C. (2018). Similarity. In C. Taylor & A. Marchi (Eds.), *Corpus approaches to discourse: A critical review* (pp. 19–37). Routledge. https://doi.org/10.4324/9781315179346

Toral, A. (2019). Post-editese: An exacerbated translationese. In *Proceedings of Machine Translation Summit XVII: Research Track* (pp. 273–281). European Association for Machine Translation. https://aclanthology.org/W19-6627

Way, A. (2018). Quality expectations of machine translation. In J. Moorkens, S. Castilho, F. Gaspari & S. Doherty (Eds.), *Translation quality assessment: From principles to practice* (pp. 159–178). Springer. https://doi.org/10.48550/arXiv.1803.08409

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 52*(3), 412–425. https://doi.org/10.7202/016728ar

Winters, M. (2009). Modal particles explained: How modal particles creep into translations and reveal translators' styles. *Target 21*(1), 74–97. https://benjamins.com/online/target/articles/target.21.1.04win

Winters, M. (2015). Hans-Christian Oeser: Translation strategies of a literary translator. In S. Egger (Ed.), *Cultural/literary translators: Selected Irish-German biographies II* (pp. 210–224). Wissenschaftlicher Verlag Trier.

Woolls, D. (2021). *Tetrapla plus*. CFL Software Ltd.

Youdale, R. (2020). *Using computers in the translation of literary style*. Routledge. https://doi.org/10.4324/9780429030345