

# Repeated reporting verbs in English novels and their Italian and Polish translations: A preliminary multifactorial study

Lorenzo Mastropiero<sup>1</sup>  and Łukasz Grabowski<sup>2\*</sup> 

<sup>1</sup> University of Insubria, Italy

<sup>2</sup> University of Opole, Poland

## ORIGINAL RESEARCH PAPER

Received: December 13, 2023 • Revised manuscript received: March 14, 2024 • Accepted: July 5, 2024

© 2024 The Author(s)



### ABSTRACT

In this explanatory study, we aim to determine the predictors of repetition avoidance or reproduction in the translation of reporting verbs from English into Italian and Polish, using a sample of 11 novels. We fit multiple negative binomial regression models with fixed effects to assess the impact of five predictor variables (i.e. frequency of a source-text verb, number of its translation equivalents in lexical databases, its number of senses, semantic-pragmatic category of verb, and length in characters) on the response variable, i.e., the number of different target language verb types into which a source text reporting verb is translated. The overall model fit per the lowest Akaike Information Criterion (AIC) value obtained through backward elimination reveals that verb category, frequency of a source-text reporting verb, and number of translation equivalents make the largest individual contributions to explaining the proportion of variation in the response variable in the Italian data; for the Polish translations the corresponding variables are verb category, frequency, the number of senses, and the interaction between the number of translation equivalents and the number of senses. The summaries of the final models provide a detailed multifactorial picture of when repetition of reporting verbs is maintained or avoided in literary translation.

### KEYWORDS

repetition, reporting verbs, literary translation, multifactorial statistics, parallel corpus

\* Corresponding author. E-mail: lukasz@uni.opole.pl

## 1. INTRODUCTION

In recent times, the field of corpus-based translation and interpreting studies has experienced a surge in popularity (Baker, 1995; Dayter, 2021; Delaere & De Sutter, 2013; Grabowski, 2013; Kotze, 2019; Laviosa, 2002; Mastropiero, 2020). Upon closer examination of the latest research, it becomes evident that researchers are increasingly aware of the limitations inherent in conventional research methodologies, both quantitative (descriptive statistics) and qualitative (concordance analysis), which have proven inadequate for addressing the nuanced complexities of linguistic variation in translation and interpreting. Recent discussions (e.g., De Sutter & Lefer, 2020) emphasise that monofactorial descriptive studies conducted by corpus linguists, such as those examining frequency differences of specific linguistic features between source and target texts or between translational and non-translational texts, do not substantially contribute to our understanding of translation. Beyond mere frequency, there exists a multitude of other factors influencing language use in translation. Consequently, there is a growing necessity to employ advanced statistical methodologies, such as regression models, to account more effectively for multifaceted empirical phenomena (Calzada Pérez & Laviosa, 2021; De Sutter & Lefer, 2020; De Sutter, Lefer, & Vanroy, 2023; Kajzer-Wietrzny & Grabowski, 2021; Maekelberghe & Delaere, 2023; Mastropiero, 2022). In this paper, we employ multifactorial statistics to study the phenomenon of repetition in the source text and how it is dealt with in translation. More precisely, focusing primarily on the ways reporting verbs found in a selection of English literary texts are translated from English into Italian and Polish, we aim to pinpoint precisely the factors that to the greatest degree impact translational decisions and, consequently, the level of repetition and/or linguistic variation in translation.

## 2. REPETITION IN LANGUAGE, TRANSLATION AND BEYOND

Repetition is a central element in many aspects of human social and cultural expression (Resina & Wulf, 2019: x-xi), including linguistic behaviours. In language, repetition can manifest itself in a multitude of different patterns, on several linguistic levels (e.g., morphological, syntactic, lexical, etc., as discussed by Dobaczewski, 2018), fulfilling a variety of functions. The pervasive nature of linguistic repetition has been extensively demonstrated by corpus linguistics research. Building on Sinclair's (1991, 2004) seminal work, corpus linguistics has shown that language is highly patterned and that the repetition of linguistic features plays a fundamental part in language production. One of the underlying assumptions of corpus linguistics is that the repetition of a linguistic feature is a reflection of its functional relevance (Mahlberg, 2010: 297); therefore, studying repeated patterns can shed light on how texts enact their communicative functions and convey meaning.

An area of research that has prominently emphasised the functional relevance of repetition in language is stylistics, especially in its applications to literary texts. In stylistics, repetition is considered as one of the main devices through which foregrounding, an “artistically motivated deviation” (Leech & Short, 2007: 39), can be created. Foregrounding through repetition strikes readers' attention because of its violation of standard rules of usage – overuse in this case. Moreover, repetition can be seen as establishing networks of “echoically linkable” structures that achieve a “focusing and depth of texture [...] atypical in non-literary discourse” (Too-lan, 2012: 21). Repetition can therefore not only convey meanings, but can also be used to create



aesthetic and stylistic effects, particularly in literary texts. Overall, as [Wales \(2011: 366\)](#) explains, “it is impossible not to appreciate the significance of repetition, on all linguistic levels, in literary language.”

In the context of translation studies, the investigation of repetition comes with an extra layer of complexity compared to monolingual studies, due to the comparative and multilingual nature of translation research. In addition to exploring the role of reiterated patterns in the source text (ST), studying repetition in translation also involves establishing a relationship between repetition in the ST and its reproduction in the target text (TT). This is necessary to answer questions such as whether or not repetitive patterns have been reproduced in translation, to what extent, what the implications are of the potential divergences between ST and TT, and what factors may have influenced, limited, or encouraged the recreation of reiterated patterns. However, establishing a relationship between repetition in the ST and repetition in the TT is far from straightforward. In fact, in addition to the choices made by translators at the level of translation strategies and techniques, there are many factors that can affect such a relationship, like differences in stylistic conventions between languages and cultures ([Nádvorníková, 2020](#); [Piotrowski, 1994](#); [Urzha, 2018](#)), typological divergences between source and target language ([Comrie, 1981/1989](#)), as well as the translator’s own style ([Saldanha, 2011a; 2011b](#)). For example, by analysing repetitions in the Polish-to-French translation of the novel *Dukla* by Andrzej Stasiuk, [Mitura \(2019\)](#) recognises the existence of repetitions predetermined by the typological characteristics of a language, in addition to supra-language repetitions, which reflect translator’s creativity/idiolect/style as independent from a specific language system. Different stylistic conventions are discussed by [Piotrowski \(1994\)](#), who points out differences in the perception of repetitive features between English and Slavic languages’ speakers. He argues that, in contrast to speakers of English, speakers of Russian, Polish, and other Slavic languages tend to choose lexemes with specific meanings, rather than ones with a broad meaning, which is why the former regard texts with multiple repetitions as having plain, simple, or even bad style ([Piotrowski, 1994: 95–96](#)). Moreover, the growing use of computer technologies for translation is becoming an additional important factor affecting the relationship between repetition in the ST and its reproduction in the TT, as automated procedures may reproduce in the TT the repeated features of the ST without the direct input of the human translator. For example, computer-assisted translation tools (CATs), in widespread use since the mid-1990s, take advantage of repeatedly used text fragments in the ST recorded in translation memories to automate the translation process as far as possible. As [Garcia \(2015: 72\)](#) puts it, “[t]he more internal repetition, the better, since as the catchcry says ‘with TM one need never translate the same sentence twice’.”

Despite its importance and complexity, there is surprisingly little research that explores repetition in translation, notably in the English-to-Polish and English-to-Italian language pairs. Rather, what the existing literature focuses on are the consequences of repetition avoidance on the style and interpretation of individual target texts. In fact, according to [Ben-Ari \(1998: 3\)](#), “[o]ne of the most persistent and inflexible norms in translation [...] is that of avoiding repetitions.” This tendency is probably the result of a persistent assumption that sees repetition as synonymous with an impoverished style. [Wales \(2011: 366\)](#) explains that repetition is often associated with unsophistication, with the lack of premeditation that is typical of ordinary language; therefore, it is frequently avoided in favour of variation by synonymy or substitution. This negative perception is then reflected in translation, where – to use the words of influential Czech-French writer Milan [Kundera \(2015: 164, cited in Elbanowski, 2019: 76\)](#) – “repetitions



drive translators mad” (our translation). Kundera (2015, cited in Elbanowski, 2019) maintains that translators should preserve all repetitions (of particular words and phrases) and restrain themselves from “synonymization” (i.e. using synonyms in an attempt to avoid repetition). However, studies such as Mastropierro (2020), Čermáková (2015, 2018), Čermáková and Mahlberg (2018), Čermáková and Fárová (2010), Bednarczyk (2015), and Zupan (2006) show that reiterated patterns in the ST are seldom reproduced in the TT. Building on the stylistics literature on the role of repetition in the creation of stylistic effects, these studies illustrate how replacing repetition with variation can impact the recreation of the original’s style in the TT, with consequences for the aesthetic and literary effects that repetition may be contributing to. The focus on the consequences of avoiding repetition in translation offers important insights into the significance of translator choices, emphasising once more the manipulative nature (Hermans, 2014) of translation as a form of rewriting (Lefevere, 1992). However, such approaches shed little light on repetition as a linguistic and translational phenomenon in its own right. Scarce attention has been paid, for example, to when repetition avoidance occurs, with what types of items, and in what textual contexts. Hence, more research is needed that explores repetition and its causes in translation – as opposed to its consequences – across a wider variety of texts and language pairs. This provides motivation for us to undertake a comparative multifactorial study like this one.

### 3. REPORTING VERBS IN TRANSLATION: BETWEEN REPETITION AND LEXICAL VARIETY

There are two main interconnected reasons why reporting verbs are an ideal unit of analysis in the study of repetition in translation: their stylistic significance and their frequency. Reporting verbs play an important stylistic function, that is, they contribute to characterisation (Ruano San Segundo, 2016). In fact, reporting verbs not only attribute reported speech to a speaker, but also gloss the reported speech with additional prosodic and evaluative meanings that can be completely unrelated to the content of the utterance. For example, in a sentence like “*I won’t tell him*”, *he shouted*, the reported verb *shouted* adds prosodic, emotional, and interpretative meanings to the utterance “*I won’t tell him*”. Such an utterance would be perceived completely differently if it was instead followed by *he whispered*. Patterns of reporting verbs in fiction have been shown to contribute to character building (Ruano San Segundo, 2016, 2017, 2018) and development (Mastropierro, 2020). Moreover, reporting verbs play a similarly important role in translation. As Nádvořníková (2020: 213) explains, the potential of reporting verbs to “directly or indirectly *explicitate* the emotions or intentions of the speaker [...] may likewise be reflected in translation” (emphasis in the original). However, given “their dispersed distribution in the text”, it is not unlikely that translators “[overlook] the characterising value of [these] verbs” (Ruano San Segundo, 2017: 110). The existing literature on the translation of reporting verbs (e.g. Čermáková & Mahlberg, 2018; Mastropierro, 2020; Nádvořníková, 2020; Urzha, 2018, 2019) consistently shows that translators tend to alter the ST patterns of reporting verbs, with consequences for the style and the role that these patterns play in the TT. Also, stylistic conventions related to the use of reporting verbs vary across languages and cultures: for instance, Garbovskiy (2011: 16, cited in Urzha, 2018: 119) puts forward a hypothesis that English stylistic norm gives preference to the verb *say* when reporting dialogues of literary characters, while in



Russian we find a wide variety of reporting verbs performing the same function. Overall, by focusing on reporting verbs, we can confidently assume to be analysing an item the repetition of which is stylistically and functionally significant, not only in the original, but in translation too. There are some (in fact, very few, e.g., Lubocha-Kruglik & Malysa, 2019) studies on the translation of reporting verbs into Polish or from Polish, but they are merely qualitative and concern source or target languages other than English. There is a similar dearth of studies on the Italian translation of reporting verbs, especially in registers different from the academic ones (as in Masi, 2007). Hence, our contrastive research aims to fill this gap by providing new knowledge regarding the treatment of repeated reporting verbs in translation from English into Polish and Italian. In addition to being a functionally relevant form of repetition, reporting verbs are a suitable unit of analysis also because they are particularly common in literary language and occur frequently in most novels. On the contrary, linguistic items that are relatively rare in texts do not easily lend themselves to a large-scale multifactorial statistical investigation as the present one.

#### 4. METHODOLOGY: RESEARCH MATERIAL, STUDY STAGES AND RESEARCH QUESTIONS

This study investigates repetition in translation. More specifically, it explores the effect that a set of linguistic factors can have on the reproduction or avoidance of repetition of reporting verbs in the Italian and Polish translations of 11 English novels. The factors are (i) the frequency of the repeated ST verb, (ii) the number of senses of the ST verb, (iii) the number of translation equivalents of the verb in the target language, (iv) the semantic-pragmatic category of the ST verb, and (v) the length of the verb measured in number of characters. Thus, all the measures and variables used in the study are “linguistically interpretable” (Egbert, Larsson, & Biber, 2020), that is, they account for real-world language phenomena and have clear operational definitions (see also Larsson & Biber, 2024).

Reporting verbs were retrieved from 11 English novels and their Polish and Italian translations, accessed through InterCorp version 15 (Čermák & Rosen, 2012), a multilingual aligned parallel corpus freely available as part of the Czech National Corpus suite of tools (<https://www.korpus.cz/>). The novels were selected on the basis of the availability of both a Polish and an Italian translation from English: at the time of the data collection, there were 11 English novels which had both an Italian and a Polish translation, and these novels were selected for the analysis. The novels are the following: J.K. Rowling’s *Harry Potter and the Philosopher’s Stone*, *Harry Potter and the Chamber of Secrets*, *Harry Potter and the Prisoner of Azkaban*, *Harry Potter and the Order of the Phoenix*, and *Harry Potter and the Half-Blood Prince*; Dan Brown’s *The Da Vinci Code*; John Steinbeck’s *The Grapes of Wrath*; Douglas Adams’s *The Hitchhiker’s Guide to the Galaxy*; and J. R. R. Tolkien’s *The Fellowship of the Ring*, *The Two Towers*, and *The Return of the King*. Table 1 shows summary statistics for each text.

The reporting verbs were retrieved with a query that combined regular expressions and CQL syntax, using the Czech National Corpus interface tool, KonText, and language-specific tagsets<sup>1</sup>

<sup>1</sup>The tagsets are available at the following link: [https://wiki.korpus.cz/doku.php/en:cnk:intercorp:verze15#morphosyntactic\\_annotation](https://wiki.korpus.cz/doku.php/en:cnk:intercorp:verze15#morphosyntactic_annotation).



Table 1. Summary statistics for the 11 novels

Novel	Author	ST tokens	IT translator	IT TT tokens	PL translator	PL TT tokens
<i>Harry Potter and the Philosopher's Stone</i>	J. K. Rowling	98,656	M. Astrologo	102,533	A. Polkowski	86,495
<i>Harry Potter and the Chamber of Secrets</i>	J. K. Rowling	109,006	M. Astrologo	109,303	A. Polkowski	95,682
<i>Harry Potter and the Half-Blood Prince</i>	J. K. Rowling	216,036	B. Masini	197,144	A. Polkowski	184,806
<i>Harry Potter and the Prisoner of Azkaban</i>	J. K. Rowling	139,893	B. Masini	130,753	A. Polkowski	120,071
<i>Harry Potter and the Order of the Phoenix</i>	J. K. Rowling	326,326	B. Masini	298,920	A. Polkowski	284,225
<i>The Da Vinci Code</i>	D. Brown	170,886	R. Valla	172,852	K. Mazurek	163,628
<i>The Grapes of Wrath</i>	J. Steinbeck	239,308	C. Perroni	229,633	A. Liebfeld	203,663
<i>The Hitchhiker's Guide to the Galaxy</i>	D. Adams	58,548	L. Serra	58,880	P. Wiczorek	51,111
<i>The Fellowship of the Ring</i>	J. R. R. Tolkien	224,050	V. di Villafranca	218,742	M. Skibniewska	194,172
<i>The Two Towers</i>	J. R. R. Tolkien	187,603	V. di Villafranca	181,070	M. Skibniewska	166,875
<i>The Return of the King</i>	J. R. R. Tolkien	161,766	V. di Villafranca	209,467	M. Skibniewska	141,277

implemented in the InterCorp corpus. We searched for the following patterns in the STs: “closing quotation marks + noun phrase or *he* or *she* + past tense verb” and “closing quotation marks + past tense verb + noun phrase or *he* or *she*”. This means that we focused on finite verbs used in direct speech in the 3rd person singular, which are used to report “speech, thoughts and perceptions” and for this reason they are also called ‘quotative verbs’ (Klamer, 2000, p. 69). These patterns were matched by equivalent patterns in the target languages, taking into consideration specific orthographic conventions of recording dialogues in Italian and Polish, generating aligned search results like the ones shown in Fig. 1, through which we were able to identify and record the translation of each ST verb.<sup>2</sup> Using matching ST and TT queries we

<sup>2</sup>The complete CQL queries are available on a data repository at the following link: <https://osf.io/bj2m5/>.



InterCorp v15 - English <input checked="" type="checkbox"/>		InterCorp v15 - Italian <input checked="" type="checkbox"/>	
"Let Hedwig out, " <b>he</b> told Ron .	rowlingova-hpot_tajemn	'Libera Edvige' <b>disse</b> a Ron . 'Può seguirci in volo .	
"I'll pack these for you, " <b>Hermione said</b> brightly, taking Harry's presents out of his arms as the three of them headed back upstairs .	rowlingova-hpot_relik	« Te li metto via io » <b>disse</b> allegramente Hermione , prendendo i regali dalle braccia di Harry e i tre si avviarono di nuovo di sopra .	
"I ca n't stand much more of this, " <b>Ron muttered</b> , his teeth chattering, as the orchestra ground back into action and the ghosts swept back onto the dance floor .	rowlingova-hpot_tajemn	'Non credo che riuscirò a resistere ancora ' <b>bofonchiò</b> Ron battendo i denti , mentre l' orchestra rientrava in azione e i fantasmi tornavano a occupare la pista da ballo .	
"I'm sorry to hear that, " <b>Fache said</b> , watching Langdon closely .	brown-sifra	« Mi dispiace » <b>disse</b> , osservandolo con attenzione .	
"They would n't mind, " <b>Harry reassured</b> her .	rowlingova-hpot_tajemn	'Non credo che gliene importerebbe molto ' <b>la rassicurò</b> Harry .	
"We got to git along, " <b>Ma said</b> .	Steinbeck-Hrozny_hnevu	"Dobbiamo spicciarci, " <b>disse</b> Ma ' .	
"Yes, sir, " <b>he said</b> .	Steinbeck-Hrozny_hnevu	"Sissignore, " <b>disse</b> .	
InterCorp v15 - English <input checked="" type="checkbox"/>		InterCorp v15 - Polish <input checked="" type="checkbox"/>	
"Wait, " <b>said</b> Trillian .	adams-zivot_vesmir	- Czekajcie ! - <b>krzyknęła</b> Trillian .	
"Hactar ! " <b>called</b> Trillian .	adams-zivot_vesmir	- Haktar ! - <b>krzyknęła</b> Trillian .	
"Hactar ? " <b>she called</b> again .	adams-zivot_vesmir	- Haktar ? ! - <b>krzyknęła</b> znowu .	
"Hello ? " <b>she said</b> again .	adams-zivot_vesmir	Halo ? - <b>zawołała</b> ponownie .	
"Why, " <b>said</b> Trillian , " do you feel you have to destroy the Universe ? "	adams-zivot_vesmir	- Dlaczego uważasz , że musisz zniszczyć wszechświat ? - <b>zapytała</b> Trillian .	
"All right, " <b>said</b> Trillian firmly .	adams-zivot_vesmir	- Świetnie - <b>powiedziała</b> Trillian pewnym głosem .	
"So, did you make, " <b>asked</b> Trillian again, " the model of the wrecked spacecraft ? "	adams-zivot_vesmir	- Czy zrobił eś - znów <b>zapytała</b> Trillian - model rozbitego statku ?	

Fig. 1. Query samples for English-to-Italian and English-to-Polish pairs of ST and TT reporting verbs

ensured that the translations of the ST reporting verbs retrieved were also reporting verbs in the TTs. It is crucial for our study, where we explore repetition or lexical variety within reporting verbs in STs and TTs, taking into consideration language-specific (English, Italian and Polish) stylistic conventions related to the use of this class of linguistic items. It means that multiple reporting verb equivalents would indicate lexical variety while few reporting verb equivalents would indicate repetition. To illustrate the former, the 3rd person singular past tense verb *muttered* (20 occurrences) in the English-original novel *Harry Potter and the Chamber of Secrets* was translated into Polish using 9 verbs: *warknął*, *wymamrotał*, *bełkotał*, *jęknął*, *napisał*, *powtarzał*, *mruknął*, *wyjakał*, and *mruczał*, which indicates lexical variety and avoidance of repetition in translation. Conversely, the English reporting verb form *whispered* (23 occurrences) was translated into Polish in the same novel with 4 verbs: *szeptnął*, *wyszeptał*, *oznajmił*, *zapytał*, which indicates lower lexical variety and more repetition within Polish translation equivalents.

We are aware that ST reporting verbs could also be rendered with different classes of items (nominalizations, multiword units, etc.) or omitted altogether in translation, but we consider such translational solutions as separate instances of lexical variety that do not contribute to the stylistic effect of reporting verb patterns in the STs and TTs. As mentioned earlier, Toolan (2002: 21) explains that repetition creates networks of “echoically linkable” structures, which build on the cohesive connections that verbatim reiteration establishes (Halliday & Hasan, 1976). Mastropiero and Mahlberg (2017) and Mastropiero (2018) show that these cohesive networks are disrupted when repeated terms in the ST are replaced with different classes of linguistic items in the TTs. Finally, it is worth pointing out that all of our novels are narrated in the third-person



and in the past tense, therefore we can assume that instances of reporting verbs occurring in patterns different from the ones searched (e.g. first-person or present-tense narration) are a fraction compared to the verbs that our queries retrieved. Once we excluded hapax legomena,<sup>3</sup> we collected 16,742 verb tokens in the original novels (with 23–79 ST reporting verb types across the novels, which yields the average of 47.45 per ST), as well as all of their translations in Italian and in Polish.

The obtained data was downloaded from InterCorp in the XLSX format and prepared for statistical data analysis. At the stage of data preparation, the Polish masculine and feminine past tense forms of reporting verbs were reduced to the single masculine form (e.g., *powiedziała*, “she said” was reduced to *powiedział*, “he said”), a procedure that was facilitated using custom-designed Python scripts and undertaken to ensure comparability between Polish and Italian data.

The linguistic features of the repeated verbs were annotated in terms of the five factors listed above. The first factor is “freq”, the frequency of occurrence of the reporting verb in the STs, directly retrieved from InterCorp. Given the fact that the frequency data is skewed (i.e. verb forms such as *said*, *asked*, *told* or *replied* have considerably higher frequency than the remaining verbs in each ST), we applied the log transformation to the “freq” factor, as suggested by West (2022). The second factor is “trans”, that is, the number of translation equivalents of the ST verb as indicated in Treq (Škrabal & Vavřín, 2017), a translation equivalents database (<https://treq.korpus.cz/#>) that employs data from InterCorp to provide a list of equivalents for a query word in any of the language pairs available in the corpus. To maximise the retrieval of potential equivalents in Treq, we searched for lemmas, but then recorded only those translation equivalents that were verbs and accounted for at least 4% of the translation options. For example, Fig. 2 shows the results of a search for the lemma *shout*. *Gridare* and *urlare* are shown to be used as translation equivalents of *shout* in Italian in 49.9% and 24.6% of the cases respectively, while *grido*, even though it occurs in 9% of cases, is not a reporting verb but a noun. Hence, the value of “trans” for *shout* would be 2.

The third factor is “senses”, which represents the polysemy of the ST verb and was operationalised by retrieving the number of senses of each verb from the lexical database WordNet 3.1 (<https://wordnet.princeton.edu/>, Fellbaum, 1998). For example, Fig. 3 shows a query for the verb form *screamed*, which is reported to have three distinct senses; hence, the value of “senses” for *screamed* would be 3.

The fourth factor is “verbtype”, which indicates the semantic-pragmatic category of the reporting verb based on Caldas-Coulthard’s (1987) classification. This taxonomy divides reporting verbs into seven main categories on the basis of their discursal function, but also taking into account semantic as well as paralinguistic features of the verb. For example, whereas “structuring” verbs like *asked* and *replied* are defined in terms of their role in structuring an exchange pair (e.g., *she asked*, *he replied*), “prosodic” verbs like *shouted* or *yelled* are defined as such because they describe the manner in which the utterance is uttered rather than the content of the utterance. The other verb categories are shown in Table 2 (see Caldas-Coulthard (1987) for a

<sup>3</sup>The full list of ST verbs (including hapax legomena) retrieved and their TT equivalents are downloadable from a data repository at the following link: <https://osf.io/bj2m5/>.





**Treq** | TRANSLATION  
EQUIVALENTS  
DATABASE

Source language: English | Target language: Italian | Restrict to: Collection(s): 7

shout [Search]

Lemma  Multiword  RegEx  A = a

▲ Frequency ▼	▲ Proportion ▼	▲ English ▼	▲ Italian ▼
967	49.9	shout	<a href="#">gridare</a>
476	24.6	shout	<a href="#">urlare</a>
175	9.0	shout	<a href="#">grido</a>
64	3.3	shout	<a href="#">urlo</a>
42	2.2	shout	<a href="#">esclamare</a>

Fig. 2. Query for shout in Treq

WordNet Search - 3.1  
- [WordNet home page](#) - [Glossary](#) - [Help](#)

Word to search for: screamed [Search WordNet]

Display Options: (Select option to change) [Change]

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations  
Display options for sense: (gloss) "an example sentence"

**Verb**

- **S: (v)** [shout](#), [shout out](#), [cry](#), [call](#), [yell](#), [scream](#), [holler](#), [hollo](#), [squall](#) (utter a sudden loud cry) "she cried with pain when the doctor inserted the needle"; "I yelled to her from the window but she couldn't hear me"
- **S: (v)** [yell](#), [scream](#) (utter or declare in a very loud voice) "You don't have to yell--I can hear you just fine"
- **S: (v)** [scream](#) (make a loud, piercing sound) "Fighter planes are screaming through the skies"

Fig. 3. Query for screamed in WordNet

complete discussion of the taxonomy). The categories that are relevant in our analysis will be discussed in the analysis sections.

The fifth and final factor is “wordlength” and it indicates whether the verb is longer or shorter than the median length of all words in each novel. More precisely, “wordlength” was retrieved as follows: first, the length in characters of all 3rd person past tense reporting verbs in



**Table 2.** Reporting verb taxonomy (based on [Caldas-Coulthard, 1987](#))

Category	Sub-category	Examples
Neutral		<i>say, tell</i>
Structuring		<i>ask, inquire, reply, answer</i>
Metapropositional	Assertive	<i>exclaim, proclaim, agree</i>
	Directive	<i>urge, instruct, order</i>
	Expressive	<i>accuse, lament, swear</i>
Metalinguistic		<i>narrate, quote, recount</i>
Prosodic		<i>cry, shout, scream</i>
Paralinguistic	Voice qualifier	<i>whisper, murmur, mutter</i>
	Voice qualification	<i>laugh, sigh, groan</i>
Signalling discourse		<i>repeat, add, go on, hesitate</i>

each ST novel was calculated; then, the median word length was calculated; if the number of characters of a reporting verb was equal or higher than the median, then that verb was labelled as “long”; verbs that were shorter than the median were labelled as “short”. Since word length correlates with word familiarity and with the degree of polysemy ([Grzybek, 2014](#)), it is reasonable to assume that repetition of longer words is more likely to be avoided in translation (as the translator is more likely to notice their repetition).

The final annotated data was saved in the CSV file format. Next, a negative binomial regression model with fixed effects was fitted to the data to verify whether these five factors had a significant effect on “types”, that is, the number of different translations each ST verb was translated into. For example, a value of “types” of 11 indicates that the ST reporting verb was translated into 11 different verbs in the TT. A value of 1 indicates that the repetition of the verb in the ST was maintained in the TT, while a value higher than 1 indicates that the repetition of the ST verb was replaced with lexical variety: the higher the value, the higher the variety of different verbs that were used in the TT to translate the same repeated ST verb. Usually, when an outcome variable is a count variable measured in non-negative integer numbers (as it is the case in this study), the Poisson regression is used, which is a type of generalised linear model ([Winter, 2019](#), p. 247). However, in our study we observed overdispersion in both Italian and Polish data, that is, the variance was found to be higher than the mean. For this reason, we used negative binomial regression instead, as recommended by [Winter \(2019\)](#) and [Scherber \(2017, 2019\)](#), among others. Moreover, [Hair, Black, Babin, and Anderson \(2009\)](#), p. 176 recommend that for multiple regression at least 15–20 observations (i.e. reporting verb types in English source-texts) per predictor should be used to yield reliable findings. In view of the fact that in our novels we had between 23 and 79 ST reporting verb types (i.e. circa 47 per novel), we decided to collate all the data instead of running the analysis on individual texts. In this way, we also ensured that the response variable does not have an excessive number of zeros (in fact, none), so that we avoided the problem of zero-inflation ([Tu, 2002](#)).



Our goal was to identify the best fitting model (i.e. the model with the lowest value of Akaike Information Criterion (AIC), which is a measure of the model quality (see Brezina, 2018, pp. 124–125), that is, the one that reaches significance with as few variables as possible through their “backward selection” (Winter, 2019, p. 310). This is a type of stepwise regression using  $p$  values: we start with a full model with all potential predictor variables and iteratively remove those that are not significant statistically, i.e. have  $p$ -values lower than 0.05 (Winter, 2019, p. 310). With the final model in mind, we pinpointed statistically significant predictors of our response variable “types” and determined which semantic-pragmatic categories of reporting verbs were more or less likely translated into different TT verbs. Our analyses were conducted in the R environment using the following packages: *car* (Fox & Weisberg, 2019), *MASS* (Venables & Ripley, 2002) and *pscl* (Jackman, 2020). The final data and the programming scripts are available on the Open Science Framework repository to ensure reproducibility and replicability of our study.<sup>4</sup>

This methodology will allow us to answer the following research questions: (i) What linguistic factors have a significant effect on the avoidance or reproduction of repetition of reporting verbs in the Italian and Polish novels? (ii) Are there any language-specific differences in the way these factors impact repetition of reporting verbs in the TTs in Polish and Italian? Or can a more generalisable trend in how translators deal with repetition be recognised? By answering these questions, this study will offer important cross-linguistic insights into the translation of repeated linguistic items.

Finally, it is worth pointing out that the present study builds on Mastropiero (2022), addressing the limitations of our previous study by encompassing more verbs, texts, languages, and translators. Whereas Mastropiero (2022) focuses on the translation of reporting verbs in the *Harry Potter* series in Italian, in this study we also include an additional target language (Polish) with its accompanying stylistic conventions, an additional factor (“wordlength”), multiple texts with different authors and translators, which altogether increases substantially the number of ST and TT verbs taken into account in our fixed effect model. With a larger evidential pool, we aim to improve the generalisability of the results, making the study findings more widely relevant for the exploration of repetition in translation.

## 5. RESULTS: ENGLISH-TO-ITALIAN TRANSLATION

In order to assess the influence of the five predictor variables on the response variable “types”, we fitted a series of negative binomial regression models using backward selection. An optimal quality model with the lowest AIC value (1629.2), a measure of goodness of fit (Brezina, 2018, pp. 124–125), was the following:  $types \sim verbtype + logfreq + trans$  (deviance table in Fig. 4). This model shows that the verb type, the frequency, and the number of translation equivalents of a ST verb contribute significantly to the model’s fit, although the significance of “trans” is borderline. In other words, the deviance table shows the incremental contribution of each

<sup>4</sup>See the link: <https://osf.io/bj2m5/>.



	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
NULL			462	1659.25	
verbtype	7	1092.27	455	566.98	< 2e-16 ***
logfreq	1	320.12	454	246.86	< 2e-16 ***
trans	1	3.22	453	243.64	0.07262 .
---					
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					

Fig. 4. Optimal model (anova) for Italian data: a deviance table

variable to the model’s fit and their individual contributions in explaining variance<sup>5</sup> in the response variable “types”. The analysis was then repeated removing *said* from the data. By far the most frequent verb and the verb with the largest number of different translations, *said* was clearly an outlier; by removing it from the data, we tested whether *said* was skewing our models. Our results did not change, however: the optimal model was the same, as were the significant factors. This indicates that the outlier nature of *said* in our data does not significantly affect results.

In order to evaluate the overall fit of the final model (*said* included), with all significant predictor variables, a summary is shown in Fig. 5, with coefficients that estimate effect size of each predictor (including all levels of categorical variables such as “verbtype”) on the response variable “types”. As the deviance table and final model summary display different aspects of model performance, we may note slight differences between the two. Statistically significant predictors are the ones where *p*-value is lower than 0.05, while the estimates reveal the magnitude of the effect and its direction, which may be a positive or negative association between a predictor and the response variable.

Overall, based on *p*-values and positive estimates, we can conclude that how often a ST reporting verb occurs (“logfreq”), whether it is a “neutral” verb (e.g., *said*, *tell*) and the number of translation equivalents it has (although the last two are borderline predictors) influence the chances of seeing that verb translated into multiple Italian TT verbs. For example, for a one-unit change (increase) in “logfreq”, i.e. the frequency of a ST reporting verb, the number of “types” (TL equivalents) increases  $e^{0.48}$ -fold,<sup>6</sup> that is, by 0.48 units, while holding the rest of the predictor variables constant.<sup>7</sup> In the same way, for a one-unit change (increase) in “trans”, i.e. the number of TL equivalents of a ST verb, “types” increases  $e^{0.055}$ -fold, that is, by 0.055 units. Conversely, “signalling discourse” verbs (e.g., *repeat*, *add*) are more likely to be consistently reproduced with the same target-text equivalents in the Italian translations.

<sup>5</sup>According to McFadden’s pseudo R-squared, the proportion of explained variance in the model is 28.47% compared to a model with only an intercept while maximum likelihood R-squared (r2ML) estimates the explained variance at 74%, indicating moderate and good model fit respectively. However, negative binomial regression does not have an equivalent metric to R-squared used in linear or logistic regression, so both metrics used here should be interpreted with caution (UCLA: Statistical Consulting Group, 2023).

<sup>6</sup>This can be also read as follows: if a (log)frequency of a ST reporting verb increases by 1, then the log number of “types” increases by 0.48.

<sup>7</sup>According to the guidelines provided by UCLA: Statistical Consulting Group (2023: online): “we interpret negative binomial regression coefficients as follows: for a one-unit change in the predictor variable, the difference in the logs of expected counts of the response variable is expected to change by the respective regression coefficient”.



```

Call:
glm.nb(formula = types ~ verbytype + logfreq + trans, data = data,
       init.theta = 13.6301528, link = log)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.02452   0.10289   0.238  0.8117
verbytypeN   0.30647   0.16639   1.842  0.0655 .
verbytypePros -0.17199  0.11635  -1.478  0.1393
verbytypeSdis -0.22725  0.11531  -1.971  0.0488 *
verbytypeStr -0.17462  0.15643  -1.116  0.2643
verbytypeVier -0.15484  0.11431  -1.355  0.1755
verbytypeVion -0.06859  0.08700  -0.788  0.4305
verbytypeMlin -0.77609  1.03921  -0.747  0.4552
logfreq      0.48366   0.02844  17.005 <2e-16 ***
trans        0.05505   0.03059   1.800  0.0719 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(13.6302) family taken to be 1)

Null deviance: 1659.25  on 462  degrees of freedom
Residual deviance: 243.64  on 453  degrees of freedom
AIC: 1629.2

Number of Fisher Scoring iterations: 1

```

**Fig. 5.** Summary of the final model for Italian data (summary)

## 6. RESULTS: ENGLISH-TO-POLISH TRANSLATION

In a similar way, in order to assess the influence of the five predictor variables on the number of different Polish translations each ST English verb was translated into, we fitted a series of negative binomial regression models using backward selection. An optimal quality model with the lowest AIC value (1578.5) was the following:  $types \sim verbytype + logfreq + trans * senses$  (deviance table in Fig. 6). This model indicates that the verb type, the frequency of a ST reporting verb as well as the interaction between the number of translation equivalents and the number of senses of a ST reporting verb contribute significantly to the model's fit. In other words, we can see the incremental contribution of each variable to the model's fit and their individual contributions in explaining variance<sup>8</sup> in the response variable "types" in Polish translations. In this case too, the analysis was repeated removing *said* from the data: likewise, the optimum model did not change after removing it.

In order to evaluate the overall fit of the final model (*said* included), with all predictor variables considered together, a summary is shown in Fig. 7, with coefficients that estimate effect size of each predictor (including all levels of categorical variables such as "verbytype") on the

<sup>8</sup>According to McFadden's pseudo R-squared, the proportion of explained variance in the model is 38.78% compared to a model with only an intercept while the maximum likelihood R-squared (r2ML) estimates the explained variance at 89%, indicating moderate and good model fit respectively, and better than in the case of Italian data.



	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
NULL			452	4700.7	
verbytype	7	3652.3	445	1048.4	< 2e-16 ***
logfreq	1	833.8	444	214.6	< 2e-16 ***
trans	1	2.3	443	212.3	0.12715
senses	1	4.9	442	207.3	0.02617 *
trans:senses	1	6.5	441	200.9	0.01083 *

---  
 Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Fig. 6. Optimal model (anova) for Polish data: a deviance table

```
Call:
glm.nb(formula = types ~ verbytype + logfreq + trans * senses,
       data = data, init.theta = 57.8587743, link = log)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-0.251172	0.149488	-1.680	0.092916 .
verbytypeN	0.455537	0.133869	3.403	0.000667 ***
verbytypePros	-0.153613	0.099913	-1.537	0.124180
verbytypeSdis	-0.251094	0.104425	-2.405	0.016193 *
verbytypeStr	-0.409590	0.129129	-3.172	0.001514 **
verbytypeVier	-0.194335	0.105486	-1.842	0.065435 .
verbytypeVion	-0.119051	0.084120	-1.415	0.156996
verbytypeMlin	0.182861	0.597757	0.306	0.759671
logfreq	0.591217	0.023145	25.544	< 2e-16 ***
trans	0.097861	0.031347	3.122	0.001797 **
senses	0.029168	0.017024	1.713	0.086651 .
trans:senses	-0.008737	0.003373	-2.591	0.009580 **

---  
 Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(57.8588) family taken to be 1)

Null deviance: 4700.74 on 452 degrees of freedom  
 Residual deviance: 200.86 on 441 degrees of freedom  
 AIC: 1578.5

Number of Fisher Scoring iterations: 1

Fig. 7. Summary of the final model for Polish data (summary)

response variable “types”. As it has been explained above, the deviance table and final model summary show different aspects of model performance, hence the differences between the two. Statistically significant predictors are the ones where *p*-value is lower than 0.05, while the estimates reveal size effects of each predictor on the response variable and its direction (a positive or negative one).

Overall, based on *p*-values and positive estimates, we can see that how often a ST reporting verb occurs (“logfreq”), whether it is a “neutral” verb (e.g., *said*, *tell*), and the number of translation equivalents it has influence the chances of seeing that verb translated into multiple Polish TT verbs. For example, for a one-unit change (increase) in “trans”, “types” increases  $e^{0.098}$ -fold, that is, by 0.098 units, while holding the rest of the predictor variables constant. On the contrary,



other types of reporting verbs (“structuring”, e.g., *ask*, *reply*; “signalling discourse”, e.g., *repeat*, *add*), as well as the interaction between the number of senses of a ST verb and the number of its translation equivalents increase the chances of seeing more repetition in the Polish translation.

## 7. DISCUSSION AND CONCLUSIONS

The optimum data models obtained using negative binomial regression with fixed effects, per the lowest AIC value yielded through backward selection, show the variables that have a significant effect on the number of different translations each ST verb was translated into. The deviance tables in Figs 4 and 6 indicate which predictors have the largest incremental contribution to the model’s fit and the largest individual contribution in explaining the proportion of variation in the response variable “types”. For the Italian data these predictors are the type and the frequency of ST reporting verbs, as well as the number of different translation equivalents; while, for the Polish data, they are the type and the frequency of ST reporting verbs, the number of senses of ST reporting verbs plus the interaction between the number of translation equivalents and the number of senses.

The summaries of the final models (i.e. the coefficients) in Figs 5 and 7 provide estimates of size effects of each significant predictor and its direction (positive or negative) regarding the response variable. For both Italian and Polish, the frequency, being a “neutral” verb, and the number of translation equivalents (a borderline predictor for Italian) of the ST verb are positively correlated with lexical variation in translation, which implies fewer repetitions. Here, some differences between the final Italian and Polish models can be noticed. As for the former, only the “discourse signalling” verbs are negatively correlated with lexical variation in the Italian translations, while in the Polish model “structuring” verbs and the interaction between the number of senses of a ST verb and the number of its translation equivalents also contribute to more repetition in translation.

As for the shared results, our findings indicate that frequency and verb type of the ST verb are significant predictors of the response variable in both the Italian and the Polish data. As mentioned in Section 2, repetition is often seen negatively and as a symptom of “bad style”, especially within the remit of literary language. Therefore, it does not surprise us to see that the frequency of the ST verb correlates positively with the degree of lexical variety in translation. It is reasonable to assume that the more often the same verb occurs in the ST, the more likely the translator notices its repetition and avoids it in the TT. The significance of “freq” in both languages seems to suggest that avoiding the repetition of reporting verbs in translation is a conscious strategy translators adopt. This may be obvious for *said*, the repetition of which is less tolerated in languages other than English (Mastropierro, 2020, pp. 244–245). However, the same tendency can be seen at work with other frequent verbs too, as the models without *said* show the same correlation between “freq” and “types”.

This conscious strategy can also explain the significance of “verbtype”, at least as far as “neutral” verbs are concerned. It is in fact the repetition of “neutral” verbs that is more likely to be avoided in translation, and “neutral” verbs comprise mainly *said* (Caldas-Coulthard, 1987, p. 152). As for “signalling discourse” and “structuring” verbs, the coefficients in Figs 5 and 7 show the opposite picture; that is, these types of verbs correlate negatively with the response variable “types”. The reason why the repetition of these verbs is less likely to be avoided in



translation may be related to the function these items perform in the ST. “Signalling discourse” verbs mark “the relationship of the quote to other parts of the discourse” (e.g., *repeated* or *echoed*) or “mark the development of the discourse”, like *paused* or *broke in* (Caldas-Coulthard, 1987, pp. 163–164). “Structuring” verbs, such as *answered* or *asked*, “describe the way in which a given speech act [...] fits into a sequence of speech acts” (Caldas-Coulthard, 1987, p. 155). Both verb types, then, organise the structure and sequence of the reported dialogue, rather than indicating how something is said. Translating a verb like *broke in* with a verb, let’s say, like *explained*, which does not necessarily mark how dialogue lines interact with each other, may affect the very organisation of the reported exchange, and this may be the reason why translators tend to keep the repetition of “signalling discourse” and “structuring” verbs unaltered. It is worth pointing out that only “discourse signalling” verbs are negatively correlated with lexical variation in the Italian translations, while in the Polish model “structuring” verbs also correlate negatively with the response variable. As for the significance of “trans”, it may suggest that Italian and Polish translators tend to use different target language equivalents if these are available to them: the wider the range of different dictionary equivalents at their disposal, the wider the degree of lexical variety adopted to translate the repetition of the same ST verb. It is important to stress, though, that the significance of “trans” in the Italian data is borderline, which may signal the need for additional data to confirm or confute this finding. Finally, in both Italian and Polish translations, the word length of the ST verb turned out to be an insignificant predictor.

The main difference between the Italian and Polish data concerns the negative correlation between the interaction of “trans” and “senses” with the number of TT equivalents in the latter dataset. It may seem to be counterintuitive that those verbs that have multiple senses and have many translation equivalents in a translation equivalents database such as Treq are associated with fewer repetitions in translation. However, we have to express two important reservations here. First, Treq records forms rather than senses and this limitation also pertains to the verbs recorded as equivalents there. Second, polysemy (along with homonymy) may be realised differently across languages (Srinivasan & Rabagliati, 2015; Zuercher, 2019), which means that the interaction between the two predictors is not straightforward. In other words, as there are no direct correspondences between languages when it comes to the degree of polysemy of words that refer to the same concept, the number of equivalents in the target language for a given sense of the ST reporting verb may vary. Hence, the interaction between translation equivalents and polysemy may result in the negative association (more repetition) that we observed in the Polish data. This finding too signals the need for additional data to understand this phenomenon better. Overall though, the differences noticed between the two languages are minor compared to the similarities, which indicates a more generalisable trend in how translators deal with repetition. Of course, this study has a number of limitations, which call for caution when invoking generalisability. It takes into account only two languages, Polish and Italian, only one genre (literary texts), a limited sample of novels (11) and a narrow selection of five potential predictors. However, our preliminary results clearly suggest that the way translators deal with repetition is influenced by the features of the repeated items. Understanding what features influence the translation of repetition can have important implications for translation practice and training: by learning what can prompt the translation of repeated items into a wider lexical variety, we can improve translation strategies to deal with repetition, making them more sensitive to the stylistic effects of the original. We strongly believe that this study is a first step in this direction.





There are several other steps that can be taken next to develop this project further. As it seems reasonable that translators may avoid repetition the more they notice it, the span of repetition could be considered as an additional independent variable. For example, testing whether repetition is more likely to be avoided if it occurs within a limited text fragment (e.g., a paragraph) compared to whole chapters or texts could bring further evidence to support the hypothesis that repetition avoidance is an intentional strategy. Other potential predictors could be the gender of the speaker of the reported speech, the date of publication of the TTs, or individual translators. With the latter specifically, considering the fact that our sample of texts is, in theory, selected from a potentially vast number of novels translated from English into both Italian and Polish, taking into account the translators as a random intercept would allow us to conduct a mixed-effects negative binomial regression to provide a more comprehensive explanation of the phenomenon. In the same vein, the impact of individual source text may be another factor impacting the results and as such can be also used as a random intercept. It is also possible to consider another operationalization of the response variable, e.g. the proportion of top-frequency TL equivalent in the total number of equivalents of a ST reporting verb, which would include the number of types and tokens in the metric. Moreover, the findings for literary texts could be compared to other text types or genres where reporting verbs occur (e.g., press articles). With other text types, notably specialised ones, it may be justified to use binary independent variables related to whether the translation was proofread, produced using a machine-translation or other AI-assisted translation at some stage. To further explore the impact of language-specific stylistic conventions on translation of reporting verbs, a similar study should be conducted in the reverse direction, that is, in the Italian-to-English and Polish-to-English translation. Finally, it is particularly interesting how these and other technological advancements would affect literary translation in the future, especially the degree of repetition found in it, which we attempted to explore in this preliminary study.

## ACKNOWLEDGEMENTS

This research was funded by the National Science Centre (NCN), Poland, grant number: 2023/51/B/HS2/00697.

## REFERENCES

- Baker, M. (1995). Corpora in translation studies: An overview and some suggestions for future research. *Target*, 7(2), 223–243. <https://doi.org/10.1075/target.7.2.03bak>.
- Bednarczyk, A. (2015). Idiolekt w przekładzie. *Między Oryginałem a Przekładem*, 29, 57–68.
- Ben-Ari, N. (1998). The ambivalent case of repetitions in literary translation. Avoiding repetitions: A “universal” of translation? *Meta*, 43(1), 68–78. <https://doi.org/10.7202/002054ar>.
- Brezina, V. (2018). *Statistics for corpus linguistics*. Cambridge University Press.
- Caldas-Coulthard, C. R. (1987). In M. Coulthard (Ed.), *Reported speech in written narrative texts. Discussing discourse* (pp. 149–167). University of Birmingham.



- Calzada Pérez, M., & Laviosa, S. (2021). Twenty-five years on: Time to pause for a new agenda for CTIS. *MonTI. Monografías de Traducción e Interpretación*, *TI*, (13), 7–32. <https://doi.org/10.6035/MonTI.2021.13.01>.
- Čermák, F., & Rosen, A. (2012). The case of InterCorp, a multilingual parallel corpus. *International Journal of Corpus Linguistics*, *17*(3), 411–427. <https://doi.org/10.1075/ijcl.17.3.05cer>.
- Čermáková, A. (2015). Repetition in John Irving's novel *A Widow for one year*. A Corpus stylistic approach to literary translation. *International Journal of Corpus Linguistics*, *20*(3), 355–377. <https://doi.org/10.1075/ijcl.20.3.04cer>.
- Čermáková, A. (2018). Translating children's literature: Some insights from corpus stylistics. *Ilha do Desterro: A Journal of English Language, Literatures in English and Cultural Studies*, *71*(1), 117–134. <https://doi.org/10.5007/2175-8026.2018v71n1p117>.
- Čermáková, A., & Fárová, L. (2010). Keywords in *Harry Potter* and their Czech and Finnish translation equivalents. In F. Čermák, P. Corness, & A. Klégr (Eds.), *InterCorp: Exploring a multilingual corpus* (pp. 177–188). NLN.
- Čermáková, A., & Mahlberg, M. (2018). Translating fictional characters – Alice and the queen from the wonderland in English and Czech. In A. Čermáková, & M. Mahlberg (Eds.), *The corpus linguistics discourse. In honour of Wolfgang Teubert* (pp. 223–253). John Benjamins Publishing Company. <https://doi.org/10.1075/scl.87.10cer>.
- Comrie, B. (1981/1989). *Language universals and linguistic typology: Syntax and morphology*. Oxford: Basil Blackwell.
- Dayter, D. (2021). Strategies in a corpus of simultaneous interpreting. Effects of directionality, phraseological complexity, and position in speech event. *Meta*, *65*(3), 594–617. <https://doi.org/10.7202/1077405ar>.
- De Sutter, G., & Lefer, M.-A. (2020). On the need for a new research agenda for corpus-based translation studies: A multi-methodological, multifactorial and interdisciplinary approach. *Perspectives*, *28*(1), 1–23. <https://doi.org/10.1080/0907676X.2019.1611891>.
- De Sutter, G., Lefer, M.-A., & Vanroy, B. (2023). Is linguistic decision-making constrained by the same cognitive factors in student and in professional translation? Evidence from subject placement in French-to-Dutch news translation. *International Journal of Learner Corpus Research*, *9*(1), 60–95. <https://doi.org/10.1075/ijlcr.22005.des>.
- Delaere, I., & De Sutter, G. (2013). Applying a multidimensional, register-sensitive approach to visualize normalization in translated and non-translated Dutch. *Belgian Journal of Linguistics*, *27*, 43–60. <https://doi.org/10.1075/bjl.27.03del>.
- Dobaczewski, A. (2018). *Powtórzenie jako zjawisko tekstowe i systemowe. Repetycje, reduplikacje i quasi-tautologie w języku polskim*. Toruń: Wydawnictwo UMK.
- Egbert, J., Larsson, T., & Biber, D. (2020). *Doing linguistics with a corpus: Methodological considerations for the everyday user. Cambridge elements in corpus linguistics*. Cambridge University Press.
- Elbanowski, A. (2019). Przekład literacki z perspektywy pisarzy. *Language and Literary Studies of Warsaw*, *9*, 67–85.
- Fellbaum, C. (Ed.) (1998). *WordNet: An electronic lexical database*. MIT Press.
- Fox, J., & Weisberg, S. (2019). *An R Companion to applied regression* (3rd ed.). Thousand Oaks CA: SAGE. <https://socialsciences.mcmaster.ca/jfox/Books/Companion/>.
- Garbovskiy, N. (2011). Перевод и “переводной дискурс”. [Translation and translational discourse]. *Вестник Московского университета. Серия 22, Теория перевода [Science Journal of Moscow State University. Series 22, Theory of Translation]*, *4*, 3–19.



- Garcia, I. (2015). Computer-aided translation: Systems. In C. Sin-wai (Ed.), *The Routledge Encyclopedia of translation technology* (pp. 68–87). Routledge.
- Grabowski, Ł. (2013). Interfacing corpus linguistics and computational stylistics: Translation universals in translational literary Polish. *International Journal of Corpus Linguistics*, 18(2), 254–280. <https://doi.org/10.1075/ijcl.18.2.04gra>.
- Grzybek, P. (2014). Word length. In J. Taylor (Ed.), *The Oxford handbook of the word* (pp. 1–25). Oxford University Press.
- Hair, Jr., J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2009). *Multivariate data analysis* (7th ed.). Pearson Prentice Hall.
- Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. Longman.
- Hermans, T. (Ed.) (2014). *The manipulation of literature: Studies in literary translations*. Routledge.
- Jackman, S. (2020). *pscl: Classes and methods for R developed in the political science computational laboratory*. United States Studies Centre. University of Sydney, Sydney, New South Wales, Australia. R package version 1.5.5.1.
- Kajzer-Wietrzny, M., & Grabowski, Ł. (2021). Formulaicity in constrained communication: An intermodal approach. *MonTI. Monografías de Traducción e Interpretación*, TI, (13), 148–183. <https://doi.org/10.6035/MonTI.2021.13.05>.
- Klamer, M. (2000). How report verbs become quote markers and complementizers. *Lingua*, 110(2), 69–98. [https://doi.org/10.1016/S0024-3841\(99\)00032-7](https://doi.org/10.1016/S0024-3841(99)00032-7).
- Kotze, H. (2019). Converging what and how to find out why: An outlook on empirical translation studies. In L. Vandevoorde, J. Daems, & B. Defranq (Eds.), *New empirical perspectives on translation and interpreting* (pp. 333–371). Routledge.
- Kundera, M. (2015). *Sztuka powieści*. Warszawa: Grupa Wydawnicza Foksal.
- Larsson, T., & Biber, D. (2024). On the perils of linguistically opaque measures and methods: Toward increased transparency and linguistic interpretability. In P. Crosthwaite (Ed.), *Corpora for language learning: Bridging the research-practice divide* (pp. 131–141). Taylor & Francis.
- Laviosa, S. (2002). *Corpus-based translation studies: Theory, findings, applications*. Rodopi.
- Leech, G., & Short, M. (2007). *Style in fiction. A linguistic introduction to English fictional prose* (2nd ed.). Pearson Education.
- Lefevere, A. (1992). *Translation, rewriting, and the manipulation of literary fame*. Routledge.
- Lubocha-Kruglik, J., & Malysa, O. (2019). Glagoly reči i strategija avtora: (na materiale pol'skogo perevoda “Žizni nasekomyh” Viktora Pelevina). In A. Banaszek-Szapowalowa (Ed.), *Słowiński krąg: Słowo – Myśl – Obraz w tradycji i współczesności* (pp. 154–164). Katowice: Wydawnictwo Uniwersytetu Śląskiego.
- Maekelberghe, Ch., & Delaere, I. (2023). Functional hybridity in translation: A multifactorial perspective on the English gerund in the language pairs English–German and English–Dutch. *Languages in Contrast*, 23(2), 252–275. <https://doi.org/10.1075/lic.00029.mae>.
- Mahlberg, M. (2010). Corpus linguistics and the study of nineteenth-century fiction. *Journal of Victorian Culture*, 15(2), 292–298. <https://doi.org/10.1080/13555502.2010.491667>.
- Masi, S. (2007). The dynamics of intersubjectivity as a stance-shaping device: English vs. Italian verbs of report in argumentative texts. *Texus*, 20(1), 181–203.
- Mastropierro, L. (2018). Key clusters as indicators of translator style. *Target*, 30(2), 240–259.
- Mastropierro, L. (2020). The translation of reporting verbs in Italian: The case of the *Harry Potter* series. *International Journal of Corpus Linguistics*, 25(3), 241–269. <https://doi.org/10.1075/ijcl.19124.mas>.



- Mastropierro, L. (2022). The avoidance of repetition in translation: A multifactorial study of repeated reporting verbs in the Italian translation of the *Harry Potter* series. In L. Defang, & R. Moratto (Eds.), *Advances in corpus applications in literary and translation studies* (pp. 138–157). Routledge.
- Mastropierro, L., & Mahlberg, M. (2017). Key words and translated cohesion in Lovecraft's at the Mountains of Madness and one of its Italian translations. *English Text Construction*, 10(1), 78–105.
- Mitura, K. (2019). Powtórzenie w oryginale, powtórzenie w przekładzie Uwagi o zjawisku repetycji na materiale Dukli Andrzeja Stasiuka i jej wersji francuskiej. *Między Oryginałem a Przekładem*, 1(51), 87–108.
- Nádvorníková, O. (2020). Differences in the lexical variation of reporting verbs in French, English and Czech fiction and their impact on translation. *Languages in Contrast*, 20(2), 209–234. <https://doi.org/10.1075/lic.00016.nad>.
- Piotrowski, T. (1994). *Problems in bilingual lexicography*. Wrocław: Wydawnictwo Uniwersytetu Wrocławskiego.
- Resina, J. R., & Wulf, C. (Eds.) (2019). *Repetition, recurrence, returns: How cultural renewal* Wors. Lexington Books.
- Ruano San Segundo, P. (2016). A corpus-stylistic approach to Dickens' use of speech verbs: Beyond mere reporting. *Language and Literature*, 25(2), 113–129. <https://doi.org/10.1177/0963947016631859>.
- Ruano San Segundo, P. (2017). Reporting verbs as a stylistic device in the creation of fictional personalities in literary texts. *Atlantis. Journal of the Spanish Association of Anglo-American Studies*, 39(2), 105–124. <https://doi.org/10.28914/Atlantis-2017-39.2.06>.
- Ruano San Segundo, P. (2018). Dickens's hyperbolic style revisited: Verbs that describe sounds made by animals used to report the words of male villains. *Style*, 52(4), 475–493. <https://doi.org/10.1353/sty.2018.0047>.
- Saldanha, G. (2011a). Translator style: Methodological considerations. *The Translator*, 17(1), 25–50. <https://doi.org/10.1080/13556509.2011.10799478>.
- Saldanha, G. (2011b). Style of translation: The use of foreign words in translations by Margaret Jull Costa and Peter Bush. In A. Kruger, K. Wallmach, & J. Munday (Eds.), *Corpus-based translation studies: Research and applications* (pp. 237–258). Bloomsbury Publishing.
- Scherber, C. (2017). Using R to interpret interaction effects in statistical models. *Software Developer's Journal*. Online version. [https://www.researchgate.net/profile/Christoph\\_Scherber/publication/312093784\\_Using\\_R\\_to\\_Interpret\\_Interaction\\_Effects\\_in\\_Statistical\\_Models/links/586f67ad08ae329d6215fc4c/Using-R-to-Interpret-Interaction-Effects-in-Statistical-Models.pdf](https://www.researchgate.net/profile/Christoph_Scherber/publication/312093784_Using_R_to_Interpret_Interaction_Effects_in_Statistical_Models/links/586f67ad08ae329d6215fc4c/Using-R-to-Interpret-Interaction-Effects-in-Statistical-Models.pdf).
- Scherber, C. (2019). *An introduction to generalized linear models*. Online version. <http://www.christoph-scherber.de/content/PDF%20Files/Generalized%20linear%20models.pdf>.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.
- Sinclair, J. (2004). Trust the text. In: *Language, corpus and discourse*. Routledge.
- Škrabal, M., & Vavřín, M. (2017). Databáze překladových ekvivalentů Treq. *Časopis pro Moderní Filologii*, 99(2), 245–260.
- Srinivasan, M., & Rabagliati, H. (2015). How concepts and conventions structure the lexicon: Cross-linguistic evidence from polysemy. *Lingua*, 157, 124–152. <https://doi.org/10.1016/j.lingua.2014.12.004>.
- Toolan, M. (2012). Poems: Wonderfully repetitive. In R. Jones (Ed.), *Discourse and creativity* (pp. 17–34). Routledge.
- Tu, W. (2002). Zero-inflated data. In A. H. El-Shaarawi, & W. W. Piegorisch (Eds.), *Encyclopedia of environmetrics* (pp. 2387–2391). John Wiley & Sons.



- UCLA: Statistical Consulting Group. (2023). Introduction to SAS. Available at: <https://stats.oarc.ucla.edu/stata/output/negative-binomial-regression/> (Retrieved 30 November 30, 2023).
- Urzha, A. (2018). Стратегии интерпретации глаголов, вводящих речь, в современных русских переводах художественной прозы. [Strategies of reporting verbs' interpretation in modern Russian translations of fiction.]. *Vestnik Volgogradskogo gosudarstvennogo universiteta., Seriya 2, Yazykoznanie [science Journal of Volgograd State University. Linguistics]*, 17(4), 117–128.
- Urzha, A. (2019). “Сказал он” или “вздыхнул он”? Интерпретация глаголов, вводящих речь, в русском художественном переводе. In J. Lubocha-Kruglik, O. Małysa, & G. Wilk (Eds.), *Przestrzenie Przekładu* (pp. 145–165). Katowice: Wydawnictwo Uniwersytetu Śląskiego.
- Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.). Springer.
- Wales, K. (2011). *A dictionary of stylistics*. Pearson Education.
- West, R. M. (2022). Best practice in statistics: The use of log transformation. *Annals of Clinical Biochemistry*, 59(3), 162–165. <https://doi.org/10.1177/00045632211050531>.
- Winter, B. (2019). *Statistics for linguists: An introduction using R*. Routledge.
- Zuercher, B. (2019). Why does polysemy vary across languages? An explanation in the framework of the sign theory of language. *Canadian Journal of Linguistics*, 64(2), 281–325.
- Zupan, S. (2006). Repetition and translation shifts. *ELOPE: English Language Overseas Perspectives and Enquiries*, 3(1–2), 257–268. <https://doi.org/10.4312/elope.3.1-2.257-268>.

---

**Open Access statement.** This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium for non-commercial purposes, provided the original author and source are credited, a link to the CC License is provided, and changes – if any – are indicated.

