

# Between repetition and lexical variety: How do English-to-Russian translators of literary texts deal with recurrent reporting verbs?

**Łukasz Grabowski**  
University of Opole

**Daniel Borysowski**  
University of Opole

## Abstract

In this exploratory study, we seek to identify the predictors of repetition or lexical variety in the translation of English reporting verbs into Russian. Using a sample of 20 literary novels from the InterCorp corpus (v.15), we fitted multiple negative binomial regression models with a random intercept. The goal was to assess how selected predictor variables—namely, the frequency of a source-text verb, its number of senses, semantic type, length in characters, date of translation, and translator—affect the response variable: the number of Russian target-text reporting verbs an English source-text (ST) reporting verb is translated into. The findings showed that the semantic category of a ST reporting verb, its frequency and translation date as well as the translator as a random intercept have the largest individual contributions to explaining the proportion of variation in the response variable. More precisely, the model allows us to explain 73% of the variation (per conditional r-squared) in the number of distinct target text (TT) reporting verb types a ST verb is translated into. Viewed in the context of prevailing stylistic norms in Russian, the findings offer an attempt at explanation for the translator’s choices in rendering recurring reporting verbs following dialogues, which play an important stylistic effect in literary texts.

**Keywords:** multifactorial methods, corpus linguistics, parallel corpus, translation, reporting verbs

## 1 Introduction

This paper is positioned at the interface of corpus stylistics, literary translation and multifactorial statistics. As defined by Mahlberg (2013, 2018), corpus stylistics involves applying quantitative and qualitative methods from corpus linguistics to enhance the linguistic description of literary works, particularly concerning their literary appreciation or rhetorical effects on readers. In short, corpus stylistic methods—such as wordlist analysis, keyword analysis, and concordance analysis—offer insights into both frequent and rare linguistic features found in literary texts. They thus provide a rigorous and complementary approach to exploring literary style, where style is understood as “a property of texts constituted by an ensemble of formal features which can be observed quantitatively or qualitatively” (Hermann et al. 2015: 44). Mahlberg (2013: 9) also claims that corpus linguistic methods allow to identify prominent linguistic features that are

used in texts with unusual frequency or the ones that simply “stand out” in comparison with some norms, for example a general language use. Mahlberg (2013: 9) refers to this phenomenon as “foregrounding” and distinguishes between two main types: deviations (the untypical use of certain linguistic features) and repetitions (the overly frequent use of certain features). These patterns may influence the reception or rhetorical effects of translated texts, as shown in Mastropiero’s (2019) corpus stylistic comparison of Joseph Conrad’s *Heart of Darkness* and its four Italian translations. In this study, as explained later in the paper, we will deal with repetition in literary texts.

Corpus linguistic methods have been also popular in descriptive research on translation, notably on translator’s style, translation style (Saldanha 2011) or translation universals (Baker 1993, 2000, 2004; Chesterman 2004; Halverson 2003; Laviosa 1998, 2001; Mauranen 2008), often referred to as corpus-based descriptive translation studies (e.g. Calzada Pérez & Laviosa 2021). According to the premises of the cognitive model of translational activity (Halverson 2003, 2010, 2017), which provides conceptual clarification of the notion of translation universals in the spirit of cognitive linguistics, translational choices are governed by three forces: gravitational pull, magnetism and connectivity (Olejniczak 2023). In short, gravitational pull posits that the source language significantly shapes translation output, inducing translators to favour target language equivalents that are highly frequent or salient in the source text. This can lead to the overproduction of specific linguistic features (cf. Mauranen 2008). In contrast, magnetism describes a scenario where translator’s choices are predominantly informed by target language norms, often resulting in stylistic standardization (cf. Toury 1995), normalization (cf. Baker 1996), or so-called “domesticated” translation (Shuttleworth & Cowie 1997; Venuti 1995). The interplay of dominant stylistic norms and conventions in both source and target languages appears pivotal in these cases. Lastly, connectivity denotes a phenomenon specific to a given language pair (e.g. English-Polish, cf. Olejniczak 2023), whereby certain source language items or constructions possess a more or less established, preconceived, or typical equivalents in the target language. In this study, we will make ample use of a parallel corpus, i.e. a collection of English source texts and their Russian translations, and multifactorial statistics to explore potential factors, both source-language and target-language related, that impact the ways the translators handle repetition in source texts, taking reporting verbs signalling utterances as a case in point.

## 2 Repetition and reporting verbs in translation from or into Russian

In studies on the Russian language, repetition is typically examined either within the literary works of specific authors (e.g., Černyševa 2014; Petrunina 2012) or across different text genres (Šutěmova 2010; Žilina-Èls 2020). Generally, it is regarded as a discourse-organizing or stylistic device and can manifest at various levels of textual organization, from phonetic to semantic (Moskvin 2000; Žilina-Èls 2020). A particularly salient form of repetition involves the recurrent use of reporting verbs. These have been studied in relation to their functional classifications (cf. Arsahanova & Zekieva 2024; Ermolaeva 2017), as well as within the context of translation both into and out of Russian (Filatova 2023; Urža 2018).

Recent studies (e.g., Filatova 2023; Urža 2018) indicate that repetition of reporting verbs tends to be reduced in translation. Translators often employ a broader range of target-language reporting verbs to render a repeatedly used source-language verb (e.g.

*said*), thereby introducing greater lexical variation in the translated text. Arsahanova and Zekieva (2024) found that translations of the English verb *say* into Russian vary depending on the content of the utterance, its form, and particular speech act. They also observe that English texts tend to rely more heavily on neutral reporting verbs with broad meaning (such as *say* or *tell*), whereas Russian texts more frequently employ reporting verbs with narrower, more specific meanings (e.g. заметить *zametit'* 'to point out'), particularly those with emotional or expressive colouring (e.g. воскликнуть *voskliknut'* 'to exclaim'). This tendency aligns with observations made by Urža (2018), who investigated the translation of reporting verbs in children's literature. Urža (2018) and Filatova (2023) also observe that the translation of reporting verbs is closely linked to the choice of translation strategy. When translators adopt a domesticating approach, they tend to avoid repetition, aligning with the norms of Russian stylistic tradition. In contrast, a foreignizing strategy preserves the stylistic features of the source text, including limited variation in reporting verbs, thereby maintaining the repetition present in the original.

All this offers an opportunity to investigate the factors that drive translators to avoid repetition or—if the opposite is the case—to preserve it in translated texts (i.e. by repeatedly using one or a small number of target-language reporting verbs in translation). In this context, it is also essential to take into consideration varying norms and stylistic conventions across languages and cultures with respect to tolerance of repetition. For example, in some languages, repetition is generally viewed as poor, monotonous or unsophisticated style (e.g. in Polish or Russian), but in others (e.g. in English), it may be viewed favourably as contributing to consistency, coherence and precision of narration (cf. Filatova 2023; Piotrowski 1994; Salciute-Civiliene 2016). Anecdotal evidence suggests that in Russian (and other Slavic languages) stylistic conventions dictating the avoidance of repetition predominate, notably in the pedagogical landscape at schools.

This study attempts to uncover the predictors of repetition or lexical diversity in the translation of reporting verbs—following dialogues—from English into Russian. Reporting verbs, which are lexical verbs primarily serving a referential function, namely attributing an utterance to its speaker, perform important stylistic functions in literary texts, ranging from contribution to characterization of literary personae (Ruano San Segundo 2016, 2017), to character development (Mastropierro 2020), to coherence and expressiveness of character's utterances (Witoszowa 1985; Kowalska 2020), to sound symbolism as well as to supporting the writer's or poet's tone (Abdulla 2001), among others. For this reason, authors of literary novels use a wide array of reporting verbs following dialogues (e.g., *said, muttered, murmured, whispered, sighed, gasped*), which—due to their high frequency in such texts—are particularly well-suited to wide-range statistical analyses. By selecting reporting verbs as the unit of analysis, we focus on a functionally significant type of repetition in literary novels. The findings will complement earlier preliminary research (Mastropierro & Grabowski 2024; Grabowski et al. 2026), where—to some extent—similar methods were used, on the English-to-Italian and English-to-Polish translation.

Hence, in this small-scale study, we aim to answer the following research questions: (i) What linguistic factors have a significant effect on the avoidance or reproduction of reporting verbs' repetition in the Russian novels? (ii) What is the proportion of variance explained by fixed and/or random effects? The methodology that we used in the study to address these questions is described in the following section.

### 3 Methods

Examining recurrent language patterns is a key aspect of corpus linguistic research, including translation studies. Apart from traditional techniques, such as word frequency list analysis, keyword analysis, concordance analysis etc. (cf. Buts & Jones 2021), in many corpus studies conducted nowadays researchers also employ machine learning techniques popular in the area of natural language processing (NLP), also known as computational linguistics (Hirschberg & Manning 2015), as well as data science, which has become, in a sense, a separate discipline devoted to data analysis (cf. Grabowski 2023). We can also observe the rise in the use of multifactorial statistical methods in research on translation and interpreting (De Sutter & Lefer 2019; Kajzer-Wietrzny & Grabowski 2021; Levshina 2021; De Baets & De Sutter 2022; Chmiel et al. 2024; Wang & Xin 2024), which enable translation scholars to test not only descriptive or interpretative hypotheses, but also explanatory or predictive ones (cf. Chesterman 2011). For example, without using appropriate regression methods suited to the type of a dependent variable, it would not have been possible to accurately test hypotheses explaining the presence of specific linguistic features in translation (Gries & Wulff 2012; Levshina 2015). As translation is a complex and multifaceted phenomenon, conditioned by external (e.g., language pair, function of translation, tools) and internal factors, including linguistic ones (e.g., translator's agency, habits, idiolect; cf. Chesterman 2011: 77), putting forward explanatory hypotheses may help us better understand translatorial decisions and hence language use in translation. A key element in such research lies in selecting well-founded predictors, informed by prior findings and cumulative knowledge, that play a crucial role in explaining the presence of certain linguistic features in translated texts.

This small-scale multifactorial study aims to identify the linguistic predictors that impact lexical variety or repetition in the English-to-Russian translation of reporting verbs in literary texts. Using custom-designed bilingual CQL queries, we extracted pairs of aligned reporting verbs from 20 novels (*Harry Potter and the Chamber of Secrets*, *Harry Potter and the Half-Blood Prince*, *Harry Potter and the Order of the Phoenix*, *Harry Potter and the Philosopher's Stone*, *Harry Potter and the Prisoner of Azkaban*, *The Da Vinci Code*, *The Fellowship of the Ring*, *The Grapes Of Wrath*, *The Hitch Hiker's Guide to the Galaxy*, *The Return of the King*, *The Two Towers*, *1984*, *Lolita*, *Lord of the Flies*, *Girl with a Pearl Earring*, *The Partner*, *The Hobbit or There and Back Again*, *The Adventures of Tom Sawyer*, *The War of the Worlds*, *Three Men in a Boat*). This selection was based primarily on the availability of texts translated from English into Russian, found in the InterCorp corpus version 15 (Rosen et al. 2022), which is a large annotated<sup>1</sup> multilingual parallel corpus (Čermák & Rosen 2012; Čermák 2019).<sup>2</sup> The texts represent various literary genres, such as fantasy novels, realist novels, science fiction novels, erotic novels, detective fiction and comedy novels, as for multifactorial statistics the priority was to collect as many texts with as many reporting verbs following dialogues as possible.

---

<sup>1</sup> For tagsets used in InterCorp (v. 15) see:

[https://wiki.korpus.cz/doku.php/en:cnk:intercorp:verze15#morphosyntactic\\_annotation](https://wiki.korpus.cz/doku.php/en:cnk:intercorp:verze15#morphosyntactic_annotation)

<sup>2</sup> More details regarding the novels, with comprehensive descriptive statistics and metasituational information (see the files: *RU\_metadata.csv* and *RU\_1.csv – RU\_20.csv*), as well as CQL queries are specified in the data repository at: <https://osf.io/357g6/>

More precisely, we searched for the following patterns in the source texts (henceforth ‘ST’): “closing quotation marks + *he* or *she* + past tense verb” and “closing quotation marks + past tense verb + *he* or *she*”, and these patterns were matched by equivalent patterns in Russian, taking into consideration specific orthographic conventions of recording dialogues in each language. Thus, the English and Russian reporting verbs were retrieved with a parallel query that combined regular expressions and CQL syntax, using the Czech National Corpus interface tool, KonText. From the obtained parallel concordances in each novel, we retrieved the translation of each ST verb into a Russian reporting verb, where a ST verb had a minimum frequency of 2, as we were interested in exploring how translators deal with repeatedly used English reporting verbs. More precisely, we retrieved 14,044 verb tokens and 110 verb types in the original novels, as well as all of their translations in Russian as reporting verbs (633 unlemmatized and 480 lemmatized types, the latter of which were used in the study. To ensure the compatibility of the English and Russian data, the Russian 3rd person past tense feminine and masculine forms were lemmatized to their base forms. The lemmatization was conducted using `lemmagen3`<sup>3</sup>, a Python wrapper developed by Podpečan (2024) for Lemmagen lemmatizer (ver. 2.2), which supports 19 languages, including Russian (Juršič et al. 2010).

We annotated reporting verbs from 20 novels based on seven linguistic features, identifying potential predictors for their reproduction or avoidance of repetition. We began with recording the frequency of each ST reporting verb type in the studied literary novels (e.g. *said* was used 815 times in the novel *Return of the King* while *exclaimed* was used twice). Due to such considerable differences, we transformed the frequencies into a logarithmic scale and coded the variable as “logfreq”.

The next potential predictor, coded as “senses”, refers to the number of senses of each ST reporting verb as recorded in the semantic-relational lexical database Princeton WordNet® 3.1 (Fellbaum 1998), in the semantic domain of communication. This decision was guided by the fact that we focus on reporting verbs following dialogues. For example, the verb *say* has 11 distinct senses in the Princeton WordNet, and 10 of them belong to the domain of communication; in the case of the verb *add*, it has 6 distinct senses, and only 2 of them belong to the domain of communication.

The third variable indicates semantic-functional category of the reporting verb based on Caldas-Coulthard’s (1987) categorisation, which distinguished between neutral verbs (e.g., *say*, *tell*), structuring verbs (e.g., *ask*, *reply*), metapropositional (e.g., *exclaim*, *instruct*, *swear*), metalinguistic (e.g., *narrate*, *quote*), prosodic (e.g., *cry*, *shout*), verbs signalling discourse (e.g., *repeat*, *add*) and paralinguistic verbs, which include voice qualifiers (e.g., *whisper*, *murmur*) and voice quantification (e.g., *laugh*, *groan*). Hence, the nominal variable “semverbtype” has multiple levels, which were coded as “N”, “Str”, “Mprop”, “Mlin”, “Pros”, “Sdis”, “Vier” and “Vion” respectively. Given the insights from earlier investigations (e.g., Filatova 2023; Urža 2018), we anticipate that neutral reporting verbs will yield a diverse repertoire of target-language equivalents.

As 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 in source texts), the fourth potential predictor is length—per number of characters—of the ST reporting verb. The fifth one, namely “translation\_time”, corresponds to whether

<sup>3</sup> <https://pypi.org/project/lemmagen3/>

the Russian translation was completed either before the year 1990 (“pre1990”) or in 1990 and later (“post1990”), a timeline that marks the beginning of social and political developments leading to the collapse of communism in the USSR.

The last factor, which is “translator\_id”, indicates the translator of each of the 20 novels coded by acronyms. For example, “YS” stands for Elena Suric (Елена А. Суриц), who translated the novel *Lord of the Flies* (Повелитель мух *Povelitel' muh*) in 1981 into Russian. This information was retrieved from metadata files available in the InterCorp corpus. Given our assumption that individual translators’ choices influenced the translation of reporting verbs—and considering that our study is based on a sample of 20 literary novels drawn from InterCorp, a subset of a theoretically unlimited pool of texts—we included “translator\_id” as a random intercept in our regression model. According to the metadata found in InterCorp, the 20 novels under scrutiny were translated by 18 different translators. More precisely, Marina Litvinova translated 2 novels, namely *Harry Potter and the Chamber of Secrets* and *Harry Potter and the Prisoner of Azkaban*; 3 novels were produced by teams of 2 or 3 translators; the author of one translation (*Harry Potter and the Philosopher’s Stone*) is unknown and hence marked as NA<sup>4</sup>. This way we met one of the assumptions for using mixed regression models, where random effects should have at least 5 or 6 levels (Bentz & Winter 2013).

Thus, the predictors used in the study include three numerical variables, “logfreq”, “senses” and “wordlength”, and four nominal variables, including “semverbtype” (with seven levels), “translation\_time” (with two levels) and “translator\_id” (18 IDs) as a random intercept. The dependent variable, i.e. the outcome or response, is called “types”. More precisely, it is a count variable representing the number of Russian target text (henceforth ‘TT’) reporting verb types (lemmas) used as translation equivalents of English ST reporting verbs. For instance, a value of “types” of 4 indicates that the ST reporting verb (e.g. *shouted*, a prosodic verb according to the Caldas-Coulthard (1987) typology) was translated into 4 different verbs in Russian (закричал *zakričal* ‘started shouting’, воскликнул *voskliknul* ‘exclaimed’, кричал *kričal* ‘was shouting’, крикнул *kriknul* ‘shouted’) in the novel *The Fellowship of the Ring* in the position following dialogues (i.e. attributing an utterance to the utterer). Thus, the value 1 shows that the verb repetition in the original (ST) was preserved in translation (TT), while the value greater than 1 indicates lexical variety: the higher the value, the wider the range of different reporting verbs used in the TT as equivalents of a ST reporting verb. Data preparation was carried out using custom-designed Python scripts. The final dataset was saved as 20 comma-separated value (CSV) files, each corresponding to one novel, and deposited in an open-access public repository to ensure the study’s replicability and reproducibility.

Given our objective—to identify statistically significant predictors of the number of Russian translation equivalents (i.e. types of reporting verbs) for English source-text reporting verbs, as well as to assess the extent to which these predictors explain variance—the dependent variable “types” is treated as count data measured in non-negative integers. Accordingly, and in line with recommendations in the relevant literature (Scherber 2017, 2019; Winter 2019: 247; Kabacoff 2015: 312; Zeileis et al. 2008), we initially opted for Poisson regression. This form of generalized linear model (GLM) is commonly employed for modelling count data and contingency tables, particularly when predictors include

<sup>4</sup> See RU\_metadata.csv file in the repository for detailed information.

a mix of continuous/numerical and categorical/nominal variables, which is the case in our study. However, as we observed overdispersion in the data, that is, the variance for “types” was found to be considerably higher than the mean (mean of 5.26 and variance of 202.02, cf. Bentz & Winter 2013), we used negative binomial regression instead of Poisson regression. This adjustment follows recommendations by Winter (2019), Scherber (2017, 2019), and Hair et al. (2009). Using Poisson regression in such a context would risk underestimating standard errors and confidence intervals, potentially leading to spurious findings of statistical significance (Kabacoff 2015: 315). This was later confirmed using the Likelihood Ratio Test (see Results section). In order to avoid the problem of zero-inflation, we also checked that our dependent variable does not have an excessive number of zeros (in fact, none). As a rule, the best fitting model is the one with the lowest AIC (Akaike Information Criterion) and BIC (Bayesian information criterion), that is, the one that reaches significance with as few variables as possible through their “backward selection” (Winter 2019: 310). This is the type of stepwise regression: we start with a full model with all potential predictor variables and iteratively remove those that are not significant statistically, i.e. that have p-values lower than 0.05 (ibid.). The lowest AIC shows that a given model fits the data best, so it is a measure of the quality of the model (cf. Brezina 2018: 124–125), while BIC penalizes overly complex models (with too many variables etc.). In other words, BIC measures the simplicity of the model (Brezina 2018; Winter 2019).

The analyses were conducted in the R environment using the following packages: *car* (Fox & Weisberg 2019), *MASS* (Venables & Ripley 2002), *MuMIn* (Bartoń 2024), *glmmTMB* (Brooks et al. 2017) and *performance* (Lüdtke et al. 2021).

## 4 Results

To evaluate the influence of six predictor variables—along with a random intercept for “translator\_id”—on the number of distinct Russian reporting verbs used as translation equivalents (response variable “types”), we fitted a series of negative binomial regression models. Model selection was carried out using a backward elimination procedure (Winter 2019: 310), whereby predictors with p-values exceeding 0.05 were iteratively removed to retain only statistically significant variables. The full dataset, comprising all 20 novels, contained a total of 462 observations. As mentioned earlier, in order to verify the appropriateness of using a negative binomial regression model—which accounts for an additional dispersion parameter—we fitted both Poisson and negative binomial models including all predictors. A Likelihood Ratio Test (LRT)<sup>5</sup> confirmed that the negative binomial model provided a significantly better fit to the data ( $p < 0.05$ ) (Figure 1).

---

<sup>5</sup> We run `anova()` function with two nested models in R. LRT is described, for example in Lewis et al. (2011), and it can only be used in cases where one model is a special case of another (then we deal with so-called nested models), which is the case in this study (negative binomial regression has an additional dispersion parameter).

```

Models:
glm1p: types ~ logfreq + semverbtype + senses + wordlength + translation_time + , zi=~0, disp=~1
glm1p:      (1 | translator_id), zi=~0, disp=~1
glm1nb: types ~ logfreq + semverbtype + senses + wordlength + translation_time + , zi=~0, disp=~1
glm1nb:      (1 | translator_id), zi=~0, disp=~1
      df    AIC    BIC logLik deviance Chisq chi Df Pr(>Chisq)
glm1p  13 1751.4 1805.1 -862.68  1725.4
glm1nb 14 1708.6 1766.5 -840.30  1680.6 44.765      1 2.221e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

**Figure 1.** Likelihood Ratio Test with nested models: results

The inclusion of a random intercept for “translator\_id” was further supported by the results of a Likelihood Ratio Test comparing negative binomial regression models with and without the random intercept. The model including the random intercept demonstrated better fit, as indicated by lower AIC (1708.6 vs. 1743.4) and BIC (1766.5 vs. 1797.2) values. Also, we addressed the potential problem of collinearity, when the predictors are highly correlated so that “one predictor can be predicted by other predictors” (Winter 2019: 112). To this end, we used performance R package (Lüdtke et al. 2021) and calculated variance inflation factors (VIF) for all the predictors in the base model, where the values larger than 10 are commonly accepted as indicating serious collinearity issues (e.g. Montgomery & Peck 1992, cited in Winter 2019: 114). The findings revealed that our predictors show low (VIF<5) to moderate collinearity (5<VIF<10), which is not worrisome for our study.<sup>6</sup>

Our experiments—conducted via backward selection—revealed that an optimal model with the lowest AIC value (1706.5) and BIC value (1756.1) was the following: *types ~ logfreq + semverbtype + translation\_time + (1|translator\_id)* (Figure 2). It shows that the semantic category of a ST reporting verb, its frequency and translation date as well as the translator as a random intercept have the largest individual contributions to explaining the proportion of variation in the response variable “types” in the Russian translations. More precisely, the model allows us to explain almost 73% of the variation (per conditional r-squared) in the response variable, that is, the number of different verb types a ST verb is translated into. This was computed with the help of *r.squaredGLLM* function in R using a delta method (Nakagawa et al. 2017; Bartoń 2024). Without taking into consideration “translator\_id” treated as a random intercept, we would have explained only 66% of variation. Although the variable “translation\_time (pre1990)” yields a p-value of 0.09—exceeding the conventional significance threshold of 0.05 but falling below 0.1, which indicates weak evidence of statistical significance (cf. Ganesh & Cave 2018)—we opted to retain it in the model. The reason for this was that its elimination resulted in the increase of AIC value to 1707.1, which indicates that the model has lower quality. For the same reason, we did not remove the variable “semverbtype”, although most of its levels were found not to be statistically significant.

<sup>6</sup> The obtained VIF values are as follows: logfreq (4.12), senses (3.02), wordlength (2.74), translation\_time (1.02) and semverbtype (8.27).

```

Family: nbinom2 ( log )
Formula:
types ~ logfreq + semverbtype + translation_time + (1 | translator_id)
Data: data

      AIC      BIC   logLik -2*log(L)  df.resid
1706.5   1756.1   -841.3   1682.5     450

Random effects:

Conditional model:
  Groups      Name      Variance Std.Dev.
translator_id (Intercept) 0.04875  0.2208
Number of obs: 462, groups: translator_id, 18

Dispersion parameter for nbinom2 family (): 26.4

Conditional model:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)      0.08516   0.10854   0.785   0.4327
logfreq          0.60896   0.02574  23.662 < 2e-16 ***
semverbtypeN     0.18059   0.13025   1.386   0.1656
semverbtypePros  0.09012   0.09437   0.955   0.3396
semverbtypesdis -0.16901   0.10166  -1.663   0.0964 .
semverbtypeStr  -0.77122   0.15632  -4.934 8.07e-07 ***
semverbtypevier -0.05950   0.10141  -0.587   0.5574
semverbtypevion  0.07350   0.08679   0.847   0.3971
semverbtypeMlin  0.01044   0.74341   0.014   0.9888
translation_timepre1990 -0.21062   0.12643  -1.666   0.0957 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

**Figure 2.** Negative binomial regression: final model summary

When we look at the predictors' individual contributions to explaining variance in the response variable "types", a summary<sup>7</sup> (at the bottom of Figure 2) reveals that the frequency of ST verb impacts its chances of being translated into multiple Russian TT reporting verbs. For example, with a one-unit change (increase) in "logfreq", the log of the expected number of translation equivalents increases by 0.609, while holding the rest of the predictor variables constant. For this count, this corresponds to a multiplicative increase of  $\exp(0.609) \approx 1.84$  times (or 84% increase), which means that higher frequency has a strong positive effect on the count of "types". This finding confirms our intuition: if a ST reporting verb is frequently used, then translators notice its repetition, and in order to avoid reproducing this repetition in translation they resort to using a wide array of TT reporting verbs.<sup>8</sup> This finding also aligns with prevailing stylistic conventions, according to which repetition and limited lexical variation are often associated with an impoverished or less sophisticated style. Also, it is possible that the larger number of translation equivalents for higher frequency verbs might be explained by larger numbers of alternative equivalents for more common words. This claim, however, should be verified further in the future using data extracted from bilingual dictionaries or lexical databases of translation equivalents.

<sup>7</sup> In bold (Figure 2), we present p-values and estimates (positive or negative) that indicate the direction and effect size of each predictor (including all levels of variables).

<sup>8</sup> For example, the verb *said* used 473 times in the English-original novel *Harry Potter and the Chamber of Secrets* has as many as 140 distinct Russian reporting verb equivalents, the most frequent ones being *сказал skazal* 'said', *спросил sprosil* 'asked', *ответил otvetil* 'answered'; the prosodic verb *shouted*, used 17 times in the same English novel, has 10 distinct Russian equivalents, the most frequent ones being *крикнул kriknul* 'shouted-PERF', *кричал kričal* 'shouted-IMPERF', *воскликнул voskliknul* 'shouted-PERF' (see the RU\_1 file in the repository).

In contrast, when we look at negative estimates of statistically significant predictors, we can see that structuring verbs (e.g., *ask*, *reply*) and—to a lesser degree—discourse signalling verbs (e.g., *repeat*, *add*) are more likely to be consistently rendered by a narrow range of the same Russian equivalents, which means more repetition in translation. For example, in the novel *Girl with a Pearl Earring* (1999) by Tracy Chevalier, the structuring verb *asked* is used 10 times and it has two lemmatized Russian equivalents in the translation by R. Bobrova, namely *спросил* *sprosil* ‘asked’ (used 9 times, cf. Examples (2–5)) and *попросил* *poprosil* ‘requested’ (used once, cf. Example (1)). When employed in the sense of ‘asked’, the verb is invariably translated as *спросил* *sprosil* ‘asked’, as demonstrated in Examples (2–5).

- (1) ‘Will you help me?’ she **asked** me as she tried to climb on to a barrel the neighbours had left out in the street.  
 «Помоги мне», — **попросила** она, пытаясь вскарабкаться на бочку, которую соседи оставили на улице.  
 «Pomogi мне», — **poprosila** ona, pytaâs' vskarabkat'sâ na bočku, kotoruiû sosedi ostavili na ulice.
- (2) ‘How is it?’ she **asked**, not even saying hello.  
 — Ну как тебе там? — **спросила** она, позабыв поздороваться.  
 — Nu kak tebe tam? — **sprosil** ona, pozabuv pozdorovat'sâ.
- (3) ‘May I come with you?’ she **asked**, smiling at me with well-practised innocence.  
 — Можно мне с тобой? — **спросила** она со своей обычной невинной улыбкой.  
 — Možno mne s toboj? — **sprosil** ona so svoej obyčnoj nevinnoj ulybkoi.
- (4) ‘What were you wearing up there?’ he **asked**.  
 — Что у тебя было на голове? — **спросил**.  
 — Čto u tebâ bylo na golove? — **sprosil**.
- (5) ‘And what is it that he wants today?’ he **asked**.  
 — Ну и что ему сегодня нужно? — **спросил** он.  
 — Nu i čto emu segodnâ nužno? — **sprosil** on.

In a similar vein, the structuring verb *replied*, with 17 occurrences in the novel *Da Vinci Code* by Dan Brown (2003) has only 4 Russian lemmatized equivalents in the translation by N. Rein, namely *ответил* *otvetil* ‘replied’ (12 occurrences), *сказал* *skazal* ‘said’ (used 3 times), *заметил* *zаметil* ‘noticed’ (used only once), *поправил* *popravitil* ‘corrected’ (used only once). In the overwhelming majority of cases, the verb *reply* is translated as *ответил* *otvetil* ‘replied’, consistently attributing the utterance to various characters within the novel, as in Examples (6–10).

- (6) “Mr. Saunière did,” Langdon **replied** as they entered the tunnel.  
 — Вообще-то мистер Соньер, — ответил Лэнгдон, входя в туннель.  
 — Voobše-to mister Son'er, — **otvetil** Lèngdon, vhotâ v tunnel'.

- (7) "Sophie Neveu," *Fache replied*.  
 — Софи Невё, — **ответил** Фаш.  
 — *Sofi Nevě, — otvetil Faš.*
- (8) "My grandfather gave it to me," *Sophie replied, watching the man closely*.  
 — Мой дед передал его мне, — **ответила** Софи, не сводя с банкира пристального взгляда.  
 — *Moj ded preдал ego mne, — otvetila Sofi, ne svodâ s bankira pristol'nogo vzglâda.*
- (9) "That, my dear," *Teabing replied,* "is Mary Magdalene."  
 — Она, моя дорогая, — **ответил** Тибинг, — не кто иная, как Мария Магдалина.  
 — *Ona, moâ dorogaâ, — otvetil Tibing, — ne kto inaâ, kak Mariâ Magdalina.*
- (10) "Let Sophie and Leigh leave the church," *Langdon replied*.  
 — Прежде отпусти Софи. Пусть они с Тибингом выйдут из церкви,  
 — **ответил** Лэнгдон.  
 — *Prežde otpusti Sofi. Pust' oni s Tibingom vyjdut iz cerkvi, — otvetil Lèngdon.*

Such a limited set of translation equivalents suggests a substantial degree of preservation of repetition patterns for structuring verbs from the original novels in their Russian translations, which is also supported by our regression analysis. We hypothesize that the more specific meanings of structuring verbs—which correspond to particular speech acts (e.g., asking a question, requesting something, answering someone)—necessitate more precise and well-established equivalents in Russian (e.g., спросить *sprosit'* 'ask', ответить *otvetit'* 'reply'). This functional clarity may constrain the translator's lexical choices, encouraging consistency rather than variation, especially when the goal is to preserve the speech act's illocutionary force. A second possible interpretation relates to textual cohesion: maintaining a limited set of equivalents could reflect the translator's strategic choice to prioritize transparency and clarity over stylistic embellishment. Overall, we may hypothesize that preservation of ST repetitions in translation may be justified when such repetitions are pragmatically motivated in TT. We also found that this trend correlates with translations published before 1990, although the evidence for this is rather weak, as we mentioned earlier ( $p$ -value < 0.1).

Finally, we reported a low level of variance (0.048) and standard deviation (0.22) for the entire model, the values that indicate the spread of random effect. Although the 20 novels used in the study were translated by 18 different translators or translation teams, the tendencies that the model shows are independent from individual translator's habits and choices with respect to rendition of reporting verbs. In other words, there is some variability between the translators, but it is rather small (unlike in the case when variance is in the region of 0.2-0.3). This means that no single translator significantly influenced the observed patterns, and the effect was spread across 18 translators.

## 5 Conclusions

Our findings indicate that Russian translators' decisions—to either reproduce repetition patterns using a limited set of reporting verbs or avoid them by employing a wide variety of verbs—are influenced by the frequency of use and specific semantic categories (here structuring verbs) of source text (ST) reporting verbs.<sup>9</sup> This was observed across 20 translations of English-original literary novels under scrutiny. Based on earlier research involving another Slavic language, that is, in the English-Polish language pair (Mastropierro & Grabowski 2024), and observations made in studies by Filatova (2023) and Urža (2018), we expected that neutral reporting verbs (e.g., *said*, *told*), which have broad meaning in English, would be primed for being translated with the whole variety of Russian equivalents. However, our regression analysis did not identify them as a statistically significant predictor, thus not supporting our initial hypothesis.

The obtained findings have a number of implications for translation research and translation training. First, our study showed that both fixed and random effects should be considered when attempting to identify the predictors of translatorial decisions that manifest themselves into either lexical variety or preservation of source-text repetitions in target texts. Second, the predictors of lexical variety or repetition should include both source and target language related predictors as well as sociolinguistic predictors, such as the date of translation and the identity of the translators. This might be of utmost importance in more comprehensive future studies encompassing various text types, genres and language pairs. Finally, the knowledge of what prompts avoidance of repetition or its preservation in translations may be useful in translator training. For example, the findings may inform data-driven translator training by challenging oversimplified assumptions such as the notion that repetition is inherently undesirable or is a marker of poor style. When accompanied by illustrative examples from bilingual concordances, the findings can also support translation students in critically evaluating their strategies or techniques used to handle repetitions, in particular when such repetitions are semantically or stylistically motivated in literary texts. This may help make future translators aware of translation strategies or techniques addressing repeated linguistic patterns in source texts in general. This critical awareness may be important in the future given the expanding integration of technology in translation, including neural machine translation (NMT), NMT post-editing, AI-assisted post-editing or AI-assisted translation in general. The extent to which these technologies would contribute—in the long term—to either increased homogeneity (i.e. more repetitions) of translated texts or the propagation of pre-existing patterns from training data remains currently unknown, though their expected future impact on translation style is rather undeniable.

This study has a number of limitations that need to be addressed in the future. As the findings suggest that the ways translators deal with repetition are influenced by the features of repeated items in the original, more predictors could be taken into consideration, e.g., gender of the referent, the number of translation equivalents in lexical databases or dictionaries as well as the span of repetition. In this study, we used a formal equivalence approach as we focused on translating reporting verbs as reporting verbs, but it may also happen that ST reporting verbs are rendered with other stylistic means (nominalizations,

---

<sup>9</sup> Importantly, similar results were obtained when individual texts (*text\_id*) were included as a random intercept in the model.

omissions etc.), which also needs to be addressed (although our provisional inspection of bilingual concordances using only English CQL queries revealed that other strategies are marginal). Furthermore, comparisons across text types (e.g. press articles) and languages (i.e. other language pairs) as well as selection of units of analysis other than reporting verbs could provide a more comprehensive picture of what dictates translatorial decisions with respect to repetition. Although our study was limited to English-to-Russian translations of specific literary texts, it is possible that the findings were additionally impacted by the broad spectrum of literary genres present in the selected parallel texts, including fantasy, detective fiction, erotic novels, and science fiction. It seems plausible that focusing on a collection of texts representing a single genre could yield different results and offer more fine-grained insights into the predictors of lexical variety or repetition as compared to other genres or text types.

We hope that this small-scale descriptive and exploratory study will inspire further explanatory research into translators' attitudes toward source texts, and more specifically, into the factors that predict translatorial choices shaping translator style or translation style across different text types, genres, and language pairs.

## Acknowledgements

This research was funded by the National Science Centre (NCN), Poland, grant number: 2023/51/B/HS2/00697. We would like to thank Lorenzo Mastropiero for his profound inspiration and encouragement to conduct this study in the English-to-Russian language pair. Finally, we thank anonymous reviewers and editors of FJL for their time and constructive comments, which substantially improved the earlier version of this paper.

## References

- Abdulla, Adnan K. 2001. Rhetorical Repetition in Literary Translation. *Babel* 47(4). 289–303. <https://doi.org/10.1075/babel.47.4.02abd>
- Arsahanova, Malûthan & Zekieva, Petimat. 2024. Leksiko-semantičeskie gruppy glagolov reči v russkom i anglijskom âzykah [Lexical-semantic groups of verbs of speech in Russian and English languages]. *Mir nauki, kul'tury, obrazovaniâ* 3(106). 468–471.
- Baker, Mona. 1993. Corpus Linguistic and Translation Studies: Implications and applications. In Baker, Mona & Francis, Gill & Tognini-Bonelli, Elena (eds.), *Text and Technology: In Honour of John Sinclair*, 233–250. Amsterdam: John Benjamins.
- Baker, Mona. 1996. Corpus-based Translation Studies: The Challenges That Lie Ahead. In Somers, Harold (ed.), *Terminology, LSP and Translation: Studies in Language Engineering. In Honour of Juan C. Sager*, 175–186. Amsterdam/Philadelphia: John Benjamins.
- Baker, Mona. 2000. Towards a Methodology for Investigating the Style of a Literary Translator. *Target* 12(2). 241–266.
- Baker, Mona. 2004. A corpus-based view of similarity and difference in translation. *International Journal of Corpus Linguistics* 9(2). 167–193.
- Bartoń, Kamil. 2024. *MuMIn: Multi-Model Inference. R package version 1.48.4*. <https://CRAN.R-project.org/package=MuMIn> (Accessed: 2025-01-10)
- Bentz, Christian & Winter, Bodo. 2013. Languages with More Second Language Learners Tend to Lose Nominal Case. In Wichmann, Soren & Good, Jeff (eds.), *Quantifying Language Dynamics: On the Cutting edge of Areal and Phylogenetic Linguistics*, 96–124. Leiden: Brill.
- Brezina, Vaclav. 2018. *Statistics for Corpus Linguistics*. Cambridge: Cambridge University Press.
- Brooks Mollie E. & Kristensen, Kasper & van Benthem, Koen J. & Magnusson, Arni & Berg, Casper & Nielsen, Anders & Skaug, Hans J. & Maechler, Martin & Bolker, Benjamin. 2017. glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. *The R Journal* 9(2). 378–400.

- Buts, Jan & Jones, Henry. 2021. From text to data: Mediality in corpus-based translation studies. *MonTI. Monografías de Traducción e Interpretación* 13, 301–329. <https://doi.org/10.6035/MonTI.2021.13.10>
- Caldas-Coulthard, Carmen R. 1987. Reported speech in written narrative texts. In Coulthard, Malcolm (ed.), *Discussing Discourse*, 149–167. Birmingham: University of Birmingham.
- Calzada Pérez, María & Laviosa, Sara. 2021. Twenty-five years on: Time to pause for a new agenda for CTIS. *MonTI. Monographs in Translation and Interpreting* 13, 7–32. <https://doi.org/10.6035/MonTI.2021.13.01>
- Čermák, Petr. 2019. InterCorp: A parallel corpus of 40 languages. In Doval, Irene & Sanchez-Nieto, M. Teresa (eds.), *Parallel Corpora for Contrastive and Translation Studies: New resources and applications*, 93–101. Amsterdam: John Benjamins.
- Čermák, Frantisek & Rosen, Alexandr. 2012. The case of InterCorp, a multilingual parallel corpus. *International Journal of Corpus Linguistics* 17(3). 411–427.
- Černyševa, Mariâ. 2014. O probleme peredači zvukovyh povtorov pri perevode poëtičeskikh tekstov (na materiale tvorčestva G. Bajrona) [On the problem of dealing with sound repetitions in the translation of poetic texts (based on G. Byron's work)]. *Vestnik Čerepovetskogo gosudarstvennogo universiteta* 5. 75–78.
- Chesterman, Andrew. 2004. Hypothesis about translation universals. In Hansen, Gyde & Malmkjær, Kirsten & Gile, Daniel (eds.), *Claims, Changes and Challenges in Translation Studies*, 1–13. Amsterdam: John Benjamins.
- Chesterman, Andrew. 2011. The Significance of Hypotheses. *TTR* 24(2). 65–86. <https://doi.org/10.7202/1013395ar>
- Chmiel, Agnieszka & Kajzer-Wietrzny, Marta & Koržinek, Danijel & Jakubowski, Dariusz & Janikowski, Przemysław. 2024. Syntax, stress and cognitive load, or on syntactic processing in simultaneous interpreting. *Translation, Cognition & Behavior* 7(1). 22–47.
- De Baets, Pauline & De Sutter, Gert. 2022. How do translators select among competing (near-)synonyms in translation? A corpus-based approach using random forest modelling. *Target-International Journal of Translation Studies* 35(1). 1–33. <https://doi.org/10.1075/target.20095.deb>
- De Sutter, Gert & Lefer, Marie-Aude. 2019. 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.
- Ermolaeva, Irina. 2017. Semantičeskaâ klassifikaciâ glagolov reči v russkom âzyke [Semantic classification of verbs of speech in the Russian language]. *Vestnik SPBGU. Âzyk i literatura* 14(3). 362–375.
- Filatova, Ganna. 2023. Stilističeskaâ interpretaciâ glagolov rečevogo dejstviâ pri perevode fantastičeskogo teksta [Stylistic interpretation of verbs of speech in the translation of a fantastic text]. *Novyj filologičeskij vestnik* 4(67). 58–69.
- Fellbaum, Christiane (ed.) 1998. *WordNet: An electronic lexical database*. MIT Press.
- Fox, John & Weisberg, Sanford. 2019. *An R Companion to Applied Regression*. 3rd edn. Sage: Thousand Oaks.
- Ganesh, Siva & Cave, Vanessa. 2018. P-values, p-values everywhere! *New Zealand Veterinary Journal* 66 (2). 55–56. <https://doi.org/10.1080/00480169.2018.1415604>
- Grabowski, Łukasz. 2023. Statistician, Programmer, Data Scientist? Who is, or Should Be, a Corpus Linguist in the 2020s? *Journal of Linguistics/Jazykovedný časopis* 74(1). 52–59.
- Grabowski, Łukasz & Olejniczak, Jędrzej & Mastropierro, Lorenzo. 2026. (In print.) A multifactorial study of English-to-Polish translation of reporting verbs in literary novels: A negative binomial regression with mixed effects. *Research in Corpus Linguistics* 15(1).
- Gries, Stefan & Wulff, Stefanie. 2012. Regression analysis in translation studies. In Oakes, Michael & Li, Meng (eds.), *Quantitative Methods in Corpus-Based Translation Studies: A practical guide to descriptive translation research*, 35–52. Amsterdam: John Benjamins.
- Hair, Joseph F. Jr. & Black, William C. & Babin, Barry J. & Anderson, Rolph. 2009. *Multivariate Data Analysis*. 7th edn. Upper Saddle River: Pearson Prentice Hall.
- Halverson, Sandra. 2003. The cognitive basis of translation universals. *Target. International Journal of Translation Studies* 15(2). 197–241. <https://doi.org/10.1075/target.15.2.02hal>
- Halverson, Sandra. 2010. Cognitive translation studies. In Shreve, Gregory & Angelone, Erik (eds.), *Translation and Cognition*, 349–369. Amsterdam: John Benjamins.
- Halverson, Sandra. 2017. Gravitational pull in translation. Testing a revised model. In De Sutter, Gert & Lefer, Marie-Aude & Delaere, Isabel (eds.), *Empirical Translation Studies: New Methodological and Theoretical Traditions*, 9–46. Berlin: De Gruyter.
- Hermann, Berenike & van Dalem Oskam, Karina & Schoech, Christoph. 2015. Revisiting Style, a Key Concept in Literary Studies. *Journal of Literary Theory* 9(1). 25–52.

- Hirschberg, Julia & Manning, Christopher D. 2015. Advances in natural language processing. *Science* 349(6245). 261–266.
- Juršič, Matjaž & Mozetič, Igor & Erjavec, Tomaž & Lavrač, Nada. 2010. LemmaGen: Multilingual Lemmatisation with Induced Ripple-Down Rules. *Journal of Universal Computer Science* 16(9). 1190–1214.
- Kabacoff, Robert. 2015. *R in Action. Data analysis and graphics with R*. Shelter Island: Manning Publications Co.
- Kajzer-Wietrzny, Marta & Grabowski, Łukasz. 2021. Formulaicity in constrained communication: An intermodal approach. *MonTI* 13. 148–183.
- Kowalska, Danuta. 2020. Funkcja powtórzeń językowych w Uchu Igielnym Wiesława Myśliwskiego [The function of linguistic repetition in Wiesław Myśliwski's *Ucho igielne*]. *Studia Językoznawcze* 19. 219–245.
- Laviosa, Sara. 1998. Core Patterns of Lexical Use in a Comparable Corpus of English Narrative Prose. *Translators' Journal* 43 (4). 557–570.
- Laviosa, Sara. 2001. Universals of translation. In Baker, Mona & Malmkjær, Kirsten (eds.), *Routledge Encyclopedia of Translation*, 288–291. London: Routledge.
- Levshina, N. 2015. *How to Do Linguistics with R. Data exploration and statistical analysis*. Amsterdam: John Benjamins.
- Levshina, N. 2021. Conditional inference trees and random forests. In Paquot, M. & Gries, T. (eds.), *Practical Handbook of Corpus Linguistics*, 611–643. New York: Springer.
- Lewis, Fraser & Butler, Adam & Gilbert, Lucy. 2011. A unified approach to model selection using the likelihood ratio test. *Methods in Ecology and Evolution* 2. 155–162.
- Lüdecke Daniel & Ben-Shachar, Mattan & Patil, Indrajeet & Waggoner, Phillip & Makowski, Dominique. 2021. performance: An R Package for Assessment, Comparison and Testing of Statistical Models. *Journal of Open Source Software* 6(60). 3139. <https://doi.org/10.21105/joss.03139>
- Mahlberg, Michaela. 2013. *Corpus Stylistics and Dicken's Fiction*. London: Routledge.
- Mahlberg, Michaela. 2018. Corpus Stylistics. In Sotirova, Violeta (ed.), *The Bloomsbury Companion to Stylistics*, 139–156. London: Bloomsbury.
- Mastropiero, Lorenzo. 2019. *Corpus Stylistics in Heart of Darkness and its Italian Translations*. London: Bloomsbury.
- Mastropiero, Lorenzo. 2020. The translation of reporting verbs in Italian: The case of the Harry Potter series. *International Journal of Corpus Linguistics* 25(3). 241–269.
- Mastropiero, Lorenzo & Grabowski, Łukasz. 2024. Repeated reporting verbs in English novels and their Italian and Polish translations: A preliminary multifactorial study. *Across Languages and Cultures* 25(2). 310–330.
- Mauranen, Anna. 2008. Universal tendencies in translation. In Anderman, Gunilla & Rogers, Margaret (eds.), *Incorporating Corpora: The Linguist and the Translator*, 32–48. Clarendon: Multilingual Matters.
- Montgomery, Douglas C. & Peck, Elizabeth A. 1992. *Introduction to linear regression analysis*. New York: Wiley.
- Moskvin, Vasilij. 2000. Tipologija povtorov kak stilističeskoj figury [Typology of repetition as a stylistic figure]. *Russkij ŗzyk v škole* 5. 81–85.
- Nakagawa, Shinchi & Johnson, Paul C. D. & Schielzeth, Holger. 2017. The coefficient of determination R<sup>2</sup> and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of the Royal Society Interface* 14. 20170213. <https://doi.org/10.1098/rsif.2017.0213>
- Olejniczak, Jędrzej. 2023. *Interlingual Interference in English-Polish Literary Translations: Corpus-based Study of Literary Translator's Style*. Wrocław: University of Wrocław. (PhD Dissertation.)
- Petrunina, Nadežda. 2012. Ob ěkvivalentnosti edinic povtora v russkoŗzyčnyh perevodah liriki U.B. Jejtca [On the equivalence of repetition units in Russian-language translations of W.B. Yeats's lyrics]. *Vestnik Tomskogo gosudarstvennogo universiteta* 362. 17–19.
- Piotrowski, Tadeusz. 1994. *Problems in Bilingual Lexicography*. Wrocław: Wydawnictwo Uniwersytetu Wrocławskiego.
- Podpečan, Vid. 2024. *Lemmagen3. Python library*. <https://pypi.org/project/lemmagen3/> (Accessed: 2024-12-03)
- R Core Team. 2023. *R: A language and environment for statistical computing*. Vienna: The R Foundation for Statistical Computing. <https://www.R-project.org/> (Accessed: 2024-11-02)
- Rosen, Alexandr & Vavřin, Martin & Zasina, Jan A. 2022. *The InterCorp Corpus – Czech, version 15 of 11 November 2022*. Prague: Institute of the Czech National Corpus, Charles University.
- Ruano San Segundo, Pablo. 2016. A corpus-stylistic approach to Dickens' use of speech verbs: Beyond mere reporting. *Language and Literature* 25(2). 113–129.

- Ruano San Segundo, Pablo. 2017. Reporting verbs as a stylistic device in the creation of fictional personalities in literary texts. *Journal of the Spanish Association of Anglo-American Studies* 39(2). 105–124.
- Salciute-Civiliene, Gabriele. 2016. Relative and dynamic aspects of variation in response to lexical repetition. A corpus-based case study of the translations of Faulkner's *The Sound and the Fury* into Lithuanian, Polish and Russian. London: King's College London. (PhD Dissertation.) <https://kclpure.kcl.ac.uk/portal/en/studentTheses/relative-and-dynamic-aspects-of-variation-in-response-to-lexical--2> (Accessed: 2025-04-20)
- Saldanha, Gabriela. 2011. Translator style. Methodological considerations. *The Translator* 17(1). 25–50.
- Scherber, Christoph. 2017. Using R to interpret interaction effects in statistical models. (Online version: <http://www.christoph-scherber.de/content/PDF%20Files/Scherber2017InteractionEffects.pdf>) (Accessed: 2025-05-28)
- Scherber, Christoph. 2019. *An introduction to generalized linear models*. (Online version: <http://www.christoph-scherber.de/content/PDF%20Files/Generalized%20linear%20models.pdf>) (Accessed: 2024-12-30)
- Šutěmova, Nataliá. 2010. Povtor v stihe kak ob"ekt perevodčeskoj refleksii [Repetition in poetry as an object of translational reflection]. *Vestnik Permskogo universiteta* 3(9). 74–79.
- Shuttleworth, Mark & Cowie, Moira. 1997. *Dictionary of Translation Studies*. Manchester: St Jerome Publishing
- Toury, Gideon. 1995. *Descriptive Translation Studies and Beyond*. Amsterdam/Philadelphia: John Benjamins.
- Urža, Anastasiá. 2018. Strategii interpretacii glagolov vvodáših reč' v sovremennyh russkih perevodah hudožestvennoj proze [Interpretation strategies of verbs introducing speech in modern Russian translations of prose fiction]. *Vestnik VolGU. Seriá 2, Ázykoznanie* 17(4). 117–128.
- Venables, William & Ripley, Brian D. 2002. *Modern Applied Statistics with S*. 4th edn. New York: Springer.
- Venuti, L. 1995. *The Translator's Invisibility*. London: Routledge.
- Wang, Guofeng & Xin, Yihang. 2024. An analytical framework for corpus-based translation studies. *Humanities and Social Sciences Communications* 11, 1709. <https://doi.org/10.1057/s41599-024-04250-4>
- Winter, Bodo. 2019. *Statistics for Linguists: An Introduction Using R*. London: Routledge.
- Witoszowa, Bożena. 1985. Rola powtórzeń w tekście literackim stylizowanym na tekst mówiony: Na przykładzie monologu wypowiedzianego [The role of repetition in a literary text styled as a spoken text: On the example of a spoken monologue]. *Język Artystyczny* 3. 95–109.
- Zeileis, Achim & Kleiber, Christian & Jackman, Simon. 2008. Regression Models for Count Data in R. *Journal of Statistical Software* 27(8). 1–25.
- Žilina-Èls, Tat'ána. 2020. Povtor kak funkcional'nyj èlement v russkoj volšebnoj skazke i problemy ego perevoda na anglijskij ázyk [Repetition as a functional element in Russian fairy tales and the problems of its translation into English]. *Art Logos* 1(10). 129–140.

## Contact information:

Łukasz Grabowski  
University of Opole  
Institute of Linguistics  
e-mail: lukasz@uni.opole.pl

Daniel Borysowski  
University of Opole  
Institute of Linguistics  
e-mail: dborysowski@uni.opole.pl