From ‘said’ to ‘said differently’: modelling repetition versus lexical variation in English-to-Slovak translation of reporting verbs in literary novels

Łukasz Grabowski[1*], Filip Kalaš[2], Daniel Borysowski[1]

[1] University of Opole, Opole, Poland
[2] Bratislava University of Economics and Business, Bratislava, Slovakia
[*] Corresponding author’s email: lukasz@uni.opole.pl

DOI: https://doi.org/10.53482/2026_60_431

ABSTRACT

This corpus-based multifactorial study aims to explore potential predictors of repetition versus lexical variety in translation of repeated reporting verbs from English into Slovak in literary novels. First, we provide a theoretical overview of research on repetition and reporting verbs in Slovak and Czech translation studies. Next, using a sample of 14 literary novels extracted from InterCorp corpus (v.15), we fit multiple negative binomial regression models with mixed effects to assess the effect that several predictor variables (frequency, semantic category, verb length, number of senses, translators) have on the response variable, i.e. the number of Slovak target-text reporting verbs an English source-text (ST) reporting verb is translated into. The findings revealed that factors such as frequency of use of ST reporting verbs, the semantic category of neutral ST reporting verbs, as well as the translators as a random effect, influence Slovak translators’ decisions of using a wide variety of Slovak reporting verbs instead of preserving the originals’ patterns of repetition. More precisely, the model allowed us to explain 70% of the variation (per conditional r-squared) in the response variable. Against the backdrop of prevailing stylistic norms in Slovak, the findings shed light on the translator’s choices in rendering recurrent reporting verbs introducing direct speech, a stylistically salient feature of literary texts.

Keywords: literary translation, parallel corpus, repetition, corpus-based analysis, regression modelling

1 Introduction

Repetition is a common literary device (Klinger, 2019), which affects coherence, or stylistic integrity of the text (Peprník, 1969). The poetic function of repetition (Patáková, 1987) further highlights its aesthetic and structural significance in literary texts: it is often linked to rhythm, emphasis, and emotional depth. All this means that handling repetition is not a trivial task in translation and its treatment varies depending on text type, cultural norms, and translator preferences, among others. For example, Ben-Ari (1998) asserts that translation rules frequently prohibit repetition, particularly in writing, where diversity

is encouraged by target language stylistic traditions. Furthermore, the treatment of repetition in translation has implications for equivalence, a central concept in translation studies (Gromová, 2003; Pym, 2009).[1] The two have a complex relationship in that repetition may be either a problem or a strategy for achieving equivalence according to linguistic and cultural norms (Nida and Taber, 1982), e.g. literary translations may require adaptation of repetition to meet target language aesthetics while preserving the stylistic and emotional impact of the original. The notion of equivalence is intrinsically linked to the concept of shifts (cf. Catford, 1965; Zupan, 2006), emerging as a consequence of the translator's interpretive process.[2] These shifts contradict the notion of repetition, but it may happen that the avoidance of repeated lexical items contributes to the flow, readability, and naturalness of a text.

Repetition in translation has been also linked to so-called translation universals such as explicitation and simplification (Baker, 1993, 1996). Blum-Kulka's (1986) explicitation hypothesis implies that translators can manipulate repetitions to enhance textual clarity and render the target text more explicit than the original. Similarly, Hoey's (1991, 2005) lexical priming theory emphasises the cognitive function of repetition in the shaping of readers' expectations and text coherence, and implies that translators' choice of repetition might be conditioned by their own language exposure and stylistic preference. In addition, Slovak translation and stylistic studies, particularly by Miko (1970, 1978) and Popovič (1976), underscore the interaction between textual organisation and stylistic variation, which can possibly throw more light on how repetition is managed in Slovak translations.

The handling of repetition in translated texts has been thoroughly examined by Slovak and Czech translation scholars, who have identified a number of strategies, including synonymization, omission, and structural restructuring. While some contend that in order to maintain rhetorical function, repetition should be preserved in translation, others draw attention to the target language's stylistic conventions, which frequently value stylistic diversity over direct repetition. In other words, some scholars advocate for synonymization and reduction of source text repetitions to improve readability of translated texts (cf. Bečka, 1992; Levý et al., 2011; Nádvorníková, 2020) while others (cf. Augustinská, 1985; Grepl, 1967; Patáková, 1987) argue that repetition serves essential rhetorical, communicative and poetic functions that should be preserved in translation. Consequently, repetition is not only viewed as a linguistic feature but

[1]The Slovak school of translation is grounded in the principle of functional equivalence, which (Vilikovský, 1984, p. 39) defines as follows: "The identity of function serves as a common denominator across all levels of communication, irrespective of specific linguistic expressions. The task of translation is not to replicate the linguistic means themselves but rather to convey the function they fulfill within a broader context".

[2](Gromová, 2003, p. 44) highlights that shifts in translation are inevitable due to linguistic, cultural, or literary differences between the source and target languages. As (Popovič, 1975, p. 112) asserts, the translator's primary responsibility is to comprehend and interpret the original text. However, interpretation is influenced not only by the source text itself but also by the translator's perspective and other contextual factors, accordingly leading to modification. The expressive structure of the source language cannot be translated to the target language without some degree of transformation or loss. Consequently, translation shifts must be acknowledged as an inherent aspect of the process.

also as a strategic tool in translation, where its preservation or omission can significantly alter the impact of the original text. For instance, the well-known repetition of *I have a dream* is occasionally reduced or altered in Czech translations of Martin Luther King's *I Have a Dream* (*Mám sen, věřím* – 'I have a dream, I believe'), which changes its rhythmic and persuasive power. The findings of the study on simplification hypothesis in translated Czech (Cvrček and Chlumská, 2015) suggest that repetition is often reduced or eliminated in translations to enhance clarity and readability. This is especially noticeable in scholarly and journalistic texts, where the intended audience may find excessive repetition to be unnecessary. In a similar vein, the findings of Čermáková's (2015) corpus study confirm a tendency to neutralize repetition in Czech and Slovak translations. Gresty (2012) also looks at how Slovak writing rules allow for more repetition and verbosity than English: his study focuses on instances when Slovak expressions, such as *On sa naozaj naozaj veľmi snažil* ('He really, really tried hard'), are translated into English (as *He made a great effort*), indicating a sacrifice of repetition in favour of conciseness. Timing and space limits are also important variables in reducing repetition, notably in audiovisual translation, which presents a major problem in preservation of source text repetitions (Gromová and Janecová, 2013). To preserve brevity, repeated lines, such as *No, no, no!*, are frequently shortened to a single *Nie*! in Slovak dubbing and subtitling, illustrating how technical constraints and translation mode influence translatorial decisions.

However, other scholars advocate preserving repetition as a rhetorical device in translation. Abdulla (2001) contends that repetition, especially in persuasive words, is intentional and significant rather than just redundant. Kundera (1998) argues that translators should maintain all repetitions of words and phrases in the original and strive to avoid using synonyms instead. In a similar vein, noting that rhetorical repetition in speeches and political documents tends to underline important points, Abdulla (2001) criticizes the propensity to misuse synonyms in translation.

Thus, repetition is an important stylistic phenomenon yet it remains relatively underexplored in English-to-Slovak translation. Most of the studies conducted so far are primarily descriptive in nature, which means that based on their findings it is difficult to precisely pinpoint those linguistic factors (predictors) that are responsible for either preservation of source-text repetition or its avoidance, that is, opting for more lexical variety in translation. To this end, our empirical explanatory research is positioned on the interface of corpus stylistics, translation studies, literary translation and multi-factorial statistics. According to Mahlberg (2018), corpus stylistics, which employs corpus linguistic methods (such as wordlist analysis, keyword analysis, and concordance analysis) to develop a more comprehensive description of language use in literary works, has strengthened methodological rigour and offered deeper insights into both frequent and rare features of literary texts. For example, Semino and Short (2004) demonstrated how corpus linguistic approaches can uncover stylistic patterns in literary dialogue, including the role of repetition in constructing narrative voice. As translation is inherently a complex and multifaceted phenomenon

(i.e. no single factor, be it linguistic, social or cultural, is solely responsible for particular translatorial decisions), we can currently observe a growing popularity of inferential and multifactorial statistics as well as machine learning methods (e.g., regression models, decision trees, random forest) in research on potential predictors of translatorial choices and, thus, on language use in translation (e.g. De Baets and De Sutter, 2022; De Sutter and Lefer, 2019; De Sutter et al., 2023; Dupont and Zufferey, 2017; Grabowski and Borysowski, 2025; Kajzer-Wietrzny and Ivaska, 2020; Kang and Zhang, 2025; Kruger, 2019; Mastropierro, 2022; Wang and Xin, 2024; Zufferey, 2016). These methods have been successfully used in research on authorship attribution, translatorial attribution, and translator's style (e.g. Eder, 2011; Grieve, 2023; López-Escobedo et al., 2013; Rybicki and Heydel, 2013; Seroussi et al., 2014; Škorić et al., 2022), making them an appealing choice for studying how translators deal with repetitions in source texts.

Thus, our general goal is to identify those factors that impact the ways repeatedly used source-text English reporting verbs that introduce direct speech (i.e. following utterances) are translated into Slovak in selected literary novels. More precisely, we aim to identify the predictors of preservation or avoidance of repetition in translation (see Section 3 of this paper for a detailed description of methodology). Since reporting verbs serve as our unit of analysis, in what follows we present a more detailed overview of their treatment in Slovak and Czech linguistics, including their role and function in translation.

2 Reporting verbs and their treatment in translation: a view from the Slovak and Czech linguistics

In specialized literature, we may find two similar terms related to verbs related to the act of speaking, namely reporting verbs and verba dicendi. The relationship between them can be considered to be hierarchical: in short, all verba dicendi are reporting verbs but not all reporting verbs are verba dicendi. More precisely, reporting verbs are used to introduce reported discourse, whether direct or indirect speech or thought (e.g. *say, tell, report, remark, exclaim, think, believe, assume, admit, warn, beg, know, comment*, and others (Huddleston and Pullum, 2017; Quirk et al., 1985). A common feature of these is their position immediately preceding or following the direct speech. Based on their semantic-pragmatic characteristics and discourse functions, reporting verbs may be further classified into several subgroups. These include verbs of saying, e.g. *say, tell, report, remark, exclaim, comment*, verbs of thinking, e.g. *think, believe, assume, know*, and other verbs that introduce the speaker's stance, emotion, interpersonal positioning, or speech act, e.g. *admit, warn, beg*. However, this typology is not unified as various scholars categorize reporting verbs from different theoretical perspectives (cf. Caldas-Coulthard, 1987; Levin, 1993; Searle, 1976; Thompson and Yiyun, 1991). Among reporting verbs, *say* is regarded as the most frequent and neutral one (Quirk et al., 1985).

In Slovak and Czech linguistics, the Latinized term *verba dicendi* (verbs of saying) has become well established, although native equivalents such as *slovesá hovorenia* (Slovak) and *slovesa mluvení/hovoření* (Czech) are also frequently used. Nemcová (2012) characterizes verba dicendi as a group of verbs defined by a common semantic feature, namely, the description of speech activity.[3] Accordingly, any verb semantically associated with speech is typically categorised among verba dicendi. It should be noted that within the Czech and Slovak linguistic tradition, the distinction between reporting verbs and verba dicendi is often blurred. Research tends to focus predominantly on verbs of saying, while other types of reporting verbs are frequently overlooked. This tendency explains why numerous scholarly articles on verba dicendi are referenced in the present study, as they often encompass broader analysis of reporting verbs despite focusing primarily on speech-related verbs.

Research on verba dicendi, as the most frequently investigated group of reporting verbs, has been extensive in Slovak and Czech linguistics, highlighting the verbs' important roles in semantics (Daneš, 1973; Hirschová, 2017a, 2017b; Knápek, 2019; Nemcová, 2012; Preislerová, 2015; Svobodová, 2007), discourse (Hirschová, 1982, 1988; Šoltys, 1983), syntactic function (Bauer and Grepl, 1972), stylistic variation (Hoffmannová, 2024; Patáková, 1987; Pisárčiková, 1978; Samlerová, 2010) as well as in translation (Fárová, 2016; Fialová, 2020; Nádvorníková, 2017, 2020; Staroňová, 2023). In general, verba dicendi hold particular significance as they form an essential part of reported speech structures. Mistrík (1993, 2021) reinforced this by analysing their role in fictional prose, where verba dicendi bridge narrative elements with direct speech. Mistrík (2021) examined their placement in sentences introducing direct speech, revealing that such structures often contain verba dicendi, such as *povedal* 'said', *vravel* 'said', *odvetil* 'answered' and *spýtal sa* 'asked'. These verbs not only introduce direct speech but also act as synonymous paraphrases of the speech act itself. Thus, their syntactic and pragmatic function is deeply intertwined with situational context rather than explicit linguistic encoding. This feature of verba dicendi is particularly evident in what is referred to as authorial speech, which functions as a connective textual element within prose fiction (Mistrík, 1993). In this context, the most typical representative is the verb *hovoriť*/*povedať* 'say' in Slovak or *říkat*/*říci* 'say' in Czech. Additionally, Knápek (2019) noted that while verba dicendi primarily appear in authorial speech, they also frequently occur in direct speech from fictional characters. Knápek's (2019) study of verba dicendi in Karel Čapek's prose revealed that his detective fiction utilizes

[3]While verba dicendi primarily indicate information transmission, they may also overlap with verba sentiendi (e.g., *vidieť* 'see', *počuť* 'hear', *cítiť* 'feel' and verba cogitandi (e.g., *myslieť* 'think', *uvažovať* 'consider'), though these appear less frequently in introductory sentences (Šoltys, 1983). Hoffmannová (2024, p. 132) reaffirmed that verba dicendi emphasize the act of transmission rather than information acquisition. Beyond their primary reporting function, verba dicendi are also subject to metaphorical extensions. Drawing from the SYN2005 corpus of Czech, Samlerová (2010) identified 24 metaphorical extensions of verba dicendi. These include metaphors of animal sounds (*bručet* 'grumble', *kňučet* 'moan"), natural phenomena (*kvílet* 'howl', *zašumět* 'rustle'), speech mannerisms (*mumlat* 'mumble', *žvanit* 'chatter'), and evaluative speech (*kritizovat* 'criticize', *chvástat* se 'brag'). The findings underscore how verba dicendi go beyond their basic reporting function to convey additional semantic nuances through metaphorical associations, which highlights the interplay between semantics and pragmatics.

repetitive, straightforward reporting verbs for clarity, while his fairy tales incorporate archaic and stylized forms to evoke folkloric elements, which highlights how authors strategically deploy verba dicendi to shape narrative tone and reader perception.

Repetition in translation, particularly in the use of reporting verbs, reveals significant linguistic and stylistic differences across languages. One of the most significant differences in reporting verb usage concerns lexical diversity. Vilikovský (1984, p. 215) speaks at this point of the ingrained principle of replacing the stereotypical verbs of the original with a wider repertoire of variations in Slovak.[4] Nádvorníková (2017, 2020) compared the type-token ratio (TTR) of reporting verbs in English, French, and Czech fiction, demonstrating that Czech exhibits the highest degree of lexical variation. While French retains a relatively neutral verb *répondit* (‘answered’), Czech introduces a stylistically richer verb *zamumlal* (‘muttered’), which adds more information about the tone of speech. Similarly, Fialová (2020) analysed German-Czech translations and found that Czech prefers synonymous expansions over direct repetition. The German *sagen* (‘say’) is frequently rendered in Czech with context-sensitive verbs, such as *sdělit* ‘inform’, *upozornit* ‘warn’, *zašeptat* (‘whisper’). Pisárčiková (1978) notes that Slovak literature, even in modern texts, retains a strong preference for expressive reporting verbs, introducing synonyms or restructuring sentences to avoid repetition, e.g. *“To je strašné!” zvolal* – (‘That’s terrible! he exclaimed’). In a corpus-based analysis, Čermáková (2015) showed that translators usually neutralize repetition by replacing repeated words with synonyms or restructuring phrases. For example, the Czech translation of the English phrase *He said, he said, he said* is *“Řekl, poznamenal, dodal”* (‘He said, he remarked, he added’), which lessens the original’s stylistic impact. This result supports Levý’s (1963) theory that Czech and Slovak translators typically choose lexical variation over exact repetition to preserve fluidity. According to Levý’s analysis of Czech translations of Shakespeare, English conversations that contain repeated verbs like *speak, speak, speak* are frequently substituted with other phrases, e.g. *mluv, prav, řekni* ‘speak, say, say’ (Jettmarová, 2008; Schultze, 2015).

Translation studies have also provided valuable insights into the variability of verba dicendi across languages. Staroňová (2023), examining translations between Slovak and English, found that Slovak translators exhibit greater lexical diversity, using semantically rich verbs (i.e. with narrow meaning) to align with dominant literary norms. English translations, conversely, favour simplification, frequently reducing nuanced Slovak reporting verbs to verbs with broad meaning, such as *say*. These findings

[4]According to Vilikovský, the accumulation of interpretive variations can damage the author’s style and frequently have an unintended humorous effect, particularly if rare and emotional verbs are sought as equivalents. Interestingly, writers of the older generation use a wide repertoire of introductory verbs with a low repetition index; modern writers, on the other hand, use a smaller number of verbs with a high repetition index; in more than half of the cases, dialogue is introduced with the verb tell (Vilikovský, 1984, p. 218).

support the asymmetry hypothesis proposed by Klaudy and Károly (2005), which suggests explicitation is more common in translation of reporting verbs than implicitation.

All these findings make reporting verbs introducing direct speech a valuable unit of analysis for investigating broader translation patterns related to dealing with repetition (cf. Mastropierro, 2020, 2022; Mastropierro and Grabowski, 2024). Although we hypothesize that Slovak translators would rather avoid reproduction of repeatedly used reporting verbs introducing direct speech in the English originals (e.g. translating the verb *said* with a wide range of reporting verb equivalents in Slovak), we do not precisely know the rationale behind such decisions. What about reporting verbs with narrower meanings in the English originals, e.g. *answered, replied?* Would they be handled in the same way as the verbs with broader, neutral meanings, like *said or told*? Is the treatment of reporting verbs in translation conditioned by other factors (e.g., word length, the number of senses or the translator's idiolect)? Thus, in this corpus-based multifactorial study we attempt to provide answers to the following research questions: (i) What linguistic factors have a significant effect on the avoidance or reproduction of reporting verbs' repetition in the selected Slovak novels translated from English? (ii) What is the proportion of variance explained by fixed and/or random effects? We hope that our findings will contribute new perspectives to the discussion on repetition in translation into Slovak, enhancing our understanding of this phenomenon. In what follows, we describe the methodology in greater detail.

3 Methodology

The methods used in this study are grounded in corpus linguistics, as we use bilingual concordances extracted from the InterCorp parallel corpus (Čermák and Rosen, 2012) as a research material, and multifactorial statistics. More precisely, we conducted regression modelling in order to identify those linguistic predictors that impact repetition or lexical variety in the English-to-Slovak translation of reporting verbs signalling direct speech (i.e. character's utterances) in literary texts. Using custom-designed bilingual English-Slovak CQL queries, we extracted pairs of aligned reporting verbs from 14 novels (*Winnie the Pooh, The Jungle Book – other, The Jungle Book – Mowgli, the House at Pooh Corner, The Hobbit or There and Back Again, The Hitch Hiker's Guide to the Galaxy, The Fellowship of the Ring, The Da Vinci Code, Harry Potter and the Philosopher's Stone, For Whom the Bell Tolls, Dracula, Catch-22, Alice in Wonderland, 1984*). This selection was based primarily on the availability of texts translated from English into Slovak, found in the InterCorp corpus (ver. 15), which is a large annotated multilingual parallel corpus (Rosen et al., 2022).

As mentioned earlier, reporting verbs in source (ST) and target texts (TT) were selected as the unit of analysis. Therefore, using the CQL queries we searched for the following patterns in the STs: "closing quotation marks + he or she + past tense verb" and "closing quotation marks + past tense verb + he or she"

(Mastropierro and Grabowski, 2024), and these were matched by equivalent patterns in Slovak, taking into consideration specific orthographic conventions of recording dialogues in each language. From the obtained parallel concordances in each novel, we retrieved the translation of each ST verb into a Slovak reporting verb. For the analysis, we used only those ST verbs with a frequency of 2 or higher, as otherwise we would not deal with repetition. This way we retrieved 5,298 verb tokens and 130 verb types in the original novels, as well as all of their translations in Slovak as reporting verbs (530 unlemmatized and 411 lemmatized types). In further analyses, we used lemmatized types in order to ensure compatibility of the English and Slovak data, i.e., the Slovak 3rd person past tense feminine, neutral (in 1 novel) and masculine forms were lemmatized to the single base form. The lemmatization was conducted using lemmagen3, a Python wrapper developed by Podpečan (2024) for Lemmagen lemmatizer (ver. 2.2), which supports 19 languages, including Slovak (Juršic et al., 2010).[5]

All reporting verb types were annotated for six linguistic features that constitute potential predictors of reproduction or avoidance of repetition, starting with the frequency of each ST reporting verb type in each literary novel (e.g. *asked* was used 53 times in the English-original novel "Da Vinci Code" while *yelled* was used 3 times). 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 "wnet_senses_com", 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, that is, introducing direct speech. For example, the verb *asked* has 7 distinct senses in the Princeton WordNet, and 5 of them belong to the domain of communication; in the case of the verb *snapped*, it has 13 distinct senses, and only 2 of them belong to the domain of communication.

The third variable indicates semantic-functional category of the English ST reporting verb based on Caldas-Coulthard's (1987) categorisation, who 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*), 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 "sem_verb_type" has several levels coded as "N", "Str", "Mprop", "Mlin", "Pros", "Sdis", "Vier", "Vion" respectively. Consequently, the potential predictors "sem_verb_type" represents the factor with 8 levels. The reason for selecting this typology was, first, that we used the

[5]It is available at the following link: https://pypi.org/project/lemmagen3/.

English reporting verbs as the unit of analysis and, second, that the typology is well-suited for reporting verbs found in literary texts rather than more formulaic genres or text types.

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 a ST), the fourth potential predictor is "verb_length", measured in characters.

The last factor, which is "translator_id", indicates a Slovak translator of each novel coded as using acronyms (e.g. "PF" stands for P. Frank, who translated *The Hitch Hiker's Guide to the Galaxy* into Slovak. All in all, 14 novels were translated by 12 individual translators (J. Samcová translated two texts, namely *The Jungle Book – other* and *The Jungle Book – Mowgli*) and 1 novel, namely *Alice in Wonderland*, was translated by a pair of translators, J. Vojtek and V. Vojtková, coded as JV_VV). This information was retrieved from metafiles available in the InterCorp corpus ver. 15. As we argue that individual translators' choices impacted the ways in which reporting verbs were translated, and that we analysed only 14 literary novels out of a potentially infinite pool of texts, we decided to include "translator_id" as a random effect in our regression model (cf. Gries, 2015). This also allowed us to meet one of the assumptions for using mixed regression models, where random effects should have at least 5 or 6 levels (Bentz and Winter, 2013), and we have as many as 13 levels.

Hence, the predictors employed in the study include three numerical variables ("log_freq", "wnet_senses_com" and "verb_length") and two nominal variables ("sem_verb_type" and "translator_id") treated as factors with multiple levels, which include both fixed and random effects. The dependent variable is called "type_count", which is a count variable representing the number of Slovak TT reporting verb types (lemmas) used as translation equivalents of English ST reporting verbs. For instance, the value of "types_count" of 7 indicates that the ST reporting verb was translated into 7 different reporting verbs in Slovak. As an illustration, the verb *cried*, a prosodic verb according to the Caldas-Coulthard (1987) typology, was translated into Slovak as *zvolal* 'exclaimed', *kričal* 'shouted', *skríkol* 'screamed', *zavolal* 'called', *ukázal* 'pointed', *volal* 'called', *vykríkol* 'cried out' in the novel *The Hobbit or There and Back Again*. Thus, a value of 1 shows that the verb repetition in the original (ST) was preserved in translation (TT), while a value greater than 1 indicates lexical variety: the higher the value, the wider the range of distinct reporting verbs used in the TT as equivalents of a ST reporting verb. The process of data preparation was conducted using custom-designed Python scripts, and the final data was stored in 14 comma delimited files (csv), corresponding to each novel. These files have been made publicly accessible in an open data repository to ensure replicability and reproducibility of the study.[6]

[6]All the data and metadata, including CQL queries, used in this study is made available in an open data repository under the following link: https://osf.io/t2c6h/.

In order to identify statistically significant predictors of the number of Slovak translation equivalents (reporting verb types) of ST English reporting verbs as well as the predictors’ contribution to explaining variance, we initially intended to use Poisson regression, which is a type of a generalized linear model (GLM) typically employed to model count data and contingency tables, as recommended in specialized literature (Coxe et al., 2009; Scherber, 2017, 2019; Kabacoff, 2015, p. 312; Winter, 2019, p. 247; Winter and Bürkner, 2021, p. 1). The dependent variable “types_count” represents count data measured in non-negative integers, and the predictors in Poisson regression can be a mixture of numerical/continuous and nominal/categorical variables (Kabacoff, 2015, p. 312), which – as previously explained – is the case in this study. However, as we observed overdispersion in the data, that is, the variance for “type” is found to be considerably higher than the mean (mean of 4.65 and variance of 122.50, cf. Bentz and Winter, 2013), we used negative binomial regression instead of Poisson regression, as recommended by Scherber (2017, 2019) and Winter (2019) or Hair et al. (2009). This was later confirmed using the Likelihood Ratio Test (see Section 4). 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 the fewest variables through their “backward selection” (Winter, 2019, p. 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 statistically significant, i.e. have p-values higher than 0.05 (ibid.). As such, our approach and selected methods helped us ensure comparability of our findings with research conducted in English-to-Polish and English-to-Russian language pairs (Grabowski and Borysowski, 2025; Grabowski et al., 2026).

The analyses were conducted in the R environment using the following packages: car (Fox and Weisberg, 2019), MASS (Venables and Ripley, 2002), MuMIn (Bartoń, 2024) and glmmTMB (Brooks et al., 2017). In what follows, we present our findings.

4 Results

In order to assess the influence of all predictor variables, including a random intercept (“translator_id”), on the number of different Slovak reporting verbs used as translation equivalents (variable “types”), we fitted a series of negative binomial regression models using “backward selection” (Winter, 2019, p. 310), which means that statistically insignificant variables (with the p-value higher than 0.05) were iteratively removed. 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, p. 310). In order to double check whether we should proceed with a negative binomial regression model, which takes into consideration an additional dispersion parameter, we fitted both Poisson and negative binomial models with all the predictors and run

Likelihood Ratio Test (LRT)[7], which confirmed that the latter one is significantly better (p-value < 0.05, Figure 1). Both AIC and BIC values are lower for the negative binomial regression model (glm1nb), and this improvement in model fit is statistically significant (p-value < 0.05). This finding further confirms overdispersion in the data.

```
Models:
glm1p: types_count ~ logfreq + sem_verb_type + wnet_senses_com + verb_length + , zi=~0, disp=~1
glm1p:     (1 | translator_id), zi=~0, disp=~1
glm1nb: types_count ~ logfreq + sem_verb_type + wnet_senses_com + verb_length + , zi=~0, disp=~1
glm1nb:     (1 | translator_id), zi=~0, disp=~1
       Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
glm1p  13 1083.7 1131.1 -528.86   1057.7
glm1nb 14 1067.0 1118.0 -519.51   1039.0 18.702      1  1.528e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 1: Likelihood Ratio Test with nested models: results

Our experiments revealed that an optimal model with the lowest AIC value (1064.7) and BIC value (1108.5) was the following: types_count ~ logfreq + sem_verb_type + (1 | translator_id) (Figure 2). It shows that the frequency of a ST reporting verb, semantic category of neutral reporting verbs (e.g. *said*, *told*) as well as the translator as a random effect make the largest individual contributions to explaining the proportion of variation in the response variable “types_count” in the Slovak translations. The model allows us to explain almost 70% of the variation (per conditional r-squared) in the response variable, that is, the number of different Slovak verb types an English ST verb is translated into. This was computed with the help of r.squaredGLLM function in R using a delta method (Bartoń, 2024; Nakagawa et al., 2017). Without taking into consideration “translator_id” treated as a random effect, we would have explained only around 60% of variation, which is 10% less.

As for the predictors’ individual contributions to explaining variance in the response variable “types”, a summary (at the bottom of Figure 2, with p-value and positive or negative estimates indicating the direction and effect size of each predictor, including all levels of variables) reveals that the frequency of an ST reporting verb significantly influences its likelihood of being translated into multiple Slovak TT reporting verbs. For example, with a one-unit change (increase) in “logfreq”, the log of the expected number of translation equivalents increases by 0.56, while holding the rest of the predictor variables constant. For this count, this corresponds to a multiplicative increase of $\exp(0.56) \approx 1.75$ times (or 75% increase), which means that higher frequency has a strong positive effect on the count of “types_count”. This finding confirms our intuition: if a ST reporting verb is frequently used, then translators notice its

[7]We ran anova() function with two nested models in R. LRT is described, for example in Lewis et al. (2011), and it can only be applied 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).

```
 Family: nbinom2  ( log )
Formula:          types_count ~ logfreq + sem_verb_type + (1 | translator_id)
Data: data

     AIC      BIC   logLik deviance df.resid
  1064.7   1108.5   -520.4   1040.7      271

Random effects:

Conditional model:
 Groups        Name        Variance Std.Dev.
 translator_id (Intercept) 0.07934  0.2817
Number of obs: 283, groups:  translator_id, 11

Dispersion parameter for nbinom2 family (): 24.1

Conditional model:
                 Estimate Std. Error z value Pr(>|z|)
(Intercept)       0.08661    0.12813   0.676 0.499112
logfreq           0.56172    0.03662  15.341  < 2e-16 ***
sem_verb_typeN    0.49798    0.14852   3.353 0.000799 ***
sem_verb_typeSdis -0.16616   0.12950  -1.283 0.199457
sem_verb_typeStr  -0.24816   0.16105  -1.541 0.123339
sem_verb_typeVion -0.06464   0.13692  -0.472 0.636864
sem_verb_typePros  0.16834   0.12370   1.361 0.173563
sem_verb_typeVier -0.17722   0.16213  -1.093 0.274377
sem_verb_typeMProp 0.16916   0.47710   0.355 0.722922
sem_verb_typeMlin  0.00878   0.40322   0.022 0.982628
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 2: Negative binomial regression: final model summary

repetition and in order to avoid reproducing this repetition in translation they resort to using a wide array of TT reporting verbs. In the same vein, neutral verbs (e.g. *said*, *told*), which are broad in meaning and are frequently used as reporting verbs following dialogues, are also consistently translated with a wide range of Slovak translation equivalents, which translates into more lexical variety in translation as compared with the original novels. For example, the verb *said* occurs 131 times following dialogues in the novel *Alice in Wonderland* by Lewis Carroll (1865) and it has 33 lemmatized reporting verb equivalents in the Slovak translation performed by J. Vojtek and V. Vojtková (2004), and 37 unlemmatized types[8]. These findings also accord with a dominant stylistic convention in Slovak (see Section 1), where repetition and lack of lexical variety are synonymous with poor style, as in:

‘As wet as ever,’ ***said*** *Alice in a melancholy tone: ‘it doesn’t seem to dry me at all.’ → "Taká mokrá ako predtým," smutno* ***odpovedala*** *Alica. "Tie tvoje reči ma akosi nesušia."*

[8]These equivalents are as follows: *povedala* ’she said’, *spýtala sa* ’she asked’, *odpovedala* ’she answered’, *povedal* ’he said’, *namietla* ’she objected’, *odsekla* ’she retorted’, *ozvala sa* ’she responded’, *súhlasila* ’she agreed’, čudovala sa ’she wondered’, bránila sa ’she resisted’, hrešila ’she scolded’, maznala sa ’she caressed’, nepočula ’she didn’t hear’, *oborila sa* ’she snapped’, *obrátil sa* ’he turned around’, *obrátila sa* ’she turned around’, *obzerala sa* ’she looked around’, *ohradila sa* ’she protested’, *opakovala* ’she repeated’, *opýtala sa* ’she asked’, *otriasol sa* ’he shivered’, *ozval sa* ’he responded’, *pokračovala* ’she continued’, *počkala* ’she waited’, *preľakla sa* ’she got scared’, *prisvedčila* ’she nodded’, *prosila* ’she begged’, *riekla* ’she said’, *spýtal sa* ’he asked’, *tíšila* ’she calmed’, *vravela* ’she said’, *vydýchla si* ’she had a rest’, *vyhŕkla* ’she burst out’, *vyčítala si* ’she reproached’, *zahundral* ’he grumbled’, *začala* ’she began’, *zvolala* ’she shouted’. 31 variants are feminine forms and 6 variants are masculine forms.

‘*I beg your pardon,’* ***said*** *Alice very humbly: ‘you had got to the fifth bend, I think?’* → *"Prepáč, "****povedala*** *Alica skrúšene. "Ak sa nemýlim, máš už za sebou štyri zákruty."*

‘*I wish I hadn’t cried so much!’* ***said*** *Alice, as she swam about, trying to find her way out.* → *"Nemala som toľko plakať,"****vyčítala si*** *Alica, ako plávala a usilovala sa dostať z kaluže.*

‘*Poor little thing!’* ***said*** *Alice, in a coaxing tone, and she tried hard to whistle to it.* → *"Drobček môj!"* ***maznala sa*** *s ním, ba chcela naň aj zapískať.*

‘*What can all that green stuff be?’* ***said*** *Alice* → *"Čo je to za zeleň?"* ***čudovala*** *sa Alica.*

‘*I couldn’t help it,’* ***said*** *Five, in a sulky tone."* → *"Ja za to nemôžem,"* ***zahundral*** *Pätorka.*

Finally, we reported a low level of variance (0.079) and standard deviation (0.28) for the entire model, the values that indicate the spread of random effect. Although the 14 novels used in the study were translated by 12 different translators and one translation team of 2 translators (J. Vojtek & V. Vojtková), the tendencies that the model shows remain 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 (e.g. unlike in the case when variability is in the region of $0.2 - 0.3$): no single translator significantly influenced the observed patterns and the effect was spread across the translators.

5 Discussion and conclusion

The study findings revealed that that factors such as frequency of use as well as a select semantic category (neutral verbs) of ST reporting verbs influence Slovak translators’ decisions of using a wide variety of Slovak reporting verbs, thus avoiding repetition, in the 14 translations of English-original literary novels under scrutiny. These findings, which address the research question (i), partly align with the findings obtained for the English-to-Polish language pair (Grabowski et al., 2026), where broad-meaning neutral reporting verbs (e.g. *said*, *told*), among others, were also found to have been primed for being translated with the whole variety of Polish equivalents, which confirms our intuition. Moreover, our findings indicate that both fixed and random effects should be accounted for when attempting to identify the predictors of translatorial decisions with respect to the translation of reporting verbs. Without including the random intercept, i.e. the translators of the studied English novels into Slovak, we would have explained 10% less variation in the dependent variable. More precisely, the full model explained almost 70% of the variation, which provides an answer to the research question (ii). In an additional experiment, we used individual novels (text_id) instead of the translators as a random intercept and the results did not change. Overall, understanding the factors that lead to either the avoidance or preservation of repetition in translations can be valuable for translator training, for example the findings may inform pedagogical approaches by helping translators recognize patterns in lexical variation and develop strategies for dealing with repetitions in

original texts. These insights could also open up opportunities for reflecting upon these findings within the broader context of stylistic conventions in both source and target languages.

This descriptive and explanatory study has a number of limitations, though. 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 and sociolinguistic background of the referent (i.e. literary protagonist), the number of translation equivalents in lexical databases or dictionaries, as well as the span of repetition. Furthermore, a narrative perspective and point of view (first-person vs. third-person narration) could be explored even further, as subjective perspectives may involve using more expressive or evaluative reporting verbs. Other aspects worth considering include cognitive load and translator expertise, notably if the study is extended to interpreting. For example, it might reveal whether the translator's expertise influences the likelihood of lexical diversification or adherence to repetition. On top of that, the emotional intensity of the surrounding discourse could be assessed by applying sentiment analysis, which could help determine whether stronger emotions correlate with greater lexical variation in translation. In this exploratory 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 through alternative stylistic means (nominalizations, omissions, periphrastic constructions etc.), which also needs to be addressed in the future. This would provide a broader picture of how lexical repetition is managed beyond verb-to-verb equivalence perspective adopted in this study.

Furthermore, in this research, we focused on selected English-to-Slovak translations of literary novels only, yet comparisons across other text types, genres or modalities could provide a more comprehensive picture of what impacts translatorial decisions with respect to repeated reporting verbs or verba dicendi. For example, Preislerová (2015) examined verba dicendi in Czech fiction and journalism, noting stylistic differences between the two genres.[9] Hoffmannová (2024), who explored the frequency of verba dicendi in contemporary spoken Czech, revealed broader linguistic shifts favouring simplified, frequently used reporting verbs in informal speech.[10] Although both studies were conducted on the Czech language material, it may be tempting to verify if they also apply to Slovak texts. For instance, a diachronic study comparing older and contemporary Slovak translations could reveal whether stylistic norms around repetition have shifted over time. We therefore assume that comparisons across various genres and text types, which could be treated as additional independent variables, could offer more fine-grained insights into the predictors of lexical variety or repetition in translation. Expanding the analysis to other Slavic

[9]Fictional texts use these verbs to enhance narrative depth and provide character insights, often employing evaluative and expressive reporting verbs. In contrast, journalism prioritizes neutral, standardized verba dicendi to maintain factuality and clarity.

[10]Hoffmannová's (2024) findings indicate that Czech verbs *říct*/*říkat* 'say' dominate spoken usage, while archaic and literary lexemes (e.g., *blahořečit* 'bless', *vyřknout* 'utter') have become largely obsolete.

languages (e.g., South Slavic languages) could additionally determine whether the tendencies in lexical diversification are language-specific or broadly shared across the Slavic language family.

Another future avenue that would allow further verification of the obtained results would be conducting a similar study in the opposite direction, namely in the Slovak-to-English translation. This way we could see whether the tendency of normalization applies in the case of translating Slovak reporting verbs into a narrower range of English reporting verbs, notably neutral ones. From a methodological perspective, integrating neural machine translation or AI-assisted translation (i.e. conducted using selected large language models) into the study could tell whether automated translation systems replicate or diverge from human translator tendencies in handling repetition.

Summing up, irrespective of the aforementioned limitations, this is one of the first multifactorial corpus-based study on translation conducted in the English-to-Slovak language pair with the use of multifactorial methods. As a first step, we hope that this research will inspire broader future investigations into the factors influencing translatorial decisions used when dealing with repeated linguistic items or constructions (lexical items, multi-word units, syntactic structures etc.) in source texts, extending beyond reporting verbs as the unit of analysis, beyond literary texts as a research material, and beyond the language pair or translation direction under scrutiny in this paper.

Acknowledgments

This research was funded by the National Science Centre (NCN), Poland, grant number: 2023/51/B/HS2/00697 (Repetition in translation: a multifactorial descriptive and explanatory study, 2024–2026, grant holder: Łukasz Grabowski).

Data availability statement

The data associated with this research are available in a data repository under the following link: https://osf.io/t2c6h/.

References

Abdulla, A. (2001). Rhetorical repetition in literary translation. *Babel*, 47, pp. 289–303. https://doi.org/10.1075/babel.47.4.02abd

Augustinská, D. (1985). Opakovanie ako výrazový prostriedok vo vecnom texte. *Slovenská reč*, 50, pp. 286–291.

Baker, M. (1993). Corpus linguistics and translation studies: Implications and applications. In M. Baker, G. Francis, E. Tognini-Bonelli (Eds.), *Text and Technology: In honour of John Sinclair* (pp. 233–250). John Benjamins.

Baker, M. (1996). Corpus-based translation studies: The Challenges that lie ahead. In H. Somers (Ed.), *Terminology, LSP and Translation in Language Engineering in Honour of Juan C. Sager* (pp. 175–186). John Benjamins.

Bartoń, K. (2024). *MuMIn: Multi-Model Inference. R package version 1.48.4*. Retrieved January 10, 2025, from https://CRAN.R-project.org/package=MuMIn

Bauer, J., Grepl, M. (1972). *Skladba spisovné češtiny*. Státní pedagogické nakladatelství.

Bečka, J. (1992). *Česká Stylistika*. Academia.

Ben-Ari, N. (1998). The ambivalent case of repetitions in literary translation. *Target*, 10(2), pp. 381–409.

Bentz, C., Winter, B. (2013). Languages with More Second Language Learners Tend to Lose Nominal Case. In S. Wichmann, J. Good (Eds.), *Quantifying Language Dynamics: On the Cutting edge of Areal and Phylogenetic Linguistics* (pp. 96–124). Brill.

Blum-Kulka, S. (1986). Shifts of cohesion and coherence in translation. In S. Blum-Kulka, J. House (Eds.), *Interlingual and intercultural communication* (pp. 17–35). Gunter Narr Verlag.

Brooks, M. E., Kristensen, K., Van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Mächler, M., Bolker, B. M. (2017). Glmmtmb balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal*, 9(2), pp. 378–400.

Caldas-Coulthard, C. R. (1987). Reported speech in written narrative texts. In M. Coulthard (Ed.), *Discussing Discourse* (pp. 149–167). University of Birmingham.

Catford, C., John. (1965). *A Linguistic Theory of Translation: An Essay in Applied Linguistics*. Oxford University Press.

Čermák, F., Rosen, A. (2012). The case of InterCorp, a multilingual parallel corpus. *International Journal of Corpus Linguistics*, 17(3), pp. 411–427.

Čermáková, A. (2015). Repetition in John Irving's novel a widow for one year: A corpus stylistics approach to literary translation. *International Journal of Corpus Linguistics*, 20(3), pp. 355–377.

Coxe, S., West, S. G., Aiken, L. S. (2009). The analysis of count data: A gentle introduction to Poisson regression and its alternatives. *Journal of Personality Assessment*, 91(2), pp. 121–136.

Cvrček, V., Chlumská, L. (2015). Simplification in translated Czech: A new approach to type-token ratio. *Russian Linguistics*, 39(3), pp. 309–325.

Daneš, F. (1973). Verba dicendi a výpovědní funkce. *Studia Slavica Pragensia*, pp. 115–124.

De Baets, P., De Sutter, G. (2022). How do translators select among competing (near-) synonyms in translation? A corpus-based approach using random forest modelling. *Target*, 35(1), pp. 1–33.

De Sutter, G., Lefer, M.-A. (2019). On the need for a new research agenda for corpus-based translation studies: A multi-methodological, multifactorial and interdisciplinary approach. *Perspectives*, 28(1), pp. 1–23.

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), pp. 61–96.

Dupont, M., Zufferey, S. (2017). Methodological issues in the use of directional parallel corpora: A case study of English and French concessive connectives. *International journal of corpus linguistics*, 22(2), pp. 270–297.

Eder, M. (2011). Style-markers in authorship attribution: A cross-language study of the authorial fingerprint. *Studies in Polish Linguistics*, 6, pp. 99–114.

Fárová, L. (2016). Uvozovací slovesa v překladech třech různých jazyků. *Jazykové paralely*, pp. 145–161.

Fellbaum, C. (1998). *Wordnet: An electronic lexical database*. MIT press.

Fialová, E. (2020). Deutsche Verba dicendi und ihre tschechischen Entsprechungen. Eine korpusbasierte translatologische Analyse [Master’s thesis, Masaryk University].

Fox, J., Weisberg, S. (2019). *An R companion to applied regression* (3rd). Sage.

Grabowski, L., Borysowski, D. (2025). Between repetition and lexical variety: How do English-to-Russian translators of literary texts deal with recurrent reporting verbs? *Finnish Journal of Linguistics*, 38, pp. 195–210.

Grabowski, L., Olejniczak, J., Mastropierro, L. (2026). A multifactorial study of English-to-Polish translation of reporting verbs in literary novels: A negative binomial regression with mixed effects [In press]. *Research in Corpus Linguistics*, 15(1).

Grepl, M. (1967). O funkci záměrného opakování částí výpovědi ve výstavbě promluvy. *Naše řeč*, 50(2), pp. 77–87.

Gresty, J. (2012). On wordiness in written production of Slovak users of English. In *English matters III* (pp. 25–28). Prešovská univerzita v Prešove.

Gries, S. (2015). The most under-used statistical method in corpus linguistics: Multi-level (and mixed-effects) models. *Corpora*, 10(1), pp. 95–125.

Grieve, J. (2023). Register variation explains stylometric authorship analysis. *Corpus Linguistics and Linguistic Theory*, 19(1), pp. 47–77.

Gromová, E. (2003). *Teória a didaktika prekladu*. Univerzita Konštantína Filozofa.

Gromová, E., Janecová, E. (2013). Audiovisual translation-dubbing and subtitling in Slovakia. *World Literature Studies*, 5(4), pp. 61–71.

Hair, J., Black, W., Babin, B., Anderson, R. (2009). *Multivariate data analysis* (9th). Pearson.

Hirschová, M. (1982). K některým otázkám reprodukování cizích výpovědí. *Acta Universitatis Palackianae, Studia Bohemica II, Facultas Philosophica Philologica*, 46, pp. 97–102.

Hirschová, M. (1988). *Česká verba dicendi v performativním užití: Příspěvek ke zkoumání komunikativních funkcí výpovědi*. Univerzita Palackého.

Hirschová, M. (2017a). *Verbum dicendi*. Retrieved April 21, 2026, from https://www.czechency.org/slovnik/VERBUM%20DICENDI

Hirschová, M. (2017b). *Verbum sentiendi*. Retrieved April 21, 2026, from https://www.czechency.org/slovnik/VERBUM%20SENTIENDI

Hoey, M. (1991). *Patterns of lexis in text*. Oxford University Press.

Hoey, M. (2005). *Lexical priming: A new theory of words and language*. Routledge.

Hoffmannová, J. (2024). Verba dicendi v současné mluvené češtině. *Prace Filologiczne*, (1), pp. 131–143.

Huddleston, R., Pullum, G., K. (2017). *The Cambridge Grammar of the English Language*. Cambridge University Press.

Jettmarová, Z. (2008). Czech and Slovak translation theories: The lesser-known tradition. In J. Králová, Z. Jettmarová (Eds.), *Tradition versus Modernity. From the Classic Period of the Prague School to Translation Studies at the Beginning of the 21st Century* (pp. 15–46). Charles University in Prague.

Juršic, M., Mozetic, I., Erjavec, T., Lavrac, N. (2010). Lemmagen: Multilingual lemmatisation with induced ripple-down rules. *Journal of Universal Computer Science*, 16(9), pp. 1190–1214.

Kabacoff, R. (2015). *R in action: Data analysis and graphics with R*. Manning Publications Co.

Kajzer-Wietrzny, M., Ivaska, I. (2020). A multivariate approach to lexical diversity in constrained language. *Across Languages and Cultures*, 21(2), pp. 169–194.

Kang, H., Zhang, Y. (2025). Socio-cognitive influences on translating forward causal connectives: A multivariate analysis of English translations of Tao Te Ching. *Lingua*, 314, p. 103872.

Klaudy, K., Károly, K. (2005). Implicitation in translation: Empirical evidence for operational asymmetry in translation. *Across Languages and Cultures*, 6(1), pp. 13–28.

Klinger, S. (2019). Repetition: Translating the interplay between its linguistic form and its literary function. *Babel*, 65(2), pp. 316–332.

Knápek, D. (2019). Funkce a frekvence verb dicendi ve vybraných prózách Karla Čapka [Bachelor’s thesis]. Palacky University in Olomouc.

Kruger, H. (2019). That again: A multivariate analysis of the factors conditioning syntactic explicitness in translated English. *Across Languages and Cultures*, 20(1), pp. 1–33.

Kundera, M. (1998). *The Art of the Novel*. HarperCollins.

Levin, B. (1993). *English verb classes and alternations: A preliminary investigation*. University of Chicago Press.

Levý, J. (1963). *Uměni překladu*. Československý spisovatel.

Levý, J., Corness, P., Jettmarová, Z. (2011). *The Art of Translation*. John Benjamins.

Lewis, F., Butler, A., Gilbert, L. (2011). A unified approach to model selection using the likelihood ratio test. *Methods in Ecology and Evolution*, 2(2), pp. 155–162.

López-Escobedo, F., Méndez-Cruz, C.-F., Sierra, G., Solórzano-Soto, J. (2013). Analysis of stylometric variables in long and short texts. *Procedia-Social and Behavioral Sciences*, 95, pp. 604–611.

Mahlberg, M. (2018). Corpus Stylistics. In V. Sotirova (Ed.), *The Bloomsbury Companion to Stylistics* (pp. 139–156). Bloomsbury.

Mastropierro, L. (2020). The translation of reporting verbs in Italian: The case of the Harry Potter series. *International Journal of Corpus Linguistics*, 25(3), pp. 241–269.

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 *Advances in corpus applications in literary and translation studies* (pp. 138–157). Routledge.

Mastropierro, L., Grabowski, Ł. (2024). Repeated reporting verbs in English novels and their Italian and Polish translations: A preliminary multifactorial study. *Across Languages and Cultures*, 25(2), pp. 310–330.

Miko, F. (1970). Textová výstavba literárneho diela. *Slovenská literatúra*, 17(1), pp. 3–17.

Miko, F. (1978). *Štylistika a textová lingvistika*. SPN.

Mistrík, J. (1993). *Encyklopédia jazykovedy*. Obzor.

Mistrík, J. (2021). *Štylistika*. VEDA.

Nádvorníková, O. (2017). Les proportions des verbes say/dire/říci dans les propositions incises et leurs équivalents en traduction: Étude sur corpus parallèle. *Linguistica Pragensia*, 28(2), pp. 35–57.

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), pp. 209–234.

Nakagawa, S., Johnson, P. C., Schielzeth, H. (2017). The coefficient of determination R^2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of the Royal Society Interface*, 14(134), p. 20170213.

Nemcová, E. (2012). Niekol’ko poznámok k spracovaniu verb dicendi. In K. Buzássyová, B. Chocholová, N. Janočková (Eds.), *Slovo v slovníku Aspekty lexikálnej sémantiky – gramatika – štylistika (pragmatika) Na počesť Alexandry Jarošove* (pp. 192–208). VEDA.

Nida, E., Taber, C. (1982). *The theory and practice of translation*. Brill.

Patáková, M. (1987). Poetizačná a kontaktová funkcia opakovania v autorskej reči. *Slovenská reč*, 52(6), pp. 321–334.

Peprník, J. (1969). Reporting phrases in English prose. *Brno Studies in English*, 8(1), pp. 145–151.

Pisárčiková, M. (1978). Synonymia slovies v uvádzacích vetách. *Slovenská reč*, 43, pp. 210–216.

Podpečan, V. (2024). *Lemmagen3. Python library.* Retrieved December 3, 2024, from https://pypi.org/project/lemmagen3/

Popovič, A. (1975). *Teória umeleckého prekladu: Aspekty textu a literárnej metakomunikácie*. Tatran.

Popovič, A. (1976). *A dictionary for the analysis of literary translation*. University of Alberta Press.

Preislerová, A. (2015). Verba dicendi jako metapragmatický komentář. Slovesa mluvení v uvozovacích větách v beletrii a publicistice na materiálu SYN2010 [Master's thesis, Palacky University].

Pym, A. (2009). *Exploring translation theories*. Routledge.

Quirk, R., Greenbauch, S., Leech, G., Svartvik, J. (1985). *A comprehensive grammar of the English language*. Longman.

Rosen, A., Vavřín, M., Zasina, A. J. (2022). *The InterCorp Corpus – Czech, version 15 of 11 November 2022*. https://kontext.korpus.cz/

Rybicki, J., Heydel, M. (2013). The stylistics and stylometry of collaborative translation: Woolf's night and day in polish. *Literary and Linguistic Computing*, 28(4), pp. 708–717.

Samlerová, L. (2010). Metaforický charakter verb dicendi [Master's thesis, Technická univerzita v Liberci.].

Scherber, C. (2017). Using R to interpret interaction effects in statistical models. *Software Developer's Journal*. 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*. http://www.christoph-scherber.de/content/PDF%20Files/Generalized%20linear%20models.pdf

Schultze, B. (2015). Jiří Levý's contribution to translation studies as represented in the De Gruyter Encyclopedia Übersetzung, Translation, Traduction. *Acta Universitatis Carolinae Philologica*, 3, pp. 105–112.

Searle, J. R. (1976). A classification of illocutionary acts. *Language in Society*, 5(1), pp. 1–23.

Semino, E., Short, M. (2004). *Corpus stylistics: Speech, writing and thought presentation in a corpus of English writing*. Routledge.

Seroussi, Y., Zukerman, I., Bohnert, F. (2014). Authorship attribution with topic models. *Computational Linguistics*, 40(2), pp. 269–310.

Škorić, M., Stanković, R., Ikonić Nešić, M., Byszuk, J., Eder, M. (2022). Parallel stylometric document embeddings with deep learning based language models in literary authorship attribution. *Mathematics*, 10(5), p. 838.

Šoltys, O. (1983). *Verba dicendi a metajazyková informace*. Československá akademie věd, Ústav pro jazyk český.

Staroňová, K. (2023). Operational Asymmetry in Translation of English and Slovak Reporting Verbs [Master's thesis, Masaryk University].

Svobodová, J. (2007). Sémantika sloves mluvení v uvozovacích větách uměleckého textu. *Studia Bohemica*, 10, pp. 153–160.

Thompson, G., Yiyun, Y. (1991). Evaluation in the reporting verbs used in academic papers. *Applied linguistics*, 12(4), pp. 365–382.

Venables, B., Ripley, B. (2002). *Modern Applied Statistics with S* (4th). Springer.

Vilikovský, J. (1984). *Preklad ako tvorba*. Slovenský spisovateľ.

Wang, G., Xin, Y. (2024). An analytical framework for corpus-based translation studies. *Humanities and Social Sciences Communications*, 11(1), p. 1709.

Winter, B. (2019). *Statistics for linguists: An introduction using R*. Routledge.

Winter, B., Bürkner, P.-C. (2021). Poisson regression for linguists: A tutorial introduction to modelling count data with brms. *Language and Linguistics Compass*, 15(11), e12439.

Zufferey, S. (2016). Discourse connectives across languages: Factors influencing their explicit or implicit translation. *Languages in Contrast*, 16(2), pp. 264–279.

Zupan, S. (2006). Repetition and translation shifts. *ELOPE: English Language Overseas Perspectives and Enquiries*, 3(1–2), pp. 257–268.