
Translation Direction in the *Gesta regum Sclavorum*: A Quantitative and Qualitative Approach

Dragana Kunčer¹[\[0000-0002-4688-4775\]](#)

¹ Institute of History, Belgrade, Serbia, dragana.kuncer@iib.ac.rs

Scientific paper

DOI: 10.18485/judig.2026.2.ch24

Abstract

The paper examines the claim that the Latin chronicle *Gesta regum Sclavorum*, a text concerning the medieval history of the South Slavs, was not originally composed in Latin but was translated from Italian into Latin as part of a deliberate forgery. The question of translation direction between the two versions is approached through quantitative, philological, and pragmatic analyses. The quantitative analysis is based on concepts such as translation universals, laws of translation, and translationese, according to which translated texts may be distinguished from texts originally composed in the target language by certain linguistic features. The hypotheses concerning the distinctive features of translated texts were operationalised for quantitative analysis. The analysis was performed on preprocessed versions of the two texts using Python scripts generated with the assistance of ChatGPT from prompts designed for this purpose. The philological analysis was conducted through a manual comparison of the parallel texts. The pragmatic analysis was based on Ivan Lučić's comments on the content of the chronicle, as expressed both in his historiographical work *De regno Dalmatiae et Croatiae libri sex* and in letters to his friends. Lučić was the first editor of the Latin text and, according to the theory under examination, is presumed to have been its translator and one of the two figures responsible for the alleged forgery.

Keywords: translation direction, translation universals, translationese, *Gesta regum Sclavorum*, spaCy, LatinCy, Python scripts

1. Introduction

The chronicle *Gesta regum Sclavorum* (GRS), also known as *The Chronicle of the Priest of Dioclea*, is a source for the medieval

history of the Balkans. It describes events associated with the dynasty of the South Slavic kings who are said to have ruled along the eastern Adriatic coast and its hinterland, covering the period from the end of the fifth century to the second half of the twelfth century. The work was first printed in Italian in the early seventeenth century, as an appendix to Mavro Orbin’s historiographical work *Il regno de gli Slavi* (Orbin, 1601). The Latin version was first published sixty-five years later, as an appendix to Ivan Lučić’s *De regno Dalmatiae et Croatiae libri sex* (Lučić, 1666). Until recently, the GRS was generally regarded in historiography as an authentic medieval work (Papageorgiou, 2012: 14). In 2008, however, Solange Bujan proposed that it was in fact a modern forgery (Bujan, 2008).

Bujan argues that the earlier author, Orbin, first composed the original text in Italian, drawing on numerous sources available to him, and then falsely presented it in his book as a translation from Latin into Italian. According to her theory, a second author, Lučić, later translated Orbin’s Italian text into Latin and falsely presented it as an original Latin text. She further maintains that the manuscript now preserved in the National Library of Serbia (R–570), the so-called Belgrade manuscript, is in fact the first “draft” of the translation from Italian. This, in her view, is shown by the fact that the manuscript is “dependent on the Italian version”, for which she provides three examples in total (Table 1).

Table 1: Parallel excerpts from the GRS manuscripts and Orbin.

GRS (<i>Cod. Vat.</i>)	GRS (<i>Cod. Belgr.</i>)	Orbin
Replevit multitudine Sclavorum, & quievit terra in diebus illis	Replevit, et terram multitudine Sclavorum, et quievit terra in diebus illis	Peroche empi la terra di molto Slavi, e a di suoi il paese si riposo
maximè à iuvenibus nostrae Civitatis	maximè vero à iuvenibus nostrae Civitatis	ma principalmente da giovani della nostra città
in Sylloduxiam Provinciam	venerunt Syllodusiam Provinciam	vennero nella provincia Silodusia

If Bujan’s examples are examined, it can indeed be seen that Orbin’s Italian version contains words that are present in the Latin text of the Belgrade manuscript but absent from the Vatican

manuscript: *terram, vero, and venerunt*. It is not clear, however, why this should indicate the direction of translation. The presence of these additional words in both the Belgrade manuscript and Orbin's Italian version does not, in itself, establish the direction of translation.

It should also be noted that Bujan formulated this argument before the publication of the critical edition of the GRS (Kunčer, 2009), in which the so-called Belgrade manuscript was systematically examined alongside the Vatican manuscript for the first time. This edition shows, among other things, that the Belgrade manuscript cannot have been the exemplar of the Vatican manuscript, since it contains numerous lacunae in comparison with it. In addition, when compared with the Vatican manuscript, the Belgrade manuscript displays a greater number of lexical or syntagmatic forms with inadequate or nonsensical meanings.

The aim of this case study is therefore to subject Bujan's argument to systematic scrutiny and, by applying both quantitative and qualitative methods, to assess the validity of her theory. The two hypotheses advanced by Bujan—one concerning the falsification of the content of the GRS, and the other concerning the falsification of its linguistic form—do not receive equal attention in her work. The first hypothesis, which concerns the creation of the content of the GRS by Orbin, constitutes the greater part of her study. By contrast, the second hypothesis, which concerns Lučić's hypothetical translation of Orbin's work from Italian into Latin, occupies only a few pages in her first article (Bujan, 2008) and none in the second (Bujan, 2011). Given the importance of this second hypothesis, the limited attention devoted to it seems unjustified. If this hypothesis is refuted—that is, if it can be confirmed that the Latin text is the original and not a translation from Italian, and that the reverse is true, namely that the Italian text is a translation from Latin, as was previously believed—then Bujan's entire theory would be undermined. For this reason, the present study deals only with assessing the validity of the second hypothesis.¹

¹ This paper presents an abridged English version of the argument developed in the author's Serbian-language monograph *Primena korpusne lingvistike u istoriografiji (O autentičnosti Gesta regum Sclavorum)* (Kunčer, In press). The discussion of different lexical-diversity measures and the section "Further Research" are new to the present publication.

Considered more closely, this second hypothesis reveals that Bujan proposes a non-traditional understanding of the direction of translation between the two versions of the text. Traditionally, the Latin text has been regarded as the source text, whereas Bujan’s hypothesis classifies it as the target text. Conversely, the Italian text, traditionally regarded as the target text, is classified by Bujan as the source text. The validity of Bujan’s principal argument, and indeed of her entire theory, therefore depends on establishing the actual direction of translation between the two versions, that is, on clearly identifying which text is the original and which is the translation.

This question is approached here through quantitative, philological, and pragmatic analyses.

2. Quantitative Analysis

2.1. Translation Universals

The study draws on the theoretical concepts of translation universals (Baker, 1993; Laviosa, 2009), laws of translation (Toury, 1995 [2012]), and translationese (Gellerstam, 1986; Volansky et al., 2015; Rabinovich and Wintner, 2015). These approaches rest on the assumption that translated texts in a given language differ in certain respects from texts originally composed in that language. It has been proposed that translations tend towards simplification and explicitation, that they adhere more closely to linguistic norms than original texts, that they differ less from one another than original texts do among themselves, and, finally, that the language of the source text—both lexically and structurally—“shines through” in the translated text (Dutta Chowdhury, 2024).

For the quantitative analysis in the present study, only the simplification and explicitation hypotheses are tested.

2.1.1. Simplification Hypothesis

Within the simplification hypothesis, translated texts are assumed to be simpler than texts originally written in the target language, with simplification potentially occurring at the lexical, syntactic, and stylistic levels. Lexical simplification has been described as “the process and/or result of making do with less words” (Blum and Levenston, 1978: 399) and may involve, among other strategies, the use of more general or more frequent words and paraphrases.

At the syntactic and stylistic levels, proposed tendencies include the shortening of complex structures, the reduction of repetition and redundant information, and the omission of modifiers (Baker, 1993; Laviosa, 2009).

The simplification hypothesis has been operationalised for quantitative analysis in a number of corpus-based studies, among other ways, through measures such as lexical diversity, mean sentence length, and lexical density, understood here as the content-to-function-word ratio (Laviosa, 1998; Corpas Pastor et al., 2008; Lembersky et al., 2012; Volansky et al., 2015).

2.1.2. Explicitation Hypothesis

Within the explicitation hypothesis, information that is implicit or underspecified in the source text is assumed to become more explicit in translation, usually through the addition of lexical material (Blum-Kulka, 1986). Proposed strategies include expanding condensed expressions, adding cohesive markers, introducing descriptive modifiers, specifying pronouns, names, and concepts more precisely, and repeating information already given in the preceding context (Baker, 1993; Laviosa, 2009).

The explicitation hypothesis was operationalized, among other ways, through the ratio of personal pronouns to proper nouns (Volansky et al., 2015).

2.2. Methodology

Although hypotheses concerning translation universals are generally tested by comparing translated texts with texts originally composed in the target language, using so-called comparable corpora, they were originally conceptualised on the basis of comparisons between translations and their source texts, that is, using parallel corpora (Laviosa-Braithwaite, 1998: 288). Since the GRS corpus is itself a parallel corpus, comprising two versions of the same text, one of which is a translation and the other its source, the present study proceeds from the assumption that the textual features described by these hypotheses may also be identified in the translated version within this corpus and may therefore indicate the most probable direction of translation.

Here, the simplification and explicitation hypotheses are operationalised in the following way:

1. the lexical diversity of a translated text is lower than that of the source text;
2. the mean sentence length in a translation is greater than that in the source text;
3. the ratio of content words to function words in a translation is lower than in the source text;
4. the ratio of personal names to person-referring pronouns is higher in the translation than in the source text.

The measurements were carried out by means of Python scripts that used the spaCy NLP library ([Honnibal et al., 2020](#)) to process the Italian text and the LatinCy language models ([Burns, 2023](#)) to process the Latin text. The scripts were written on the basis of prompts provided to ChatGPT ([OpenAI, 2026](#)). The files and scripts used in the quantitative analysis are available on Zenodo ([Kunčer, 2026](#)).

To enable quantitative analysis, both texts were converted into machine-readable plain-text (.txt) format. Of the two texts compared, the GRS survives in manuscript form, whereas Orbin’s version was transmitted in print.

For the GRS, the digital text of the Kunčer (2009) critical edition was used. Since this edition incorporates readings from both the Vatican and Belgrade manuscripts, the Vatican readings were manually restored where necessary. This was done because Bujan’s theory concerns Lučić’s edition, which was based solely on the Vatican manuscript; accordingly, the GRS text analysed here corresponds to that manuscript. Orbin’s text was manually transcribed from digital images of the 1601 printed edition into plain-text format.

3. Results

Lexical Diversity (TTR, Type-token ratio) is given in Table 2. TTR is sensitive to text length: “As a text becomes longer (i.e., by adding words to it), the number of overall words in the text (or tokens) increases. But although the token increase is linear (one new word = one new token), the rate of increase of the number of different words in the text (or types) steadily slows” ([McCarthy and Jarvis, 2010: 382](#)).

Table 2: Lexical diversity statistics for the GRS and Orbin corpora.

	Measure	GRS	Orbin
	Total number of tokens	12,077	13,858
	Number of distinct lemmas	1,645	1,667
(1)	Lexical diversity (TTR)	0.136	0.120

To assess the possible effect of this factor, the result obtained in this study was checked against additional measures of lexical diversity: Guiraud’s R, Herdan’s C, Maas’s measure, HD-D with the parameter set at 42, the Moving-Average Type-Token Ratio (MATTR) with windows of 100, 500, and 1,000 tokens, the Mean Segmental Type-Token Ratio (MSTTR) with segments of 100, 500, and 1,000 tokens, and the Measure of Textual Lexical Diversity (MTLD), with a threshold of 0.72. Special attention was paid to HD-D, MATTR, and MTLD, which have been discussed as measures less sensitive to differences in text length (Bestgen, 2024). These three control measures, together with the remaining measures, yielded the same direction of difference as TTR.

Moreover, since the present study compares lexical diversity (Table 3) in two typologically different languages, it can be represented more precisely by the ratio of the number of lemmas to the total number of tokens than by the ratio of distinct word forms to the total number of tokens (Treffers-Daller et al., 2018). Strictly speaking, the corresponding abbreviation would be LTR; here, however, TTR is retained in accordance with common practice, with the term type understood to refer to a lemma rather than an individual word form.

Table 3: Comparison of linguistic features: mean sentence length (2), lexical density (3), and personal name-to-pronoun ratio (4) in the GRS and Orbin corpora.

	Measure	GRS	Orbin
(2)	Mean sentence length (words)	18.66	24.24
	Content words	7,347 (60.74%)	7,203 (51.98%)
(3)	Function words	4,748 (39.26%)	6,654 (48.02%)
	Personal names	558 (44.15%)	577 (44.35%)
(4)	Person-referring pronouns	706 (55.85%)	724 (55.65%)

The difference in the fourth measure is very small and should therefore be interpreted cautiously. Taken together, however, these quantitative results are consistent with the traditional theory of the direction of translation between the two texts, according to which the Italian text is the translation and the Latin text the source text.

4. Philological Analysis

The philological analysis was based on a comparison of the two textual versions, Italian and Latin, translation unit by translation unit.

a) It showed that, in a number of passages, the Latin version contains information absent from the Italian version and not recoverable from the surrounding Italian context (Table 4).

Conversely, certain passages present in the Italian version are absent from the Latin version, although these are fewer in number (Table 5).

The additional information preserved in the Latin text is interpreted as material belonging to the original version, or source text, which was omitted in the course of translation into Italian; it therefore provides an argument in favour of the traditional theory. By contrast, the additional information found in the Italian text, which might at first sight be used to support Bujan's theory in the same way, must be considered in light of several other facts. On pages 208, 213, 215, 222, and 233, Orbin includes marginal notes that may indicate that he had access to other exemplars of the text (e.g. *altri esemplari dicono*, *altri esemplari hanno*). If this is the case, the additional material found in his Italian version may derive from one or more manuscripts containing these passages that were not contained in the manuscript or manuscripts used by Lučić for his edition.

The same analysis also identified a number of patterns in the Italian text that would be consistent with its character as a translation.

b) For example, in several passages, specific expressions in the Latin text correspond to more general expressions in the Italian version. This tendency towards generalisation may be interpreted in light of the simplification hypothesis. The parallel passages in the Italian and Latin versions that illustrate this pattern are given in Table 6.

Table 4: Parallel passages, examples with absent information from the Italian version.

Orbin	GRS
<p>[...] Del che auisato egli e fatto un buon apparato di gente, venne e assaltò il campo Greco. Doue molti Greci caderono, e altri furono presi viui.</p> <p>Dopo questo nacquero à Sebeslau due figliuoli gemelli [...]</p>	<p>[...] Audiens autem Sebeslavus rex congregavit gentem multam et venit et irruit in castra eorum et ceciderunt Graeci et multi gladio perierunt et plurimi capti sunt, alii vero in fugam conversi sunt.</p> <p><i>Dum haec agerentur Attila, rex Hungarorum, cum exercitu intravit in terram Sebeslavi regis et depraedavit, destruxit atque incendit maximam partem terrae illius et reversus est in locum suum. Inde abiit contra occidentalem plagam. Sebeslavus autem rex reversus non invenit eum coepitque recuperare terram suam.</i></p> <p>Nati sunt ei duo filii gemini [...]</p>
<p>[...] La qual cosa concesse Iddio à lui morto, per dimostrare, che egli innocente, e in verità martire era passato da questa vita.</p> <p>Il che vedendo Vuladislauo, se ne sbigottì molto. Onde concesse alla sua moglie, qual dì, e notte senza cessar mai, lo piangeua, che leuasse il suo corpo, e lo sepelisse doue le piaceua. [...]</p>	<p>[...] Ut autem Dominus declararet merita beati martyris Vladimiri, multi diversis languoribus vexati intrantes ecclesiam orantesque ad eius tumulum sanati sunt.</p> <p><i>Nocte vero videbatur ibi ab omnibus lumen divinum et quasi plurimas ardere candelas.</i></p> <p>Uxor vero beati Vladimiri flevit cum fletu magno ultra quam dici potest diebus multis. Videns autem imperator mirabilia, quae ibi Deus operaretur, paenitentia ductus satis timuit consequenterque consobrinae suae tollere corpus eius et sepelire honorifice quocumque vellet. [...]</p>
<p>[...] E così l’huomo pessimo, il quale assentatosi prima sul desinare, haueua fatto morire il Beato Vuladimiro, poscia cennando, fu egli morto da l’Angelo del Signore.</p> <p>La onde Draghimir zio di Vuladimiro udendo questo, amassò molta gente, per ricuperare il Regno de’ suoi Aui, che à lui di ragione atteneua. [...]</p>	<p>[...] Sicque factum est ut nequissimus homicida, qui sedens ad prandium beatum Vladimirim decollari iusserat et martyrem fecerat, ipse hora cenae percuteretur ut angelus Satanae efficeretur.</p> <p><i>Quantas et quales virtutes et prodigia Deus operare dignatus est per beatum Vladimirim, famulum suum, qui scire desiderat, librum gestorum eius relegat quo acta eius per ordinem scripta sunt et agnoscat profecto quod ipse vir sanctus unus spiritus cum Domino fuit et Deus habitavit cum eo, cui honor etc.</i></p> <p>Praeterea Dragimirus, frater beati Vladimiri, audiens mortem imperatoris congregavit populum et exercitum ut terram et regnum patrum suorum sibi vindicaret. [...]</p>

Table 5: Parallel passages, examples with absent information from the Latin version.

Orbin	GRS
<p>[...] Né contentandosi di questa sola ambasciaria, da tutto il Regno concoreua gran moltitudine per vedendo, e principalmente gli habitatori della contrada di Trebine.</p> <p><i>Paulimiro fabricato adunque il castello, e postoui dentro buona guardia con il restante d'Epidaurio si pose in viaggio.</i></p> <p>E primieramente venne à Trebine, doue fu riceuuto con gran honore: e dimorando quiui per alcun tempo, tutti li Principi, e Baroni del suo Regno vennero à lui; [...]</p>	<p>[...] Coeperunt undique ad eum confluere, imprimis habitatores regionis Tribuniae, venientes cum magno honore duxerunt illum in Tribuniam.</p> <p>Postea bani venientes in Tribuniam cum iupanis et satnicis susceperunt illum honorifice et in die ascensionis Domini constituerunt illum regem. [...]</p>

Table 6: Parallel passages illustrating more general expressions in the Italian.

Orbin	GRS
[...] attendeua all'orationi, <i>et ad altre diuerse macerationi del corpo</i> [...]	[...] vacabat orationibus <i>atque ieuniis</i> [...]
[...] sonando trombe, <i>et altri instrumenti da guerra</i> [...]	[...] coeperunt hinc inde <i>buccinas et tubas</i> clangere et vociferare [...]
[...] facesse morire Branislauo <i>con gli altri, che teneua prigioni</i> [...]	[...] iussit decollari kinesum Branislavum <i>cum fratre et filio</i> [...]
[...] e trouatasi gran abbondanza <i>di tutte le cose pertinenti al vitto humano</i> [...]	[...] Dedit enim in diebus illis Deus abundantiam <i>frumenti et vini</i> [...]

c) Several passages in the Italian version explicitly define the family relationships between characters; these may be understood as instances of explicitation. Some of these passages are given in Table 7.

Table 7: Parallel passages with instances of explicitation.

Orbin	GRS
<p>[...] Perciò mi paressere ben fatto, <i>che Michaglia, Predemir, e Sagance, miei figliuoli</i> vadino con una parte dell'essercito [...]</p> <p>[...] <i>Et io con gli altri figliuoli Goislauo</i>, e il restante dell'essercito, <i>restaremo</i> qui [...]</p>	<p>[...] Faciamus ergo sic: <i>remaneant duo ex vobis, Goyslavus et Radoslavus</i>, ceteri vero tres accipite unusquisque X strenuos [...]</p>

These passages form part of King Dobroslav’s orders to his sons at a time of grave military danger. In the Italian text, the king addresses them by name but also adds the apposition *miei figliuoli* (*Michaglia, Predemir, e Sagance, miei figliuoli*), a clarification that appears pragmatically redundant in this context. The formulation *Et io con gli altri figliuoli Goislauo* is similarly incongruous, since the plural is used although only one son is named. By contrast, the Latin text reads more naturally, referring to the sons by name or addressing them directly in the second person, without explicit reference to their family relationship.

d) The Italian text also contains a number of syntactic and semantic anomalies that appear to reflect errors arising in translation from Latin into Italian. Some of these passages are presented in Table 8.

Table 8: Parallel passages with syntactic and semantic anomalies.

Orbin	GRS
[...] e comandò che <i>tutto il suo Regno</i> si congregasse in quella medesima campagna di Dalma. Et mentre stauano <i>questi</i> à radunarsi [...]	[...] Tunc rex iussit congregari in eadem planitie Dalmae omnes <i>populos</i> terrae et regni sui. Inter haec, dum <i>populi</i> congregarentur [...]
[...] e in quella <i>pose il Vescouo, che fino al dì d’hoggi dura</i> [...]	[...] Ecclesiam autem supradictam statuit episcopatum fieri <i>ordinavitque ibi episcopum et episcopatum usque in praesentem diem</i> [...]

In the first passage, there is no plural noun in the Italian text that could function as the antecedent of the pronoun *questi*, since the only potential antecedent is the singular phrase *tutto il suo Regno*. The use of *questi* may therefore be interpreted as a trace of the source (Latin) text, in which the corresponding noun occurs in the plural (*populi*).

In the second passage, the Latin text yields a semantically coherent sentence: it refers to an ecclesiastical institution (the bishopric) that “still endures (to the present day)”. By contrast, the Italian text states that the “bishop” appointed by the king “still endures”, which results in a semantically problematic statement. This passage therefore appears to reveal another inadvertent error in the translation from Latin into Italian.

e) Finally, an etymological wordplay is identified that further supports the traditional theory. This passage is given in Table 9.

Table 9: Parallel passages with etymological wordplay examples.

Orbin	GRS
[...] Onde Paulimiro fatto già grandetto marauigliosamente si portaua in quei tumulti; che non solo alli suoi parenti, ma ancora à gli altri Romani per il suo valore e prudenza era molto caro, <i>e come valente guerriero, era chiamato Bello</i> [...]	[...] Paulomirus iam iuuenis effectus coepit esse valde robustus et fortis bellator ita ut in civitate Romana nullus ei esset similis. Unde parentes eius, nec non et alii Romani, coeperunt eum valde diligere immutaveruntque <i>nomen eius et imposuerunt ei nomen Bello, eo quod bellum facere valde delectabatur</i> [...]

This wordplay, which associates the name of the brave warrior Belo (Bello) with the Latin word *bellum*, “war”, can be understood only in Latin.

5. Pragmatic Analysis

The pragmatic analysis focused on Lučić’s commentary on the GRS. It showed that Lučić, both in his own historiographical work and in his letters to friends, repeatedly emphasised the unreliability of the chronicle’s content. He described the text, for example, as “rich in fables” (*fabulosum itidem est, quod Diocletianus Presbyter refert*) (Lučić, 1666: 39), and warned a historian friend that, if he wished to use this source, “he would not follow the right path” (*va fuori della buona strada*) (Poparić, 1905: 304). Elsewhere, after describing the considerable effort he had devoted to comparing different versions of the text, Lučić stated that he did not intend to make any use of them (*delle quali ... [dopo] tanta fatica, non mi voglio valere in nulla*) (Poparić, 1905: 301). This, too, may be interpreted as an argument in favour of the traditional theory, since Lučić’s open rejection of the work’s historical reliability is difficult to reconcile with the role of forger attributed to him by Bujan.

6. Discussion

Taken together, the quantitative, philological, and pragmatic analyses all support the traditional theory: the relationship between the two versions is best explained as a translation from Latin into Italian, rather than from Italian into Latin, thereby strengthening the case for the medieval origin, and hence the authenticity, of the Latin text.

7. Further Research

The research could be further developed by formulating additional hypotheses concerning translation universals and operationalising them for quantitative analysis in the same manner as the hypotheses examined in the present study.

Another avenue might involve neural-network training. Such an approach would seem to depend on the availability of a sufficiently large and balanced corpus of reliably identified Latin–Italian translation pairs in both directions, preferably with the Italian texts dating from the early seventeenth century and with texts in both languages belonging to the historiographical genre.

Another line of research, based on the methodology employed in this study, would involve testing the claim made in the GRS itself that the chronicle is a translation from a Slavic language into Latin. Since the extant Slavic version of the chronicle cannot be the presumed original—it is considerably shorter and has a different ending (Kunčer, 2009)—such research would have to await the discovery of a Slavic text corresponding closely in content to the Latin version. Quantitative analyses of the kind employed in this study could then be applied.

Language Resources

Kunčer, Dragana. 2026. *The Latin and Italian Versions of the Gesta regum Sclavorum: Preprocessed Texts and Corpus Analysis Scripts*. Zenodo. Language resource.

OpenAI. 2026. *ChatGPT (GPT-5.6 Sol, July 2026)*. OpenAI. Accessed in 2026.

References

Mona Baker. 1993. *Corpus Linguistics and Translation Studies—Implications and Applications*. In Mona Baker, Gill Francis, and Elena Tognini-Bonelli, editors, *Text and Technology: In Honour of John Sinclair*, pages 233–250. Amsterdam.

Yves Bestgen. 2024. *Measuring Lexical Diversity in Texts: The Twofold Length Problem*. *Language Learning*, 74(3):638–671.

- Shoshana Blum and Edward A. Levenston. 1978. [Universals of Lexical Simplification](#). *Language Learning*, 28(2):399–415.
- Shoshana Blum-Kulka. 1986. Shifts of Cohesion and Coherence in Translation. In Juliane House and Shoshana Blum-Kulka, editors, *Interlingual and Intercultural Communication: Discourse and Cognition in Translation and Second Language Acquisition Studies*, volume 35, pages 17–35. Tübingen.
- Solange Bujan. 2008. [La Chronique du prêtre de Dioclée: Un faux document historique](#). *Revue des études byzantines*, 66:5–38.
- Solange Bujan. 2011. [Orbinijevo izdanje „Ljetopisa popa Dukljanina”: povijesni falsifikat](#). *Radovi Zavoda za hrvatsku povijest*, 43(1):65–80.
- Patrick J. Burns. 2023. [LatinCy: Synthetic Trained Pipelines for Latin NLP](#).
- Gloria Corpas Pastor, Ruslan Mitkov, Naveed Afzal, and Viktor Pekar. 2008. [Translation Universals: Do They Exist? A Corpus-Based NLP Study of Convergence and Simplification](#). In *Proceedings of the 8th Conference of the Association for Machine Translation in the Americas: Research Papers*, pages 75–81, Waikiki, USA. Association for Machine Translation in the Americas.
- Koel Dutta Chowdhury. 2024. [A Representation Learning-Based Approach to the Study of Translationese](#). Ph.D. thesis, Saarland University, Saarbrücken.
- Martin Gellerstam. 1986. Translationese in Swedish Novels Translated from English. In Lars Wollin and Hans Lindquist, editors, *Translation Studies in Scandinavia: Proceedings from the Scandinavian Symposium on Translation Theory (SSOTT) II*, pages 88–95. Lund.
- Matthew Honnibal, Ines Montani, Sofie Van Landeghem, and Adriane Boyd. 2020. [spaCy: Industrial-Strength Natural Language Processing in Python](#).
- Dragana Kunčer, editor. 2009. *Gesta regum Sclavorum*, volume 1. Institute of History and Ostrog Monastery, Belgrade and Nikšić. Edited and translated by Dragana Kunčer.
- Dragana Kunčer. In press. *Primena korpusne lingvistike u istoriografiji (o autentičnosti Gesta regum Sclavorum)*.

- Sara Laviosa. 1998. Core Patterns of Lexical Use in a Comparable Corpus of English Narrative Prose. *Meta*, 43(4):557–570. ISBN: 9780415369305.
- Sara Laviosa. 2009. Universals. In Mona Baker and Gabriela Saldanha, editors, *Routledge Encyclopedia of Translation Studies*, 2 edition, pages 306–310. London.
- Sara Laviosa-Braithwaite. 1998. Universals of Translation. In Mona Baker, editor, *Routledge Encyclopedia of Translation Studies*, pages 288–291. London.
- Gennadi Lembersky, Noam Ordan, and Shuly Wintner. 2012. [Language Models for Machine Translation: Original vs. Translated Texts](#). *Computational Linguistics*, 38(4):799–825.
- Ivan Lučić. 1666. *De regno Dalmatiae et Croatiae libri sex*. Amsterdam.
- Philip M. McCarthy and Scott Jarvis. 2010. [MTLD, vocd-D, and HD-D: A Validation Study of Sophisticated Approaches to Lexical Diversity Assessment](#). *Behavior Research Methods*, 42(2):381–392.
- Mavro Orbin. 1601. *Il regno de gli Slavi*. Pesaro.
- Angeliki Papageorgiou. 2012. *To Xρονικό του Ιερέα της Διόκλειας*. Athens.
- Bare Poparić. 1905. Pisma Ivana Lučića Trogiranina. *Starine*, 31:276–320.
- Ella Rabinovich and Shuly Wintner. 2015. [Unsupervised Identification of Translationese](#). *Transactions of the Association for Computational Linguistics*, 3:419–432.
- Gideon Toury. 1995 [2012]. *Descriptive Translation Studies—and Beyond*, revised edition. Amsterdam and Philadelphia.
- Jeanine Treffers-Daller, Patrick Parslow, and Shirley Williams. 2018. [Back to Basics: How Measures of Lexical Diversity Can Help Discriminate between CEFR Levels](#). *Applied Linguistics*, 39(3):302–327.
- Vered Volansky, Noam Ordan, and Shuly Wintner. 2015. [On the Features of Translationese](#). *Digital Scholarship in the Humanities*, 30(1):98–118.

Смер превода *Gesta regum Sclavorum*: квантитативни и квалитативни приступ

Драгана Кунчер

Сажетак

У раду се испитује теорија Соланж Бујан, према којој латинска хроника *Gesta regum Sclavorum* (GRS), дело од значаја за средњовековну историју Јужних Словена, није изворно написана на латинском. Бујан сматра да хроника представља превод с италијанског на латински, објављен 1666. године у делу историчара Ивана Лучића, као и другу фазу кривотворења текста. Прва фаза била би састављање хронике на италијанском, објављене 1601. године од стране историчара Мавра Орбина. У раду се, међутим, полази од тога да је за проверу ове теорије довољно испитати хипотезу која се тиче друге фазе претпостављеног кривотворења, односно одредити смер превода између две верзије: уколико анализа потврди да је латински текст оригинал, а не превод, а да је, напротив, италијански текст превод, а не оригинал, самим тим хипотеза о првој фази, односно састављању хронике на италијанском, постаје неодржива. Како би се дошло до могућег решења проблема, урађена је квантитативна, филолошка и прагматичка анализа.

Квантитативна анализа полази од појмова као што су универзалије преводјења, закони преводјења и преводилачки језик (*translationese*), односно од претпоставке да се преведени текстови по појединим језичким особинама разликују од оригинално написаних текстова. Теоријске претпоставке о особинама преведеног текста операционализоване су тако да могу бити испитане квантитативно. На основу прегледа сличних истраживања издвојене су оне квантитативне мере које су се показале као могући показатељи преведених текстова: 1) лексичка разноврсност у преводу мања је од оне у оригинално написаном тексту; 2) просечна дужина реченица у преведеном тексту већа је од оне у оригинално написаном тексту; 3) лексичка густина, односно однос пунозначних и функцијских речи, у преведеном тексту мања је од оне у оригиналном тексту; 4) однос личних имена и заменица које се односе на лица већи је у преведеном тексту од оног у оригиналном тексту.

Текстови су претходно припремљени за рачунарску обраду, а мерења су извршена помоћу Python скрипти које су за италијански

текст користиле библиотеку spaCy, а за латински текст језичке моделе LatinCy. Скрипте су написане уз помоћ ChatGPT-а, на основу специјално дизајнираних промптова. Резултати прве три мере иду у прилог традиционалној теорији. Што се тиче четврте мере, разлика у односу личних имена и заменица које се односе на лица веома је мала, па њено тумачење захтева опрезност.

Филолошка анализа упоредних места у два текста показује, између осталог, да латински текст на више места садржи додатни материјал који се не може реконструисати на основу италијанске верзије. То се тумачи као последица поједностављивања у преводу с латинског на италијански, у којем су та места изостављена. С друге стране, примећено је постојање додатног материјала и у италијанској верзији, који се не може реконструисати на основу латинске верзије. Иако се ова појава, на први поглед, може на сличан начин тумачити као аргумент у прилог теорији Бујан, то јест као последица скраћивања италијанског текста у преводу на латински, она се ипак тумачи као аргумент у прилог традиционалној теорији. Наиме, сматра се да је тај материјал унет због тога што је Орбин за превод на италијански користио више латинских рукописа и да се налазио у рукопису који Лучић није познавао. Филолошка анализа такође показује да се у италијанској верзији јављају уопштавања, појашњавања породичних односа, синтаксичке и семантичке необичности, као и игра речи која је разумљива само на латинском. Ове појаве се најлакше објашњавају као последице превода с латинског на италијански.

Прагматичка анализа Лучићевих коментара показује да је он отворено доводио у питање историјску поузданост хронике, што је у нескладу са улогом кривотворитеља која му је приписана у теорији Бујан, те се може тумачити као аргумент у прилог традиционалној теорији.

Посматрани заједно, резултати квантитативне, филолошке и прагматичке анализе одговарају схватању да је италијанска верзија превод с латинског, а не обротно, чиме се додатно оснажује теза о средњовековном пореклу латинске верзије *Gesta regum Sclavorum*, а самим тим и о њеној аутентичности.

Кључне речи: смер превода, универзалије превођења, преводачки језик, *Gesta regum Sclavorum*, spaCy, LatinCy, Python скрипте

Works are licensed under the terms and conditions of the Creative Commons Attribution 4.0 International License.

<https://creativecommons.org/licenses/by/4.0/>

