Balcania et Slavia

Studi linguistici | Studies in Linguistics

e-ISSN 2785-3187

Vol. 4 – Num. 2 December 2024

Edizioni
Ca'Foscari

e-ISSN 2785-3187

Balcania et Slavia

Studi linguistici | Studies in Linguistics

Edizioni Ca' Foscari - Venice University Press
Fondazione Università Ca' Foscari
Dorsoduro 3246, 30123 Venezia
URL https://edizionicafoscari.unive.it/en/edizioni4/riviste/balcania-et-slavia/

Balcania et Slavia
Studi linguistici | Studies in Linguistics

Semestral journal

Head office Università Ca' Foscari Venezia | Dipartimento di Studi Linguistici e Culturali Comparati | Ca' Bernardo | Dorsoduro 3199, 30123 Venezia, Italia | balcaniaetslavia@unive.it

Publisher Edizioni Ca' Foscari | Fondazione Università Ca' Foscari | Dorsoduro 3246, 30123 Venezia, Italia | ecf@unive.it

Certificazione scientifica delle Opere pubblicate da Edizioni Ca' Foscari: tutti i saggi pubblicati hanno ottenuto il parere favorevole da parte di valutatori esperti della materia, attraverso un processo di revisione anonima sotto la responsabilità del Comitato scientifico della rivista. La valutazione è stata condotta in aderenza ai criteri scientifici ed editoriali di Edizioni Ca' Foscari.
Scientific certification of the works published by Edizioni Ca' Foscari: all essays published in this issue have received a favourable opinion by subject-matter experts, through an anonymous peer review process under the responsibility of the Advisory Board of the journal. The evaluations were conducted in adherence to the scientific and editorial criteria established by Edizioni Ca' Foscari.

Balcania et Slavia
Vol. 4 – Num. 2 – Dicembre 2024
e-ISSN 2785-3187

URL https://edizionicafoscari.unive.it/it/edizioni/riviste/balcania-et-slavia/2024/2/
DOI http://doi.org/10.30687/BES/2785-3187/2024/02

e-ISSN 2785-3187

Balcania et Slavia
Vol. 4 — Num. 2 — Dicembre 2024

Introduction

Iliyana Krapova, Svetlana Nistratova,
Luisa Ruvoletto, Giuseppina Turano
Università Ca' Foscari Venezia, Italy

The second issue of *Balcania et Slavia* for 2024 brings together a rich collection of articles exploring diverse linguistic phenomena across the Balkan and Slavic regions, offering fresh insights into both historical and contemporary developments.

This issue opens with Beatrice Bindi's study, *"I corpora diacronici per la lingua russa: caratteristiche, modalità di funzionamento e possibilità di impiego."* The study focuses on the growing role of diachronic corpora in Russian linguistics, examining eight key resources to highlight their design, functionality, and potential applications, while also addressing their limitations.

Hysnie Haxhillari's contribution "*Genitival phrases in Albanian and Romanian. A comparative approach*" delves into the unique structures of the Genitival Phrase in Albanian and Romanian, comparing their morphological and syntactic features to uncover shared traits and distinctions within the Balkan Sprachbund. The study highlights both the origin and syntactic role of the genitival article in these two languages, which exhibit notable similarities and differences in their morphological and syntactic structures.

Finally, a corpus-based study of the Timok variety of Serbian by Mirjana Mirić and Boban Arsenjević investigates complementizer drop and clitic climbing, offering new perspectives on the structural properties of complement clauses in Torlakian dialects. Through statistical analysis, the authors reveal the conditions under which these phenomena occur, contributing to our understanding of syntactic variation in non-standard varieties and opening new avenues for research into the so-called "Balkan subjunctive."

This issue also includes an exploration of the linguistic impact of the COVID-19 pandemic, analyzing the rapid emergence of neologisms in Russian related to the virus and the disease. The study presented by Julija Nikolaeva and entitled "Лингвистические мутации коронавируса, или об истории означивания новой болезни" traces the evolution of these terms, their integration into everyday language, and the stylistic differentiation that emerged across various registers of contemporary Russian.

Complementing this, a comparative analysis of destructive nominations in Russian and Italian journalism sheds light on how language reflects societal change, identifying both forward and inverse models of lexical-semantic transformation in journalistic discourse. At the same time, Yuliya Rohava's article "Модели деструктивных субстантивов русского и итальянского языков" underscores the importance of comparative studies in journalistic discourse, illustrating how lexical semantics encodes destructive events of the 21st century and shapes public perception.

Together, these articles showcase the dynamic interplay between language, society, and cognition, making valuable contributions to Slavic and Balkan linguistics. They offer deeper insights into key language-internal mechanisms of change while also demonstrating how language adapts to new realities and shapes public perception. This issue underscores the significance of comparative linguistic approaches in examining the structural and cognitive functions of language across different cultural and historical contexts.

e-ISSN 2785-3187

Balcania et Slavia
Vol. 4 — Num. 2 — Dicembre 2024

I *corpora* diacronici per la lingua russa: caratteristiche, modalità di funzionamento e possibilità di impiego

Beatrice Bindi
Università G. D'Annunzio di Chieti-Pescara, Italia

Abstract Although still at an early stage of development, diachronic *corpora* of the Russian language are at the centre of the attention of the scientific community and seem destined to play an increasingly important role in the panorama of contemporary Slavistics. In the light of these considerations, the present article examines a selection of eight of these resources, describing in detail their design criteria, modes of operation and possible uses, with the aim of providing an overview that highlights their potentialities and shortcomings.

Keywords Corpus Linguistics. Historical corpora. Corpus-based approach. Old Church Slavonic. Old Russian.

Edizioni Ca'Foscari

Peer review

Submitted 2024-02-24
Accepted 2024-11-21
Published 2024-12-18

Open access

Citation Bindi, Beatrice (2024). "I *corpora* diacronici per la lingua russa: caratteristiche, modalità di funzionamento e possibilità di impiego". *Balcania et Slavia*, 4(2), 111-162.

DOI 10.30687/BES/2785-3187/2024/01/001

1 Introduzione

La Russia ha imboccato la strada della «*corpus revolution*» (Rundell, Stock 1992) all'inizio del nuovo millennio,[1] e da allora vi procede speditamente. L'ultimo decennio, infatti, è stato caratterizzato da un rapido sviluppo di *corpora* per le lingue slave. Un simile incremento di attività, sostiene T.I. Reznikova (2009, 402-3),

> è dovuto principalmente a un ripensamento del ruolo del corpus nella ricerca linguistica. Se i primi esperimenti di raccolte elettroniche di testi erano finalizzati all'analisi statistica della lingua e alla pratica lessicografica, rimanendo per questo alla periferia degli interessi della comunità linguistica, adesso, con il riconoscimento dell'importanza del corpus come strumento autonomamente efficace, in grado di cambiare radicalmente le possibilità di ricerca del linguista in un'ampia gamma di campi scientifici, la compilazione di *corpora* è divenuta una mansione urgente di una grande cerchia di specialisti in differenti paesi.[2]

I *corpora*, di fatto, modificano i compiti e gli scopi della linguistica. Nel noto volume di M.V. Kopotev (2014), *Vvedenie v korpusnuju lingvistiku* (Introduzione alla linguistica dei *corpora*), viene ripetutamente sottolineato come, grazie alla comparsa delle raccolte elettroniche, il testo sia divenuto insieme oggetto, soggetto e fine delle ricerche del linguista (cf. Baranov 2015, 41): lontana da quell'approccio prescrittivo, che si era rivelato fondante nei secoli XIX-XX, la linguistica dei *corpora* fonda la propria metodologia sulla centralità dell'uso rispetto alla norma, determinando una nuova centralità dell'elemento testuale, 'habitat naturale' dell'elemento linguistico.

Il concetto di corpus, inteso come

> collezione di dati linguistici di grandi dimensioni, conservata in formato elettronico, standardizzata, strutturata, annotata e filologicamente attendibile, destinata alla risoluzione di problemi linguistici concreti (Zacharov, Bogdanova 2013, 5)

è un prodotto della modernità e dello sviluppo delle nuove tecnologie informatiche. Alla raccolta di più opere, pubblicate con lo scopo di fornire serie ordinate e complete di testi su cui condurre analisi

1 Il risultato più significativo dell'approdo della linguistica dei *corpora* in Russia è costituito dalla creazione del *Nacional'nyj Korpus Russkogo Jazyka* (NKRJa) (Corpus nazionale della lingua russa) nel 2004 (cf. *infra*).

2 Se non diversamente indicato, le traduzioni sono di chi scrive.

di diverso tipo, fanno ricorso sin dagli inizi anche gli studiosi che si occupano di linguistica storica.

Negli ambiti della linguistica storica e della medievistica, infatti, il testo in sé, o il raggruppamento di date fonti testuali, ha sempre occupato una posizione centrale: il documento scritto è e rimane l'unico oggetto disponibile all'osservazione, dal quale è possibile estrapolare i dati relativi all'argomento di studio, realizzarne la comparazione quantitativa, chiarirne la combinabilità con altre unità, metterne in luce le caratteristiche di distribuzione. In sostanza, sostiene V.A. Baranov (2015, 41),

> le fonti scritte, scelte per un dato studio, rappresentano l'analogo di un sotto-corpus, mentre, per la loro metodologia di lavoro con il materiale testuale, i tradizionali studi di linguistica storica, che si indirizzano a un certo documento, o a un gruppo di documenti, e hanno modo di selezionare manualmente gli esempi rilevanti per il problema posto, sono assimilabili alle tecnologie dei *corpora*, dove un documento o un gruppo di documenti costituisce il corpus, mentre gli esempi estratti un campione.

Data l'impossibilità di fare riferimento all'esperienza linguistica del ricercatore, la realizzazione di raccolte tradizionali si è a lungo presentata come l'unica strategia possibile per lo studio di fonti antiche.

All'inizio degli anni Novanta, poi, con l'avvento delle nuove tecnologie informatiche, i paleoslavisti si sono resi conto delle potenzialità dell'utilizzo del computer per la collezione dei monumenti scritti in un unico corpus elettronico, strutturato in modo tale da poter essere sfruttato per estrarre qualsiasi informazione sulla più ampia gamma possibile di fonti.[3]

Da quel momento, nei centri di slavistica della Russia, della Bulgaria, della Macedonia, degli Stati Uniti, della Germania, della Finlandia, della Norvegia e di alcuni altri paesi sono state via via preparate, e costantemente perfezionate, numerose edizioni informatiche di codici manoscritti slavi, e di programmi per la loro elaborazione digitale, finché, con l'avvento del nuovo millennio, le raccolte elettroniche di testi medievali slavi hanno preso a moltiplicarsi.

Molti dei *corpora* storici della tradizione scritta slava, come vedremo (§3), si trovano ancora in una fase di elaborazione iniziale, eppure sono al centro dell'attenzione della comunità scientifica, avendo innescato una vera e propria rivoluzione nelle modalità in cui il ricercatore si relaziona ai dati linguistici.

3 Simili temi sono stati discussi per la prima volta nel contesto della *First International Conference on the Application of Computer Technology to the Study of Mediæval Slavonic Manuscripts*, tenutasi a Blagoevgrad nel luglio 1995.

L'impiego di risorse digitali basate su documenti storici, difatti, non si limita alla ricerca di materiale fattuale per l'analisi: l'offerta di strumenti per la ricerca, la visualizzazione e l'organizzazione dei dati, nonché la disponibilità di un *markup* offrono nuovi modi di estrarre, raggruppare, classificare e confrontare gli esempi, il che, in ultima analisi, fornisce allo studioso nuove modalità di lavoro con i fatti testuali, consentendo di vagliare rapidamente le ipotesi, e di utilizzare tecniche *corpus-based* per l'analisi dei dati (cf. Baranov 2015, 41-2).

2 *Corpora* diacronici: caratteristiche specifiche e problematiche comuni

Si definiscono diacronici, o storici, i *corpora*

> che consentono di analizzare lo sviluppo della lingua in un certo periodo di tempo (in genere piuttosto lungo), oppure di studiare una lingua nei suoi stadi precedenti. (Zacharov 2015, 11)

I *corpora* storici differiscono sensibilmente da quelli della lingua moderna, e meritano, perciò, una trattazione separata.

Esistono una serie di peculiarità nell'elaborazione di simili risorse, di cui è necessario rendere conto, prima di passare in rassegna alcuni dei principali *corpora* diacronici per la lingua russa.[4]

In primo luogo, la modesta disponibilità di documenti conservati dall'Alto e dal Basso Medioevo rende pressoché impossibile creare un corpus bilanciato e rappresentativo. Laddove i *corpora* di lingua moderna sono chiamati a offrire

> una rappresentazione statisticamente affidabile di una lingua o di una parte di essa, ottenuta grazie a un numero adeguato e a un'adeguata diversità di genere di testi (Kopotev 2014, 9)

le raccolte di tipo storico non possono che limitarsi a fare tesoro del materiale che ha sconfitto l'usura del tempo per giungere sino ai nostri giorni.

In seconda istanza, l'eventuale inaccessibilità delle copie originali definisce la necessità di attingere a fonti secondarie: sulla base dei loro obiettivi, gli sviluppatori dei *corpora* diacronici devono scegliere se ricorrere a riproduzioni fotografiche, pubblicazioni fototipiche, o edizioni critiche a stampa dei manoscritti.

4 Nell'elencazione dele problematiche e delle peculiarità dei *corpora* storici, qui offerta, si fa riferimento a Baranov (2015, 42-8). Per una riflessione critica sulle difficoltà di un approccio *corpus-based* allo studio delle lingue antiche cf. Afanasiev (2020).

e-ISSN 2785-3187

A tutto questo, poi, si aggiungono le complesse condizioni per la preparazione di copie leggibili a macchina: l'unicità di ciascun manoscritto, di ciascuna copia, e di ciascun frammento conservatosi implica la presentazione elettronica non soltanto del testo in sé, come generalmente accade nei *corpora* di lingua moderna, ma anche la trasmissione del suo esatto posizionamento all'interno di uno specifico supporto fisico. Si tratta, di fatto, di preparare due edizioni elettroniche: quella della copia manoscritta, e quella del testo in essa contenuto. Spetta agli sviluppatori del corpus decretare se, e in quale misura, restituire le caratteristiche grafico-ortografiche dell'originale, o se concentrare l'attenzione sul dato testuale, livellando le differenze formali nella struttura e nella composizione del manoscritto.

La presenza di due unità fondamentali nei *corpora* storici – il testo e il suo supporto – comporta altresì la necessità di realizzare un'annotazione su un duplice livello: da una parte, il *markup* strutturale deve permettere di visualizzare le caratteristiche essenziali della trasmissione e della disposizione del testo all'interno del manoscritto; dall'altra, l'annotazione linguistica deve poter rendere conto delle caratteristiche morfo-sintattiche dei singoli elementi di ogni testo. La mancata disponibilità di dizionari grammaticali elettronici per la lingua paleoslava e slavo-ecclesiastica, purtroppo, complica quest'ultima operazione: numerosi specialisti stanno lavorando all'elaborazione di differenti strategie di automazione delle procedure di annotazione linguistica dei *corpora* storici, ma nessuna di queste si è ancora rivelata pienamente soddisfacente.[5]

5 Uno dei più recenti tentativi in questo senso è il progetto *QuantiSlav - Quantitative Methods in Slavonic Studies,* al quale lavora un gruppo di studiosi dell'Accademia bavarese delle Scienze di Monaco e del Dipartimento di Slavistica dell'Università di Friburgo: E. Arnold, A. Jouravel, P. Lendvai, A. Rabus, E. Renje. Il progetto mira allo sviluppo di metodi quantitativi per l'analisi di testi slavi ecclesiastici al fine di una loro categorizzazione automatica sulla base di criteri cronologici e geografici. Poiché un simile obiettivo richiede una chiara identificazione delle proprietà linguistiche (ortografiche, grammaticali, morfologiche) delle varie forme di parola, i ricercatori stanno cercando di ottimizzare il sistema di segmentazione del testo, correggendo manualmente le singole unità, così da riaddestrare i modelli. Se questa strategia si rivelasse efficace, il progetto *QuantiSlav* offrirebbe alle nuove generazioni di studiosi la possibilità di lavorare su grandi quantità di dati testuali con un alfabeto non latino, una morfologia complessa e un'ortografia non normalizzata. Per ottenere maggiori informazioni sull'iniziativa, se ne rimanda alla pagina di presentazione: `https://quantislav.badw.de/en/project.html`; cf. anche Rabus et. al (2023). Esistono, inoltre, alcuni *tagger* statistici liberamente accessibili al ricercatore che intenda sperimentare le possibilità di annotazione grammaticale e morfo-sintattica dei testi della tradizione paleoslava e slava orientale: *Stanza* (`https://stanfordnlp.github.io/stanza/`) e *UDPipe* (`https://lindat.mff.cuni.cz/services/udpipe/`). Si tratta di *pipeline* di dati per l'elaborazione del linguaggio naturale (NLP), che prendono in input il testo grezzo e producono annotazioni comprensive di tokenizzazione, lemmatizzazione, *Pos tagging*, *parsing* delle dipendenze e riconoscimento di entità denominate (NER). Le dipendenze universali (UD) per il paleoslavo si basano sul *treebank* PROIEL (`https://dev.syntacticus.org/proiel.html#downloads`). Per i testi della tradizione slava orientale, invece, si distinguono

Ulteriori difficoltà, infine, si riscontrano nella programmazione dei motori ricerca e di visualizzazione dei dati linguistici. Il modulo di interrogazione dei *corpora* storici, infatti, deve permettere di inserire una maschera di forme di parola, che tenga conto dei differenti gradi di normalizzazione e unificazione di queste stesse. Si tratta, insomma, di eliminare la variazione grafica e ortografica delle forme testuali, utilizzando una forma del dizionario accettata come invariante o inserendo, ad esempio, gli equivalenti moderni nel campo di interrogazione.

3 Rassegna dei *corpora* diacronici per la lingua russa

A seguito delle osservazioni preliminari condotte sin qui, ci proponiamo di offrire di seguito una presentazione dell'attuale stato di elaborazione dei *corpora* diacronici per la lingua russa, con l'intenzione di mettere in luce potenzialità e difetti di una serie di risorse che sembrano destinate a occupare un ruolo sempre più significativo all'interno del panorama scientifico della slavistica contemporanea.

3.1 I *corpora* storici inclusi nel *Nacional'nij korpus russkogo jazyka*

L'indiscusso protagonista dei *corpora* di lingua russa è il *Nacional'nij korpus russkogo jazyka* (NKRJa) (Corpus nazionale della Lingua Russa),[6] sviluppato nell'ambito del progetto *Filologija i informatika* (Filologia e informatica) presso l'Accademia Russa delle Scienze, e reso liberamente accessibile sul web a partire dell'aprile 2004.

Oltre a includere una selezione testuale rappresentativa della moderna lingua letteraria russa, il Corpus nazionale comprende un'ampia sezione storica (*istoričeskie korpusa*), che al momento conta un totale di circa quattordici milioni di parole, distribuite all'interno di cinque differenti sotto-*corpora*.

tre modelli di dipendenze universali, basati su tre differenti *treebanks*: TOROT per lo slavo orientale antico (`https://torottreebank.github.io/`), RNC per il 'medio-russo' (`https://universaldependencies.org/treebanks/orv_rnc/index.html`), e *Old East Slavic Birchbark* (`https://universaldependencies.org/treebanks/orv_birchbark/index.html`), specificamente pensato per la lingua delle iscrizioni su corteccia di betulla (*berestjanye gramoty*, cf. *infra*). Simili strumenti non consentono ancora una totale automatizzazione delle procedure di annotazione grammaticale e morfo-sintattica dei testi slavi antichi, tuttavia sembrano essere utili ausili, su cui è opportuno mantenere uno sguardo vigile. Per una presentazione maggiormente dettagliata di queste risorse, cf. Dozat, Zhang, Manning (2018); Dozat et al. (2020); Straka, Hajič, Strakova´ (2016).

6 `https://ruscorpora.ru/`.

Il primo di questi, definito per mezzo dell'etichetta «*drevnerusskij*» (slavo orientale antico),[7] contiene testi originali o tradotti dal greco, redatti nella Rus' a partire dall'XI secolo: si tratta di cronache, testi agiografici, opere giuridiche, scritti di diritto canonico, alcuni esempi di letteratura di pellegrinaggio, e una selezione di testi di carattere pedagogico-morale.

Le procedure di elaborazione del sotto-corpus *drevnerusskij* si sono svolte in seno all'Istituto di lingua russa V.V. Vinogradov dell'Accademia Russa delle Scienze, sotto la direzione di A.A. Pičchadze.

La creazione di copie elettroniche delle composizioni originali ivi incluse si è basata sui più recenti studi sulle fonti storiche, mentre per le opere tradotte è stato fatto riferimento ad uno studio della stessa Pičchadze (2011), dedicato alla considerazione dell'aspetto linguistico dell'attività traduttoria nella Rus' nel periodo pre-mongolo.

I testi del sotto-corpus *drevnerusskij* sono riportati conservando l'antica ortografia originale dei manoscritti (compresi quelli che vengono ripresi dalle edizioni); in alcune fonti i segni di accentuazione sono stati conservati nell'annotazione.[8]

La risorsa dispone di un'annotazione a più livelli: in primo luogo, tutti i testi sono dotati di etichette morfologiche con rimozione dell'omonimia. Questo tipo di annotazione è stata effettuata manualmente, sulla base di un dizionario aggiornato in automatico per mezzo del *tagger Morphy*, sviluppato da T.A. Archangel'skij.[9]

Il corpus presenta inoltre un'annotazione di tipo testuale: ogni fonte è etichettata attraverso metadati, che indicano il genere, il carattere originale o tradotto dell'opera, la data di composizione del testo, e quella della copia manoscritta di riferimento.

7 La traduzione dell'aggettivo *drevnerusskij* con l'etichetta 'slavo orientale antico' ci sembra maggiormente opportuna rispetto alla resa per mezzo della locuzione letterale 'russo antico'. Come sottolineato in Garzaniti (2019, 187, 222), infatti, la prima definizione è da considerarsi ad oggi preferibile, alla luce di una rinnovata cautela terminologica, legata al riconoscimento di una compresenza delle componenti russa, ucraina e bielorussa nel contesto culturale della Rus' di Kiev.

8 Per comodità dell'utente, è stato sviluppato un sistema di ricerca a tre stadi, che consente di selezionare diverse modalità ortografiche, in base agli scopi e interessi della ricerca. Con la modalità «ortografia esatta» (*točnaja orfografija*) è possibile cercare una forma di parola, o parte di essa, tenendo conto della forma grafica originale (lettera per lettera). Nel regime «ortografia semplificata» (*uprošennaja orfografija*), la *query* viene impostata tenendo conto delle varianti ortografiche nella stessa posizione (ad esempio, alternanza tra o/e, e/ѣ, o/ѡ i/ї). È infine possibile interrogare il corpus con una «ortografia moderna» (*modernizirovannaja orfografija*), la quale non equipara soltanto le varianti ortografiche, ma anche alcune lettere, che indicano fonemi diversi (ad esempio, e/ѣ/и, e tutte le loro varianti ortografiche) (Mišina, Pičchadze 2015, 102).

9 Si sono occupati in modo particolare dell'annotazione della risorsa G.S. Barankova, M.V. Ermolova, N.P. Iordani, D.S. Krylov, I.I. Makeeva, E.A. Mišina, M.S. Mušinskaja, P.V. Petruchin, A.V. Ptencova, D.V. Sičinava, V.K. Skripka, A.A. Fitiskina e I.S. Jur'eva.

Il secondo sotto-corpus della sezione storica del NKRJa viene definito «*starorusskij*» (medio-russo)[10] e raccoglie opere di genere vario, composte nello stato moscovita dal XV al XVII secolo, tra cui cronache e racconti, documenti di carattere commerciale, corrispondenze private, opere di letteratura religiosa, composizioni poetiche.

In verità, si tratta di un corpus dal carattere ibrido: i testi qui contenuti, prodotti in una fase di transizione della lingua, sono frutto della combinazione di diversi livelli grammaticali e lessicali. Da una parte, essi riflettono tratti morfo-sintattici dello slavo orientale antico (XI-XIV sec.) (ad esempio, forme duali, come крылома in luogo del plurale крылы). Dall'altra parte, sono numerose le influenze lessicali e grammaticali dello slavo ecclesiastico.[11] Le opere ivi raccolte sono inoltre soggette a variazioni di tipo dialettale (ad esempio, исподка, 'camicia da donna' nei dialetti settentrionali), e all'instabilità delle norme ortografiche (come la confusione tra е/ѣ/и) (Gavrilova, Šalganova, Ljaševskaja 2016, 9).

Lo sviluppo della suddetta collezione si deve a un gruppo di studiosi della Scuola Superiore di Economia di Mosca, guidato da O.N. Ljaševskaja.[12]

Ai fini della compilazione della risorsa sono state utilizzate edizioni digitalizzate delle fonti, con particolare riferimento alla collana *Biblioteka literatury Drevnej Rusi* (I-XX, 1997-2020) (Biblioteca della letteratura dell'antica Rus'), e alla serie *Polnoe sobranie russkich letopisej* (I-XLIII, 1841-2004) (Raccolta completa delle cronache russe).

Purtroppo, non tutte queste edizioni presentano i testi in modo linguisticamente corretto, semplificando, in un modo o nell'altro, l'ortografia. Per questo motivo, il corpus dispone di un'annotazione meta-testuale, la quale, oltre a riportare i dati dell'edizione del testo, fornisce anche una rappresentazione delle informazioni di archivio, cosicché l'utente, in caso di necessità, possa ricevere contezza dell'originale.

I testi del sotto-corpus *starorusskij* sono stati inoltre sottoposti a procedure di marcatura grammaticale e lemmatizzazione, orientata sulle norme dello *Slovar' russkogo jazyka XI-XVII vv.* (I-XXX, Moskva

10 All'interno della pagina dedicata sul sito del NKRJa, il sotto-corpus «*starorusskij*», viene alternativamente indicato come «*sredne(veliko)russkij*», 'medio(grande)russo' (`https://ruscorpora.ru/corpus/mid_rus?search=CgQyAggRMAE=`). In assenza di una terminologia univoca, ci proponiamo di usare, qui e in seguito, la locuzione 'medio-russo' per indicare la lingua letteraria russa in formazione nella sua fase diacronica intermedia (XV-XVII sec.).

11 La lingua 'russa', ricorda B.A. Uspenskij (1994, 4), «in virtù della sua natura non normata (non codificata) era aperta a vari tipi di influenze esterne, sia dallo slavo ecclesiastico, che da altre lingue (in particolare il tataro, e altre). L'opposizione stessa tra russo e slavo ecclesiastico, tuttavia, persiste nei diversi stadi evolutivi.»

12 Fanno parte del gruppo di lavoro A.A. Dudin, D.V. Sičinava, A.Ja. Pen'kova, A.G. Israeljan e V.I. Legkich.

1975-2015) (Dizionario della lingua russa dei secoli XI-XVII) per mezzo di un *tagger* automatico (*Juni-tagger*), sviluppato, ancora una volta, da Archangel'skij.[13]

Segue dunque il terzo sotto-corpus storico, «*berestjanye gramoty*» (iscrizioni su corteccia di betulla), che contiene brevi testi di carattere economico-commerciale, scritti su corteccia di betulla tra l'XI e il XV secolo in area slava orientale: si tratta della raccolta maggiormente completa e aggiornata delle informazioni linguistiche relative a tali testimonianze sinora note. La collezione è sincronizzata con il database *Drevnerusskie berestjanye gramoty* (Iscrizioni su corteccia di betulla slave orientali antiche):[14] entrambe le risorse vengono annualmente aggiornate e ampliate.

La compilazione del sotto-corpus *berestjanye gramoty*, svoltasi sotto la direzione del già citato Archangel'skij, ha attinto alle informazioni linguistiche sulle iscrizioni su corteccia di betulla, disponibili in una sezione specificamente dedicata, che viene regolarmente pubblicata all'interno della rivista *Voprosy Jazykoznanija* (Questioni di Linguistica).[15]

La risorsa è dotata di un'annotazione di tipo morfologico, realizzata automaticamente sulla base di un noto studio di A.A. Zaliznjak (2004) sul cosiddetto «antico dialetto di Novgorod» (*drevnenovgorodskij dialekt*). I risultati della marcatura automatica, comunque, sono stati corretti e integrati, e i testi non inclusi nel volume di Zaliznjak sono stati etichettati manualmente, seguendo il medesimo schema.[16]

Tutti i testi digitali delle lettere su corteccia di betulla sono annotati tramite metadati, che riportano informazioni circa la data convenzionale di redazione delle testimonianze, il tipo di testo, la sua conservazione, il sito archeologico di ritrovamento, e la datazione relativa (stratigrafica). Sulla base di questi parametri è possibile creare sotto-*corpora* personalizzati. All'interno della finestra contenente i metadati, che si apre cliccando il nome del documento nei risultati della ricerca, è inoltre presente un collegamento ipertestuale, rimandante alla pagina specificamente dedicata alla *gramota* in questione sul già citato database *Drevnerusskie berestjanye gramoty* (cf. *supra*).[17]

13 Lo svolgimento delle procedure di annotazione grammaticale sono state curate dalla stessa Ljaševskaja, mentre Dudin si è occupato della preparazione dei testi e della realizzazione dell'annotazione meta-testuale.

14 `http://gramoty.ru/birchbark`.

15 `https://vja.ruslang.ru/ru/keywords/225`.

16 La realizzazione dell'annotazione manuale del corpus si deve a Mišina e Sičinava.

17 D.V. Sičinava e A.N. Dyškant si sono occupati della sincronizzazione del corpus con il suddetto database; cf. Sičinava, Dyškant (2021).

Dal 2021 il corpus *berestjanye gramoty* è divenuto parallelo (cf. Sičinava 2022): ai testi disponibili nelle versioni di Zaliznjak e A.A. Gippius sono state aggiunte traduzioni russe, e addirittura rese accessibili le traduzioni in inglese a cura di R. Kovalev e J. Schaken, che coprono in totale più di trecento *gramoty*.

Di recentissima pubblicazione (2023), il quarto sotto-corpus storico del NKRJa, «*vostočnoslavjanskaja epigrafika*» (epigrafia slava orientale), comprende iscrizioni dei secoli XI-XV rivenute su pareti di chiese, pietre, croci e altri oggetti ad uso ecclesiastico o domestico. Tali reperti sono generalmente brevi didascalie, che tuttavia possono contenere nomi unici, o fenomeni linguistici interessanti. Molte iscrizioni in slavo orientale, infatti, hanno una datazione molto precoce, antecedente non solo ai manoscritti canonici giunti sino a noi, ma anche alla maggior parte delle lettere su corteccia di betulla.

Il lavoro congiunto di D.V. Sičinava e D.D. Ignat'ev ha consentito l'esportazione dei dati per la creazione del sotto-corpus *vostočnoslavjanskaja epigrafika* a partire dal database *Drevnerusskaja epigrafika* (Epigrafia antica russa),[18] sviluppato sotto la guida di A.A. Gippius, grazie a una sovvenzione della Fondazione russa per la scienza (RNF), assegnata alla Scuola Superiore di Economia di Mosca. Per il corpus sono stati selezionati solo testi precedentemente pubblicati, scritti in lingue slave e contenenti caratteri alfabetici e numerici: si tratta di più di seicento iscrizioni, per un totale di oltre cinquemila parole.[19]

La raccolta è dotata di un sistema di annotazione morfologica con rimozione dell'omonimia, che si basa sullo stesso standard di presentazione dei lemmi e di *markup* del corpus *drevnerusskij*, con l'aggiunta di tag speciali per l'etichettatura di frammenti di parole, e altre parti mancanti.

Tutti i testi sono dotati di una marcatura meta-testuale a più livelli, che indica: il supporto dell'iscrizione, il tipo di materiale del supporto, la tecnica di iscrizione, la categoria e il genere del testo, il luogo in cui l'iscrizione è stata rinvenuta, il tipo di alfabeto, il grado di conservazione dell'iscrizione, e la datazione esatta o approssimativa. Dai metadati di ogni fonte, consultabili all'interno della finestra dei risultati della ricerca, si apre un collegamento ipertestuale, che rimanda direttamente al database *Drevnerusskaja epigrafika*, dove l'utente può ottenere informazioni più dettagliate sul testo (ad esempio, le varie interpretazioni concorrenti, i commenti, i dati bibliografici), nonché vedere le fotografie e i disegni delle iscrizioni.

18 http://epigrafika.ru/epigraphy/.

19 Alla realizzazione della risorsa hanno collaborato anche S.V. Micheev, A.N. Dyškant, M.M Dobryševa, D.S. Krylov, S.A. Lašuk, K.D. Pokrovskaja, N.A. Sedukova e A.A. Fitiskina.

Come la collezione *berestjanye gramoty*, anche il corpus *vostočnoslavjanskaja epigrafika* esiste nella sua versione parallela: per la maggior parte delle iscrizioni vengono riportate le traduzioni in russo, disponibili altresì per la ricerca morfologica.

Conclude l'elenco il quinto, e ultimo, sotto-corpus storico, denominato «*cerkovnoslavjanskij*» (slavo ecclesiastico), e preposto alla raccolta di testi rappresentativi di tutti i principali generi della tradizione scritta slava ecclesiastica (testi liturgici, testi patristici, testi di carattere ascetico-morale, sacre scritture, manuali di diritto ecclesiastico), i quali, venendo ancora oggi parzialmente impiegati nello svolgimento del servizio liturgico della Chiesa russa, appartengono tanto al passato, quanto alla pratica linguistica contemporanea.

L'elaborazione della raccolta si deve ad A.E. Poljakov (in qualità di principale sviluppatore), A.G. Kraveckij ed E.R. Dobrušina.[20]

I testi del sotto-corpus *cerkovnoslavjanskij* sono stati acquisiti dalla *Biblioteka svjatootečeskoj literatury* (Biblioteca della letteratura dei santi padri),[21] la più grande raccolta elettronica di monumenti della tradizione scrittoria cristiana, approntata presso il Ginnasio Ortodosso intitolato a Sergio di Radonež di Novosibirsk. Tutte le fonti, tuttavia, hanno subito un processo di revisione e integrazione al fine di soddisfare le esigenze proprie del corpus (Poljakov 2014, 246).

A ogni testo sono state aggiunte una serie di informazioni metatestuali che ne indicano il titolo, il numero di parole, il genere, la data di composizione e la natura di opera originale o tradotta.

La classificazione tematica e di genere delle fonti è orientata maggiormente all'utente comune, che allo specialista, e pertanto si fonda sui principi di convenienza pratica e comprensibilità. In particolar modo, si è pervenuti all'individuazione delle seguenti forme testuali su base funzionale: Bibbia, testi per il servizio liturgico (*Acatisto, Irmologion, Menee, Octoechos, Ieratikon, Eucologio, Triodion, Libro delle ore*), *Typikon*, letteratura patristica, opere di diritto canonico (*Canoni apostolici*).

Per la maggior parte, i testi vengono riportati nelle loro moderne edizioni (XIX-XX sec.). La loro data di creazione, tuttavia, è di molto precedente, e spesso definibile solo convenzionalmente (Poljakov 2014, 247). In assenza di criteri cronologici esatti, si è giunti a una periodizzazione generica delle fonti incluse nel corpus *cerkovnoslavjanskij,* che si suddividono in: testi «standard» (XIX-XX sec.),

20 È opportuno sottolineare, comunque, che le procedure di elaborazione che hanno condotto alla nascita del sotto-corpus *cerkovnoslavjanskij*, sono il frutto del lavoro congiunto di un più ampio numero di specialisti, tra cui si annoverano: V.A. Plungjan, A.I. Zobnin, T.Ju. Ivanova-Allenova, A.V. Žirova, A.A. Pletneva, L.I. Marševa, *svjaš.* Feodor Ljudogovskij, R.N. Krivko, I.V. Segalovič; cf. Dobrušina, Poljakov (2013, 32 nota 2).

21 http://www.orthlib.ru/.

e-ISSN 2785-3187

testi «arcaici» (XVII-XVIII sec.), «testi del XX secolo» e testi «ibridi»[22] (Dobrušina, Poljakov 2013, 37).

Il sotto-corpus *cerkovnoslavjanskij* è annotato grammaticalmente: ciascuna forma di parola è ricondotta al lemma di riferimento e corredata dalle relative informazioni grammaticali. Il modello descrittivo della declinazione delle parole in slavo ecclesiastico è stato derivato empiricamente, sulla base dell'analisi degli stessi dati del corpus. Questo ha portato alla creazione di una versione pilota di un dizionario grammaticale, che ha gradualmente conferito un carattere maggiormente automatico alle procedure di lemmatizzazione. L'omonimia delle forme grammaticali, invece, non è ancora stata risolta.

Ai fini della pubblicazione dei testi all'interno del corpus *cerkovnoslavjanskij* è stato elaborato un sistema ortografico semplificato, che preserva gli elementi distintivi da un punto di vista linguistico, senza tuttavia cercare di riprodurre l'esatto aspetto tipografico del testo. Le forme di parola all'interno del corpus sono presentate così come le si incontrano nei testi, mentre i lemmi sono riportati in una grafia unificata, che non tiene conto di elementi quali lettere sovrascritte e *titulus* (ad esempio, la forma млсть si riferisce al lemma милость).

Tutti i testi slavo ecclesiastici del corpus sono rappresentati attraverso il sistema di codifica standard Unicode. I caratteri slavi non presenti nello standard Unicode sono stati semplificati dagli sviluppatori della raccolta, che hanno considerato la leggibilità del testo un criterio di importanza maggiore rispetto alla fedele riproduzione del sistema ortografico.

Il carattere eterogeneo dei *corpora* storici del NKRJa, sopra presentati, tiene conto della presenza di differenti livelli evolutivi all'interno della tradizione scrittoria slava orientale, nonché della reciproca influenza che andò costantemente instaurandosi tra lo standard di una lingua scritta che voleva essere comune a tutto il mondo bizantino slavo, e le vive parlate locali.

I testi storici e il corpus principale (*osnovnoj korpus*), le cui fonti più antiche risalgono al periodo a cavallo tra i secoli XVII e XVIII, sono poi collegati da un modulo di ricerca comune, definito «*panchroničeskij korpus*» (corpus pancronico), il quale consente di condurre ricerche linguistiche relative a differenti secoli della storia della lingua letteraria russa, su un totale di più di trecento milioni di forme di parola.

22 Secondo quanto sostenuto da Dobrušina e Poljakov (2013, 38), rientrano nella categoria degli «ibridi» alcuni testi, scritti in uno slavo ecclesiastico di carattere ibrido, che vennero pubblicati inizialmente senza l'intervento delle tipografie sinodali, ma vennero in seguito ristampati da queste. Delle fonti in questione, tra cui si ricordano, ad esempio, *Alfavit duchovnyj* (Alfabeto spirituale) di Dimitrij Rostovskij (1651-1709) e la *Filocalia* slava (*Dobrotoljubie*), non è stata effettuata una definitiva standardizzazione linguistica.

La scelta di creare un corpus pancronico (2022) è stata dettata, in primo luogo, dalla presenza di differenti principi di presentazione dei lemmi, delle norme orografiche, e di quelle grammaticali all'interno dei *corpora* storici e di lingua moderna, che rendono impossibile trasferire automaticamente una *query* da un corpus all'altro. In secondo luogo, un certo numero di testi è incluso nell'una o nell'altra delle suddette raccolte senza osservare chiari limiti cronologici: le date di confine stabilite tra i diversi *corpora* (1400 tra i sotto-*corpora drevnerusskij* e *starorusskij*, 1700 tra il corpus storico e quello di lingua moderna), sono del resto più convenzionali, che reali.

I singoli *corpora* storici e di lingua moderna, comunque, si conservano nella loro interezza, venendo regolarmente integrati e aggiornati, mentre il corpus pancronico viene sincronizzato per tenere conto di tali implementazioni.

Dal punto di vista del concreto funzionamento, tutti e cinque i sotto-*corpora* storici del NKRJa presentano due distinti moduli di ricerca: la ricerca «per forme esatte» (*poisk točnych form*) funziona tramite semplice inserimento di stringhe, quella «per forme lessico-grammaticali» (*leksiko-grammatičeskij poisk*), invece, consente di specificare una sequenza di lessemi e/o forme di parola con determinate caratteristiche grammaticali e/o semantiche.

All'interno di questo secondo modulo si aprono differenti finestre: nella sezione «lemma» (*lemma*) deve venire digitato un lessema (necessariamente in forma base), oppure una forma di parola tra virgolette. Si può ricorrere all'impiego del simbolo '*' per indicare una qualsiasi sequenza di caratteri all'inizio o alla fine del dato lessema/forma di parola. L'utilizzo degli operatori logici, 'OR' (rappresentato dal simbolo |), 'NOT' (rappresentato dal simbolo -) e 'AND' (rappresentato dal simbolo &), inoltre, consente all'utente di definire con precisione classi di stringhe (sequenze di caratteri o altro) complesse e articolate, permettendo numerose varianti di implementazioni a partire da una specifica sintassi.

Nel campo «forma di parola» (*slovoforma*) può invece venire inserita la parola, così come appare nel testo (dunque, in forma flessa). È possibile, anche in questo caso, impiegare il simbolo '*' per indicare una qualsiasi sequenza di caratteri e gli operatori logici per affinare la *query*.

Nella sezione «caratteristiche grammaticali» (*grammatičeskie priznaki*) l'utente ha modo di specificare i valori morfologici del lessema e/o della forma di parola ricercata, selezionandoli da una finestra preimpostata, che si apre dal link «scegli» (*vybrat'*), oppure digitandoli manualmente sulla tastiera. È possibile ricorrere, ancora una volta, all'impiego degli operatori logici 'OR' (|), e 'AND' (&), mentre le parentesi tonde consentono di impostare espressioni complesse.

Tutte le etichette delle caratteristiche grammaticali necessarie alla digitazione manuale della *query* sono elencate nella sezione

«morfologia» (*morfologija*), a cui si accede dalla pagina principale del NKRJa.[23] Se si sceglie di impiegare la finestra preimpostata, invece, è sufficiente spuntare le caselle relative ai significati grammaticali desiderati, e lasciare che il sistema generi in automatico la stringa complessa della *query*.

Sfruttando la casella «parametri aggiuntivi» (*dopolnitel'nye priznaki*), è possibile ricercare parole, collocate in una determinata posizione (ad esempio, prima e dopo un certo segno di punteggiatura, all'inizio e alla fine di una frase). La lista dei possibili valori selezionabili si apre attraverso il link «scegli».

Le soprammenzionate sezioni del modulo di ricerca per forme lessico-grammaticali sono comuni a tutti e cinque i sotto-*corpora* storici del NKRJa. Alcuni di essi, poi, si distinguono per la presenza di altri campi specifici.

Il modulo per la ricerca lessico-grammaticale dei sotto-*corpora drevnerusskij, berestjanye gramoty* e *vostočnoslavjanskaja epigrafika* include altresì la casella «interpretazione» (*tolkovanie*), che consente di specificare il significato di parole ambigue, contribuendo a distinguere eventuali omonimi nella ricerca, mentre i sotto-*corpora berestjanye gramoty* e *starorusskij* sono accomunati dalla sezione «lemma nella sua forma originaria» (*rannjaja lemma*), nella quale può venire digitato un lemma nella forma fonetica più arcaica (prima della caduta delle vocali ridotte).

Meritano ancora una menzione separata i sotto-*corpora berestjanye gramoty* e *vostočnoslavjanskaja epigrafika*, che, come anticipato sopra, esistono anche nella versione parallela. Al corpus parallelo si accede dal modulo di ricerca per forme lessico-grammaticali, selezionando l'opzione «cerca: nelle *gramoty*[/iscrizioni] e nelle traduzioni» (*iskat': v gramotach[/nadpisjach] i perevodach*). Tale modalità consente di rintracciare corrispondenze slave orientali di lessemi e costrutti russi (e viceversa).

Per la raccolta *berestjanye gramoty* è addirittura contemplata la possibilità di impostare la ricerca sul lessico per campi semantici. All'interno del modulo di ricerca per forme lessico-grammaticali della collezione, infatti, si trova la sezione «semantica» (*semantika*), che consente di specificare le caratteristiche semantiche desiderate per un dato lessema, selezionandole da una lista di valori preimpostati, oppure digitandole manualmente sulla tastiera.

L'elenco automaticamente generato, che si apre dal link «scegli», contiene una selezione di caratteristiche semantiche suddivise per categorie (tassonomia, mereologia, topologia, accezione, tipo di formazione delle parole). L'utente deve semplicemente spuntare le caselle di controllo relative alle caratteristiche desiderate per la forma

23 https://ruscorpora.ru/page/instruction-morph/.

di parola ricercata. Le caselle di controllo che indicano valori semantici all'interno di una medesima categoria vengono automaticamente trattate come operatore logico 'OR', mentre tra le caratteristiche semantiche di categorie differenti si instaura il rapporto 'AND'.

Al momento di impostare una *query* su ciascuno dei cinque *corpora* storici del NKRJa, si ha l'opportunità di scegliere se interrogare le suddette raccolte nella loro interezza, o se invece creare ulteriori sotto-*corpor*a personalizzati, sulla base di fonti testuali scelte, o in relazione a specifici criteri cronologici.

Di fronte alla necessità di interrogare tutte le collezioni contemporaneamente, invece, si può fare ricorso al già menzionato corpus *panchroničeskij*, che dispone dei due stessi moduli di ricerca per parole esatte, e per forme lessico-grammaticali.

È possibile impostare una *query* per singole forme di parola, oppure per sequenze di queste. Al fine di ottenere un ventaglio di risultati massimamente ampio, è necessario inserire le forme di parola nella loro ortografia moderna (senza l'utilizzo, cioè, delle lettere storiche, degli *jer* duri in fine di parola, del *titulus*). Il sistema di ricerca del corpus *panchroničeskij*, infatti, è appositamente impostato per far sì che, alla generica richiesta di ricerca di una stringa quale 'сам еси', ne vengano rintracciate anche tutte le possibili varianti ortografiche storiche (самъ ѥси, са(ѝ) еси, самъ ес[и]).

Nella suddetta raccolta è possibile ordinare i risultati, oppure creare un sotto-corpus, sulla base di due parametri: la data di composizione dell'originale, da una parte, e quella della copia manoscritta o dell'edizione di riferimento, dall'altra. L'ordinamento dei risultati in base alla data della copia di riferimento si rivela particolarmente utile ai fini dell'individuazione di eventuali alterazioni delle caratteristiche linguistiche (ortografiche, fonetiche, morfologiche) di un originale per mano di copisti di epoche successive.

Per mezzo del corpus *panchroničeskij*, l'utente può ottenere anche utili dati statistici. I risultati di una determinata *query* su questa raccolta, infatti, non si limitano a poter venire visualizzati sotto forma di concordanze (*konkordans*), oppure nel formato KWIC (*Key-Word In Context*), ma consentono addirittura la creazione di grafici di frequenza per l'intero intervallo cronologico dell'interrogazione effettuata. Selezionando la voce «grafico» (*grafik*) dalla pagina di presentazione dei risultati, si ottiene infatti una rappresentazione schematica della distribuzione di un certo lemma anno per anno, la cui frequenza di occorrenza viene statisticamente calcolata sul milione di forme di parola.

L'impiego dei *corpora* storici del NKRJa per la ricerca linguistica riscuote grande successo nell'ambito scientifico-accademico della Federazione Russa.

Si può anzitutto ricordare un vivo filone di studi, sviluppato in seno all'Università statale dell'Udmurtia, che si occupa dell'analisi delle

collocazioni verbo-nominali all'interno di testi della Rus', sulla base delle fonti contenute nelle sezioni storiche del NKRJa: L.S. Kilina, S.R. Zajnullina (2017), ad esempio, hanno considerato le collocazioni verbo-nominali con la componente *životŭ* ('vita') nel loro aspetto funzionale e semantico-strutturale, per riconoscerne l'eventuale carattere idiomatico. Successivamente, Kilina (2020) ha affrontato la questione dello status delle collocazioni della lingua russa in riferimento a differenti periodi del suo sviluppo, facendo emergere una certa criticità di identificazione per la fase antica; la studiosa si è altresì occupata dell'analisi del funzionamento delle associazioni verbo-nominali con la componente *činiti* ('riparare' 'sistemare'), assieme a O.S. Rusanova (Kilina, Rusanova 2021). Di quest'ultima, infine, possono essere menzionati una serie di contributi incentrati sulla disamina di combinazioni lessicali di differente composizione: si spazia dallo studio delle possibilità associative del verbo *dějati* ('fare') (Rusanova 2020), a quelle della componente *iměti* ('avere') (Rusanova 2022), passando per l'analisi delle peculiarità del funzionamento delle collocazioni del nucleo verbale *tvoriti* ('creare') all'interno dei testi di carattere ufficiale dei secoli XIII-XVII (Rusanova 2021).

Ricerche *corpus-based* di tipo morfo-sintattico sono invece condotte da Ja.A. Pen'kova (Istituto di lingua russa V.V. Vinogradov): l'interesse principale della studiosa è rappresentato dall'analisi dei costrutti perifrastici ad espressione del valore di futuro anteriore nella varietà linguistica 'medio-russa' dei secoli XIII-XVI (Pen'kova 2018; 2019; 2020a; 2020b; 2021a). Della ricercatrice si ricordano inoltre studi dedicati a costrutti perifrastici con diversi valori (ad esempio, *stati* + infinito, *iměti* + infinito) nel periodo 'medio-russo' (Pen'kova 2021c; 2021d), e una ricerca incentrata sui relativi liberi in '*ni budi/ni est*'', che concorrono nella lingua russa dei secoli XVII-XVIII (Pen'kova 2021b).

Ancora nell'ambito degli studi morfo-sintattici devono essere menzionati una serie di contributi singoli, dedicati a una pluralità di aspetti, quali la sintassi degli ausiliari verbali nella lingua slava orientale (Cimmerling 2019), o il tracciamento dell'evoluzione diacronica di parti del discorso: laddove Kilina ed E.I. Kolosova (2020) hanno ripercorso la storia della formazione del contenuto semantico degli aggettivi *čelověčĭ* e *čelověčĭskŭ* ('umano', 'di uomo') nelle cronache della Rus', E.V. Budennaja (2020) ha trattato il percorso di evoluzione diacronica (XI-XVII sec.) dell'espansione referenziale dei pronomi russi, mentre O.E. Pekelis (2020) ha considerato le proprietà semantiche e sintattiche della particella *že* nelle interrogative totali e parziali delle fonti dei secoli XVIII-XIX sec, confrontandole infine con l'uso contemporaneo.

Si individuano quindi alcuni studi *corpus-based* di linguistica storica, che avanzano considerazioni sul lessico slavo orientale (Zagrebina 2022), e di lessicologia storica, che invece propongono la storia evolutiva di determinati lemmi (Pen'kova 2017).

N.V. Nikolenkova dell'Università Lomonosov di Mosca e T.V. Kusennaja dell'Università federale baltica «Immanuel Kant» di Kaliningrad hanno invece fatto ricorso ai *corpora* storici del NKRJa per realizzare ricerche specificamente incentrate sul lessico agiografico (Nikolenkova 2017; Kusennaja 2021).

Nell'ambito degli studi di semantica storica si può menzionare il contributo di L.V. Popova (2022), dell'Università di Čeljabinsk, dedicato all'analisi della struttura semantica del concetto di *prestuplenie* ('delitto') nell'universo linguistico-culturale russo nei secoli XI-XIX, mentre un recente lavoro di O.V. Fel'de ed E.A. Birjulina (2022), dell'Università federale siberiana di Krasnojarsk, apre la strada all'impiego dei *corpora* storici del NKRJa in studi di etnolinguistica, prendendo in esame la dinamica evolutiva del contenuto semantico dell'universale culturale *gumannost'* ('humanitas') nella cultura linguistica russa.

La sezione storica del NKRJa, poi, riscontra un impiego particolarmente attivo anche nel panorama scientifico-accademico italiano, all'interno del quale è possibile individuare due principali filoni di ricerca *corpus-based* in prospettiva diacronica: da una parte, si annoverano studi di carattere morfosintattico, dall'altra ricerche lessicali.

Del primo gruppo fanno parte svariati contributi di L. Ruvoletto (Università Ca' Foscari, Venezia), dedicati alla categoria dell'aspetto verbale nello slavo orientale antico (Ruvoletto 2018; 2019; 2020).

Il secondo gruppo, invece, viene tracciato sulla base di alcuni lavori di M.C. Ferro (Università degli Studi «G. D'Annunzio» di Chieti-Pescara) e F. Romoli (Università degli Studi di Pisa), i quali si inscrivono all'interno di un più ampio progetto per la creazione di un «Lexicon paleoslavo-slavo ecclesiastico-russo-italiano del lessico religioso e filosofico-teologico» (cf. Ferro 2018b; Romoli 2018). Si deve infine menzionare uno specifico raggruppamento di contributi di Ferro (2018a; 2019; 2020; 2021), i quali sfruttano le fonti presenti nelle sezioni storiche del NKRJa come materiali di confronto per la conduzione di un'analisi di tipo lessicale-semantico sull'eredità letteraria di Massimo il Greco (ca. 1470-1555/1556).

Come evidente, il NKRJa offre una rappresentazione più che soddisfacente della produzione scritta di area slava orientale, coprendo un vasto arco temporale e presentando un'ampia varietà di generi e forme delle fonti selezionate, le quali risultano peraltro caratterizzate da un eccellente livello di elaborazione tecnica.

La raccolta, tuttavia, trascura tutta la tradizione paleoslava, che, sebbene presenti una matrice meridionale, costituisce un patrimonio comune delle lingue slave nel loro complesso. Una simile mancanza viene almeno parzialmente supplita da una serie di altri *corpora* diacronici, su cui proponiamo di spostare adesso l'attenzione.

3.2 *Corpus Cyrillo-Methodianum Helsingiense*

Il *Corpus Cyrillo-Methodianum Helsingiense* (CCMH) (Corpus Cirillo-Metodiano di Helsinki)[24] è una raccolta di alcuni dei più importanti testi in paleoslavo, approntata presso l'Università di Helsinki tra il 1986 e il 2017 sotto la guida di J. Lindstedt prima e M. Wahlström poi.[25]

La collezione offre una selezione molto ristretta, con cinque testimoni del cosiddetto 'canone paleoslavo' (*Codice Assemaniano*, *Codice Mariano*, *Codice Supraliense*, *Codice Zografense* e *Libro di Savva*, X-XI sec.) e le Vite dei santi Cirillo e Metodio (*Vita Constantini* e *Vita Methodii*), le quali, pur compilate in un periodo arcaico (fine IX-inizio X sec.), presentano una tradizione manoscritta più tarda (XV sec. per la Vita di Costantino, XIII sec. per la Vita di Metodio).

La digitalizzazione delle suddette testimonianze non si è basata sui manoscritti originali, bensì su loro edizioni critiche. Nello specifico, per il *Codice Assemaniano* si è fatto riferimento all'edizione di J. Vajs e J. Kurz,[26] per il *Codice Mariano* a quella di V. Jagić,[27] per il *Libro di Savva* a quella di V.N. Ščepkin.[28] Il testo del *Codice Supraliense* si è basato principalmente sull'edizione di S. Severjanov,[29] integrata con l'edizione in facsimile a cura dell'Accademia bulgara delle Scienze. La trascrizione del *Codice Zografense* si è fondata sulla versione di V. Jagić,[30] unita alle correzioni di L. Moszyński.[31] Per le Vite dei santi Cirillo e Metodio, invece, si rimanda all'edizione curata da F. Grivec e F. Tomšič.[32]

24 `https://korp.csc.fi/download/ccmh-src/`. Il corpus non è consultabile online, ma può venire scaricato in forma di cartella compressa.

25 Nel corso degli anni, la risorsa si è avvalsa della collaborazione di numerosi specialisti nel campo della slavistica, tra cui si ricordano in modo particolare J. Hallaaho, E. Jurkiewicz-Rohrbacher, e J. Sarsila. Le procedure di elaborazione dei metadati del corpus si devono invece a M. Lennes, T. Jauhiainen, U. Dieckmann, e S. Pöyhönen.

26 *Evangeliarium Assemani. 3 Codex Vaticanus slavicus glagoliticus. Editio phototypica cum prolegomenis, textu litteris cyrillicis transcriptio, analysi, annotationibus paleographicis, variis lectionibus, glossario.* Ediderunt Vajs Josef, Josef Kurz, 2 voll., Pragae 1929, 1955.

27 V. Jagić, *Quattuor Evangeliorum versionis palaeoslovenicae Codex Marianus Glagoliticus*, Berlin 1883.

28 V.N. Ščepkin, *Rassuždenie o jazyke Savvinoj knigi*, Sankt-Peterburg 1899.

29 S. Severjanov, *Suprasl'skaja rukopis'. (Editiones monumentorum slavicorum veteris dialecti)*, Sankt-Peterburg 1904 (reprint 1956).

30 V. Jagić, *Quattuor evangeliorum codex glagoliticus olim Zographensis nunc Petropolitanus. Characteribus cyrillicis transcriptum, notis criticis, prolegomenis, appendicibus auctum adiuvante summi ministerii borussici liberalitate edidit*, Berlin 1879.

31 L. Moszyński, *Ze studiów nad rękopisem Kodeksu Zografskiego*, Wrocław-Warszawa-Kraków 1961.

32 *Constantinus et Methodius Thessalonicenses: Fontes*. Recensuerunt et illustraverunt Franciscus Grivec et Franciscus Tomšič, in *Radovi Staroslavenskog Instituta*, 4, Zagreb 1960.

I documenti elettronici contenuti nel corpus non sono pensati per essere completamente auto-esplicativi: le procedure di controllo dei testi, infatti, non sono state ultimate, e alcuni di essi potrebbero contenere errori. Per questo motivo, si raccomanda sempre la consultazione delle fonti elettroniche unitamente alle edizioni tradizionali.

Nessuna delle fonti contenute nel corpus dispone di alcun tipo di *PoS tagging*, né è programmato lo sviluppo di una qualche forma di annotazione grammaticale o morfo-sintattica.

In quanto raccolta di edizioni digitali di singole fonti scrittorie, inoltre, il CCMH consente la sola ricerca di parole e combinazioni di lettere (stringhe) all'interno di ciascuna di queste.

Dati i motivi sopra elencati, nel panorama scientifico contemporaneo il corpus in questione appare una risorsa piuttosto primitiva, e probabilmente ormai superata.

3.3 I *corpora* di paleoslavo e russo antico del *Thesaurus Indogermanischer Text-und Sprachmaterialien*

La collezione di testi paleoslavi del CCMH è ricompresa anche nel *Thesaurus Indogermanischer Text-und Sprachmaterialien* (TITUS) (Thesaurus dei Testi e dei Materiali Linguistici Indoeuropei),[33] progetto scientifico internazionale per la raccolta coordinata di testi originali considerati rilevanti nelle antiche lingue indoeuropee in atto dagli anni '90 del secolo scorso grazie a una collaborazione tra l'Istituto di Linguistica Comparata dell'Università J.W. Goethe di Francoforte sul Meno, l'Istituto del Vicino Oriente antico dell'Università Carolina di Praga, il Dipartimento di Linguistica Generale e Applicata dell'Università di Copenaghen, e il Dipartimento di Filologia Classica e Romanza dell'Università di Oviedo.

Il vasto database TITUS comprende una sezione specificamente dedicata al gruppo delle lingue slave, all'interno della quale sono inclusi due *corpora* storici per la lingua russa: il corpus di paleoslavo (*Old Church Slavonic*) e quello del russo antico (*Old Russian*).[34]

33 https://titus.uni-frankfurt.de/indexe.htm.

34 Poiché questo corpus riunisce in un unico insieme fonti che spaziano notevolmente dal punto di vista cronologico (XI-XVIII sec., cf. *infra*), si può parlare genericamente di 'russo antico'. Con questa espressione ci riferiamo alla lingua scritta utilizzata nelle terre della Rus', prima, e della Moscovia, poi, sottintendendone sia la fase arcaica (*drevnerusskij*, 'slavo orientale antico'), che quella intermedia (*starorusskij*, 'medio-russo') (§3.1.). Quando si parla della storia della lingua letteraria russa, del resto, è sempre necessario tenere presente che il concetto di 'russo' (*russkij*) è definito da quei confini culturali che sono correlati ai nomi 'Rus'' (*Rus'*) o 'Russia' (*Rossija*), pertanto, nel corso del tempo, questa nozione cambia il suo contenuto. Così, per il periodo più antico si può parlare di una lingua letteraria comune a tutto il territorio degli slavi orientali.

Il corpus di paleoslavo, compilato dai medesimi sviluppatori del CCMH,[35] ne contiene gli stessi testi, con la sola aggiunta dei *Fogli di Kiev*, inclusi a propria volta nel canone paleoslavo, e dei *Fogli di Praga*, che del suddetto canone non fanno parte, rappresentando tuttavia l'unica, preziosa, testimonianza diretta della sopravvivenza della tradizione paleoslava nel mondo boemo.

Il corpus del russo antico, sviluppato per merito di un nutrito gruppo di lavoro internazionale,[36] include invece una selezione di scritti composti nelle terre slave orientali (e, in seguito, russe). La collezione è particolarmente varia in relazione alle tipologie testuali, raccogliendo testimoni di generi differenti, tra cui: cronache, narrazioni epiche, trattati di carattere economico-commerciale, testi di carattere giuridico, composizioni di natura didattico-pedagogica. Il corpus spazia notevolmente anche dal punto di vista della cronologia: si va dalla *Povest' vremennych let* (Cronaca degli anni passati), la più antica testimonianza annalistica delle terre slave orientali (XII sec.), allo *Junosti čestnoe zercalo* (Lo specchio onesto della gioventù), un manuale di comportamento per giovani nobili composto per ordine di Pietro I (1717).

Allo stato presente, i testi sono accessibili sul server TITUS nella sola versione HTML, codificati in formato Unicode (UTF-8). Per facilitare la lettura dei dati da parte di ciascun utente, i font Unicode di TITUS sono scaricabili dalla stessa pagina web della risorsa.[37]

Come nel caso del CCMH, i documenti elettronici inclusi nei *corpora* di paleoslavo e russo antico del sistema TITUS non si basano sui testi originali, bensì su loro edizioni critiche; allo stesso modo, le procedure di controllo dei testi non sono state ultimate, perciò i file digitali non sono da considerarsi perfettamente auto-esplicativi, necessitando di un costante raffronto con le loro edizioni tradizionali. Ancora, nessuna delle fonti contenute nelle raccolte menzionate dispone di alcun tipo di *PoS tagging*.

A differenza del CCMH, però, il database TITUS offre maggiori opportunità di impiego per la ricerca. Poiché i *corpora* ivi caricati

Successivamente, invece, il termine 'russo' viene impiegato per riferirsi principalmente all'area granderussa (cf. Uspenskij 2002, 24).

35 Accanto al già citato Lindsted, deve qui essere menzionato J. Schaeken, dell'Università di Leida.

36 S. Pálsson dell'Università d'Islanda, S. Häusler dell'Università di Halle-Wittenberg, V. Alexeev, dell'Università di Francoforte, P. Hendriks e il già menzionato Schaeken dell'Università di Leida, D. Müller and J. Tischler dell'Università di Dresda, I.A. Gerasimov, A.V. Sannikov, M. Jasinskaja da Mosca, e V.A. Baranov, titolare della cattedra di Linguistica dell'Università tecnica statale di Iževsk e responsabile del progetto «*Manuskript». Slavjanskoe pis'mennoe nasledie* («Manuskript». Patrimonio scritto slavo) (cf. *infra*).

37 `https://titus.uni-frankfurt.de/indexe.htm?/unicode/unitest2.htm#TITUUT`.

presentano un'annotazione a livello dei singoli testi, che suddivide le fonti per capitoli, paragrafi, versi, pagine, e righe, è contemplata la possibilità di consultazione di un dato passaggio testuale, per mezzo di un apposito modulo di inserimento (*text location query form*). All'utente viene richiesto innanzitutto di scegliere la lingua di interesse;[38] vengono dunque fatti indicare il testo desiderato e uno specifico passaggio all'interno di questo.

Il sistema TITUS permette inoltre di interrogare i *corpora* tramite un motore di ricerca per parole (*word search engine*), il quale è accessibile in differenti modi e a differenti livelli.

In primo luogo, ogni parola all'interno di un dato testo è collegata al motore di ricerca: facendo doppio clic su essa, il motore di ricerca ne rintraccia tutte le occorrenze (comprese le varianti ortografiche minori) all'interno del dato testo. I risultati vengono visualizzati in stile KWIC all'interno di una tabella, che può essere utilizzata come punto di partenza per ulteriori *query*.

Tutte le parole generate all'interno della tabella dei risultati, infatti, sono a propria volta collegate al motore di ricerca: cliccando su di esse, ne viene avviata la ricerca automatica all'interno di tutti i testi disponibili sul database di TITUS e in tutte le varietà linguistiche corrispondenti. I risultati sono ancora una volta presentati nel formato KWIC e ordinati in base a forme di parola e datazione del testo.

Ogni testo, poi, è dotato di un ulteriore modulo di ricerca, che può essere selezionato tramite un link nella cornice destra della finestra di testo (*switch to word index*). Il modulo di inserimento, che viene visualizzato nello stesso riquadro una volta attivato, contiene diversi pulsanti di opzione per determinare il dominio della *query* (ad esempio, varietà linguistiche, ambito dei testi da includere, condizioni di ricerca) e una casella in cui è possibile inserire la parola (o le parole) da cercare.

In modo del tutto analogo, l'accesso al motore di ricerca è altresì disponibile da una pagina speciale di interrogazione del database (*database query form*). Anche in questo caso, è anzitutto necessario selezionare la lingua (o le varianti linguistiche) a cui si applica la *query*, quindi inserire la forma linguistica desiderata.

Il database TITUS si presenta come uno strumento utile ai fini del reperimento di materiale linguistico e lessicale, che tuttavia non può venire automaticamente analizzato in termini di caratteristiche grammaticali e morfosintattiche.

38 A tal proposito, è opportuno sottolineare che per consultare i testi del corpus di paleoslavo bisogna scegliere l'opzione «lingua bulgara», mentre per quelli del corpus del russo antico è necessario selezionare l'opzione «lingua russa».

3.4 *Trondheim-Sofia Corpus of Old Slavic*

Quando si parla delle risorse elettroniche preposte alla raccolta delle testimonianze letterarie slave più antiche, bisogna menzionare, almeno per il suo suolo pionieristico, il *Trondheim-Sofia Corpus of Old Slavic* (Corpus di paleoslavo di Trondheim-Sofia).[39]

Frutto di un progetto di cooperazione tra l'Università «S. Clemente di Ocrida» di Sofia e l'Università Norvegese di Scienza e Tecnologia di Trondheim, svoltosi in un periodo compreso tra il 1996 e il 1999 sotto la direzione di R. Pavlova per parte bulgara e J. Ragnar Hagland per parte norvegese,[40] la risorsa è stata resa accessibile online dal 2002.

Il corpus include i più antichi testi copiati in territorio slavo orientale da manoscritti antico-bulgari nell'XI secolo, rappresentando uno dei primi esemplari di raccolta elettronica di queste testimonianze testuali arcaiche.

A differenza di quanto osservato per i *corpora* precedentemente citati, i membri del gruppo di lavoro del *Trondheim-Sofia Corpus* hanno proceduto a digitalizzare i manoscritti copiandoli direttamente dagli originali, così da garantire la massima precisione, e pervenire all'individuazione di eventuali refusi presenti nelle edizioni preesistenti.

Tale operazione è stata condotta sulla base di microfilm e fotocopie accessibili presso le Università di Sofia e di Trondheim, mentre le edizioni disponibili dei manoscritti sono state impiegate come ulteriore fonte per il confronto con gli originali.

Il corpus include i testi delle più antiche testimonianze slave, comprensivi di interpolazioni e glosse. Non vengono incluse nella raccolta le note a margine dei manoscritti, né vengono ricostruiti frammenti di manoscritti successivi, che erano stati rilegati insieme a quelli dell'XI secolo.

I manoscritti vengono rappresentati nel database con le loro caratteristiche codicologiche autentiche: il testo viene ricostruito per fogli, righe e colonne; ciascuna delle suddette parti è appositamente designata con numeri arabi o lettere latine.

È stata riservata un'attenzione particolare alle caratteristiche paleografiche dei testi: tutti i caratteri presenti nei manoscritti sono stati ricostruiti; le parole sovrascritte vengono ricostruite e contrassegnate dal simbolo '+' o, quando possibile, inserite direttamente

39 http://www.hf.ntnu.no/SofiaTrondheimCorpus.

40 Il gruppo di lavoro dell'Università di Trondheim includeva L. Hellan, M. Vulchanova, V. Vulchanov, di formazione umanistico-linguistica, e due ingegneri informatici, K. Rye Ramberg e B. Grønnesby. Dall'Università di Sofia, invece, hanno partecipato alla realizzazione del corpus I. Kasabov, S. Bogdanova, T. Raleva, R. Stankov, V. Željazkova e T. Doseva.

all'interno della riga e segnalate dal medesimo contrassegno (cf. Pavlova, s.d.).

Il *Trondheim-Sofia Corpus*, purtroppo, ha potenzialità di impiego molto limitate. La risorsa, infatti, non è interrogabile: i testi ivi presentati possono semplicemente essere letti e copiati, se si dispone sul proprio computer del font appropriato (che tuttavia risulta ormai obsoleto).

Nei primi anni duemila però, Pavlova e il gruppo di filologi dell'Università di Sofia hanno collaborato al progetto «*Manuskript». Slavjanskoe pis'mennoe nasledie* («Manuskript». Patrimonio scritto slavo),[41] approntato presso l'Università statale dell'Udmurtia sotto la direzione di V.A. Baranov,[42] contribuendo alla realizzazione della *kollekcija slavjanskich rukopisej XI veka* ('collezione dei manoscritti slavi dell'XI secolo'), dove sono stati trasferiti tutti i testi del *Trondheim-Sofia Corpus*. Vale la pena, dunque, spostare l'attenzione su questa risorsa.

3.5 *«Manuskript». Slavjanskoe pis'mennoe nasledie*

Il sistema *Manuskript* include le copie elettroniche di codici paleoslavi e slavi orientali, e di frammenti testuali dei secoli XI-XV. La parte più cospicua del corpus storico è costituita da frammenti di testi liturgici della Rus' (*Menee, Stichiraria, Triodi, Profetologi, Vangeli*), a cui si affiancano una serie di codici paleoslavi, non soltanto cirillici, ma anche glagolitici, per un totale di 3,5 milioni di forme di parola.

Al fine di facilitare la ricerca e l'organizzazione dei materiali, il corpus è suddiviso in differenti collezioni: la *kollekcija glagoličeskich rukopisej X-XI vv.* ('collezione dei manoscritti glagolitici del X-XI sec.') include la versione elettronica di esemplari manoscritti paleoslavi (editi e inediti) in alfabeto glagolitico; sette manoscritti glagolitici rivenuti sul Monte Sinai sono stati separatamente raggruppati nella *kollekcija sinajskich glagoličeskich rukopisej XI v.* ('collezione dei manoscritti glagolitici sinaitici dell'XI sec.'), realizzata dall'Università di Vienna col supporto della fondazione austriaca per la ricerca scientifica (FWF), mentre per i testi completi dei circa venti manoscritti superstiti prodotti sul territorio della Rus' nel XI secolo si rimanda alla sopra menzionata *kollekcija slavjanskich rukopisej XI veka*.

Si devono menzionare, poi, le ricche raccolte di testi biblici e liturgici paleoslavi e della Rus': la *kollekcija slavjanskich Evangelij XI-XIV*

41 http://mns.udsu.ru/.

42 A lui si affianca, in qualità di responsabile del gruppo dei programmatori, A.N. Mironov. Il progetto vanta anche della collaborazione dei seguenti studiosi: O.O. Žolobov dell'Università Federale di Kazan', H. Miklas dell'Università di Vienna, e A.A. Pičchadze dell'Istituto di lingua russa V.V. Vinogradov di Mosca, già incaricata dell'elaborazione del sotto-corpus *drevnerusskij* all'interno del NKRJa (cf. *supra*).

vv. (‘collezione dei *Vangeli* slavi dei secoli XI-XIV’); la *kollekcija slavjanskich Minej XI-XIV vv.* (‘collezione delle *Menee* slave dei secoli XI-XIV’); la *kollekcija slavjanskich Triodej XI-XIV vv.* (‘collezione dei *Triodi* slavi dei secoli XI-XIV’); la *kollekcija slavjanskich Stichirariej XII-XIV vv.* (‘collezione degli *Stichiraria* slavi dei secoli XII-XIV’); la *kollekcija slavjanskich Parimejnikov XII-XIV vv.* (‘collezione dei *Profetologi* slavi dei secoli XII- XIV’).

Completano il corpus le seguenti collezioni: *kollekcija slavjanskich učitel’nich proizvedenij* (‘collezione dei testi di tipo didattico’) dell’XI-XIV sec., *kollekcija drevnerusskich perevodov* (‘collezione delle traduzioni della Rus’’), *kollekcija russkich letopisej* (‘collezione delle cronache slave orientali’), *kollekcija russkich žitij* (‘collezione delle vite slave orientali’).

Nel novembre 2011 sul portale *Manuskript* è stato creato anche un corpus delle opere di M.V. Lomonosov (1711-1765), che contiene più di un milione di parole, distribuite in 1150 testi.

Tutti i testi sono immediatamente accessibili nella versione demo, che consente la visualizzazione dei primi due fogli di un manoscritto e di cinque soli frammenti testuali. Per ottenere l’accesso alla versione completa delle risorse contenute nel corpus è necessario compilare e inviare una richiesta di autorizzazione, che deve essere approvata dall’amministrazione del portale. La registrazione è gratuita.[43]

Il sistema *Manuskript* offre una rappresentazione completa dei testi oggetto di studio, preservando l’integrità delle caratteristiche essenziali dei manoscritti. Il software consente la trascrizione e traslitterazione del testo, allo scopo di creare materiali di riferimento per edizioni digitali e a stampa dei manoscritti.

Il reale vantaggio del sistema *Manuskript* consiste nella possibilità di memorizzare ed elaborare testi di qualsiasi livello di complessità grafico-ortografia, in qualsivoglia lingua. A tal fine, il lavoro sui manoscritti ha dovuto seguire alcune tappe fondamentali: in primo luogo, si è puntato all’individuazione nei testi di unità di ogni dimensione, dal carattere al frammento; quindi, è stata proposta un’organizzazione gerarchica delle unità di cui sopra (carattere, forma di parola, sintagma, testo; carattere, linea, pagina, foglio, manoscritto); infine, è stato memorizzato un elenco potenzialmente illimitato delle caratteristiche degli elementi presenti nei testi, considerati significativi sul piano degli studi testuali, linguistici, codicologici, paleografici e archeografici (Baranov et al. 2004, 11.2).

Grazie a moduli specializzati, i testi sono suddivisi in unità di diverso tipo (linguistiche, cronologiche, stilistiche ecc.), che possono essere messe in relazione le une con le altre secondo un dato vincolo

43 Il modulo per la registrazione si trova sulla pagina di accesso al portale: `http://mns.udsu.ru/mns/portal.main?p1=18&p_lid=1` (*registracija*).

di cardinalità (uno a uno/uno a molti). Ciascuna unità, inoltre, è collegata a un'entrata del dizionario: i dizionari consentono una considerevole agevolazione delle procedure di descrizione ed elaborazione delle unità, rendendo possibile, allo stesso tempo, l'esame dei testi in termini di invarianti linguistiche e testuali (Baranov et. al. 2004, 11.4).

Al fine di memorizzare i caratteri specifici della lingua paleoslava e slava orientale, sono stati adottati, in conformità con lo standard Unicode, il set di caratteri UTF8LAPREXT1 e la famiglia di fonts *Menaion*, scaricabile dall'utente.

Le collezioni del corpus sono annotate a vari livelli. In primo luogo, è stata realizzata un'annotazione di tipo analitico (*analitičeskaja razmetka*), la quale consente di individuare all'interno dei documenti (sia testi che manoscritti) frammenti che differiscono gli uni gli altri secondo date caratteristiche. Vengono distinte, ad esempio, parti copiate da copisti differenti o eventuali aggiunte successive alla forma originale del manoscritto.

A differenza del *markup* strutturale, il *markup* analitico consente la descrizione di certe peculiarità di realizzazione del manoscritto, che di solito non vengono evidenziate formalmente, ma possono influire su certe caratteristiche grafico-ortografiche e linguistiche del testo, legate al suo contenuto (si pensi, ad esempio, alle registrazioni metereologiche nelle cronache), oppure di origine testuale (come nel caso della presenza di forme di notazione musicale all'interno delle composizioni innografiche) (Baranov 2015, 46). Grazie all'annotazione analitica, si rende possibile non soltanto la trasmissione dei dettagli grafici e strutturali del manoscritto originale, ma anche la registrazione di informazioni su aggiunte, modifiche e omissioni, le quali si rivelano essenziali alla critica testuale.

Al fine di ampliare le potenzialità di ricerca del corpus, i programmatori di *Manuskript* hanno preso parte all'elaborazione di un sistema automatico di analisi morfologica delle forme di parola, che ha alla base l'*Elektronnyj Grammatičeskij Slovar' Drevnerusskogo Jazyka* (Dizionario Grammaticale Elettronico Della Lingua Antica Russa), un database interno al software contenente unità linguistiche, i loro significati e i loro rapporti reciproci. In questo modo, un editor specializzato (*OldEd*) stabilisce le relazioni tra le unità del corpus e le voci del dizionario, assegnando alle prime la propria categoria grammaticale (cf. Baranov, Gulina, Mironov 2010).

I risultati così ottenuti, però, non sono scevri da criticità. In primo luogo, bisogna sottolineare che il soprammenzionato database è ben lungi dall'essere completo: questo comporta che non tutte le parole presenti nei testi del corpus siano già state registrate nel dizionario elettronico, e che quindi non possano ricevere in automatico la propria etichetta grammaticale e morfosintattica. In secondo luogo, non si è ancora proceduto all'unificazione delle varianti ortografiche,

pertanto il sistema è in grado di riconoscere un dato termine soltanto qualora sia scritto in forma 'canonica', ossia con quell'ortografia normalizzata con la quale la suddetta forma di parola è stata registrata sul database del portale. Infine, il sistema di annotazione morfosintattica non è ancora in grado di disambiguare le forme omonime, molto numerose nella lingua paleoslava (Moldovan 2014, 29).

La ricerca e il reperimento dei materiali per l'analisi all'interno del sistema *Manuskript* è garantito da una serie di moduli specializzati, che consentono di effettuare un'accurata campionatura del materiale disponibile.

Dalla pagina principale del portale, all'interno della sezione «strumenti» (*instrumenty*) si ha accesso al vero e proprio corpus storico (*istoričeskij korpus*), composto dalla totalità delle copie elettroniche dei manoscritti paleoslavi e slavi orientali (XI-XV sec.) presenti sul database del sistema.

La ricerca viene effettuata per mezzo di due possibili moduli di interrogazione testuale: uno «multi-testo» (*mnogotekstovaja zaprosnaja forma*) e uno «mono-testo» (*odnotekstovaja zaprosnaja forma*).

Il modulo di interrogazione multi-testo consente, in primo luogo, di lavorare con frammenti di un singolo manoscritto e, in seconda istanza, di confrontare i risultati ottenuti con i dati provenienti da fogli diversi della stessa fonte, o addirittura da altri testi.

L'elaborazione di questo tipo di interrogazione muove dalla selezione delle fonti su cui effettuare l'analisi. Accedendo al corpus, l'utente viene automaticamente indirizzato sulla pagina di «ricerca semplice» (*prostoj poisk*), per cui è tenuto a inserire una parola chiave (ad esempio, *apostol, žitie* ecc.) al fine di rintracciare i materiali desiderati. Per ottenere risultati più raffinati, però, è possibile passare alla modalità di «ricerca avanzata» (*rasširennyj poisk*), che consente di indicare contemporaneamente più criteri per il reperimento delle fonti. Con questa modalità di ricerca, l'utente è anzitutto chiamato a scegliere la tipologia dell'unità da ricercare, selezionandola tra le tre possibili di «manoscritto/edizione a stampa» (*rukopis'/pečatnoe izdanie*), «testo/opera» (*tekst/proizvedenie*) e «frammento» (*fragment*); ne devono quindi venir scelte le proprietà desiderate (tra le quali si menzionano, ad esempio, autore, alfabeto, autenticità, genere, data e luogo di edizione, titolo e alcune altre). Il significato che tali proprietà devono acquisire viene suggerito dal sistema stesso: è sufficiente digitare il simbolo '%' all'interno della casella «significato» (*značenie*) per ottenere una lista di tutti i possibili valori della proprietà selezionata.

Lavorando in questa modalità, è via via possibile aggiungere alla ricerca nuovi criteri, così da ottenere risultati più precisi e conformi alle esigenze dell'utente. Si potrebbe, ad esempio, chiedere genericamente al sistema di trovare tutti i manoscritti in alfabeto cirillico contenuti nel corpus, ma si potrebbe anche affinare la ricerca, limitando i risultati sulla base di criteri quali data e luogo di composizione

del manoscritto. Seguendo una differente strategia, l'utente potrebbe inoltre scegliere di lavorare su una selezione testuale che includa differenti edizioni manoscritte di un medesimo testo. Sulla base di questo criterio, del resto, sono stati elaborati due moduli di *corpora* paralleli all'interno delle collezioni dei vangeli e delle cronache.

Ultimata la selezione dei materiali, è possibile effettuarvi sopra ricerche linguistiche. A tal proposito, è innanzitutto necessario digitare una parola (o parte di essa, affiancata dal simbolo '%') e definire se debba venire cercata come lemma, forma o espressione regolare. Poiché il corpus dispone dell'annotazione morfosintattica, è possibile scegliere il valore grammaticale della parola ricercata, selezionandolo da un apposito elenco.

All'utente si apre un ampio ventaglio di modalità di visualizzazione dei dati: si può scegliere di visualizzare i risultati come testo (sia nella forma originale, ovvero priva di suddivisione in parole, sia nella forma convertita, e cioè con la separazione delle parole); come indice di forme di parola o di lemmi, oppure come concordanze. Negli ultimi due casi viene fornito, tramite collegamento ipertestuale, l'indirizzo di collocazione della parola nel testo: una stringa numerica ne indica foglio, pagina, colonna, riga e posizione all'interno della riga.

Il modulo di interrogazione multi-testo consente di effettuare un numero illimitato di *query* e di ottenere campioni di interesse, che possono essere confrontati dal ricercatore, coinvolgendo nell'analisi un'ingente quantità di dati definiti sulla base di parametri vari.

Tutto questo rende *Manuskript* una risorsa particolarmente preziosa non soltanto nell'ambito delle ricerche linguistiche, ma anche in quello degli studi filologici. Ne forniscono un chiaro esempio alcuni lavori di O.V. Zuga (2009; 2010), dell'Università dell'Udmurtia, che hanno saputo sfruttare le potenzialità del corpus storico per offrire una lettura comparativa delle varianti linguistiche (soprattutto morfologiche e lessicali) di alcuni vangeli slavi orientali del XII-XIII secolo.

Ogni testo contenuto all'interno del sistema *Manuskript* rappresenta un database a sé stante: per questo motivo, è possibile condurre anche ricerche su singole fonti. È sufficiente accedere, dalla pagina principale del portale, alla sezione «collezioni», selezionare la raccolta desiderata e scegliere la voce «nei moduli di interrogazione dei testi» (*v zaprosnye formy tekstov*). In questo modo, sarà possibile condurre analisi linguistiche nelle modalità di cui sopra all'interno di un'unica fonte.

Grazie ai moduli di interrogazione testuale, il sistema *Manuskript* può essere facilmente impiegato per la conduzione di studi a tutti i livelli linguistici (lessico, morfologia, sintassi, ma anche ortografia e stilistica).

Le potenzialità di *Manuskript*, tuttavia, non si esauriscono qui. All'interno del sistema sono infatti disponibili due ulteriori moduli, «Statistica» (*Statistika*) e «*N-Grams*» (*N-Grammy*), che consentono di

condurre analisi di tipo quantitativo-distributivo. In particolar modo, il modulo statistico consente di identificare la distribuzione di frequenza delle unità linguistiche all'interno di uno o più documenti, mentre *N-Grams* individua le collocazioni o, al contrario, casi rari di co-occorrenza tra parole (cf. Baranov, Dubovcev 2012).

Una volta selezionati uno o più documenti del corpus, il modulo statistico permette di inserire la forma di parola desiderata (o parte di essa, affiancata dal simbolo '%'), e indicare il tipo di passo (pagina, foglio, frammento) su cui effettuare il calcolo, specificandone la lunghezza. Il risultato che si ottiene è un grafico della quantità assoluta o relativa dell'unità ricercata all'interno di uno o più testi.

Una delle modalità di funzionamento maggiormente proficue del modulo statistico è il confronto tra più manoscritti contenenti gli stessi testi (è il caso, ad esempio, dei vangeli), allineati per frammenti (versetti). I grafici sovrapposti che si ottengono come output consentono di individuare intervalli di testo con una medesima o molto simile frequenza d'uso di una data unità linguistica, oppure parti in cui le copie differiscono notevolmente tra loro, ricorrendo all'impiego di unità alternative.

Attraverso il modulo *N-Grams*, invece, l'utente può identificare le associazioni di parole più frequenti, o più statisticamente significative all'interno dei testi del corpus. In questo caso, la *query* si fonda su una serie di parametri quantitativi e statistici (misure associative, quali *T-score*, *Log-likelihood* e altre), e su caratteristiche strutturali (numero di componenti, loro distanza) e linguistiche (esclusione delle parole vuote, specifici significati grammaticali delle parole, ecc.) delle associazioni. Gli elenchi che si ottengono, ordinati sulla base dei valori della misura statistica selezionata, consentono di trarre conclusioni sul livello di produttività e stabilità dei casi di co-occorrenza presi in esame.

La presenza del modulo statistico e *N-Grams* all'interno del corpus storico del portale *Manuskript* è una caratteristica particolarmente innovativa, che permette di approcciare all'analisi dei testi manoscritti per mezzo di metodi e strumenti della linguistica dei *corpora* propriamente detti (Baranov 2015, 54). Per questo motivo, sulla base delle collezioni del sistema *Manuskript*, si è sviluppato un filone molto attivo di ricerche *corpus-based*, fondate su metodologie quantitativo-statistiche.

Hanno aperto questo fronte alcuni lavori di Baranov, già citato nel ruolo di responsabile dello stesso progetto *Manuskript.* Sfruttando il modulo *N-Grams* e le misure statistiche ivi disponibili, lo studioso ha cercato di individuare a più riprese le parole/combinazioni di parole più statisticamente significative all'interno di *corpora* scelti. La sua attenzione si è concentrata sui manoscritti dell'Apostolo (Baranov 2017a; 2017b), sulle copie slave orientali della *Parenesis Efrema Sirina* (Parenesi di Efrem Siro) (Baranov 2018), sui testimoni

delle *Menee* dell'XI secolo (Baranov 2019b), sulla collezione dei codici glagolitici (Baranov 2019c) e su uno specifico sotto-corpus di cronache della Rus': *Lavrent'evskaja letopis'* (Cronaca Laurenziana), *Ipat'evsja letopis'* (Cronaca Ipaziana) e *Radzivillovskaja letopis'* (Cronaca di Radziwiłł) (Baranov 2019a; 2020).

Uno dei più recenti progetti di Baranov (2022) è quello che riguarda l'impiego di metodi statistici sui dati linguistici dei *corpora* di *Manuskript* per la creazione di un dizionario distributivo elettronico *corpus-based*. La risorsa, fondata sul principio della dipendenza dell'ambiente lessicale di un dato termine dalle caratteristiche di quest'ultimo, offre la possibilità di valutare la compatibilità contestuale delle parole, generando grafici di campionamento degli analoghi semantici, tematici e associativi dell'unità lessicale desiderata.

Devono essere menzionati, inoltre, alcuni lavori di O.F. Žolobov dell'Università Federale di Kazan', che è ricorso all'impiego dei *corpora* di *Manuskript* per la conduzione di analisi distributivo-quantitative con fini lessicali-semantici. Nei contributi dell'autore vengono indagate le relazioni di significato esistenti tra i verbi del sapere *vědĕti, vědati, znati* ('sapere', 'conoscere') nella lingua slava orientale antica (Žolobov, Baranov 2021); viene offerta una descrizione linguistico-statistica delle trasformazioni storiche del gruppo lessicale *životŭ – žiznĭ – žitie* ('vita') nella tradizione scrittoria della Rus' (Žolobov, Baranov 2022); viene identificata la dinamica dei cambiamenti nella serie lessicale e nella struttura semantica dei lessemi *branĭ – voina – ratĭ* ('guerra') all'interno di diverse fonti testuali (Žolobov 2022).

Come si evince da questo pur breve resoconto, i *corpora* del portale *Manuskript* trovano un impiego attivo all'interno del panorama scientifico contemporaneo. Adottando un approccio che esula dalle tradizionali pratiche degli studi linguistici e filologici, a favore di un più sistematico impiego delle nuove tecnologie informatiche e dei metodi della linguistica computazionale, il sistema *Manuskript* si presenta come una delle risorse più complete e meglio strutturate nell'ambito dei *corpora* diacronici per la lingua russa.

3.6 *Regensburgskij diachroničeskij korpus russkogo jazyka*

Tentativi di elaborazione di un sistema di annotazione automatica per i testi in paleoslavo e slavo orientale, simili a quelli che hanno riscontrato successo nel sistema *Manuskript*, erano stati avanzati anche nel contesto del *Regensburgskij diachroničeskij korpus russkogo jazyka* (RRuDi) (Corpus diacronico di Ratisbona della lingua russa),[44] elaborato presso la Facoltà di Lingue Slave dell'Università di

44 https://www.slawistik.hu-berlin.de/de/member/meyerrol/subjekte/rrudi.

Ratisbona in collaborazione con l'Università Humboldt di Berlino, sotto la guida di R. Meyer.[45]

Tra il 2013 e il 2016, il corpus era stato sviluppato come database per il progetto *Korpuslinguistik und diachrone Syntax: Subjektkasus, Finitheit und Kongruenz in slavischen Sprachen* (Linguistica dei *Corpora* e Sintassi Diacronica: Caso Nominativo, Finitezza e Accordo nelle Lingue Slave), supportato dalla Fondazione tedesca per la ricerca (DFG).

A seguito di questa promettente fase di sviluppo, però, le procedure di elaborazione del corpus hanno subito un'importante battuta d'arresto, così che, ad oggi, la risorsa non risulta più accessibile.[46]

Quando ancora in funzione, la raccolta consisteva di due parti: una versione vecchia e una nuova. La vecchia versione comprendeva undici testi in slavo orientale antico (XI-XIV sec.) e 'medio-russo' (XV-XVIII sec.); la nuova contava un totale di più di 100 000 lemmi, distribuiti all'interno di testi databili in un periodo compreso tra il XII e il XVI secolo (Mitrenina 2014, 457-58).

Il corpus RRuDi doveva costituire, nell'idea dei suoi sviluppatori, uno strumento per la conduzione di ricerche sull'evoluzione diacronica di vari sottotipi di soggetti nulli e costruzioni riflessive in russo. Questo ambizioso obiettivo richiedeva che la risorsa fosse annotata in maniera piuttosto dettagliata, dal momento che il sopraccitato tipo di ricerca necessitava non solo di informazioni formali, ma anche semantiche e contestuali (si pensi, ad esempio, ai casi di coreferenzialità).

L'intero processo di annotazione era dunque stato realizzato per mezzo di *Gate*, software *Java* in grado di combinare vantaggiosamente strumenti di annotazione automatica già sviluppati per la moderna lingua russa (*tagger*), e una buona dose di annotazione manuale (soprattutto negli estratti testuali più ampi). Tale tecnologia aveva fatto sì che tutti i testi del corpus avessero una propria annotazione morfosintattica.

Da qui, possiamo dedurre che RRuDi, così strutturato, avrebbe potuto costituire un più che valido strumento per la conduzione di ricerche di tipo sintattico sui testi slavi orientali: il suo mancato sviluppo, però, lo rende una risorsa inutilizzabile.[47]

45 Si ricordano i seguenti ulteriori collaboratori per parte tedesca: E. Hansack, V. Wald, I. Parchomenko e U. Yažinova. Hanno altresì preso parte all'elaborazione della risorsa alcuni partner internazionali: S. Perić Gavrančić e M. Horvat dell'Istituto di Lingua e Linguistica croata di Zagreb; H. Heckhoff dell'Università di Tromsø; I. Maier dell'Università di Uppsala; Je. Karpilovs'ka e O. Lazarenko dell'Accademia Nazionale delle Scienze dell'Ucraina; O. Nika, N. Darčuk dell'Università statale di Kiev; W.M. Mojsijenko dell'Università Statale Ivan Franko di Žitomir.

46 Eventuali tentativi di accesso dal sito dell'Università Humboldt di Berlino restituiscono un errore.

47 Per una breve panoramica delle modalità di ricerca possibili su RRuDi si può far riferimento a Mitrenina (2014, in particolare 457-458). La studiosa, infatti, sembra aver potuto accedere al corpus prima che venisse dismesso. Cf. anche Meyer (2005).

3.7 *Diachronen korpus na bălgarski ezik (IX-XVIII vv.)* e il sistema *Histdict*

Tra le risorse elettroniche che si propongono di conservare la tradizione paleoslava è infine da annoverarsi il *Diachronen korpus na bălgarski ezik IX-XVIII vv.* (Corpus diacronico della lingua bulgara dei secoli IX-XVIII).[48] Cuore pulsante del sistema *Histdict*,[49] infrastruttura per la ricerca elettronica sul patrimonio scritto bulgaro di epoca medievale, la raccolta è stata sviluppata a partire dal 2011 presso l'Università «S. Clemente di Ocrida» di Sofia, sotto la guida di A.M. Totomanova e T. Slavova.

Il corpus, di fatto, si presenta come una raccolta dell'eredità linguistica e lessicale della tradizione scrittoria bulgara, colta nella sua evoluzione storica dal Medioevo alla prima età moderna. Al suo interno, tuttavia, si trovano materiali rilevanti nel campo degli studi di paleoslavistica, medievistica nel loro complesso.

Il corpus contiene innanzitutto le fonti scrittorie del 'secolo d'oro' della letteratura paleoslava/paleobulgara (fine IX-inizio XI sec.), rappresentate dalle opere di Clemente di Ocrida (anni Trenta del IX sec.-916), Giovanni Esarca (IX-X sec.) e Costantino di Preslav (metà IX-inizio X sec.), le quali sono da considerarsi patrimonio comune del Medioevo slavo meridionale e orientale.

Si trovano parimenti inclusi nella raccolta gli scritti del patriarca Eutimio di Tărnovo (1327 ca.-1402 ca.) e di Costantino di Kostenec (1380/90-dopo il 1432), che giocano un ruolo cruciale nella storia del mondo bizantino slavo, in veste di protagonisti di quel movimento di rinnovamento spirituale che R. Picchio (1958) definisce «rinascita slava ortodossa» (cf. Marcialis 2007, 63-4).[50]

48 `https://histdict.uni-sofia.bg/textcorpus/list`.

49 Oltre al corpus storico, il sistema include le seguenti risorse lessicografiche: *starobălgarski rečnik*, vocabolario digitalizzato di antico bulgaro per un totale di 10 500 entrate; *srednovekoven grăcko-bălgarski rečnik*, dizionario inverso greco-antico bulgaro; *istoričeski rečnik*, dizionario storico della lingua bulgara (in fase di sviluppo); *Rečnik na ezika na patriarch Evtimij* (Dizionario della lingua del patriarca Eutimio) e *Terminologičen rečnik na Joan Ekzark* (Vocabolario terminologico di Giovanni Esarca), due dizionari storici sincronici. Nel sistema sono altresì inclusi alcuni strumenti per l'elaborazione delle risorse digitali: si ricordano, in particolare, un motore di ricerca e la tastiera virtuale a esso correlata, un dizionario grammaticale e un *tagger* morfologico semi-automatico e il font Unicode per l'antico bulgaro e il relativo convertitore. Cf. Ganeva (2018); Totomanova (2012; 2021).

50 Sotto l'influsso dell'esicasmo bizantino che andava diffondendosi nei Balcani, nel XIV secolo si percepì la necessità di mettere in atto una restaurazione puristica della lingua. In area bulgara si cominciò a elaborare una forma dello slavo ecclesiastico che, «recuperando il modello più antico, identificò con maggiore sicurezza una norma linguistica unitaria, con caratteristiche slavo-meridionali, e favorì una vasta opera di "correzione dei libri" allo scopo di preservare integra l'ortodossia» (Garzaniti 2021). Questo fenomeno fu gravido di conseguenze per tutto il mondo bizantino slavo: quando lo standard linguistico 'rinnovato' e 'purificato' si diffuse in area slavo-orientale, secondo un processo tradizionalmente noto come 'seconda influenza slava meridionale', «venne a formarsi uno slavo ecclesiastico, che pur conservando alcuni tratti fonetici e

Merita una menzione separata, invece, il cosiddetto *Archivski Chronograf* ('Cronografo dell'Archivio'), una sezione speciale del corpus destinata alla conservazione della più antica versione dei testi biblici dell'Ottateuco e del Libro dei Re, tradotti in Bulgaria durante il regno di Simeone il Grande (893-927).

La prima versione del *Diachronen korpus na bălgarski ezik IX-XVIII vv.* è stata ottenuta a seguito della conversione in Unicode (*Cyrillica Bulgarian 10U*)[51] di testi già digitalizzati per mezzo di un convertitore automatico e del loro conseguente caricamento sul sistema. La risorsa è stata poi gradualmente dotata di strumenti maggiormente sofisticati: la versione del convertitore in uso al momento dispone di un numero più grande di funzionalità e consente di modificare l'aspetto dei testi che vengono preparati per la pubblicazione, se necessario (cf. Totomanova 2019).

Tutte le fonti del corpus sono etichettate per mezzo di metadati, che riportano alcune informazioni di base (autore, titolo, data della copia manoscritta di riferimento, genere e tipo di ortografia) e una serie di altri dettagli più specifici (indicazione della natura tradotta o originale del testo, data dell'eventuale traduzione o di eventuali copie successive della fonte manoscritta, nome e firma presenti sul manoscritto, luogo di conservazione del manoscritto, edizione di riferimento, pagine identificativo del documento).

Il software della risorsa dispone di diverse funzionalità per l'inserimento di commenti codicologici, paleografici e testologici: le note a piè di pagina sono segnalate in giallo, le varianti di lettura in blu, mentre le parole contrassegnate sia per le note a piè di pagina che per le varianti di lettura appaiono in verde. Simili strumenti si rivelano particolarmente preziosi, consentendo di produrre edizioni elettroniche dei testi, dotate di un apparato critico adeguato.

Il *Diachronen korpus na bălgarski ezik IX-XVIII vv.* è dotato di un proprio sistema di annotazione morfologica, sviluppato a più riprese. In primo luogo, sfruttando il già esistente *tagset* del corpus *BulTreeBank*,[52] A.M. Totomanova, T. Slavova e G. Ganeva sono pervenute all'elaborazione di un *tagset* specifico per le testimonianze letterarie bulgare di epoca medievale (cf. Totomanova, Slavova, Ganeva 2015): le più di duemila etichette ivi contenute sono in grado di rendere conto delle

morfologici slavo-orientali finì per affermarsi nel corso dei secoli come normativo in tutta l'area ortodossa (soprattutto dopo l'occupazione turca dei Balcani)» (Garzaniti 2021).

51 Il font è in formato *True Type*, può essere scaricato dall'utente comune e installato su qualsiasi dispositivo ad uso personale.

52 Il progetto *BulTreeBank*, sviluppato in seno all'Istituto di Tecnologie dell'Informazione e della Comunicazione dell'Accademia Bulgara delle Scienze, offre un ricco set di alberi sintattici (*treebank*) per la moderna lingua bulgara. Per maggiori informazioni, si rimanda alla pagina web della risorsa: `http://bultreebank.org/bg/`.

possibili varianti fonetiche in fine di parola, particolarmente frequenti nei testi slavi di diverse redazioni (Totomanova 2017, 235).

Oltre al *tagset*, poi, è stato elaborato un dizionario storico della lingua bulgara di epoca medievale (*istoričeski rečnik*, cf. *supra*), costituito da tabelle contenenti i paradigmi flessivi di nomi, pronomi e verbi, che sono state caricate sul sito di accesso al corpus alla sezione «forme di parola» (*slovoformy*). È obiettivo primario degli sviluppatori della risorsa sfruttare i materiali del dizionario grammaticale per giungere all'elaborazione di un sistema di annotazione automatica (Totomanova 2021, 8).

Il corpus non dispone di un motore di ricerca proprio, ma sfrutta quello comune a tutte le risorse del sistema *Histdict*, cui si accede dalla pagina principale, alla sezione «ricerca» (*tărsene*). Allo stato attuale, l'utente ha la possibilità di cercare esclusivamente forme di parola, attraverso l'inserimento di stringhe per mezzo di una tastiera virtuale.

In risultato all'elaborazione di una data *query*, il motore di ricerca è in grado di mostrare l'elenco dei testi in cui si trova la forma di parola desiderata. Per visualizzarne il contesto d'uso, tuttavia, è necessario aprire il testo di riferimento e utilizzare la funzionalità di ricerca interna del browser (Ctrl+F). Il software, infatti, non consente ancora di impostare ricerche sulla base di parametri specifici (ad esempio, la posizione di inizio/fine parola, oppure valori grammaticali), né è disponibile per l'utente comune la modalità interrogazione diretta del corpus.

Nonostante necessiti ancora di qualche aggiornamento tecnico, il *Diachronen korpus na bălgarski ezik IX-XVIII vv.* si rivela uno strumento prezioso per la digitalizzazione di scritti medievali e per la creazione di sussidi lessicografici di vario tipo. La risorsa può essere altresì impiegata a fini scientifici, per sviluppare studi di linguistica storica su plurimi livelli (ortografia, fonetica, fonologia, morfologia, morfosintassi, lessicologia e fraseologia).

4 *Corpora* diacronici specialistici

A conclusione della presente rassegna, ci proponiamo di presentare due risorse di differente natura, cioè un corpus diacronico specialistico e un portale che archivia i testi degli antichi *azbukovniki*.

4.1 *Sankt-Peterburgskij korpus agiografičeskich tekstov*

Il *Sankt-Peterburgskij korpus agiografičeskich tekstov* (SKAT) (Corpus di testi agiografici di San Pietroburgo)[53] è una raccolta di testi agiografici dei secoli XV-XVII, sviluppata in seno al Dipartimento di

53 https://project.phil.spbu.ru/scat/page.php?page=project.

Linguistica computazionale della Facoltà di Filologia dell'Università Statale di San Pietroburgo, sotto la direzione di I.V. Azarova.[54]

La collezione si prefigge lo scopo di riflettere il carattere esemplare della letteratura agiografica slava orientale del periodo compreso tra il XV e il XVII secolo, la quale, abbondando di espressioni fraseologiche e di combinazioni linguistiche standard, impiegate in modo creativo dagli autori medievali e moderni secondo precise norme scrittorie, si rivela determinante nella definizione del carattere della lingua letteraria russa di tale epoca e dei suoi successivi sviluppi.

Allo stato attuale, il database conta più di cinquanta manoscritti, per un totale di circa 500.000 forme di parola. Di questi, tuttavia, ne sono stati indicizzati soltanto dieci, così che le possibilità di ricerca per forme di parola risultano ancora limitate a un numero circoscritto di fonti.

Il progetto ha mosso i suoi primi passi dalla compilazione di uno schedario di vite di santi (*žitija*), encomi (*pochval'nye slova*) e racconti (*skazanija*) dei secoli XVI-XVII, finalizzato al rendiconto dei diversi studi e edizioni esistenti di tali testi, a cui è seguita una fase di definizione metodologica per la realizzazione di un fondo di copie fotografiche e fotocopie dei manoscritti selezionati, che si trovavano conservati presso varie biblioteche di San Pietroburgo. A causa di alcune imperfezioni tecniche e strutturali delle edizioni reperite, quali la presenza di refusi legati al processo editoriale e omissioni di parti di testo, la procedura di digitalizzazione delle fonti scelte, intrapresa a partire dalla fine degli anni '70, ha attinto direttamente ai manoscritti.

Per offrire un'adeguata presentazione delle fonti manoscritte all'interno del database è stato sviluppato uno specifico font per computer (AGIO),[55] in grado di restituire tutte le lettere cirilliche antiche e le loro varianti semanticamente significative. Tale tipo di carattere riproduce i principali segni diacritici (*titulus*, *paerok*, caratteri sopralineari, spiriti e accenti), mentre non consente la resa dei grafemi privi di valore distintivo, o degli elementi di valenza essenzialmente paleografica (differenze relative al modulo delle lettere, legature), per cui emerge un chiaro disinteresse nei confronti della riproduzione fototipica dei testi (cf. Gerd 2004).

Il sistema grafico, così elaborato, ha in seguito subito un processo di semplificazione, rispondente all'esigenza di ridurre a un'unica

54 Sono membri fissi del gruppo di lavoro E.L. Alekseeva, A.S. Gerd, L.A. Zacharova, K.N. Lemešev e A.S. Greben'kov. Al progetto hanno altresì preso parte importanti filologi, quali: O.V. Tvorogov, dell'Istituto di letteratura russa «Puškinskij Dom»; V.M Zagrebin, collaboratore della Biblioteca Nazionale Russa; T.V. Roždestvenskaja; S.A. Averina; L.V. Zubova, dell'Università Statale di San Pietroburgo.

55 Il font è scaricabile dalla pagina di ricerca del corpus: `http://project.phil.spbu.ru/scat/search.php`.

forma le numerose e difficilmente gestibili varianti ortografiche presenti nei manoscritti. Grazie all'omissione dei diacritici e allo spostamento dei caratteri sovrascritti al livello della riga all'interno di paretesi tonde, si è resa possibile l'unificazione delle forme di parola costituite dagli stessi grafemi, distinte sulla base dell'unico parametro di presenza/assenza di caratteri sopralineari (es. бє(з)молвии – бєзмо(л)вии – бєзмолвии: бєзмолвии). Vengono altresì ridotte le varianti grafiche dovute ai mutamenti del sistema fonetico, verificatisi nella progressiva evoluzione dello slavo ecclesiastico di redazione slava orientale, come ad esempio l'alternanza di *o-/a-* a inizio di parola (es. алтарꙗ – олтарꙗ: алтарꙗ), o delle desinenze *-ogo/-ago* (es. студєнаго – студєного: студєнаго) (cf. Gerd et al. 2006).

Non è ancora disponibile l'annotazione morfosintattica dei testi, che tuttavia gli sviluppatori del database considerano di primaria importanza nell'ottica di suoi ulteriori sviluppi. Gli studenti del dipartimento di Linguistica computazionale dell'Università Statale di San Pietroburgo, affiancati da specialisti qualificati dell'equipe di SKAT, lavorano infatti manualmente alla marcatura morfologica dei testi agiografici, che viene fornita in un file separato, al momento non accessibile agli utenti del corpus (cf. Azarova, Alekseeva 2013). Per il futuro, si conta che i testi così etichettati possano venire sfruttati come precedenti per l'avvio di procedure di annotazione automatiche.

Allo stesso modo, a partire dai primi anni duemila ha preso a svilupparsi un formato per il *markup* grammaticale del corpus (cf. Alekseev, Alekseeva, Kas'janenko 2011). L'annotazione grammaticale dei testi viene a propria volta effettuata dagli studenti del dipartimento di Linguistica computazionale tramite un foglio di calcolo *Word*, caricato poi sul sito del corpus sotto forma di database. Gli utenti della risorsa, tuttavia, non hanno ancora accesso a queste informazioni.

La mancata realizzazione dell'annotazione grammaticale e morfosintattica limita le possibilità di interrogazione del corpus alla ricerca di singole forme di parola all'interno dei testi indicizzati presenti nel database (*poisk po slovoukazatelju*).

Alla richiesta dell'utente, viene prodotto un elenco di tutte le forme di parola che soddisfano le condizioni della ricerca effettuata, assieme all'indicazione della loro collocazione all'interno dei testi nella forma di numero di foglio e riga. Cliccando su tale indicazione, viene aperto il frammento di testo manoscritto contenente la parola desiderata, la quale, tuttavia, non viene evidenziata in alcun modo, né risulta cliccabile. Il frammento fornito, comunque, è sufficientemente ampio per poter analizzare la forma ricercata all'interno del contesto linguistico di appartenenza.

Allo stato attuale, il corpus SKAT si presta principalmente alla realizzazione di studi di carattere lessicale e lessicografico. Si possono ricordare, ad esempio, alcuni contributi della già menzionata Nikolenkova dell'Università Statale Lomonosov di Mosca: la studiosa ha

preso parte a un progetto volto alla creazione di un *lexicon* slavo ecclesiastico utilizzando, tra le altre fonti, il corpus SKAT (Nikolenkova 2014); ha inoltre impiegato i dati della suddetta raccolta, congiunti a quelli estratti dai *corpora* storici del NKRJa, per integrare i materiali dei dizionari storici ai fini di una più esaustiva analisi del funzionamento di alcune parole grammaticali nelle traduzioni in slavo ecclesiastico dal latino dei secoli XV-XVII (Nikolenkova 2017).

La programmata implementazione del sistema di annotazione grammaticale dei testi contenuti in SKAT, comunque, dovrebbe garantire la possibilità di condurre analisi maggiormente varie e raffinate. Tra queste potrebbero annoverarsi ricerche *corpus-based* volte a cogliere, in una prospettiva eminentemente diacronica, la dinamica delle innovazioni grammaticali dello slavo ecclesiastico delle opere di genere agiografico (cf. Alekseev, Alekseeva, Kas'janenko 2011).

4.2 Il portale *Istorija russkoj leksikografii*

Il portale *Istorija russkoj leksikografii* (OldLexicons.ru) (Storia della lessicografia russa)[56] conserva le edizioni digitali delle fonti lessicografiche slave orientali dei secoli XIII-XVIII.

La maggior parte delle opere della tradizione lessicografica della Rus', ad oggi, non è stata studiata; allo stesso modo, numerosi dizionari del XVIII secolo non sono stati pubblicati: da questa prospettiva, il sito OldLexicons.ru, sviluppato da A.N. Levičkin, collaboratore presso l'Istituto di Studi Linguistici dell'Accademia Russa delle Scienze, si rivela uno strumento straordinariamente prezioso per i lessicografi, e, più in generale, per tutti coloro che si interessano della storia dei dizionari e dei termini in essi contenuti.

Le fonti caricate sul portale sono suddivise in quattro classi, le quali riflettono convenzionalmente lo sviluppo della tradizione lessicografica russa. La categoria «prima degli *azbukovniki*» (*do azbukovnikov*) conserva opere lessicografiche di diverso genere (in particolare raccolte di glosse), composte nella Rus' tra il XIII e il XVI secolo. Gli *azbukovniki* dei secoli XVI-XVII sono riuniti in nell'omonima sezione (*azbukovniki*), affiancata dalla collezione dei «lessici» (*leksikony*) del XVII-XVIII secolo. In un raggruppamento separato, infine, sono raccolti alcuni «frasari» (*razgovorniki*) databili tra XV e XVII secolo.

Il processo di digitalizzazione delle suddette testimonianze, spesso inedite, si è basato direttamente sui manoscritti, di cui vengono peraltro offerte la descrizione e la riproduzione fotografica (di un solo foglio) all'interno di uno specifico blocco (*rukopisi*) sulla pagina di accesso al sito. Non tutte le fonti sono disponibili nella loro versione

56 https://www.oldlexicons.ru/.

e-ISSN 2785-3187

elettronica: dei testi che non sono ancora stati digitalizzati, comunque, è fornita una breve presentazione e indicato l'elenco dei codici in cui si conservano.

Ad oggi, i criteri per la conversione delle fonti in formato elettronico non sono stati definiti in modo univoco, tuttavia per il futuro si prevede la pubblicazione di ciascuna risorsa in una doppia forma: come testo digitale, corredato della numerazione delle singoli voci del dizionario, e come tabella contente la lista di tutte le copie manoscritte disponibili per l'opera lessicografica in questione.

All'interno della finestra «novità» (*novosti*) della pagina web, si trova pubblicato un annuncio secondo il quale, a partire dal 24 febbraio 2024, «il sito è stato trasferito su una nuova versione di *Drupal 10*» – una piattaforma software di *Content Management System* (CMS) –, e dunque «è in corso il ripristino delle sezioni e dei materiali».[57] Per questo motivo, al momento presente, l'utente non può più accedere al modulo per la ricerca, che era presente sul portale prima del sopraccitato aggiornamento tecnico e che tuttavia offriva possibilità di impiego limitate, consentendo la sola interrogazione per parole chiave esatte.

Prima del febbraio 2024, a beneficio dell'utenza era disposta anche la sezione «glossario» (*slovnik*), che presentava una lista di tutte le forme di parola contenute nelle fonti lessicografiche digitalizzate presenti sul portale: di ciascun termine veniva indicata la fonte (o le fonti) e offerto un breve frammento testuale come esempio. All'interno del suddetto elenco le varianti grafico-ortografiche non erano state unificate: forme distinte sulla base dell'unico parametro di conservazione o caduta degli *jer* (es. безстыдливство – безстыдливъство) venivano riportate separatamente. Lo stesso valeva per i termini che differiscono per la riduzione o mancata riduzione delle vocali nasali (es. великаѧ – великаꙗ). Era altresì contemplata la possibilità di consultazione delle singole fonti: selezionando il testo di interesse, si apriva una lista indicizzata di tutte le parole ivi contenute, corredate dall'indicazione della posizione nel foglio all'interno della copia di riferimento. Se si cliccavano i singoli termini, se ne otteneva, anche in questo caso, un breve contesto. Quando una fonte lessicografica era testimoniata da più manoscritti, i dati linguistici venivano visualizzati all'interno di una tabella sinottica, che riportava l'attestazione di ogni parola all'interno di ciascun codice.

Quantomeno nella sua vecchia versione, il progetto `OldLexicons.ru`, non puntava, come i *corpora* storici sopra considerati, allo sviluppo di un sistema di ricerca che consenta di ottenere il maggior numero possibile di risultati per mezzo di un'unica interrogazione, prediligendo, piuttosto, una riproduzione storicamente fedele delle varietà

57 `https://www.oldlexicons.ru/novosti`.

grafiche, ortografiche e fonetiche delle forme linguistiche, quali si incontrano nelle fonti lessicografiche di differenti aree e periodi. Rimaniamo vigili, comunque, sugli sviluppi attuali e futuri della risorsa.

5 Conclusioni

Con l'avvento del nuovo millennio, e in particolare nel corso degli ultimi dieci anni, si è proceduto a una massiccia digitalizzazione e catalogazione dei testi antichi e medievali della tradizione scritta slava: numerosi sforzi sono stati incanalati nell'elaborazione di *corpora* storici per la lingua russa.

L'analisi di alcune di queste risorse ha messo in luce che l'applicazione delle moderne tecnologie informatiche allo studio delle testimonianze scritte antiche, pur mancando ancora di strumenti perfettamente finiti e di una metodologia universalmente condivisa, cela in sé grandi potenzialità. L'inclusione di fonti medievali all'interno di raccolte elettroniche interrogabili, difatti, rende le testimonianze del passato attuali nel contemporaneo mondo digitale e digitalizzato, offrendo nuove e più vantaggiose modalità di analisi dei dati.

I *corpora* diacronici per la lingua russa ad oggi esistenti si caratterizzano per differenti livelli di elaborazione tecnica, differenti modalità di preparazione e presentazione dei materiali, nonché differenti possibilità di applicazione nella ricerca.

Il riconoscimento di un certo ruolo pionieristico nel campo spetta al *Corpus Cyrillo-Methodianum Helsingiense* (CCMH) (§ 3.2.) e al *Trondheim-Sofia Corpus of Old Slavic* (§ 3.4.) elaborati già a partire dagli anni '90 del secolo scorso. Le due risorse hanno il vantaggio di includere le riproduzioni elettroniche delle più antiche testimonianze scritte slave (alcuni testimoni del canone paleoslavo nel caso della prima, una selezione di testi copiati in territorio slavo orientale da manoscritti antico-bulgari nell'XI secolo per la seconda): l'assenza di strumenti per l'interrogazione, tuttavia, limita le loro potenzialità di impiego nella realizzazione di ricerche linguistiche.

Per supplire a una simile mancanza, i testi del CCMH sono stati ricompresi nel corpus di paleoslavo del *Thesaurus Indogermanischer Text-und Sprachmaterialien* (TITUS) (§ 3.3.). Il suddetto database dispone di un proprio motore di ricerca, che tuttavia non presenta requisiti di elaborazione tecnica pienamente soddisfacenti: il software, infatti, consente di reperire facilmente materiale linguistico e lessicale, senza consentirne, però, l'analisi automatica in termini di caratteristiche grammaticali e morfosintattiche.

Il *Regensburgskij diachroničeskij korpus russkogo jazyka* (RRuDi) (§ 3.6.) avrebbe potuto rappresentare uno strumento tecnicamente all'avanguardia, disponendo integralmente dell'annotazione morfologica. Le procedure di implementazione della risorsa, però, sono state

bruscamente interrotte, mettendo in luce una seria problematica nell'ambito dello sviluppo di *corpora* storici delle lingue slave: la limitata disponibilità di fondi e sovvenzioni a disposizione dei ricercatori.

Lo strumento che meglio si presta alla realizzazione di studi linguistici sulla tradizione manoscritta paleoslava e slava orientale dei secoli XI-XIV è il sistema «*Manuskript». Slavjanskoe pis'mennoe nasledie* (§ 3.5.). Rispetto ad altri *corpora* diacronici per la lingua russa, il corpus storico di *Manuskript* ha innanzitutto il vantaggio di includere alcune tra le copie e i frammenti testuali più antichi (XI sec.) della tradizione scrittoria slava giunta sino a noi, il cui formato elettronico non sacrifica peraltro la trasmissione delle caratteristiche codicologiche autentiche. La presenza di una serie di moduli specializzati, appositamente strutturati per l'elaborazione di differenti tipi di richieste e di visualizzazione dei dati, poi, offre alla raccolta un ulteriore punto di forza. Devono venire evidenziati, in modo particolare, i due moduli statistici ivi inclusi, i quali permettono di integrare le tradizionali pratiche degli studi linguistici e filologici tramite l'impiego delle nuove tecnologie informatiche dei *corpora* e dei metodi della linguistica computazionale.

Per una rappresentazione maggiormente esaustiva della tradizione scritta slava orientale, in tutte le sue forme e generi e in un arco temporale massimamente ampio (XI-XVIII sec.), il *Nacional'nyj korpus russkogo jazyka* (NKRJa) (§ 3.1.) costituisce la risorsa migliore ad oggi esistente. Con i suoi cinque *corpora* storici e l'innovativo corpus pancronico, il NKRJa consente di condurre ricerche linguistiche relative a differenti secoli della storia della lingua letteraria russa, su un totale di più di trecento milioni di forme di parola. Disponendo di un'annotazione testuale, meta-testuale, grammaticale e morfologica, e di un motore di ricerca ben collaudato, la risorsa consente di realizzare studi a tutti i livelli linguistici (lessico, morfologia, sintassi). Grazie alle più recenti implementazioni, il Corpus nazionale consente altresì l'estrazione di utili dati statistico-distributivi. Si tenga soltanto a mente che la digitalizzazione delle fonti qui contenute non si basa sugli originali, bensì su loro edizioni: in questo senso, la risorsa può rivelarsi maggiormente utile al ricercatore che consideri la leggibilità del testo un criterio di importanza superiore rispetto alla conservazione delle caratteristiche codicologiche autentiche.

Seppur preposto alla conservazione della tradizione scrittoria bulgara, il *Diachronen korpus na bălgarski ezik IX-XVIII vv.* (§ 3.7.) può trovare ampio impiego da parte di specialisti di differenti settori della slavistica, consentendo di realizzare studi su tutto quel patrimonio medievale di matrice slava meridionale, che riveste un'importanza decisiva nella storia culturale del mondo bizantino-slavo nel suo complesso. Il motore di ricerca della raccolta necessita ancora di perfezionamento, mentre sono stati ottenuti risultati piuttosto soddisfacenti nella realizzazione delle procedure di annotazione linguistica.

La disponibilità di specifiche funzionalità per l'inserimento di commenti codicologici, paleografici e testologici, inoltre, rende possibile dotare le riproduzioni digitali dei testi di un proprio apparato critico.

È opportuno mantenere uno sguardo vigile, infine, sul *Sankt-Peterburgskij korpus agiografičeskich tekstov* (SKAT) (§ 4.1.) e sul portale *Istorija russkoj leksikografii* (`OldLexicons.ru`) (§ 4.3). Le suddette risorse non hanno ancora raggiunto una definitiva elaborazione; le continue procedure di aggiornamento cui sono sottoposte, tuttavia, le fanno gradualmente emergere come strumenti preziosi, nonché unici nel loro genere, per lo studio di due ambiti specifici della tradizione scritta slava orientale: la letteratura agiografica, nel caso della prima, e la tradizione lessicografica, per ciò che concerne la seconda.

Bibliografia

Fonti primarie

CCMH: *Corpus Cyrillo-Methodianum Helsingiense*. https://korp.csc.fi/download/ccmh-src/.

HISTDICT: *Diachronen korpus na bălgarski ezik IX-XVIII vv.* Диахронен корпус на български език (Corpus diacronico della lingua bulgara). https://histdict.uni-sofia.bg/.

MANUSKRIPT: *«Manuskript». Slavjanskoe pis'mennoe nasledie* «Манускрипт». Славянское письменное наследие. («Manuskript». Patrimonio scritto slavo). http://mns.udsu.ru/.

NKRJa: *Nacional'nyj Korpus Russkogo Jazyka.* (*Istoričeskie korpusa*) Национальный корпус русского языка. (Исторические корпуса) (Corpus Nazionale della lingua russa. [Corpora storici]). https://ruscorpora.ru/.

RRuDi: *Regensburgskij diachroničeskij korpus russkogo jazyka* Регенсбургский диахронический корпус русского языка (Corpus diacronico di Ratisbona della lingua russa). https://www.slawistik.hu-berlin.de/de/member/meyerrol/subjekte/rrudi.

SKAT: *Sankt-Peterburgskij korpus agiografičeskich tekstov* Санкт-Петербургский корпус агиографических текстов (Corpus di testi agiografici di San Pietroburgo). https://project.phil.spbu.ru/scat/page.php?page=project.

TITUS: *Thesaurus Indogermanischer Text-und Sprachmaterialien.* (*Old Church Slavonic corpus. Old Russian corpus*). https://titus.uni-frankfurt.de/indexe.htm.

TRONDHEIM-SOFIA CORPUS: *Trondheim-Sofia Corpus of Old Slavic*. http://www.hf.ntnu.no/SofiaTrondheimCorpus/.

Fonti secondarie

Afanasiev, I.A. (2020). «A Corpus-based Approach in Archaeolinguistics». *Journal of applied Linguistics and Lexicography*, 2(2), 147-59. https://www.doi.org/10.33910/2687-0215-2020-2-2-147-159.

Alekseev, V.A.; Alekseeva, E.L.; Kas'janenko, S.E. (2011). «Grammatičeskaja razmetka v korpuse SKAT» Грамматическая разметка в корпусе СКАТ (Annotazione grammaticale nel corpus SKAT). Zacharov V.P. (otv. red.), *Trudy meždunarodnoj konferencii «Korpusnaja lingvistika – 2011»* (27-29 ijunija 2011 g.), Sankt-Peterburg, 69-73.

Azarova, I.V.; Alekseeva, E.L. (2013). «Osobennosti morfo-sintaksičeskoj razmetki drevnerusskich agiografičeskich tekstov» Особенности морфосинтаксической разметки древнерусских агиографических текстов (Caratteristiche dell'annotazione morfo-sintattica dei testi agiografici russi antichi). Zacharov, V.P. (otv. red.), *Trudy meždunarodnoj konferencii «Korpusnaja lingvisika – 2013»* (25-27 ijunija 2013), Sankt-Peterburg, 157-64.

Baranov, V.A. (2015). «Istoričeskij korpus kak cel' i instrument korpusnoj paleoslavistiki» Исторический корпус как цель и инструмент корпусной палеославистики (Il corpus storico come obiettivo e strumento della

paleoslavistica *corpus-based*). *Scripta & e-Scripta. The Journal of Interdisciplinary Mediaeval Studies*, volume 14-15, 39-62.

Baranov, V.A. (2017a). «Količestvennyj i statističeskij analiz srednevekovych slavjanskich tekstov: instrumentarij korpusa "Manuskript" i metodika ego ispol'zovanija» Количественный и статистический анализ средневековых славянских текстов: инструментарий корпуса «Манускрипт» и методика его использования (Analisi quantitativa e statistica dei testi medievali slavi: strumenti del corpus *Manuskript* e metodica del loro utilizzo). Korinenko, S.I. (otv. red.), *Cifrovaja gumanitaristika: resursy, metody, issledovanija. Materialy Meždunarodnoj Naučnoj konferencii* (g. Perm, 16-18 Maja 2017 g.). Čast' 1. V 2-ch častjach, Perm, 40-9.

Baranov, V.A. (2017b). «Statističeski značimye slova kak charakteristika srednevekovogo slavjanskogo teksta (na materiale kollekcii Apostolov istoričeskogo korpusa "Manuskript")» Статистически значимые слова как характеристика средневекового славянского текста (на материале коллекции Апостолов исторического корпуса «Манускрипт») (Parole statisticamente significative come caratteristica dei testi medievali slavi [sui materiali della collezione di Apostoli del corpus storico *Manuskript*]). Baranov, V.A. (otv. red.), *Gumanitarnoe obrazovanie i nauka v techničeskom vuse. Sbornik dokladov Vserossijskoj naučno-praktičeskoj konferencii s meždunarodnym učastiem* (Iževsk, 24-27 oktjabrja 2017 g.), Iževsk, 359-69.

Baranov, V.A. (2018). «Statistical Analysis of the Slavonic Paraenesis by Ephrem the Syrian (on Three Electronic Copies of the 13-14th Centuries from the *Manuscript* Corpus)». *Journal of Siberian Federal University. Humanities & Social Sciences*, 11(8), 1211-28. https://doi.org/10.17516/1997-1370-0302

Baranov, V.A. (2019a). «Sozdanie i ispol'zovanie istoričeskich korpusov slavjanskich pis'mennych pamjatnikov» Создание и использование исторических корпусов славянских письменных памятников (Creazione e uso di *corpora* storici di testimonianze scritte slave). *Scripta & e-Scripta. The Journal of Interdisciplinary Mediaeval Studies*, volume 19, 33-57.

Baranov, V.A. (2019b). «Statističeski značimye slova četyrech slavjanskich služebnych Minej XI veka» Статистически значимые слова четырех славянских служебных Миней XI века (Parole statisticamente significative di quattro *Menee* liturgiche slave dell'XI secolo). *Slavistica Vilnensis*, 64(2), 10-30. https://doi.org/10.15388/SlavViln.2019.64(2).17.

Baranov, V.A. (2019c). «K voprosu ob ispol'zovanii statističeskich metodov dlja poiska kollokacij i kolligacij v drevnejšich slavjanskich tekstach (na materiale glagoličeskich rukopisej korpusa "Manuskript"» К вопросу об использовании статистических методов для поиска коллокаций и коллигаций в древнейших славянских текстах (на материале глаголических рукописей корпуса «Манускрипт») (Uso di metodi statistici per la ricerca di collocazioni e colligazioni nei più antichi testi slavi [sui materiali dei manoscritti glagolitici del corpus *Manuskript*]). *Slovo. Časopis Staroslavenskoga instituta u Zagrebu*, 69, 1-33. https://doi.org/10.31745/s.69.1.

Baranov, V.A. (2020), «Korpusnye issledovanija srednevekovych slavjanskich rukopisej: statističeski značimye n-grammy (kollokacii) drevnerusskich letopisej» Корпусные исследования средневековых славянских рукописей: статистически значимые n-граммы (коллокации) древнерусских летописей (Ricerche sui *corpora* di manoscritti medievali slavi: n-grammi statisticamente significativi [collocazioni] nelle cronache russe antiche).

Elektronnyj naučno-obrazovatel'nyj žurnal «Istorija», t. 11, vyp. 3(89), *Cifrovoe sodružestvo nauk: opyt primenenija informacionnych technologij v istorii i smežnych disciplinach*. https://history.jes.su/s207987840009440-3-1/; https://doi.org/10.18254/S207987840009440-3.

Baranov, V.A. (2022). «Distributivnyj slovar' istoričeskogo korpusa "Manuskript": postanovka, zadači, material, metody» Дистрибутивный словарь исторического корпуса «Манускрипт»: постановка, задачи, материал, методы (Dizionario distributivo del corpus storico *Manuskript*: impostazione, obiettivi, materiali, metodi). *Aktual'nye problemy filologii i pedagogičekoj lingvistiki*. 2. *Prikladnaja lingvistika: sovremennye resursy i perspektivy*, 94-106. https://doi.org/94–106. 10.29025/2079-6021-2022-2-94-106.

Baranov, V.A.; Dubovcev S.V. (2012). «Modul' statistiki informacionno-analitičeskoj sistemy "Manuskript": funkcii i demonstracija dannnych» Модуль статистики информационно-аналитической системы «Манускрипт»: функции и демонстрация данных (Il modulo statistico del sistema informativo-analitico *Manuskript*: funzioni e dimostrazione dei dati). Baranov, V.A.; Varfolomeev, A.G. (otv. red.), *Informacionnye technologii i pis'mennoe nasledie. Materialy IV meždunarodnoj naučnoj konferencii* (Petrozavodsk, 3-8 sentjabrja 2012 g.), Petrozavodsk, 23-6.

Baranov, V.A. et al. (2004). «Old Slavic Manuscript Heritage: Electronic Publications and Full-Text Databases». Hemsley J. (ed.), *Proceedings of the EVA 2004 London (Electronic Imaging, the Visual Arts Conference & Beyond)*, London, 11.1-11.8.

Baranov, V.A.; Gulina, O.V.; Mironov, A.N. (2010). «Morfologičeskaja paradigma i ee sostavljajušie v sisteme "Manuskript"» Морфологическая парадигма и ее составляющие в системе «Манускрипт» (Il paradigma morfologico e le sue componenti nel sistema *Manuskript*). Baranov, V.A. (otv. red.), *Informacionnye technologii i pis'mennoe nasledie. Materialy meždunarodnoj naučnoj konferencii* (Ufa, 28-31 oktjabrja 2010 g.), Ufa; Iževsk, 28-31.

Budennaja, E.V. (2020). «V poiskach triggera: knižnye i nekнižnye teksty kak markery različnych aspektov russkoj referencial'noj evoljucii» В поисках триггера: книжные и некнижные тексты как маркеры различных аспектов русской референциальной эволюции (Alla ricerca del *trigger*: testi letterari e non letterari come marcatori di diversi aspetti dell'evoluzione referenziale russa). *Slověne*, 9(2), 210-43. https://doi.org/10.31168/2305-6754.2020.9.2.11.

Cimmerling, A.V. (2019). «Svjaski pljukvamperfekta v russkom jazyke XIV-XVI vekov» Связки плюквамперфекта в русском языке XIV-XVI веков (Le costruzioni del piuccheperfetto nella lingua russa dei secoli XIV-XVI). *Vestnik Volgogradskogo gostudarstvennogo universiteta.* Serija II. *Jazykoznanie*, 18(4), 41-57. DOI: https://doi.org/10.15688/jvolsu2.2019.4.4.

Dobrušina, E.R.; Poljakov, A.E. (2013). «Korpus cerkovnoslavjanskogo jazyka: vozmožnosti, metody sozdanija, perspektivy» Корпус церковнославянского языка: возможности, методы создания, перспективы (Il corpus dello slavo ecclesiastico: potenzialità, metodi di compilazione, prospettive). *Vestnik Pravoslavnogo Svjato-Tichonovskogo gumanitarnogo Universiteta.* Serija III. *Filologija*, 1(31), 32-44.

Dozat, Qi.P.; Zhang, T; Manning, C.D. (2018), «Universal Dependency Parsing from Scratch». Zeman, D; Hajič, J. (eds), *Proceedings of the CoNLL 2018*

Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies (Brussels, Belgium, October 31 – November 1). Brussels, 160-70.

Dozat, Qi.P. et al. (2020). «Stanza: A Python Natural Language Processing Toolkit for Many Human Languages». Celikyilmaz A.; Wen, T.-H. (eds), *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations* (July 5- July 10, 2020), Stroudsburg, PA, 101-8. https://doi.org/10.18653/v1/2020.acl-demos.14.

Fel'de, O.V.; Birjulina, E.A. (2022). «Universalia gumannost' v russkoj lingvokul'ture» Универсалия гуманность в русской лингвокультуре (L'universale 'humanitas' nella linguistica culturale russa). *Vestnik Kemerovskogo gosudarstvennogo universiteta*, 24(2), 255-65. https://doi.org/10.21603/2078-8975-2022-24-2-255-265.

Ferro, M.C. (2018a). «Osobennosti ispol'zovanija intellektual'noj leksiki v proizvedenijach Maksima Greka. Postanovka voprosa» Особенности использования интеллектуальной лексики в произведениях Максима Грека. Постановка вопроса (Caratteristiche dell'uso del lessico intellettuale nelle opere di Massimo il Greco. Enunciazione del problema). Remnëva, M.L. et al. (otv. red.), *Slavjanskij mir: jazyk, literatura, kul'tura. Materialy meždunarodnoj naučnoj konferencii, posvjašennoj 100-letiju so dnja roždenija zaslužennogo professora MGU imeni M.V. Lomonosova A.G. Širokovoj i 75-letiju kafedry slavjanskoj filologii filologičeskogo fakul'teta* (28-29 nojabrja 2018 goda, Moskva), Moskva, 161-63.

Ferro, M.C. (2018b). «Sl.eccl. *razumŭ*. Studi per un lexicon plurilingue dei termini religiosi e filosofico-teologico». Ferro, M.C.; Salmon, L.; Ziffer, G. (a cura di), *Contributi italiani al XVI Congresso Internazionale degli Slavisti* (Belgrado, 20-27 agosto 2018), Firenze, 23-36. https://doi.org/10.36253/978-88-6453-723-8.05.

Ferro, M.C. (2019). «Per un'analisi lessicale delle opere di Massimo il Greco». *Studi Slavistici*, 6(2), 139-56. https://doi.org/10.13128/Studi_Slavis-7482.

Ferro, M.C. (2020). «Cerkovnoslavjanskie slova imarmenja, fatunĭ, fortuna v proizvedenijach Maksima Greka» Церковнославянские слова имарменя, фатунъ, фортуна в произведениях Максима Грека (Le parole slavo-ecclesiastiche *imarmenja*, *fatunĭ*, *fortuna* nelle opere di Massimo il Greco). *Vestnik Volgogradskogo gosudarstvennogo universiteta*. Serija II. *Jazykoznanie*, 19(6), 17-30. https://doi.org/10.15688/jvolsu2.2020.6.2.

Ferro, M.C. (2021). M.C. Ferro, «'Destino' e dintorni. Innovazioni lessicali nelle opere in slavo di Massimo il Greco». *Cyrillomethodianum*, 22, 139-50.

Ganeva, G. (2018). «Electronic Diachronic Corpus and Dictionaries of Old Bulgarian». *Studia Ceranea*, 8, 111-19. https://doi.org/10.18778/2084-140X.08.06.

Garzaniti, M. (2019). *Gli slavi. Storia, culture e lingue dalle origini ai nostri giorni*. Nuova edizione. A cura di F. Romoli. Roma: Carocci editore.

Garzaniti, M. (2021). s.v. «Slavo-ecclesiastica, lingua». *CESECOM. Centro studi sull'Europa centrale, balcanica e orientale in epoca medievale e moderna. L'Europa fra civiltà occidentale e civiltà asiatiche.* Lexicon. https://www.cesecom.it/it/lexicon/slavo-ecclesiastica-lingua/16#.

Gavrilova, T.S.; Šalganova, T.A.; Ljaševskaja, O.N. (2016). «K zadače avtomatičeskoj leksiko-grammatičeskoj razmetki starorusskogo korpusa XV-XVII vv.» К задаче автоматической лексико-грамматической разметки старорусского корпуса XV-XVII вв. (Il compito dell'annotazione

lessico-grammaticale automatica del corpus medio-russo dei secoli XV-XVII). *Vestnik Pravoslavnogo Svjato-Tichonovskogo gumanitarnogo universiteta.* Serija III. *Filologija*, 2(47), 7-25.

Gerd, A.S. (2004). «Elektronnyj korpus tekstov po pamjatnikam drevnerusskoj agiografičeskoj literatury» Электронный корпус текстов по памятникам древнерусской агиографической литературы (Un corpus di testi elettronico sulle testimonianze della letteratura agiografica russa antica). *Naučno-techničeskaja informacija (NTI).* Serija II. *Informacionnye processy i sistemy,* 16-20.

Gerd, A.S. et al. (2006). «Korpus drevnerusskich agiografičeskich tekstov SKAT: sovremennoe sostojanie i perspektivy razvitija» Корпус древнерусских агиографических текстов СКАТ: современное состояние и перспективы развития (Il corpus dei testi agiografici russi antichi SKAT: stato attuale e prospettive di sviluppo). Baranov, V.A. (otv. red.), *Materialy meždunarodnoj naučnoj konferencii* (Iževsk, 13-17 ijulija 2006 g.), Iževsk, 38-42.

Kilina, L.F. (2020). «Problemy izučenija ustojčivych glagol'no-imennych sočetanij russkogo jazyka v diachroničeskom aspekte» Проблемы изучения устойчивых глагольно-именных сочетаний русского языка в диахроническом аспекте (Problemi dello studio delle collocazioni verbo-nome nella lingua russa sotto l'aspetto diacronico). Tret'jakovaja I.Ju. (otv. red.) *Fraseologija i paremiologija v diachronii i sinchronii (ot archaizacii k neologizacii). Materialy Meždunarodnoj naučno-praktičeskoj konferencii* (g. Kostroma 24-25 sentjabrja 2020 g.), Kostroma, 85-7.

Kilina, L.F.; Kolosova, E.I. (2020). «O grammatičeskoj semantike prilagatel'nych čelověčĭ i čelověčĭskŭ v starorusskich letopisjach» О грамматической семантике прилагательных человѣчь и человѣчьскъ в старорусских летописях (Sulla semantica grammaticale degli aggettivi *čelověčĭ* e *čelověčĭskŭ* nelle cronache antiche russe). *Filologija i Kul'tura*, 3(61), 62-7. https://doi.org/10.26907/2074-0239-2020-61-3-62-67.

Kilina, L.F.; Rusanova, O.S. (2021). «Ustojčivye glagol'no-imennye sočetanija s komponentom činiti v russkich gramotach XIII-XVII vv.» Устойчивые глагольно-именные сочетания с компонентом чинити в русских грамотах XIII-XVII вв. (Collocazioni verbo-nome con la componente *činiti* nelle *gramoty* russe dei secoli XIII-XVII). *Vestnik Udmurtskogo universiteta. Serija Istorija i Filologija,* 31(3), 429-37. https://doi.org/10.35634/2412-9534-2021-31-3-429-437.

Kilina L.F., Zajnullina, S.R. (2017). «Glagol'no-imennye ustojčivye sočetanija s komponentom životŭ: funkcionirovanie i razvitie (na materiale russkich letopisej XIII-XV vv.)» Глагольно-именные устойчивые сочетания с компонентом животъ: функционирование и развитие (на материале русских летописей XIII-XV вв.) (Collocazioni verbo-nome con la componente *životŭ*: funzionamento e sviluppo [sui materiali delle cronache russe dei secoli XIII-XV]). *Vestnik Udmurtskogo universiteta. Serija Istorija i Filologija,* 27(3), 429-43.

Kopotev, M.V. (2014). *Vvedenie v korpusnuju lingvistiku* Введение в корпусную лингвистику (Introduzione alla linguistica dei *corpora*). Praha: Animedia.

Kusennaja, T.V. (2021). «Motivirovannost' funkcionirovanja iskonno russkoj i cerkovnoslavjanskoj leksiki v agiografičeskich tekstach XIV-XVII vekov» Мотивированность функционирования исконно русской и церковнославянской лексики в агиографических текстах XIV-XVII веков (Motivazione del funzionamento del lessico originario russo e di quello slavo

ecclesiastico nei testi agiografici dei secoli XIV-XVII). *Filologičeskij vestnik Surgutskogo gosudarstvennogo pedagogičeskogo universiteta*, 3(7), 7-12. https://doi.org/10.26105/PBSSPU.2021.7.3.001.

Marcialis, N. (2007). *Introduzione alla lingua paleoslava.* Firenze: Firenze University Press.

Meyer, R. (2005). «The Regensburg Diachronic Corpus of Russian: A New Source for Linguistic Research (not Only) on Modality». Hansen, B.; Karlík P. (eds.), *Modality in Slavonic Languages: New Perspectives*. München: Verlag Otto Sagner, 315-36.

Mišina, E.A.; Pičchadze, A.A. (2015). «Drevnerusskij podkorpus Nacional'nogo korpusa russogo jazyka» Древнерусский подкорпус Национального корпуса русского языка (Il sotto-corpus drevnerusskij del *Corpus Nazionale della lingua russa*). Moldovan, A.M. (otv. red.), *Trudy Instituta russkogo jazyka im. V.V. Vinogradova,* N. 2(3). Moskva: Institut russkogo jazyka im. V.V. Vinogradova RAN, 99-115.

Mitrenina, O. (2014). «The Corpora of Old and Middle Russian Texts as an Advanced Tool for Exploring an Extinguished Language». *Scrinium*, 10(1), 455–61. https://doi.org/10.1163/18177565-90000109.

Moldovan, A.M. (2014). «Problemy i zadači korpusonogo izučenija slavjanskoj pis'mennosti» Проблемы и задачи корпусного изучения славянской письменности (Problemi e compiti dello studio *corpus-based* della tradizione scritta slava). *Slavjanskij al'manach*, 1-2, 25-34.

Nikolen'kova, N.V. (2014). «Materialy k slovarju cerkovnoslavjanskogo jazyka XVI-XVII vv. na osnove elektronnogo sobranija russkich agiografičeskich tekstov, istoričeskich slovarej i novych tekstov» Материалы к словарю церковнославянского языка XVI-XVII вв. на основе электронного собрания русских агиографических текстов, исторических словарей и новых текстов (Materiali per un dizionario della lingua slava ecclesiastica dei secoli XVI-XVII sulla base di una raccolta elettronica di testi agiografici russi, dizionari storici e nuovi testi). Baranov, V.A.; Željazkova, V.; Lavrent'ev, A.M. (otv. red.), *El'manuscript 2014 – Pismenoto nasledstvo i informacionnite technologii: materiali ot V meždunarodna nauč. konf.* (Varna 15-20 septembri 2014 g.), Sofija; Iževsk, 214-16.

Nikolenkova, N.V. (2017). «Sojusy v cerkonoslavjanskich perevodach s latinskogo jazyka XV-XVII vv.: dopolnenija k istoričeskim slovarjam» Союзы в церковнославянских переводах с латинского языка XV-XVII вв.: дополнения к историческим словарям (Congiunzioni nelle traduzioni slavo-ecclesiastiche dal latino [XV-XVII sec.]: integrazioni ai dizionari storici). *Vestnik moskoskogo universiteta*. Serija IX. *Filologija,* 1, 190-203.

Pavlova R. (s.d.). «Corpus of Old Slavic Texts from the XIth C. Introduction by R. Pavlova». *Trondheim-Sofia Corpus of Old Slavic.* Throndeim: NTNU.Norwegian University of Science and Technology. Faculty of Arts; Sofia: New Bulgarian University. Bachelor's Faculty. Slavic Text Laboratory. http://www.hf.ntnu.no/SofiaTrondheimCorpus/index2.html.

Pekelis, O.E. (2020). «Ob odnom slučae pragmatikalizacii v russkom jazyke: mikrodiachroničeskoe issledovanie časticy že v sostave voprosa» Об одном случае прагматикализации в русском языке: микродиахроническое исследование частицы же в составе вопроса (Un caso di pragmaticizzazione nella lingua russa: studio microdiacronico della particella *že* nelle domande). *Slověne,* 9(1), 340-61. https://doi.org/10.31168/2305-6754.2019.9.1.12.

Pen'kova, Ja.A. (2017). «K istorii i predystorii slova budto v russkom jazyke Srednevekov'ja» К истории и предыстории слова будто в русском языке Средневековья (Sulla storia e preistoria della parola *budto* nella lingua russa medievale). Osipova, E.P. (otv. red.), *I.I. Sreznevskij i russkoe istoričeskoe jazykoznanie: k 205-letiju so dnja roždenija I.I Sreznevskogo (1812-1880). Sbornik statej Meždunarodnoj naučnoj konferencii* (21-23 sentjabrja 2017 g. Rjazan'), Rjazan', 69-76.

Pen'kova, Ja.A. (2018) «Slavjanskoe predbudušee i ego grečeskie sootvetstvija v perevodnych drevnecerkovnoslavjanskich pamjatnikach» Славянское предбудущее и его греческие соответствия в переводных древнецерковнославянских памятниках (Il futuro anteriore slavo e i suoi corrispettivi greci nelle testimonianze tradotte in antico slavo ecclesiastico). *Russkij jazyk v naučnom osvešenii,* 36(2), 9-35. https://doi.org/10.31912/rjano-2018.2.

Pen'kova, Ja.A. (2019). Ja.A. «Imu, učĭnu, stanu, budu: korpusnoe issledovanie perifraz budušego vremeni v srednerusskoj pis'mennosti» Иму, учьну, стану, буду: корпусное исследование перифраз будущего времени в среднерусской письменности (*Imu, učĭnu, stanu, budu*: uno studio corpus-based delle perifrasi del futuro nella letteratura medio-russa). *Slavistična revija,* 67(4), 569-86.

Pen'kova, Ja.A. (2020a). «Layering in the System of Middle Russian Periphrastic Future Constructions Through a Corpus Perspective». *Sprachliche Diversität: Theorien, Methoden, Ressourcen 42. Jahrestagung der Deutschen Gesellschaft für Sprachwissenschaft* (04.-06. März 2020), Hamburg, 174.

Pen'kova, Ja.A. (2020b). «Predbudušee i vid glagola v vostočnoslavjanskoj pis'mennosti XV-XVII vv. na fone vlijanija pol'skogo jazyka» Предбудущее и вид глагола в восточнославянской письменности XV-XVII вв. на фоне влияния польского языка (Il futuro anteriore e l'aspetto del verbo nella letteratura slava orientale die secoli XV-XVII sullo sfondo dell'influenza polacca). Gorbova, E.V. (otv. red.), *Vzaimodejstvie aspekta so smežnymi kategorijami. Materialy VII Meždunarodnoj konferencii Komissii po aspektologii Meždunarodnogo komiteta slavistov* (Sankt-Peterburg, 5-8 maja 2020 goda). Sankt-Peterburg, 304-14.

Pen'kova, Ja.A. (2021a). «Middle Russian Future Periphrastic Constructions in the Light of Language Contacts». *Scando-Slavica,* 67(1), 43–64. https://doi.org/10.1080/00806765.2021.1901241.

Pen'kova, Ja.A. (2021b). «K istorii neopredelennych mestoimenij: kvasireljativy na ni budi i ni est v russkom jazyke XVII-XVIII vv.» К истории неопределенных местоимений: квазирелятивы на ни буди и ни есть в русском языке XVII-XVIII вв. (Sulla storia dei pronomi indefiniti: quasi-relativi con *ni budi* e *ni est* nella lingua russa del XVII-XVIII secolo). *Vestnik Sankt-Peterburgskogo Universiteta. Jazyk i literatura,* 18(1), 114-37. https://doi.org/10.21638/spbu09.2021.107.

Pen'kova, Ja.A. (2021c). «K voprosu o grammatikalizacii glagola stati v istorii russkogo jazyka» К вопросу о грамматикализации глагола стати в истории русского языка (Sulla grammaticalizzazione del verbo *stati* nella storia della lingua russa). Pičchadze, A.A. et al. (otv. red.), *Slova, konstrukcii i teksty v istorii russkoj pis'mennosti. Sbornik statej k 70-letiju akademica A.M. Moldovana*. Moskva; Sankt-Peterburg: Nestor-Istorija, 51-72. https://doi.org/10.31912/slova_konstrukzii_2021_51–73.

Pen'kova, Ja.A. (2021d). «Infinitivnye konstrukcii s glagolom iměti v pozdrevnerusskoj pis'mennosti» Инфинитивные конструкции с глаголом имѣти в поздревнерусской письменности (Costruzioni infinitive con il verbo *iměti* nella letteratura tardomedio-russa). Ladyženskij, I.M.; Puzina, M.A. (otv. red.), *Sub specie aeternitatis. Sbornik naučnych statej k 60-letiju Vadima Borisoviča Krys'ko*. Moskva: Izdatel'skij centr Azbukovnik, 212–27. https://doi.org/10.31912/975-5-91172-215-9/212-228.

Pičchadze, A.A. (2011). *Perevodčeskaja dejatel'nost' v domongol'skoj Rusi: lingvističeskij aspekt* Переводческая деятельность в домонгольской Руси: лингвистический аспект (L'attività di traduzione nella Rus' premongola: aspetto linguistico). Moskva: Rukopisnye pamjatniki Drevnej Rusi.

Picchio, R. (1958). "Prerinascimento est-europeo e Rinascita slava ortodossa". *Ricerche Slavistiche*, 6, 185-99.

Poljakov, A.E. (2014). «Korpus cerkovnoslavjanskich tekstov, problemy orfografii i grammatiki» Корпус церковнославянских текстов, проблемы орфографии и грамматики (Il corpus di testi slavo-ecclesiastici: problemi di ortografia e grammatica). *Przegląd Wschodnioeuropejski,* 5(1), 245-54.

Popova, L.V. (2022). «Prestuplenie kak imja koncepta v istorii russkoj jazykovoj kartina mira» Преступление как имя концепта в истории русской языковой картины мира (Il crimine come nome di un concetto nella storia dell'immagine linguistica russa del mondo). *Vestnik Ufimoskogo juridičeskogo instituta MVD Rossii. Serija Obšie voprosy jazykoznanija. Jazyk i pravo. Jurislingvistgika,* 1(95), 135-41.

Rabus, A. et al. (2023). «Developing a Pipeline for Automatic Linguistic Analysis of Historical Manuscripts and Early Printings: The Pre-Modern Slavic Case». Baillot, A. et al. (eds.) *Annual International Conference of the Alliance of Digital Humanities Organizations, DH 2022* (Graz, Austria, July 10-14, 2023). *Conference Abstracts*. https://doi.org/10.5281/zenodo.8107622.

Reznikova, T.I. (2009). «Slavjanskaja korpusnaja lingvistika: sovremennoe sostojanie resursov» Славянская корпусная лингвистика: современное состояние ресурсов (La linguistica dei *corpora* slava: stato attuale delle risorse). Plungjan, V.A. (otv. red.), *Nacional'nyj korpus russkogo jazyka: 2006-2008. Novye rezul'taty i perspektivy.* Sankt-Peterburg: Nestor-Istorija, 402-61.

Romoli, F. (2018). «Sl.eccl. *mudrostĭ*. Studi per un lexicon plurilingue dei termini religiosi e filosofico-teologico». Ferro, M.C.; Salmon, L.; Ziffer, G. (a cura di), *Contributi italiani al XVI Congresso Internazionale degli Slavisti (Belgrado, 20-27 agosto 2018)*. Firenze: Firenze University Press, 37-50. https://doi.org/10.36253/978-88-6453-723-8.06pp.

Rundell, M.; Stock P. (1992). «The Corpus Revolution». *English Today*, 8(3), 21-32.

Rusanova, O.S. (2020). «Drevnerusskie ustojčivye glagol'no-imennye sočetanija s komponentom dějati: semantika i funcionirovanie» Древнерусские устойчивые глагольно-именные сочетания с компонентом дѣяти: семантика и функционирование (Collocazioni verbo-nome del russo antico con la componente *dějati*: semantica e funzionamento). *Aktual'nye voprosy filologičeskoj nauki XXI veka. Sbornik statej IX Meždunarodnoj naučnoj konferencii molodych učenych* (7 fevralja 2020 g.). Čast' 1, *Sovremennye lingvističeskie issledovanija*. V 2-ch častjach, Ekaterinburg, 83-8.

Rusanova, O.S. (2021). «Glagol'no-imennye sočetanija s komponentom tvoriti v russkich delovych tekstach XIII-XVII vv.» Глагольно-именные сочетания с компонентом творити в русских деловых текстах XIII-XVII

вв. (Collocazioni verbo-nome con la componente *tvoriti* nei testi commerciali russi dei secoli XIII-XVII). *Povolžskij pedagogičeskij vestnik*, tom 9, 4(33), 97-101.

Rusanova, O.S. (2022). «Glagol'no-imennye sočetanija s komponentom iměti v russkich gramotach XIII-XVI vv.» Глагольно-именные сочетания с компонентом имѣти в русских грамотах XIII-XVI вв. (Collocazioni verbo-nome con la componente *iměti* nelle *gramoty* russi dei secoli XIII-XIV). Chramušina, Ž.A. et al. (otv. red.), *Aktual'nye voprosy perevoda, lingvistiki, istorii literatury i fol'kloraю Sbornik statej X Meždunarodnoj naučnoj konferencii molodych učenych* (11 fevralja 2022 g.), Ekaterinburg, 45-8.

Ruvoletto, L. (2018). «Dvuvidovost' glagola běžat' v diachroničeskoj perspective» Двувидовость глагола бѣжать в диахронической перспективе (Biaspettualità del verbo *bězati* in prospettiva diacronica). Patroeva, N.V. (otv. red.), *Fortunatovskie čtenija v Karelii. Sbornik dokladov meždunarodnoj naučnoj konferencii* (Petrozavodsk, 10-12 sentjabrja 2018 g., Petrozavodsk), Petrozavodsk, 139-43.

Ruvoletto, L. (2019). «Note sul verbo *'běžati'* in slavo orientale antico». Krapova, I.; Nistratova S.; Ruvoletto, L. (a cura di), *Studi di Linguistica slava. Nuove prospettive e metodologie per la ricerca*. Venezia: Edizioni Ca' Foscari, 471-79. https://doi.org/10.30687/978-88-6969-368-7/030.

Ruvoletto, L. (2020). «Glagoly boléti i goréti v drevnerusskom jazyke» Глаголы болѣти и горѣти в древнерусском языке (I verbi *boléti* e *goréti* nello slavo orientale antico). Gorbova E.V. (otv. red.), *Vzaimodejstvie aspekta so smežnimi kategorijami. Materialy VII Meždunarodnoj konferencii Komissii po aspektologii Meždunarodnogo komiteta slavistov* (Sankt-Peterburg, 5-8 maja 2020 goda), Sankt-Peterburg, 361-70.

Sičinava, D.V. (2022). «Korpus beretjanych gramot kak parallel'nyj» Корпус берестяных грамот как параллельный (Il corpus *berestjanye gramoty* come corpus parallelo). Plungjan, V.A. (otv. red.), *Trudy Instituta russkogo jazyka RAN,* N. 2(32). Moskva: Institut russkogo jazyka im. V.V. Vinogradova RAN, 92-106.

Sičinava, D.V.; Dyškant, A.N. (2021). «Integration of the Old East Slavic Epigraphical Database. Corpora and Indices». *Scripta & e-Scripta. The Journal of Interdisciplinary Mediaeval Studies*, volume 21, 95-108.

Straka, M; Hajič, J; Strakova´ J. (2016). «UDPipe: Trainable Pipeline for Processing CoNLL-U Files Performing Tokenization, Morphological Analysis, POS Tagging and Parsing». Calzolari, N. et al. (eds), *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC-2016)*, Portorož, 4290–97.

Totomanova, A.M. (2012). «Digital Presentation of Bulgarian Lexical Heritage. Towards an Electronic Historical Dictionary». *Studia Ceranea*, 2, 221-32. https://doi.org/10.18778/2084-140x.02.17.

Totomanova, A.M. (2017). «Diachronnyj korpus bolgarskogo jazyka. Sostojanie i perspektivy» Диахронический корпус болгарского языка: состояние и перспективы (Corpus diacronico della lingua bulgara: stato e prospettive). *Filologija*, 68, 223-42. https://doi.org/10.21857/yrvgqtkj39.

Totomanova, A.M. (2019). «Diachronic Corpus of the Old Bulgarian Language – content and research opportunities». *ReIReS DH Course* [DH Course on «Digital Resources for Slavonic-Byzantine Studies», Sofia 19-21 November 2019].

Totomanova, A.M. (2021). «Electronic Research Infrastructure for Bulgarian Medieval Written Heritage: history and perspective». *Diacronia*, 14, 1-9. https://doi.org/10.17684/i14a193en.

Totomanova, A.M.; Slavova, T.; Ganeva, G. (2015). Totomanova; A.M.; Slavova, T. (otv. red.), «Morfosintaktičen tagset na starobălgarskija knižoven ezik» Морфосинтактичен тагсет на старобългарския книжовен език (Tagset morfosintattico per la lingua letteraria bulgara antica). Informatika, gramatika, leksikografija BG051-3.3-06-0024/2012». *Sbornik dokladi i materiali ot zaključitelnata konferencija,* (Sofija, 29-30.06), Sofija, 5-16.

Uspenskij, B.A. (1994). *Kratkij očerk istorii russkogo literaturnogo jazyka (XI-XIX vv.)* Краткий очерк истории русского литературного языка (XI-XIX вв.) (Breve compendio della storia della lingua letteraria russa [XI-XIX sec.]). Moskva: Gnosis.

Uspenskij, B.A. (2002). *Istorija russkogo literaturnogo jazyka (XI-XVII vv.)* История русского литературного языка (XI-XVII вв.) (Storia della lingua letteraria russa [XI-XVII sec.]). Izdanie 3-e, ispravlennoe i dopolnennoe. Moskva: Aspent Press.

Zacharov, V.P. (2015). «Istoričeskie korpusa i korpusnye diachroničeskie issledovanija» Исторические корпуса и корпусные диахронические исследования (*Corpora* storici e analisi diacroniche *corpus-based*). Dergačeva-Skop E.I. (otv. red.), *Pis'mennoe nasledie i informacionnye technologii "El'Manuscript-2015".* Novosibirsk: Gosudarstvennaja publičnaja naučno-techničeskaja biblioteka SO RAN, 11-13.

Zacharov, V.P.; Bogdanova, S.Ju. (2013). *Korpusnaja lingvistika. Učebnik dlja studentov napravlenija «Lingvistika»* Корпусная лингвистика. Учебник для студентов направления «Лингвистика» (Linguistica dei *corpora*. Manuale per gli studenti dei corsi di Linguistica). Vtoroe Izdanie. Sankt-Peterburg: Sankt-Peterburgskij gosudarstvennyj universitet.

Zagrebina, U.S. (2022). «Prilagatel'noe sil'nyj v russkich letopisjach XVI veka: osobennosti semantiki i funkcionirovanija» Прилагательное сильный в русских летописях XVI века: особенности семантики и функционирования (L'aggettivo *sil'nyj* nelle cronache russe del XVI secolo: caratteristiche semantiche e di funzionamento). Chramušina, Ž.A. et al. (otv. red.), *Aktual'nye voprosy perevoda, lingvistiki, istorii literatury i fol'kloraю Sbornik statej X Meždunarodnoj naučnoj konferencii molodych učenych* (11 fevralja 2022 g.), Ekaterinburg, 40-5.

Zaliznjak, A.A. (2004). *Drevnenovgorodksij Dialekt* Древненовгородский диалект (L'antico dialetto di Novgorod). Vtoroe izdanie. Moskva: Jazyki slavjanskoj kul'tury.

Žolobov, O.F. (2022). «Leksičeskij rjad branĭ-voina-ratĭ v staroslavjanskich i cerkovnoslavjanskich istočnikach XI-XV vv.: Interpretacija kvantitativnych formul (Na materiale istoričeskogo korpusa "Manuskript")» Лексический ряд брань-война-рать в старославянских и церковнославянских источниках XI-XV вв.: интерпретация количественных формул (на материале исторического корпуса «Манускрипт») (La serie lessicale *branĭ-voina-ratĭ* nelle fonti paleoslave e slavo-ecclesiastiche dell'XI-XV sec.: interpretazione delle formule quantitative [sui materiali del corpus storico *Manuskript*)]. *Palaeobulgarica*, 46(4), 79- 99.

Žolobov, O.F.; Baranov, V.A. (2021). «Distributivno-kvantitativnye i semantičeskie charakteristiki glagolov znanija v staroslavjanskoj i drevnerusskoj pis'mennosti» Дистрибутивно-квантитативные и

семантические характеристики глаголов знания в старославянской и древнерусской письменности (Caratteristiche distributive-quantitative e semantiche dei verbi della conoscenza nella scrittura paleoslava e russa antica). *Vestnik Sankt-Peterburgskogo universiteta. Jazyk i literatura*, 18(1), 56-76. https://doi.org/10.21638/spbu09.2021.104.

Žolobov, O.F.; Baranov, V.A. (2022). «Transformacii leksičeskogo rjada životĭ – žiznĭ – žitie: opyt lingvostatističeskogo opisanija» Трансформации лексическогорядаживотъ–жизнь–житие:опытлингвостатистического описания (Trasformazioni della serie lessicale *život' – žizn' – žitie*: esperienza di descrizione linguistico-statistica). *Voprosy Jazykoznanija*, 2, 65-101. https://10.31857/0373-658X.2022.2.65-101.

Zuga, O.V. (2009). «Iz nabljudenij nad charakterom jazykovych raznočtenij v slavjanskich spiskach Evangelija XII-XIII vv. (Na materiale "Pritči o bludnom syne")» Из наблюдений над характером языковых разночитений в славянских списках Евангелия XII–XIII вв. (На материале «Притчи о блудном сыне») (Osservazioni sulla natura delle varianti linguistiche nei manoscritti slavi del Vangelo dei secoli XII–XIII [sul materiale della Parabola del Figliol Prodigo]). *Vestnik Vjatskogo gosudarstvennogo universiteta*, 3(2), 40-6.

Zuga, O.V. (2010). «Specializirovannoe internet-izdanie «Slavjanskie Evangelija» i aktual'nye problemy izučenija istorii russkogo jazyka v inojazyčnoj auditorii» Специализированное интернет-издание «Славянские Евангелия» и актуальные проблемы изучения истории русского языка в иноязычной аудитории (L'edizione online specializzata 'Vangeli Slavi' e i problemi attuali dello studio della storia della lingua russa per un pubblico straniero). *Vestnik Udmurtskogo universiteta*. *Serija Istorija i Filologija*, 2, 3-8.

e-ISSN 2785-3187

Balcania et Slavia
Vol. 4 — Num. 2 — Dicembre 2024

Genitival Phrases in Albanian and Romanian
A Comparative Approach

Hysnie Haxhillari
Universiteti Fan S. Noli, Albania

Abstract Albanian and Romanian display a specific Genitival Phrase structure which is unique among Balkan's languages. It is characterized by the presence of a genitive marker, traditionally called 'genitival article', that follows the head noun. The goal of the present paper is twofold: on the one hand, it focuses on the origin and the syntactic role the genitival article has within GenP. On the other hand, this paper compares the Genitival Phrase of Albanian to that of Romanian in order to show similarities and differences in both their morphological and syntactic structure.

Keywords Albanian. Romanian. Genitival Phrase. Genitival marker.

Edizioni Ca'Foscari

Peer review

Submitted 2024-05-16
Accepted 2024-10-23
Published 2024-12-20

Open access

Citation Haxhillari, Hysnie (2024). "Genitival Phrases in Albanian and Romanian". *Balcania et Slavia*, 4(2), 163-174.

DOI 10.30687/BES/2785-3187/2024/01/002

1 Introduction

The Genitival Phrase of Albanian and Romanian has been a main issue in Balkan Linguistics for many years. Many scholars have studied its syntactic structure, which is very different from the other Balkan languages since it displays a genitival marker, which is traditionally called 'genitival article'.

The studies on the Genitival Phrase have been focused on two basic structures: the Saxon genitive phrase like *John's book* and the prepositional one *the book of John* (Hofherr, Dobrovie-Sorin 2005, 225). Unlike the English genitival phrases illustrated above, Albanian and Romanian display a structure containing a genitival marker (ART) which is followed by a noun marked for genitive Case. It has the following configuration:

(1) ART + NP.GEN

In both languages, the genitival marker agrees with the head noun. In Romanian, it agrees only for gender and number (*Gramatica de bază a limbii romane* 2016, 63):

(2) a. **Elev** **al profesorului**
student.$_{\text{MASC.SING. ART}}$ professor.$_{\text{GEN.SING}}$
'Professor's student'

b. **Acestei** **studente** **ale facultăţii**
these.$_{\text{FEM,PL}}$ students.$_{\text{FEM,PL ART.FEM.PL}}$ faculty
'These students of the faculty'

In Albanian, the genitival article agrees in gender, number and Case:

(3) a. **Libri** **i** **nxënësit**
book.$_{\text{MASC.SING.NOM}}$ $_{\text{ART.MASC.SING}}$ student.$_{\text{MASC.SING.GEN}}$
'The student's book'

b. **Fletorja** **e** **nxënësit/nxënëses**
notebook.$_{\text{FEM.SING.NOM}}$ $_{\text{ART.FEM.SING}}$ student.$_{\text{MASC/FEM.SING.GEN}}$
'The student's notebook'

c. **Librave** **të** **nxënësve**
books.$_{\text{PL.DAT}}$ $_{\text{ART.PL.DAT}}$ students.$_{\text{PL.GEN}}$
'To the students books'

A main difference between Romanian and Albanian is that the Romanian Genitive Phrase may be realized both by the syntactic marker *al* (2a) or by the proclitic *lui* which precedes proper nouns:

(4) **Lui Ion**
'Ion's'

e-ISSN 2785-3187

Albanian, instead, has only one structure to express the Genitival Phrase: the one represented in (1).

It is very important to highlight the role of the head noun, i.e. the noun which precedes GenP: it realizes the possessive relation between the possessum and the possessor. This is shown in (5) for Albanian and in (6) for Romanian:

(5) **Libri** **i** **studentit**
book.$_{\text{MASC.SING.NOM}}$ ART.MASC.SING.GEN student.$_{\text{MASC.SING.GEN}}$
'The student's book'

(6) **O** **carte** **al elevului**
a book $_{\text{ART}}$ student$_{\text{GEN}}$
'A student's book'

Genitival Phrases of Albanian and Romanian represent a unique structure among other Balkan languages.

2 On the 'genitival article' in Albanian and Romanian GenP

In the Albanian Grammar, the so-called 'genitival article' is historically related to other prepositional articles like:

- the adjectival article *i mirë, i bukur* ('nice', 'beautiful')
- possessive article *i ati, e ëma* ('his father', 'his/her mother'), etc.
- pronominal article *i tij, i saj, i tillë* ('his', 'her', 'such'), etc.
- the article preceding week's days *e hënë, e martë, e mërkurë* ('Monday', 'Tuesday', 'Wednesday'), etc.
- the article preceding adjectival quantifiers *i parë, i dytë, të dy, të tre, të gjithë* ('first', 'second', 'both', 'the three of them', 'all'), etc.
- the article preceding deverbal neuter nouns *të folurit, të qeshurit* ('speaking', 'laughing'), etc.

The most similar article to the 'genitival article' is the adjectival article for they display the same function and morphological form. Both these articles agree in gender, number and case with the preceding noun: *drejtori i shkollës* / *drejtori i ri* ('the school director' / 'the new director'); *drejtoresha e shkollës* / *drejtoresha e re* ('the school director' / 'the new director', feminine) (*Gramatika e gjuhës shqipe* 2002, 106).

In Romanian, on the other side, except the genitival phrase, the genitival article may appear in two other structures: adjectival possessives (7a) and ordinals (7b):

(7) a. **Al nostru**
'Ours'
b. **Al doilea**
'Second'

The Albanian genitival article originated from a demonstrative pronoun which has been used before a noun in genitive Case in order to express the meaning of the possessiveness, the relevant value of the genitive Case. This demonstrative was the only element which contributed to marking the distinction between genitive and dative Case in the Albanian language (Likaj 2003, 54).

The term 'article', when referred to the Genitival Phrase, has not the meaning of 'definite article', nor the same function, both in Romanian and Albanian. The Romanian genitival article *al* has been labelled as 'syntactic genitive marker' in *Gramatica de Bază a Limbi Române* (2016, 61) or 'genitive marker' in Giurgea (2014, 71). The elliptic use of *al* in the following example makes *al* to look like a semi-dependent pronoun:

(8) **Al mei au plecat la mare**
ART mine they leave the sea
'Mine are gone to the sea'

Nevertheless, it is not considered an article yet, as it replaces the noun. Also, in the examples where *al* is used in combination with quantifiers as in *al doilea* ('second'), it is considered a lexical affix which is not related to definiteness, the main function of the articles (*Gramatica de Bază a Limbii Romane* 2016, 87). Therefore, it is more appropriate to use the word 'marker' not 'article' when we refer to the 'genitive marker" in Albanian and Romanian GenP.

3 The syntactic role of the genitive marker

According to Abney (1987, 271), determiners are in complementary distribution with possessives: AGR in D does not co-occur with lexical determiners:

(9) *John('s) the/that/some books

Assuming that possessors only appear when there is an AGR position in D (which assigns genitive Case), Abney claims that the inability of AGR to co-occur with lexical determiners explains the inability of possessors to co-occur with determiners. Hence, the genitival marker, as a possessor, cannot co-occur with other determiners in the GenPs.

In Romanian, GenP is used only after an indefinite noun (10), meanwhile in Albanian, it may follow a definite or an indefinite noun (11):

(10) a. **Un caiet al Mariei**
a notebook ART Mary.GEN
'Mary's notebook'
b. **Caietul al Mariei*
notebook$_{.DEF}$ $_{GEN.ART}$ Mary.$_{GEN}$

(11) a. **Krahu i fluturës**
arm.$_{DEF}$ $_{ART}$ butterfly.$_{GEN.DEF}$
'The butterfly's arm'
b. **Një krah i fluturës**
an arm $_{ART}$ butterfly$_{GEN.DEF}$
'An arm of the butterfly'

There are constructions in Romanian (12) and Albanian (13) where the genitive marker is in D position, adjacent to the noun, and in a complementary distribution with other determiners:

(12) a. **Caietul acesta al Mariei**
notebook.$_{DEF}$ this $_{GEN.ART}$ Mary.$_{GEN}$
'This notebook of Mary'
b. **Caietul cel nou al Mariei**
notebook.$_{DEF}$ $_{ADJ.ART}$ new $_{ART}$ Mary.$_{GEN}$
'The new notebook of Mary'

(13) a. **Krahët e bukur të fluturës**
arms$_{.DEF}$ $_{ADJ.ART}$ beautiful $_{GEN.ART}$ butterfly.$_{GEN}$
'The beautiful arms of the butterfly'
b. **Bota jonë e kinemasë**
world.$_{DEF}$ our $_{GEN.ART}$ cinema.$_{GEN}$
'Our world of cinema'

But, there are examples in Albanian (14) and Romanian (15) in which the genitival marker may be in a Specifier position, i.e. it is not adjacent to the genitive noun. In this cases, it appears in the leftmost position in the GenP, in a non-complementary distribution with other lexical determiners like demonstratives and quantifiers:

(14) **Krahu i një /kësaj/çdo fluture**
arm.$_{DEF}$ $_{GEN.ART}$ a/this/each butterfly$_{GEN}$
'The arm of a/this/each butterfly'

(15) **Al tuturor acestor fete**
$_{GEN.ART}$ all.$_{GEN.PL}$ these.$_{GEN.PL}$ girls

'Of all these girls'

In Albanian, the Genitival Phrase displays inflectional features not only on the possessor but also on the pre-nominal genitival article, which agrees in number, gender and Case with the head noun that precedes it:

(16)	Nom.	**libri**	**i**	**studentit**
		book.$_{\text{MASC.SING.NOM}}$	$_{\text{MASC.SING.NOM}}$	student
	Gen.	**i/e librit**	**të**.	**studentit**
		book.$_{\text{MASC.SING.GEN}}$	$_{\text{MASC.SING.GEN}}$	student
	Dat.	**librit**	**të**.	**studentit**
		book.$_{\text{MASC.SING.DAT}}$	$_{\text{MASC.SING.DAT}}$	student
	Acc.	**librin**	**e**	**studentit**
		book.$_{\text{MASC.SING.ACC}}$	$_{\text{MASC.SING.ACC}}$	student
	Abl.	**librit**	**të**.	**studentit**
		book.$_{\text{MASC.SING.ABL}}$	$_{\text{MASC.SING.ABL}}$	student
		'The student's book'		

The above paradigm shows that the genitive marker morphologically agrees with the head noun, even if syntactically it is part of the genitive noun.

According to Likaj (2003, 54), both the article and the endings of the head noun originated from the same demonstrative. He assumes that the endings of the head noun, which belongs to the definite declension of the noun, come from a poly-definite construction:

(17)	*libr-i*	*i*	*nxënësit = libër*	*(a)i*	*(a)tij*	*nxënësi*
	book-the ART		student.GEN = book	this	that.$_{\text{GEN.}}$	student

In Romanian genitive phrases there is not an inflectional paradigm on the genitival marker. Giurgea (2014, 72) notes that the genitive in Romanian is a syntactic notion, not a morphological one, as there is no distinct inflectional genitive Case. Theres is a morphological element restricted to genitive environments, but this is not an inflectional element: it is the possessive stem of agreeing possessors (Giurgea 2014, 4). In both languages, the genitival marker has the same origin from the articles (originally demonstratives), they differ only in their inflectional paradigm: in Albanian the genitive marker has an inflectional Case, in Romanian, instead, it agrees only in gender and number, not in Case.

As Catasso (2011, 85) notes, the agreement with the preceding noun in number and gender displayed by this genitival article "is significant for two reasons: on the one hand, as he argues for Romanian, it indicates an asymmetry between the preceding noun and the genitival phrase in that the 'article' displays a much stronger relation

to the former than to the latter – which is not the case in other languages of the Balkan Sprachbund (except Romanian) and is not so widely diffused from a typological perspective. Secondly, this controversial element is the same found in the AP, which suggests that a demonstrative-like features may be attributed to this 'article". Moreover, Catasso hypothesizes that a correlation between the element on the Adjectival Phrase and that in the GenP (if we assume that it structurally belongs to the GenP) is definitely present. See the Albanian examples in (18):

(18) a. **Libri i tij**
book.$_{DEF}$ ART his
'His book'
b. **Libri i studentit**
book.$_{DEF}$ $_{ART}$ student.$_{GEN}$
'The student's book'
c. **Libri i kuq**
book.$_{DEF}$ ADJ.ART red
'The red book'

Catasso (2011, 85) assumes that these three instances in the use of this element reveal that *i* has a special function: it seems to particularize the noun it follows by restricting the semantic domain ('his book' < 'that of him'/ 'that book which belongs to him'). The function of *i* in the possessive construction above, taken from Catasso, really seems to fulfill the role of particularizing the noun it follows by restricting the semantic domain. But, in the adjectival phrase in (18c), *i* has not the same function. The preposed adjectival article in Albanian takes a predicative meaning into a copular construction like *libri i kuq – libri është i kuq / libri që është i kuq* ('the book red' – 'the book is red' / 'the book which is red'), as Campos (2009, 1028) notes for Greek APs to *vivlio to kokkino* ('the book the red'). Even though its historical origin is the same, it originated from a definite article or a demonstrative pronoun.

In conclusion of this section, we list some similarities and differences between the genitive marker in both languages.

a) Romanian *al* is dropped after the definite article (19), whereas in Albanian the genitival article *i/e* is obligatory (20):

(19) **Băiatu' vecinului**
boy.$_{DEF}$ neighbor.$_{DEF.OBL}$
'The neighbor boy'

(20) **Djali i fqinjit**
boy.$_{DEF}$ $_{ART}$ neighbour$_{.GEN}$

e-ISSN 2785-3187

'The neighbor's boy'

b) Romanian *al* is an obligatory constituent of GenP if the head noun is indefinite (21), which is the case in Albanian too (22):

(21) **Un băiat a vecinului**
a boy $_{GEN}$ neighbour.$_{GEN}$
'A neighbor's boy'

(22) **Një djalë i fqinjit**
a boy $_{GEN}$ neighbour.$_{GEN.DEF}$
'A neighbor's boy'

c) Genitive marker, in its main use, is in D in both languages, but it may display a specifier role preceding determiners. This is shown in the Romanian example in (23a) and in the Albanian (23b):

(23) a. **Al tuturor acestor fete**
'Of all these girls'
b. **Krahu i një fluture**
'The arm of a butterfy'

d) Both languages display an elliptic use of the head noun:

(24) **Casa Mariei e mai mare decât a Ioanei**
house Mary.$_{OBL}$ is more big than of Ioan.$_{OBL}$
'Mary's house is bigger than Ioan's'

(25) Shtëpia e Marisë është më e madhe se **e Joanës**
house.$_{DEF\ GEN}$ Mary.$_{GEN}$ is more $_{ADJ.ART}$ big than $_{GEN}$ Ioan.$_{GEN}$
'Mary's house is bigger than Ioan's

4 The Structure of the Genitival Phrase in Albanian and Romanian

What kind of constituent is the GenP in both languages, Albanian and Romanian?

In languages like English, the Possessor occupies a DP position (Alexiadou, Stavrou, Haegeman 2007). It is known as the first constituent of the phrase: 'John's book'.

In Albanian (26) and Romanian (27) things are different. Unlike many languages, the possessor in Albanian and Romanian follows the head noun, appearing in a lower position lower than that of the head noun:

(26) **Libri i Gjonit**
book.$_{DEF}$ $_{GEN}$ John.$_{GEN.DEF}$
'John's book'

(27) **Un carte al elevului**
a book $_{GEN}$ student.$_{GEN}$
'A student's book'

The prenominal position is ungrammatical in both languages:

(28) a. ***I Gjonit libri**
b. ***Al elevului un carte**

The possessor occupies a lower position, to the right of the possessum. The extraction of the GenP is possible only in a copular construction. See the Albanian example in (29) and Romanian in (30):

(29) **Libri është i Gjonit**
Book.$_{DEF}$ is $_{GEN}$ John.$_{GEN}$
'The book is John's '

(30) **Caietul este al profesorului/meu**
Book.$_{DEF}$ is $_{GEN}$ professor.$_{GEN}$/mine
'The book belongs to the professor/ to me'

The examples above show that, in both languages, the genitival phrase is derived in a lower position: it doesn't move to the left edge to occupy a specifier position. It occupies a complement position which follows the head noun.

However, in copular constructions or in why-clauses, the possessor may occupy the Spec DP position:

(31) a. **I Gjonit është libri/ ky libër?**
$_{GEN}$ Gjon.$_{GEN}$ is book.$_{DEF}$ /this book
'Is John's this book?'
b. **A lui Gjon este cartea/această carte?**
$_{GEN}$ Gjon.$_{GEN}$ is book.$_{DEF}$ /this book
'Is John's this book?

According to Cornilescu & Nicolae (2011, 38), the Romanian GenP may have another syntactic position: it may precede a head noun that has a possession meaning, like the Genitive Saxon in English 'country's flag'. In these constructions, it has the role of a definite determiner and it is always the first constituent of the DP:

(32) a. **Al ţării steag**
ART country's flag
'Country's flag'
b. **Al casei prag**
ART house threshold
'House's thresold'

Following Cornilescu, Nicolae (2011, 38), if the genitive DP is prenominal, it is a [+N] element and it automatically qualifies as the bearer of definiteness. Therefore the genitive DP should be the specifier of the projection immediately below D, ultimately moving to [Spec, DP].

GenP preceding the Nominal Expression are not common in Albanian, where the unmarked order is head noun + genitive marker + $noun_{gen.}$ However, there are examples where the GenP appears at the left edge of the NP in some emphatic uses, in poetry, but unusual in other contexts:

(33) a. **Unë, i qiejve agjitator** (Kadare, I. 2018, 158, 112, 115)
I.$_{\text{NOM ART}}$ skies agitator
'Me, of skies agitator'
b. **I mendimeve vargu nuk t'u sos**
ART thoughts verse.$_{\text{DEF}}$ not get limited
'Of thoughts the verse didn't get limited'

As Giurgea (2011, 27) notes for Romanian language, Genitive Phrases can be coordinated. In these cases, the article *al* can be realized in the second GenP:

(34) **Apartamentul** [**mamei mele**] **şi** [**al Mariei**]**a fost vândut**
apartment.$_{\text{NOM DEF}}$ mother.$_{\text{OBL}}$ mine and GEN Mary.$_{\text{GEN}}$ was sold
'My mother's car and Mary's is sold out'

In Albanian, on the other side, the genitive marker must be realized in the first GenP otherwise the construction could be ungrammatical:

(35) a. **Vetura e mamasë sime dhe Marisë është shitur**
car.$_{\text{NOM DEF GEN}}$ mother.$_{\text{GEN}}$ my$_{\text{GEN}}$ and Mary.$_{\text{GEN}}$ was sold
'My mother's car and Mary's is sold out'
b. ***Vetura mamasë sime dhe e Marisë është shitur**

5 Conclusions

In this paper, I have presented the Genitival Phrase of Albanian and Romanian. I have shown that it is unique among the other Balkan languages since its structure comprehends a head noun, a genitive marker, agreeing with the head noun, and a noun inflected for genitive Case. In both languages, the genitive marker is traditionally called 'prenominal/genitival article' and is derived from an article/demonstrative (Romanian/Albanian), but it is not related to the definite article or definiteness. Therefore, it is more appropriate to use the label 'genitive marker' instead of 'prenominal / genitival article'.

The genitive marker displays two syntactic roles inside the GenP: it is a determiner D when it is in complementary distribution with other lexical determiners, but it may have the role of Specifier too, when it has a non-complementary distribution with pronouns or other elements within GenP.

Genitive Phrases of Albanian and Romanian usually are complements: they occupy a lower position in the Nominal Expression, to the right of the head noun. The extraction of the GenP is possible only in a copular construction or in why-clauses where the GenP may display a DP role, at the left edge of the Nominal Expression.

References

Abney, S.P. (1987). *The English Noun Phrase in its Sentential Aspect* [PhD Dissertation]. Cambridge (MA): MIT Press.

Academia Română, Institutul de Lingvistică "Iorgu Iordan – Al. Rosetti" (2016). *Gramatica de bază a limbii romane*. Bucureşti: AR, IL.

Akademia e Shkencave e Shqipërisë, Instituti i Gjuhësisë dhe i Letërsisë (2002). *Gramatika e gjuhës shqipe, I*. Tiranë: ASHSH, IGJL.

Alexiadou, A.; Haegeman, L.; Stavrou, M. (2007). *Noun Phrase in the Generative Perspective*. Berlin: Mouton de Gruyter. Studies in Generative Grammar 71. https://doi.org/10.1515/9783110207491.

Campos, H. (2008). "Some Notes on Adjectival Articles in Albanian". *Lingua*, 7(119), 1009-34. https://doi.org/10.1016/j.lingua.2008.09.014.

Catasso, N. (2011). "Genitive-Dative Syncretism in the Balkan Sprachbund: An Invitation to Discussion". *SKASE Journal of Theoretical Linguistics*, 2(8), 70-93.

Cornilescu, A.; Nicolae, A. (2011). "On the History of Romanian Genitives: The Prenominal Genitive". *Bucharest Working Papers in Linguistics*, 2(13), 37-55.

Giurgea, I. (2011). "Agreeing Possessors and the Theory of Case". *Bucharest Working Papers in Linguistics*, 2(13), 5-37.

Giurgea, I. (2014). "Romanian Al and the Syntax of Case Heads". *Bucharest Working Papers in Linguistics*, 2(16), 69-98.

Hofherr, P.C.; Dobrovie-Sorin, C. (2004). "The Syntax and Semantics of Argument and Modifier Genitives". (ed.) (2005), *Contributions to the Thirtieth Incontro di Grammatica Generativa Venice*. Venice: Cafoscarina, 225-41.

Kadare, I. (2018). *Vepra poetike*. Tiranë: Onufri.

Likaj, E. (2003). *Zhvillimi i eptimit në gjuhën shqipe*. Tiranë: SHBLU.

e-ISSN 2785-3187

Balcania et Slavia
Vol. 4 — Num. 2 — Dicembre 2024

Complementizer Drop and Clitic Climbing in Torlakian: Evidence from the Timok Variety

Mirjana Mirić
Institute for Balkan Studies SASA, Serbia

Boban Arsenijević
Institute of Slavic Studies, University of Graz, Austria

Abstract The paper is based on a corpus study of the complementizer *da* drop (Comp_drop) and clitic climbing (CC) in the Timok variety of Serbian. Data are taken from the *Spoken Torlak dialect corpus 1.0* and comprise over 900 examples of sentences containing the following selecting verbs: modal verbs *moći* 'can' and *morati* 'must', light verbs *ići* 'go' and *uzeti* 'take', and the aspectual verb *početi* 'begin', and their complement clauses, with or without the complementizer *da*. Two predicted variables, Comp_drop and CC, are analyzed alongside eleven predictor variables. Statistical analyses reveal that Comp_drop is optional and significantly affected by the person of the selected verb, the availability of clitics in the structure, and the presence of the impersonal construction. On the other hand, CC is possible but rare. It occurs only when the complementizer is dropped, and is affected by the type of verb, particularly facilitated by the aspectual verb *početi* 'begin'. The results are discussed in terms of the structural size of the complement clauses, in line with the proposal of a clause size smaller than the smallest one attested in the standard variety and the observation that CC is available from functionally poor domains, i.e. from restructuring complements.

Keywords Clitic climbing. Complement clause size. Complementiser drop. Serbian. Torlakian

Edizioni Ca'Foscari

Peer review

Submitted 2024-10-24
Accepted 2025-01-27
Published 2025-03-25

Citation Mirić, Mirjana; Arsenijević, Boban (2024). "Complementizer Drop and Clitic Climbing in Torlakian: Evidence from the Timok Variety". *Balcania et Slavia*, 4(2), 175 -200.

DOI 10.30687/BES/2785-3187/2024/02/003

1 Introduction

Clausal complements in Bosnian / Croatian / Montenegrin / Serbian (BCMS) have figured prominently in the recent theorizing about complement clauses (Progovac 1993; Wurmbrand 2001; Stjepanović 2004; Arsenijević 2009; Kovačević, Milićev 2018; Todorović, Wurmbrand 2020; Wurmbrand et al. 2020; Ivanović et al. 2023; Mirić forthcoming). Central research questions include the structural size of clausal complements, the possibility and structural conditioning of clitic climbing, and the role of the (type of) selecting verb in these two regards. Wurmbrand and colleagues argue that there are three types of verbs and three corresponding sizes of their complements. Aspectual verbs, and modals like *moći* 'can' or *morati* 'must', take the smallest complements, matching a bare VP in case of infinitival complements or a *v*P in case of finite complementation, neither of which can have an own subject or tense specification, as in (1). Verbs like *odlučiti* 'decide', *planirati* 'plan', which too may take infinitives, but also may take complement clauses with their own subjects and tense specification, have structurally somewhat larger complements, of the level of the TP, as in (2). Finally, in BCMS verbs like *tvrditi* 'claim', *verovati* 'believe', cannot have infinitives as complements, or complements of similarly small structures, and only accept full-fledged CPs, as in (3).

(1)

Jovan	je	pokušao	otići /	da	(*Petar)	ode /	*da	će	otići.
J	aux	tried	go.INF	comp	P	go.3SG.PRS	comp	will	go.INF

'Jovan tried to go.'

(2)

Jovan	je	odlučio	otići /	da	(Petar)	ode /	da	će	otići.
J	aux	decided	go.INF	comp	P	go.3SG.PRS	comp	will	go.INF

'Jovan decided to go.'

(3)

Jovan	je	tvrdio	*otići /	*da	(Petar)	ode /	da	će	otići.
J	aux	claimed	go.INF	comp	P	go.3SG.PRS	comp	will	go.INF

'Jovan claimed that he'll go.'

Arsenijević (2022) argues that Torlakian manifests even smaller structures than those in (1), and that therefore the structure in (1) cannot be a bare VP, since there is no structure headed by a verb that

is smaller than VP. The argument is based on structures like (4) and (5), in which the complement lacks the complementizer, and where, he argues, as illustrated in (5), nothing can stand between the two verbs, not even clitics. Arsenijević lists a set of properties indicating the small size of the complements in this construction: a) if both verbs in isolation assign the same case to a clitic, the construction is ineffable (indicating that the case is associated with the two verbs together, not each separately), as in (4-5), b) the complement cannot have its own reference time, as in (6), and c) the complement in such configurations is not a separate prosodic or syntactic unit (it cannot bear an own intonation contour or be moved), as illustrated in (7).[1] The list of properties is summarized in Table 1 (see Arsenijević 2023 for more details).

(4)

Kad	(*Peri)	Mari	(*Peri)	obećao
when	Pera.DAT	Mara.DAT	Pera.DAT	promised
(da	Peri)	pošalje	papiri.	
comp	Pera.DAT	send.3SG.PRS	papers	

intended: 'When he promised to Mara to send the papers to Pera …'

(5)

Kad	(*mu)	joj	(*mu)	obećao
when	he.DAT.CL	she.DAT.CL	he.DAT.CL	promised
(da	mu)	pošalje	papiri.	
COMP	he.DAT.CL	send.3SG.PRS	papers	

intended: 'When he_i promised to her to send the papers to him_j …'

(6)

U	ponedeljak	si	(*me)	obećao
in	Monday	aux.2sg	me.CL	promised
(da	me)	posetiš	u	utorak.

1 We assume that the degradation is due to prosody because: a) the complement can in principle be fronted as shown by the grammaticality of the same example with *da*, and a question in Torlakian in principle can begin with clitics, as shown in (i).

(i)	Si	mu	se	javio?
	aux.2sg.cl	he.dat.cl	refl.cl	called

'Did you call him?'

comp me.CL visit in Tuesday.
'You promised on Monday to visit me on Tuesday.'

(7)

*(Da)	mu	pošalje	papiri	probao	juče?
comp	he.DAT.CL	send.3SG.PRS	papers	tried	yesterday

'Did he try to send him the papers yesterday.'

Table 1 Set of properties characterizing clauses with Comp_drop in Torlakian

	With Comp *da*	**Without Comp *da***
Can include clitics	+	/
Independent argument structure	+	/
Intervention between verbs	+	/
Separate prosodic unit	+	/
Separate syntactic constituent	+	/

This implies that the complementizer-less complement is smaller than the infinitive (the infinitive has its own arguments and can be preposed). As Wurmbrand and colleagues (Todorović, Wurmbrand 2020; Wurmbrand et al. 2020) identify the infinitive as the smallest complement in the standard BCMS, and model it as a VP – the smallest projection of a verb, the availability of a smaller, complementizer-less complement requires that their analysis be amended. The infinitive complement must be larger than the VP, while still being smaller than the projection corresponding to their middle-size complement.

In this paper, we conduct a corpus investigation to test one of the central empirical arguments for Arsenijević's (2022) claim about the smallest finite complement, namely that the absence of the complementizer requires that nothing intervenes between the two verbs. More concretely, the generalization to be tested is that clitics, if available, must climb into the matrix clause and any other material must occur either before the matrix verb or after the embedded one. Further, we also investigate whether there are finer differences between the verbs taking the smallest complements (aspectual verbs, true modals, light verbs) in licensing complementizer drop (henceforth Comp_drop) and clitic climbing (henceforth CC) as markers of the smallest structure in Torlakian. Finally, we conduct exploratory research to acquire a broader picture of the properties relevant for Comp_drop and CC beyond the predictions of particular analyses.

1.1 The Timok variety

The data for the study are taken from a sample of narratives in the Timok variety of Serbian (see Section 2.1), which is a part of the Prizren-Timok dialectal zone, spreading across Southeastern Serbia, and belonging to a larger dialectal area of Torlak (Sobolev et al. 2023). The Timok variety, like the rest of the Torlakian dialects, shares several Balkan linguistic features, such as the loss of the infinitive and its replacement with finite complementation (Sobolev et al. 2023), the grammaticalization of the future tense marker, i.e. the loss of inflectional morphology on the future tense marker, the optional omission of the complementizer *da* (Mirić 2018), analytic cases, and the presence of the postpositive article (Vuković et al. 2023), which makes it distinct from other BCMS varieties.

1.2 Complementizer *da*

In BCMS, Torlakian dialects included, the complementizer *da* is used with proposition, event and situation-type of complements (Mišeska Tomić 2004; Stjepanović 2004; Progovac 2005; Todorović 2012; Kovačević, Milićev 2018; Todorović, Wurmbrand 2020; Wurmbrand et al. 2020; Wurmbrand, Lohninger 2023). In Torlakian, only *da*-complements are possible, given that these dialects lack infinitives. Moreover, the complementizer *da* following verbs which select for situation and event-type of complements can be omitted, as in (4) and (5). As noted by Sobolev (2004: 75), "the omission of the subjunctive particle", i.e. complementizer drop, represents an innovation in the central Balkan zone, found in different dialects spoken in eastern Serbia, south-western Bulgaria and south Bulgaria and is typically observed in complements of the verbs 'can' and 'want'. The (optional) Comp_drop has also been attested in the de-volitive analytic future tense construction, not only in Torlakian dialects, but in other Balkan languages (for an overview, see Mirić forthcoming).

In our study, we exploit a wider choice of verbs selecting complement clauses in which the complementizer can be omitted. These encompass modal verbs *moći* 'can' and *morati* 'must', the aspectual verb *početi* 'begin' and light verbs *ići* 'go' and *uzeti* 'take'. The respective examples are given in (8)-(12), (a) with and (b) without a complementizer.

(8)

a.	Ne	móg	da	ódim	bolí
	NEG	can	COMP	walk.1SG.PRS	hurt.3SG.PRS

'I cannot walk, (it) hurts.' (Vuković 2020)

b. Jóš se já ne mógu oporávim
Still REFL.CL I.NOM NEG can.1SG.PRS recover.1SG.PRS
'I still cannot recover.' (Vuković 2020)

(9)

a. Móra da pláti stolícutu
must.3SG.PRS COMP pay.3SG.PRS chair.DEF
'He must pay for the chair.' (Vuković 2020)

b. sprémaš ručák móra se jedé
prepare.2SG.PRS lunch must.3SG.PRS REFL.CL eat.3SG.PRS
'You prepare lunch, one must eat.' (Vuković 2020)

(10)

a. Já ódma počé də oséčam bólovi
I.NOM immediately start.1SG.AOR COMP feel.1SG.PRS pains
'I immediately started to feel (the) pain.' (Vuković 2020)

b. pa kakó da ti póčnem príčam
well how COMP you.SG.DAT.CL begin.1SG.PRS tell.1SG.PRS
'Well, how do I start telling/narrating you?' (Vuković 2020)

(11)

a. néče d íde da bére grsnice
not_want.3SG.PRS comp go.3SG.PRS comp pick.3SG.PRS hemp
'He doesn't want to go to pick hemp.' (Vuković 2020)

b. íde čúva stóku
go.3SG.PRS look_after.3SG.PRS livestock
'He goes to look after (the) livestock.' / 'He goes and looks after (the) livestock.' (Vuković 2020)

(12)

a. zatój ság uzél da rabóti
for that reason now take.PPART.M.SG comp work.3SG.PRS
'For that reason, he has now taken to work.' (Vuković 2020)

b. úzneš omésiš kólač
take.2SG.PRS knead.2SG.PRS cake
(Context: What was done for St. George's Day?) 'You take to knead the ritual cake.' / 'You take and knead the ritual cake.' (Vuković 2020)

The position we take in this paper assumes that finite complements lack *da*, rather than the given structures (matrix clause and complement clause) being coordinated and having the same syntactic status. Besides the fact that the use of structures with a complementizer is more frequent in the corpus (see Section 3.1 for the percentage of Comp_drop), the choice of the selecting verbs in the paper is such that in the default case they select for complement clauses (*moći* 'can', *morati* 'must', *početi* 'begin') which are introduced with the Comp *da*, or can be considered pseudo-coordinated structures (see the next section concerning verbs like *ići* 'go' and *uzeti* 'take') introduced with a coordinating conjunction. In both cases, they select for the event-type complements.[2]

1.3 Light verbs *ići* 'go' and *uzeti* 'take'

As illustrated in (11) and (12), light verbs *ići* 'go' and *uzeti* 'take' keep only a component of their semantics and together with a complement verb refer to a single event. In addition to being used with the complementizer *da* or without it, these verbs are also found with the coordinating conjunctions *i*, *te*, *ta*, *pa* 'and' in Torlakian, as in (13) and (14). These examples illustrate the pseudo-coordinating constructions which characterize a restricted class of verbs, usually 'go', 'come', 'take', 'sit' and 'stand', and are found in various languages (see Carden, Pesetsky 1977; Cardinaletti, Giusti 2001; Wiklund 2007; Ross 2016; Di Caro 2017; among others). Although the coordinating conjunctions are sometimes labelled as pseudo-coordinators, we treated them as complementizers for two reasons. Firstly, clauses that follow them somewhat share properties of complement clauses introduced by modal and aspectual verbs. Typically, the lower clause shares the

2 We would like to thank the anonymous reviewer for suggesting the other possibility, that the structure with the complementizer is not necessarily basic with regard to the structure without the complementizer. Although some research on Slavic languages such as Serdobol'skaya (2017) and Letuchiĭ (2023) shows that in Russian, structures without a complementizer sometimes differ in their syntactic properties from structures with a complementizer, their research is based on a different type of complementation, i.e. structures following the verb *dumat'* 'think' (Serdobol'skaya 2017) or *dumat'* 'think' and *ponimat'* 'understand' (Letuchiĭ 2023). This also holds for the complementizers introducing the proposition-type complements, which have different syntactic properties.

same subject as the matrix clause, as well as the person features of the matrix verb,[3] and the mood and tense of the verb in the lower clause are controlled by the light verb. Since the lower verb cannot have its own tense specification, it is attested either in the tenseless form corresponding to the morphological present tense or marked for the same tense or mood as the light verb.[4] Additionally, the light verb and the lower verb cannot be inverted, unlike in plain coordinating constructions. Secondly, in the case of the Comp_drop, as illustrated in (11b) and (12b), it is not clear whether the complementizer *da* or the coordinating conjunction is omitted, since both options are possible. Cases of pseudo-coordination without the connecting element are found in other languages as well (see, for instance, Di Caro 2017 for Sicilian and Arabic dialects).

(13)

istresémo	slámu	paa	ídemo	pa	opéremo
shake_out.1PL.PRS	straw	and/then	go.1PL.PRS	and/then	wash.1PL.PRS
slámarice					
straw_bedding					

'We shake out the straw and go and wash the straw bedding.' (Vuković 2020)

(14)

já	úznem [pause]	pa	zatvórim	vráta
I.NOM	take.1SG.PRS	and	close.1SG.PRS	door

'I take and close the door.' (Vuković 2020)

1.4 Clitic climbing

CC has been widely discussed in the literature on BCMS (Progovac 1993; Bošković 2004; Stjepanović 2004; Aljović 2005; Ivanović et al. 2023), as a phenomenon in which pronominal clitics as the arguments

3 In the examples in the corpus, like (11) and (12), a change of the subject and number/person features would be ungrammatical and would even lead to a change in the meaning of the sentence; e.g. **ide čuvaš stoku* 'he goes you look after the cattle', **zatoj (on) sag uzel da rabotiš* 'for that reason he has now taken work.2sg'.

4 All these properties apply to the analyzed examples of constructions with *ići* 'go' and *uzeti* 'take' in the corpus. To be more precise, the morphosyntactic locus test shows that in (12a) the verb 'take' in the matrix clause is used in the past tense form (*uzel*), while in (12b) it is used in the present tense (*uzneš*). In both cases, the verbs in the lower clause are used in the same tenseless form corresponding to the morphological present tense (as is also the case with modal verbs), which is controlled by the selecting light verb.

of the embedded verb appear attached to the selecting modal, aspectual, or desiderative matrix verbs. In BCMS varieties which allow infinitival complements, CC is strongly preferred with infinitival complements of modal or phasal verbs, but also marginally possible out of certain types of *da*-complements (Aljović 2005; Ivanović et al. 2023). Wurmbrand and colleagues (2020) show that CC is blocked with complements of the proposition type (see also Wurmbrand, Lohninger 2023), while Mirić (forthcoming) notes that CC is not observed out of the analytic de-volitive future tense construction in the Timok variety, but it is possible from the complements of modal verbs such as *smeti* 'may', *hteti* 'want' and *ne_hteti* 'not want'.

Applying an acceptability judgment survey with speakers of Serbian, Ivanović and colleagues (2023) show that the average acceptability ratings of sentences with CC out of *da*-complements are relatively low.[5] They also emphasize the optionality of this phenomenon and its dependence on the verb type: speakers judged CC out of event-type complements more favourably than out of situation-type complements, while CC out of proposition-type complements was ranked the lowest on average.

In Torlakian dialects, more precisely in the Timok variety, CC is possible, as illustrated in examples (8b) and (10b).

1.5 Aims and research questions

As hinted in section 1.1, our goal is to empirically scrutinize some of the generalizations reported in Arsenijević (2022) about Comp_drop in Torlakian dialects, and more generally shed additional empirical light on this phenomenon. We state the following concrete research questions:[6]

1. Does Comp_drop depend on the contact position, i.e. on whether something intervenes between the selecting and the selected verb (henceforth Contact_position)?
2. Does Comp_drop depend on the type of selecting verb?
3. Does CC depend on Comp_drop?
4. Does CC depend on the Contact_position?
5. Does CC depend on the type of verb?
6. Is there an effect of other grammatical properties on Comp_drop and CC, which may reasonably be expected to affect it?

5 Their analysis is based on the evaluation of sentences containing the pronominal clitic *ga* 'him'.

6 The first two questions related to the Comp_drop are also analyzed on a sample of modal verbs *moći* 'can', *morati* 'must', *smeti* 'be allowed to', *hteti* 'want', *ne_hteti* 'not want' and the future tense construction (Mirić forthcoming).

More generally, our objective is to investigate to what extent Comp_drop and CC can serve as indicators of the structural clause size(s) in Torlakian dialects.

2 Methodology

2.1 Corpus search

The data for the study are taken from the *Spoken Torlak dialect corpus 1.0* (Vuković 2020),[7] which contains 477,696 word tokens. The corpus contains transcripts of audio-recorded semi-spontaneous narratives and it is automatically annotated and lemmatized (Vuković 2021). Currently, the corpus only includes data from the Timok variety. The data were collected in 64 locations within the Timok region in eastern Serbia and were obtained from 167 informants.

We searched the corpus using the Advanced search option – CQL (Corpus Query Language) of the NoSketch Engine and the following regular-expression template:

[lemma="moći"][word="da"]? [tag!="V.*"] {0,3} [tag="V.*"][8]

The following lemmas of the selecting verbs (first position in the template) are included in the search: *moći* 'can', *morati* 'must', *početi* 'begin/start', *ići* 'go' and *uzeti* 'take', followed by the complement clause, optionally introduced with the complementizer *da*.[9] The search included only those complement clauses containing 0 to 3 words (verbs excluded) between the selecting verb and the selected verb.[10]

In addition to the complementizer *da*, the uses of the conjunctions *i*, *pa*, *te*, *ta* were also searched (found only with the verbs *ići* and *uzeti*), and were included in the analysis as alternative complementizers,

7 The corpus is available online through the CLARIN network: https://www.clarin.si/repository/xmlui/handle/11356/1281, and for search through the open-source NoSketch Engine platform: https://www.clarin.si/ske/#dashboard?corpname=torlak.

8 In the regular expression, the question mark in [word="da"]? signals the optional use of the complementizer *da*, while the exclamation mark in [tag!="V.*"] indicates the exclusion of the word class that follows the tag. In the expression, V stands for verbs.

9 The complementizer *da* appeared in several forms (transcribed as *da*, *də*, *d*, *ə*), which were all included in the analysis.

10 Despite the absence of infinitives in Torlakian dialects, the words in the corpus are normalized and lemmatized to facilitate searches based on forms in standard Serbian. Consequently, the infinitive form was employed in the search to capture all occurrences of the specified verbs and is reported as such in the paper.

given that the clause that follows is surface identical to clauses introduced by *da*.

The search was randomized and limited to 500 examples for each lemma. The output of the search provided concordances with the left and right contexts, along with the information about the file and the SPEAKER_ID. Unintelligible sentences, incorrectly annotated items, uses of selecting verbs without complement clauses and epistemic uses of the modal verb *morati* 'must,' were excluded from analysis due to their different structural properties (see Mirić forthcoming, for details).

2.2 Database

Table (2) shows the number of examples analyzed in the study per verb.

Table 2 The number of examples analyzed in the study

VERB	NUMBER OF EXAMPLES
moći 'can'	318
morati 'must'	232
početi 'begin/start'	131
ići 'go'	206
uzeti 'take'	56
SUM	943

The output was manually annotated by both authors for the set of linguistic properties aimed for the analysis, then negotiated and re-examined multiple times. This resulted in a database titled TorlComp (Mirić, Arsenijević 2025).[11] The list of all properties and coding of specific values for each property and a brief justification for the inclusion in analysis where appropriate is provided below, following the order of columns in the database.

1. REFERENCE: unique file code from the corpus;
2. EXAMPLE_ID: a unique number of each example (1–943);
3. SPEAKER_ID: the unique code for each speaker;
4. CONTR_V: selecting verbs (*ići / uzeti / moći / morati / početi*);
5. LEFT: the left context of the concordance;
6. KWIC: the concordance (relevant example containing the selecting verb and the complement clause);

11 The database is available in the OSF repository, on the following link: `https://osf.io/fsduj`.

7. RIGHT: the right context of the concordance;
8. NEGATION: 0 – selecting verb in the affirmative form, 1 – selecting verb in the negative form; negation was included because it is a clitic, and for the possibility that it requires a propositional complement structure;
9. UNINFL: 0 – inflected selecting verb, 1 – uninflected selecting verb; inflection of the selecting verb was included because the origin and effect of finiteness in syntax make one of the central questions in the literature on the topic;
10. TRUNCATED_CTR: 0 – non-truncated form of the verb *moći* 'can', 1 – truncated form of the verb *moći* 'can'; empty cells stand for all other selecting verbs which are never found in the truncated form; truncation removes the marking of certain features, such as finiteness;
11. PRESENT_CTR: 0 – selecting verb in tenses other than the present tense, 1 – selecting verb in the present tense; as Comp_drop is only available with the selected verb in the present, interaction was possible with the present tense on the selecting verb, so other tenses were merged in a single category;
12. PERSON_CTR: 1 – the first person, 2 – the second person, 3 – the third person of the selecting verb, PPART – the past participle form (if the verb was used in the complex tense form, e.g. the perfect tense); with number, person constitutes finiteness marking;
13. PLURAL_CTR: 0 – the singular form of the selecting verb, 1 – the plural form of the selecting verb; with person, number constitutes finiteness marking;
14. SAME_SUBJECT_CMPL: 0 – different subjects in the matrix and complement clause, 1 – the same subject in the matrix and complement clause; Comp_drop and CC are generalized to only be possible when the subject is the same;
15. VERB_CMPL_DT: 0 – complement verbs beginning with any other sound except for [d] or [t], 1 – complement verbs beginning with consonants [d] or [t]; complementizer *da* is sometimes reduced as /d/, so the possibility that it contracts with the following verb needed to be controlled;
16. CLITICS: 0 – the absence of clitics in the expression, 1 – the presence of clitics in the expression;[12]

12 The following clitics were taken into account: argumental genitive, dative and accusative pronominal clitics, the reflexive clitic *se*, and the evaluative dative reflexive clitic *si*, the latter being characteristic of Torlakian (Arsenijević 2013).

17. IMPERSONAL: 0 – not an impersonal expression, 1 – an impersonal expression; included for the association with finiteness and for the possible interaction with CC;
18. CONTACT_POSITION: 0 – non-contact position, other words intervene between the selecting and selected verb, 1 – immediate contact position between the selecting and selected verb (complementizers, conjunctions and clitics are not treated as intervening words); if Comp_drop takes place in the smallest structure, a bare VP, there is expected to be limited space for other material higher than the verb;
19. CC_POS: 1 – realized CC, 0 – possible CC (i.e. there are clitics, but they do not climb); empty cells stand for the examples without clitics;
20. DA_DROP: 0 – complementizer *da* used, 1 – complementizer *da* omitted; empty cells stand for the usage of the coordinating conjunctions;
21. DA_DROP_COMMENT: the conjunctions *ta* and *te* are marked; *ta* and *te* occur as phonological variants of *da*, and are marked to make sure they do not have a special behavior.

3 Results

There were 943 relevant examples, involving the selecting verbs in focus with a clausal complement, retrieved from the corpus.

Predicted variables were two: whether the complementizer is dropped (DA_DROP: yes (1), no (0)) and whether the clitics expressing arguments of the selected verb surface to the left of the selecting verb (CC_POS: yes (1), no (0)). There were 161 examples without a complementizer between the two verbs, and 35 examples exhibiting CC.

Predictor variables. The following predictor variables were considered: CONTR_V, NEGATION, TRUNCATED_CTR, PRESENT_CTR, PERSON_CTR, PLURAL_CTR, SAME_SUBJECT_CMPL, VERB_CMPL_DT, CLITICS, IMPERSONAL, CONTACT_POSITION.

We tested each predictor variable for an effect on the two predicted variables, with special attention on the type of the selecting verb and the contact position. All statistical analyses were performed in R.

3.1 Complementizer drop

Out of 887 analyzed examples (those where the complementizer was either *da* or absent), 161 examples (18.15%) exhibited DA_DROP, which suggests that this phenomenon is optional in the Timok variety.

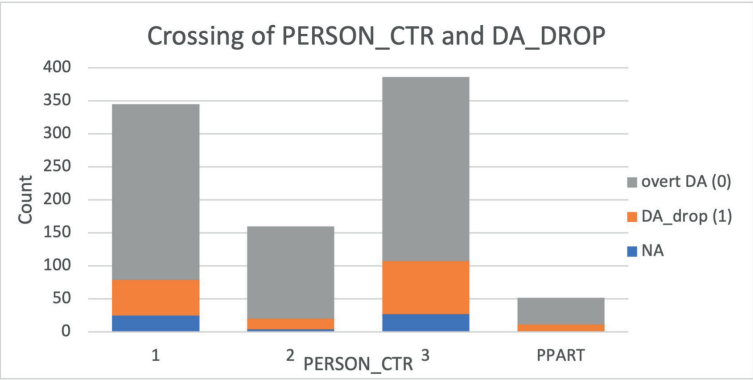
Only four variables showed a significant effect on DA_DROP: PERSON_CTR, CC_POS, CLITICS, and IMPERSONAL.[13]

Table 3 and Chart 1 summarize the results for the association between the complementizer drop and person-feature value on the selecting verb.

Table 3 Complementizer drop per person value of the selecting verb

	overt complementizer	dropped complementizer	% drop
1st person	266	54	16.88
2nd person	140	16	10.26
3rd person	279	80	22.28
participle	41	11	21.15

Chart 1 Complementizer drop per person value of the selecting verb (NA stands for the cases with complementizers other than da, which do not drop)



The Pearson's chi-squared test was conducted to examine the association between the variable representing the occurrence of complementizer drop (DA_DROP) and the variable indicating the person of the verb (PERSON_CTR). The contingency table constructed from the

13 Table A in the Appendix presents the results for the variables that did not significantly affect Comp_drop. The Appendix is available in the OSF repository, on the following link: https://osf.io/2d58b.

dataset TorlComp revealed a statistically significant association between these variables (χ^2 = 9.938, df = 3, p-value = 0.0191).

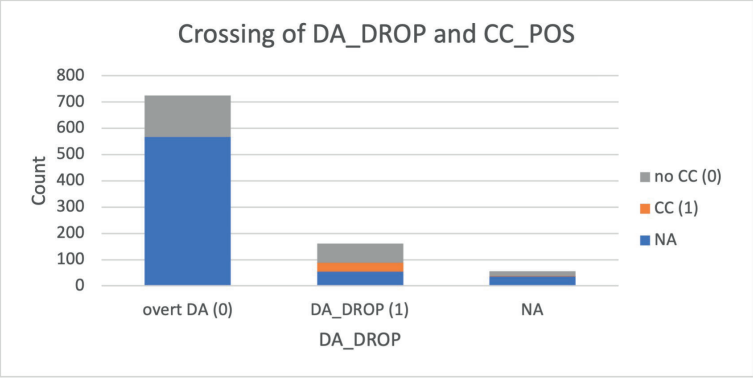
Upon examining the frequencies in the contingency table, it was observed that the occurrence of Comp_drop varied across different verb forms in a way that can be captured in terms of the presence and absence of person. First and second person forms, which are considered to correspond to the presence of person features ([speaker], [hearer]) showed a lower likelihood of Comp_drop than third person forms and the participle, which lacks person features.

Table 4 and Chart 2 summarize the results for the association between the variables CC_POS and DA_DROP.

Table 4 Clitic climbing depending on complementizer drop

	clitics in situ	**clitics climbed**	**% CC**
complementizer overt	158	0	0
complementizer dropped	73	33	31.13

Chart 2 Clitic climbing depending on complementizer drop (NA in the condition CC_POS stands for the cases in which there are no clitics in the expression)



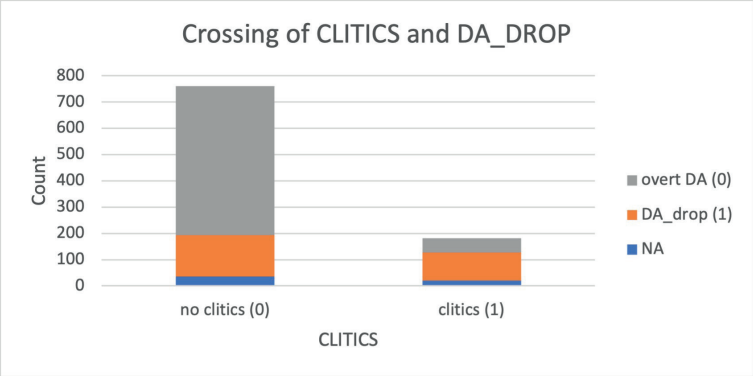
The Pearson's chi-squared test was conducted to examine the association between CC (0 for no climbing, 1 for climbing) and Comp_drop (1 for drop, 0 for non-drop) in the dataset. Strikingly, there were no instances of CC where the complementizer did not drop. The chi-squared test (χ^2 = 53.405, df = 1, p-value = 2.714e-13) indicates a significant association between CC and Comp_drop, suggesting that these two phenomena are not independent.

Table 5 and Chart 3 summarize the results for the association between the variables DA_DROP and CLITICS (which is coded with a positive value if either verb comes with clitics).

Table 5 Complementizer drop depending on the presence of clitics

	overt complementizer	dropped complementizer	% drop
no clitics	567	55	8.84
clitics present	159	106	40

Chart 3 Complementizer drop depending on the presence of clitics (NA stands for the cases with complementizers other than da)



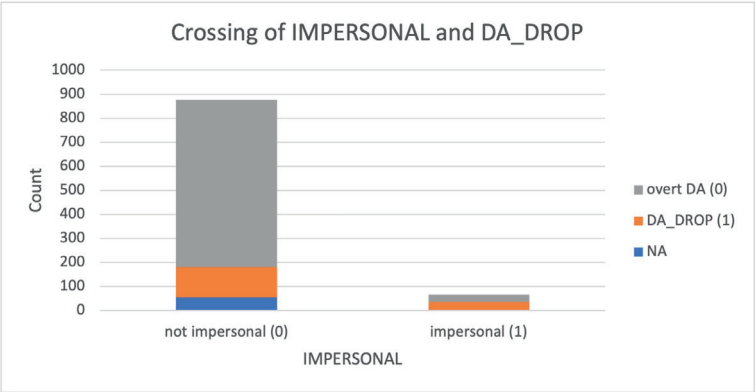
The Pearson's Chi-squared test with Yates' continuity correction was conducted to examine the association between the variables DA_DROP and CLITICS in the dataset. The test yielded a significant result (χ^2 = 119.34, df = 1, p-value = 2.2e-16), indicating a strong association between the variables. Chart 3 further illustrates the relationship between the variables CLITICS and DA_DROP, indicating that the presence of clitics facilitates complementizer drop.

Table 6 and Chart 4 summarize the results for the association between the variables DA_DROP and IMPERSONAL (which is coded with a positive value if the construction is impersonal).

Table 6 Complementizer drop depending on whether the expression is impersonal

	overt complementizer	dropped complementizer	% drop
not impersonal	696	127	15.43
impersonal	30	34	53.12

Chart 4 Complementizer drop depending on whether the expression is impersonal (NA stands for the cases with complementizers other than da)



The chi-square test was conducted to examine the association between the variables DA_DROP (presence or absence of complementizer drop) and IMPERSONAL (impersonal status of the verbal expression). A contingency table was constructed based on the frequencies of the different combinations of these variables.

The results of the chi-square test indicated a statistically significant association between the variables (χ^2 = 54.282, df = 1, p = 1.737e-13), suggesting that the presence or absence of complementizer drop is related to whether or not the verbal expression is impersonal. These findings indicate that the variables DA_DROP and IMPERSONAL are not independent, as the complementizer drop occurs more frequently with the impersonal constructions (see Table 6). However, the variable IMPERSONAL was not independent from the variable CLITICS: all impersonal constructions are marked by the reflexive clitic *se*. Therefore, we have also calculated the effect size for the chi-square test expressed in terms of Cramer's V. Its value was 0.3668 for the variable CLITICS and 0.2473 for the variable IMPERSONAL. Hence, it is likely that there is no independent effect of the variable IMPERSONAL, i.e. that the calculated effect is due to its consistent involvement of a clitic.

3.2 Clitic climbing

Out of 285 analyzed examples in which clitics were available and hence CC was possible, 35 examples (12.28%) exhibited CC, which suggests that this phenomenon is possible, although not dominant in the Timok variety.

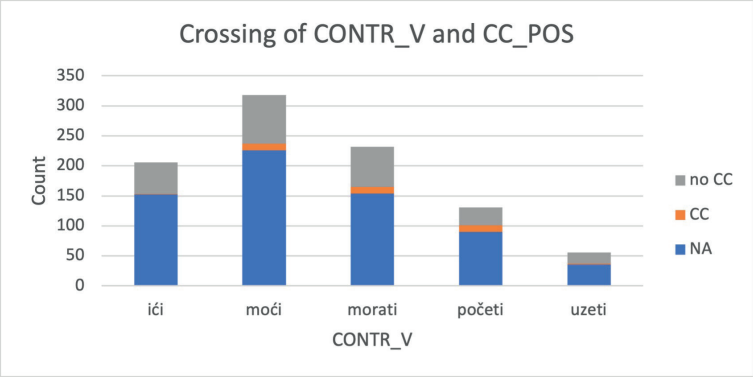
Only the verb type (factor CONTR_V), in addition to the already reported factor complementizer drop (DA_DROP) showed a significant association with CC.[14]

Table 7 and Chart 5 summarize the results for the crossing of the variables CC_POS and CONTR_V.

Table 7 Clitic climbing depending on the selecting verb

	clitics in situ	clitics climbed	% CC
ići 'go'	53	1	1.85
moći 'can'	81	11	11.96
morati 'must'	67	11	14.10
početi 'begin'	30	11	26.83
uzeti 'take'	19	1	5

Chart 5 Clitic climbing depending on the selecting verb (NA stands for the examples without clitics, where clitic climbing is not possible)



The Pearson's Chi-squared test was conducted to assess the association between the likelihood of clitics climbing and the selecting verb (CONTR_V). The contingency table was analyzed, and the test yielded a significant result ($\chi^2 = 14.741$, df = 4, p = 0.005269). This result indicates a statistically significant association between the likelihood of CC and the selecting verb (CONTR_V). However, caution is necessary in interpreting the results, as the chi-square approximation may be inaccurate due to low expected cell counts.

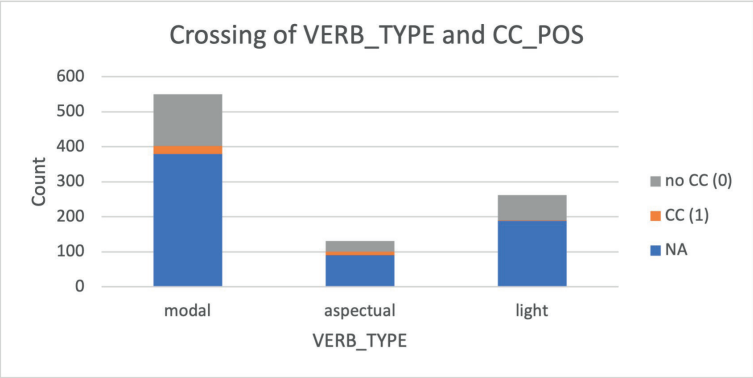
14 Table B in the Appendix presents the results for the variables that did not significantly affect CC. The Appendix is available in the OSF repository, on the following link: https://osf.io/2d58b.

Upon examination of the contingency table, it is observed that the aspectual verb *početi* 'begin' induces a higher rate of CC than modal verbs *moći* 'can' and *morati* 'must', which in turn license more CC than the two light verbs *ići* 'go' and *uzeti* 'take', which only marginally allow for CC. Indeed, when the chi-square test is applied to this grouping (a new variable VERB_TYPE with three levels, *modal*, encompassing moći and morati, *light*, encompassing ići and uzeti, and *aspectual* for početi), a stronger effect is attested: $\chi^2 = 14.426$, df = 2, p = 0.0007368 (see also the absolute values in Table 8 and Chart 6).

Table 8 Clitic climbing depending on the selecting verb type

	clitics in situ	clitics climbed	% CC
Aspectual	30	11	26.83
Modal	148	22	12.94
Light	72	2	2.70

Chart 6 Clitic climbing depending on the selecting verb type



4 Discussion

Four variables significantly affected complementizer drop: PERSON_CTR, CC_POS, CLITICS, and IMPERSONAL. In the variable PERSON_CTR, real, marked persons (i.e. those corresponding to the interlocutors: first for the speaker and second for the hearer) license less Comp_drop than the unmarked ones (third person as non-interlocutor and participles as forms that cannot bear person marking). The effect was stronger with the second person than with the first. This effect confirms Stanković's (2022) observation that matrix verbs with marked endings (including 2SG with the ending *-š* and 1PL and 2PL with endings *-mo* and *-te*) do not license Comp_drop as easily as those without such endings. To test this explanation, we

fitted a mixed linear regression model with the interaction to the factors PERSON_CTR and PLURAL_CTR. The expected interaction for the level 1st person of the former is not attested (function: glm(DA_DROP ~ PERSON_CTR * PLURAL_CTR, data = TorlComp), estimate = 0.038, Standard Error = 0.06769, t = 0.562).

The association of the Comp_drop with CC can have a twofold interpretation. One option is to observe CC as a facilitating factor for Comp_drop, and the alternative is the other way around. An informative observation is that zero cases are found where the complementizer is overt, yet the clitics climb. This confirms Arsenijević's (2022) generalization that in Torlakian CC only occurs with complementizer drop. The pattern where the complementizer is not pronounced while clitics remain in the complement clause, which are attested in the corpus, is reported by Arsenijević too, and analyzed as a phonological reduction of the complementizer to a stop pronounced left adjacent to the clitic cluster. The view that these cases do not involve complementizer drop proper needs to be empirically tested, as it predicts a phonetic contrast between the pronunciation of the clitic cluster in that context, and in the other two contexts: with an overt complementizer and having undergone climbing. Together, the results of our experiment are in line with the empirical generalizations Arsenijević makes, and thus also with his account arguing that the structure selected by the matrix verb is too small to host clitics. Otherwise, one would expect cases in which unclimbed clitics surface after the complement verb in the present tense as in (15), an option that is ungrammatical not only in Torlakian, but to our knowledge in all BCMS dialects.

(15)

*...	kad	počne	(da)	dere	ti	se…
	when	begins	(COMP)	scream.3SG.PRS	you.DAT.CL	refl.cl
	'When s/he starts screaming at you...'					

That the ungrammaticality of (15) has to do with the complementizer is obvious from the fact that infinitival event and situation-type complement clauses can host clitics after the verb, as in (16).

(16)

Sve	što	je	želeo	bilo	je
all	comp	aux	wished	been	aux
javiti	joj	se	u	snovima.	
contact	her.DAT.CL	refl.cl	in	dreams	
'All that he wanted was to talk to her in her dreams.'					

Since the complementizer can drop in the absence of clitics, but clitics cannot climb in the presence of the complementizer, we found it more appropriate to consider Comp_drop the effector, and CC its effect. Hence, we can conclude that Comp_drop is a necessary condition for CC and possibly also its trigger.

The facilitating effect of the availability of clitics realizing arguments of the complement verb on Comp_drop likely has to do with the association between Comp_drop and CC. One possible explanation, rather functional, is that since in Torlakian Comp_drop is a necessary condition for CC, climbed clitics encode the underlying presence of the complementizer. Speakers thus more readily drop the complementizer from expressions that involve clitics (see Table 5 above), as in those cases it is recoverable from the climbed clitics. A similar, rather formal, explanation is that *da* realizes the clause boundary. When an overt boundary is lacking, clitics continue moving higher in the structure in search of a proper host, i.e. a visible clause boundary. Given that CC is available only with DA-drop, it seems that even the modal *da*-complements with an overt *da* block CC. Since our observations are corpus-based, they require further testing of (un)grammaticality.

Finally, as already indicated, the facilitating effect of impersonality on Comp_drop is likely not an independent effect, but comes from the fact that impersonals are marked by the reflexive clitic: all impersonal constructions involve a clitic (as opposed to personal expressions, which may or may not involve them). As clitics facilitate Comp_drop, this effect then confounds the variable of impersonality. Impersonal constructions in BCMS have received different interpretations in the literature. Among others, Stjepanović (2004) and Aljović (2005) provide different perspectives on the optionality and obligatoriness of CC, respectively, which calls for additional syntactic analysis that goes beyond the scope of this paper.

Only one variable (in addition to Comp_drop, discussed above) had a significant effect on CC: the verb type. Aspectual and modal verbs are more likely to come with CC than light verbs like *ići* 'go' or *uzeti* 'take'. In the entire dataset, only one example for each of the two light verbs is attested, (17) for *ići* 'go' and (18) for *uzeti* 'take', where CC is exhibited, *ju* in (17) and *ga* in (18).

(17)

… i	sád	ako	čúje	da	je	néšto	godí
and	now	if	hear.3SG.PRS	comp	her.CL	something	suit.3SG.PRS
onó	ju	íde	i	gúra		gúra	gúra…
it	her.CL	go. 3SG.PRS	and	push.3SG.PRS		push.3SG.PRS	push.3SG.PRS

'… and now, if it hears that something suits her, it goes and pushes her on and on…' (Vuković 2020)

(18)

…ili	ako	dóǵe	iz	drúgo	seló
or	if	come.3SG.PRS	from	other	village
momák	ovíja	ga	úznu	pa	júre
guy	these	him.CL	take.3PL.PRS	and	chase.3PL.PRS

'… or if a guy comes from another village, these (people) go and chase him…'
(Vuković 2020)

The results suggest that CC with light verbs is impossible or close to impossible, which can be associated with the long tradition of analyzing structures with light verbs as different (often modelled in terms of coordination, note the conjunction *i* 'and' in the example (17) and *pa* 'and' in (18), as well as *pa* 'and' in (13) and (14), but see e.g. Wiklund 2007) from those involving raising or control, i.e. modal, aspectual and *try*-verbs.

Overall, the availability of Comp_drop and CC out of complement clauses has implications for the structural size of the complement clauses, in addition to the fact that selected verbs show a high degree of dependency on the selecting modal and aspectual verbs in terms of subject and tense restrictions.[15] These observations are in line with Arsenijević (2022) who proposes a clause size smaller than the smallest complement in standard BCMS and Aljović (2005) who observes that CC is available from functionally poor domains, i.e. from restructuring complements (lexical domains).

A note is due regarding our hypothesis that nothing can occur between the matrix and the complement verb when the complementizer is dropped and clitics climb. As the aforementioned examples (5), (8b), and (10b) show, Comp_drop and consequential CC typically result in immediate contact between the selecting and selected verbs. However, we also observe several examples in which other arguments do intervene between the selecting and selected verbs, as in (19), where the subject *moj unuk* 'my grandson' is overt and pronounced between the verbs. This is also the case with the example such as *ne mog ja čekam* 'I cannot wait', with an overt subject. In both cases, we observe the postverbal position of the overt subject, possibly signalling informational focus. In addition, we also find the sentences in which the adverbs occur between the verbs, e.g. *ja si poče odma*

15 As for the aspect restrictions suggested by Arsenijević (2022), some of the tested matrix verbs, such as *početi* 'begin', allow complement clauses in the imperfective aspect only, as illustrated in (10a) and (10b) above. However, modal and light verbs allow for both imperfective and perfective aspects in the embedded clause, cf. (8a) with the embedded verb *odim*.impf 'walk' and (8b) *oporavim*.pf 'recover'. See also Todorović, Wurmbrand (2020) for similar observations.

pričam 'I started talking immediately' (note, however, that the clitic *si* climbed), as well as the examples with the intervening clitics, as in *ne mog se probudim* 'I cannot wake'. This suggests that other factors may result in the presence of intervening elements, and have to be taken into account in future analyses. Nevertheless, there is still a tendency for immediate contact, which is supported by the fact that in 161 of examples of the Comp_drop, in 133 examples we find immediate contact between the matrix and the complement verb.

(19)

sa	zórana	se	móra	mój	unúk	drúži
with	Zoran	refl.cl	must.3SG.PRS	my	grandson	hang_out.3SG.PRS

'My grandson must hang out / be friends with Zoran.'

5 Conclusion

We have reported quantitative corpus-based research into the phenomena of Comp_drop and CC in the Timok variety of the Torlakian dialect group, and discussed the results in the context of the syntax of clausal complements. From a set of 11 variables included in the annotation and analysis, four have been attested to show significant effects on complementizer drop, and only one on CC. In the former set, of the four variables, which included the person value, the availability of clitics in the structure, whether or not CC takes place, and whether or not the expression is impersonal, only the first two have been argued to have an independent effect, while the effects of the other two derive from the presence of clitics. In turn, CC does not occur unless the complementizer is dropped and it also depends on the verb type. The findings are in line with analyses and models in Arsenijević (2022) and Stanković (2022), as well as with the traditional distinction between restructuring verbs on the one hand, and pseudo-coordinating ones on the other (Cinque 2006; Wiklund 2007). Expressions exhibiting Comp_drop are structurally very restricted: very little material, of very restricted types can occur between the selecting and the selected verb. Marked values of person, the first and second, inhibit the drop of the complementizer, and pseudo-coordination imposes different restrictions on the verbs involved than regular complementation, even though both have been argued to involve very little structure. The empirical picture reported is compatible with the view that Comp_drop takes place only in the smallest structures, likely a bare VP, while other types of complements, with additional options, correspond to other larger structural sizes.

Acknowledgements

The article is written within the JESH Programme (the Austrian Academy of Sciences' Joint Excellence in Science and Humanities).

References

Aljović, N. (2005). "On Clitic Climbing in Bosnian/Croatian/Serbian". Leko, N. (ed.), *Lingvistički vidici*, 34. Sarajevo: Forum Bosnae, 58-84.

Arsenijević, B. (2009). "Clausal Complementation as Relativization". *Lingua*, 119, 39-50. https://doi.org/10.1016/j.lingua.2008.08.003.

Arsenijević, B. (2013). "Evaluative dative Reflexive in Southeast Serbo-Croatian". Fernandez, B; Etxtepare, R. (eds), *Variations in Dative: A Microcomparative Perspective*. Oxford: Oxford University Press, 11-21. https://doi.org/10.1093/acprof:oso/9780199937363.003.0001.

Arsenijević, B. (2022). "Clitic Climbing, Complementizer Drop, and the Size of Verbal Complements in Torlakian Serbo-croatian". *International Conference English Language and Anglophone Literatures Today 6* (29-30 October 2022, University of Novi Sad).

Arsenijević, B. (2023). "The Size of Complements and the Height of Finiteness: The Torlakian Serbo-croatian Complementizer Less Construction". *Workshop on Finiteness, Clause Types, and Cartography* (30 June 2023, University of Graz).

Bošković, Ž. (2004). "Clitic Placement in South Slavic". *Journal of Slavic Linguistics*, 12, 37-90.

Carden, G.; Pesetsky, D. (1977). "Double-verb Constructions, Markedness, and a Fake Coordination". Beach, W.A.; Fox, S.E. (eds), *Papers from the Thirteenth Regional Meeting of the Chicago Linguistic Society*. Chicago, IL: Chicago Linguistics Society, 82-92.

Cardinaletti, A.; Giusti, G. (2001). "'Semi-lexical' Motion Verbs in Romance and Germanic". Corver, N; van Riemsdijk, H. (eds), *Semi-Lexical Categories*. Berlin: de Gruyter, 371-414. https://doi.org/10.1515/9783110874006.371.

Cinque, G. (2006). *Restructuring and Functional Heads: The Cartography of Syntactic Structures, 4*. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780195179545.001.0001.

Di Caro, V.N. (2017). "Multiple Agreement Constructions: A Macro-comparative Analysis of Pseudo-coordination with the Motion Verb Go in the Arabic and Sicilian Dialects". *Bucharest Working Papers in Linguistics*, 19(2), 71-92.

Ivanović, S.; Kovačević, P.; Milićević, N. (2023). "Clitic Climbing with Different Kinds of Da-complements in Serbian and the Status of Probabilistic Rules in Grammar". *Annual Review of the Faculty of Philosophy in Novi Sad*, (48)3, 1-21.

Kovačević, P; Milićev, T. (2018). "The Nature(s) of Syntactic Variation: Evidence from the Serbian/croatian Dialect Continuum". Lenertová, D.; Meyer, R.; Šimík, R.; Szucsich, L. (eds), *Advances in formal Slavic linguistics 2016*. Berlin: Language Science Press, 147-67.

Letuchiĭ, A. (2023). "Podchinenie v nestandartnom kontekste: nestandartnye podchinitel'nye konstruktsii i ikh sintaksicheskie svoistva". *Russkii iazyk v nauchnom osveshchenii*, 46(2), 139-79.

Mirić, M. (2018). "Upotreba/izostavljanje subjunktivnog markera da u konstrukciji futura prvog u timočkim govorima". Ćirković, S. (ed.), *Timok. Folkloristička i lingvistička terenska istraživanja 2015–2017*. Beograd: Udruženje folklorista Srbije, Knjaževac: Narodna biblioteka "Njegoš", 201-18.

Mirić, M. (forthcoming). "The Inflection Omission and the Complementizer DA Omission: Evidence from Modals in the Timok Variety of Serbian". FDSL16, University of Graz.

Mirić, M.; Arsenijević, B. (2025). "Torlcomp Database of Complementizers in Torlakian: Verbs ići, uzeti, morati, moći, početi". https://osf.io/fsduj. https://doi.org/10.17605/OSF.IO/6UN2Q.

Mišeska Tomić, O. (2004). "The Syntax of the Balkan Slavic Future Tenses". *Lingua*, 114, 517-42. https://doi.org/10.1016/S0024-3841(03)00071-8.

Progovac, Lj. (1993). "Locality and Subjunctive-like Complements in Serbocroatian". *Journal of Slavic Linguistics*, 1, 116-44.

Progovac, Lj. (2005). *A Syntax of Serbian-Clausal Architecture*. Bloomington: Slavica.

Ross, D. (2016). "Between Coordination and Subordination: Typological, Structural and Diachronic Perspectives on Pseudocoordination". Cardoso, A.; Martins, A. M.; Pereira, S.; Pinto, C.; Pratas, F. (eds), *Coordination and Subordination. Form and Meaning. Selected Papers from CSI Lisbon 2014*. Newcastle upon Tyne: Cambridge Scholars Publishing, 209-43.

Serdobol'skaya, N.V. (2017). "Bessojuznye aktantnye predlozhenija s glagolom dumat' v russkom jazyke". *Voprosy Jazykoznanija*, 5, 7-35.

Sobolev, A.N. (2004). "On the Areal Distribution of Syntactic Properties in the Languages of the Balkans". Mišeska Tomić, O. (ed.), *Balkan Syntax and Semantics*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 49-69. https://doi.org/10.1075/la.67.06sob.

Sobolev, A.N.; Mirić, M.; Konior, D.V.; Ćirković, S. (2023). "Torlak: Areal Embedding and Linguistic Characteristics". *Zeitschrift für Slavische Philologie*, 79(1), 23-46.

Stanković, B. (2022). "Complementizer Omission in Serbian Niš Dialect – a Harmonic Alignment Account". *ELALT 6*. University of Novi Sad, 30 October 2022.

Stjepanović, S. (2004). "Clitic Climbing and Restructuring with 'finite Clause' and Infinitive Complements". *Journal of Slavic Linguistics*, 12(1-2), 173-212.

Todorović, N. (2012). *The indicative and subjunctive da-complements in Serbian: A syntactic-semantic approach*. University of Illinois: PhD thesis.

Todorović, N.; Wurmbrand, S. (2020). "Finiteness Across Domains". Radeva Bork, T.; Kosta, P. (eds), *Current Developments in Slavic Linguistics. Twenty Years After (based on selected papers from FDSL 11)*. Frankfurt am Main: Peter Lang, 47-66.

Vuković, T. (2020). *Spoken Torlak Dialect Corpus 1.0 (transcription)*. Slovenian Language Resource Repository CLARIN.SI, ISSN 2820-4042. http://hdl.handle.net/11356/1281.

Vuković, T. (2021). "Representing Variation in a Spoken Corpus of an Endangered Dialect: The Case of Torlak". *Language Resources & Evaluation*, 55, 731-56. https://doi.org/10.1007/s10579-020-09522-4.

Vuković, T.; Mirić, M.; Escher, A.; Ćirković, S.; Miličević Petrović, M.; Sobolev, A.N.; Sonnenhauser, B. (2023). "Under the Magnifying Glass: Dimensions

of Variation in the Contemporary Timok Variety". *Zeitschrift für Slavische Philologie*, 79(1), 153-94.

Wiklund, A. (2007). "Chapter 6 Pseudocoordinating Verbs Are Light Verbs". *The Syntax of Tenselessness: Tense/Mood/Aspect-agreeing Infinitivals*. Berlin; New York: De Gruyter Mouton, 125-56. https://doi.org/10.1515/9783110197839.125.

Wurmbrand, S. (2001). *Infinitives: Restructuring and Clause Structure*. Berlin; New York: De Gruyter Mouton.

Wurmbrand, S.; Kovač, I.; Lohninger, M.; Pajančič, C.; Todorović, N. (2020). "Finiteness in South Slavic Complement Clauses: Evidence for an Implicational Finiteness Universal". *Linguistica*, 60(1), 119-37. https://doi.org/10.4312/linguistica.60.1.119–137.

Wurmbrand, S.; Lohninger, M. (2023). "An Implicational Universal in Complementation—theoretical Insights and Empirical Progress". Hartmann, J.M.; Wöllstein, A. (eds), *Propositional Arguments in Cross-Linguistic Research: Theoretical and Empirical Issues*. Tübingen: Narr Verlag, 183-229.

e-ISSN 2785-3187

Balcania et Slavia
Vol. 4 — Num. 2 — Dicembre 2024

Лингвистические мутации коронавируса, или об истории означивания новой болезни

Julija Nikolaeva
Università degli Studi di Roma Sapienza, Italy

Abstract The paper focuses on the neologisms that appeared in the Russian language to name «coronavirus». It describes the history of signification of a new disease and a new virus, the origin of the recent pandemic. The study analyses the main naming strategies and describes how medical terminology overstepped the limits of professional sociolect and was assimilated in a common language. Particular attention is paid to the phenomenon of multiple nominations, which developed in an extremely short time, and led to systemic relationships of synonymity and stylistic differentiation, that covered all styles of contemporary Russian literary language from scientific and religious to colloquial.

Keywords Coronavirus. Covid. Neologism. Signification. Russian language. Stylistic differentiation.

Edizioni Ca'Foscari

Peer review

Submitted 2024-10-28
Accepted 2024-11-18
Published 2024-12-18

Open access

Citation Nikolaeva, Julija (2024). "Лингвистические мутации коронавируса, или об истории означивания новой болезни". *Balcania et Slavia*, 4(2), 201-216.

DOI 10.30687/BES/2785-3187/2024/01/004

1 Вводные наблюдения

В условиях поступательного развития языка динамические процессы, инициированные неологизмами, могут затухать и оживляться, растягиваться на несколько десятилетий и не всегда приводить к формированию системных языковых отношений. Однако необычная для социума и человеческого сознания ситуация мировой пандемии и головокружительные ритмы интернет-коммуникации спровоцировали заметное ускорение языковых процессов. В эпоху пандемии слова *коронавирус* и *ковид,* родившиеся в медицинском научном дискурсе, стремительно перешагнули границы социолекта и проникли в общелитературный язык во всех его разновидностях – от юридического дискурса и публицистики до разговорной и сниженно-разговорной речи. Быстрота, с которой произошло освоение этих неологизмов в русской языковой системе, воистину поразительна. За пределами профессионального дискурса медицинские термины подверглись семантическому сдвигу, заметно расширили свою семантику и приобрели общеязыковое нетерминологическое значение. Одновременно данные неологизмы стали отправной точкой регулярных деривационных процессов (Громенко, Павлова, Приемышева 2020; Зеленин, Буцева 2020; Минеева 2020; Геккина 2021; Митурска-Бояновска 2021, 461-3; 466-7), породили развернутые словообразовательные гнезда (Левина 2020; Рацибурская 2021; Приемышева 2021а, 23; 27-8) и развили системные отношения многозначности и синонимии (Приемышева 2021а, 40-2; Приемышева 2021б, 488-91). Всего за несколько месяцев в русском языке сформировались целые синонимические ряды слов, именующих новое заболевание, в которых прослеживается четкая дифференциация по всем стилям современного литературного языка от научного до разговорно-сниженного.

Оглянемся назад, в те времена, когда вирус, спровоцировавший мировую пандемию, был малоизвестным и новая болезнь не имела еще наименования, и попробуем проследить, как в русском языке происходило означивание этого вируса и вызываемого им заболевания. Наш экскурс в «историю болезни» пройдет путь от медицинского дискурса к дискурсу юридическому, публицистике и доберется до сниженно-разговорной речи. Это позволит продемонстрировать, как шел процесс означивания нового, малопонятного и загадочного явления внеязыковой действительности, как происходило вхождение неологизмов в язык и как возникали множественные номинации заболевания. Наши дальнейшие наблюдения основаны на следующих источниках (с января 2020 г. до декабря 2021 года): медицинская профессиональная пресса, законодательные акты, центральные и областные газеты, любительская проза и поэзия, блоги и чаты, соцсети.

2 Коронавирус: от первых фиксаций до моносемичного термина

Слово *коронавирус* вошло в русский язык через английский как медицинский термин со значением '*вирусный возбудитель желудочно-кишечных и респираторных инфекций*' задолго до недавней пандемии. В английском языке слово *coronavirus* родилось буквально под микроскопом вирусологов в ходе лабораторных наблюдений. Сопоставляя штаммы загадочного вируса, выделенные за несколько лет до этого у человека в Солсбери и Чикаго, с известными уже вирусами инфекционного птичьего бронхита и мышиного гепатита, ученые на основе визуальных наблюдений пришли к выводу о том, что все эти вирусы принадлежат к одной группе. Сообщение о данном научном открытии было опубликовано 16 ноября 1968 года в журнале *Nature*:

> Вирионы имеют сферическую форму; наблюдается определенный полиморфизм. Характерным признаком является «бахрома» из округлых или лепестковидных отростков длиной около 200 Ангстрем, непохожих на заостренные отростки, типичные для микровирусов. *Подобная форма, напоминающая солнечную корону из протуберанцев,* наблюдается у вируса мышиного гепатита, а также у ряда штаммов, недавно выделенных у человека, например, B814, 229E и некоторых других. Эти вирусы обладают рядом общих свойств [...] По мнению восьми вирусологов, *эти вирусы входят в одну ранее не выделявшуюся группу. Ученые предлагают назвать ее группой коронавирусов на основании характерной формы, которую вирусы имеют под электронным микроскопом*» (курсив мой – Ю.Н.) (Virology 1968, 650).

Вскоре коронавирусы были действительно выделены в отдельную группу и внесены под этим наименованием в международный таксономический справочник (Wildy 1971), благодаря чему термин быстро проник в медицинский дискурс различных европейских языков как прямое заимствование из английского. Например, в итальянской прессе, слово *coronavirus* впервые появилось в 1970 году и сразу вошло в научные номенклатуры биологии и медицины для обозначения

> большого семейства респираторных вирусов, способных спровоцировать широкий спектр патологий – от обычной простуды до тяжелых заболеваний (Marazzini 2020, Pietrini 2020).

В русском медицинском дискурсе первые фиксации слова, видимо, относятся к 1971 году, когда *Медицинский реферативный*

журнал кратко резюмировал содержание одной англоязычной статьи о вирусологии, а *Вестник Академии медицинских наук СССР* осветил на своих страницах историю выделения новой группы вирусов, внесенной накануне в международные англоязычные каталоги под названием *корона-вирусы*. Характерно, что первые фиксации неологизма появляются в прямых переводах с английского и отличаются графической вариативностью *коронавирус / корона-вирус* (Шеболдов 1971, 17; Закстельская, Шеболдов 1971, 65-8). Вводя новый термин в русский медицинский дискурс, авторы статьи считают целесообразным пояснить его внутреннюю форму, лаконично перефразируя приведенный выше текст из журнала *Nature*:

> при электронноскопии методом негативного контрастирования изображение вириона напоминает солнечную корону. Именно эта особенность вирусов послужила основанием к обозначению их термином “корона-вирусы” (Закстельская, Шеболдов 1971, 65).

Означивание возбудителей различных коронавирусных инфекций у животных и у людей в медицинской терминологии шло рука об руку с уточнением научных представлений об их этиологии. Открытие 1968 года значительно повлияло на процессы номинации, в известной степени обернув время вспять. К этому моменту некоторые вирусы из будущего семейства *Coronaviridae* были описаны учеными как разнородные явления, взаимное родство которых еще не было установлено. Следовательно, их первоначальная номинация не могла мотивироваться сходством с солнечной короной и строилась на иных стержневых признаках, таких, как вызываемые этими вирусами заболевания и нумерация лабораторного смыва, на котором выделялись новые штаммы.

Одна из стратегий номинации – по провоцируемому заболеванию – широко применялась в ветеринарии. Именно такому принципу следовала номинация первого представителя будущего большого семейства коронавирусов, который был изолирован впервые при изучении респираторных заболеваний кур в 1931 году.[1] Он получил тогда в англоязычной научной литературе описательное наименование – *IBV Infectious bronchitis virus* (вирус инфекционного бронхита (Schalk, Hawn 1931), на протяжении почти сорока лет фигурировал под этим названием во

1 Указанные в данном и следующем абзаце даты относятся к английскому языку. В русском языке датировка вхождения этих терминов в медицинский дискурс обычно отличается на несколько лет, заимствование происходит по принципу семантического калькирования.

всех научных номенклатурах и лишь постфактум был включен в научную таксономию как *ACoV - avian coronavirus* (коронавирус птиц) (Wildy 1971). Означивание коронавируса мышей (*MCoV – Murine coronavirus*) также является иллюстрацией того, как по мере уточнения естественнонаучных знаний о природе этого вируса уточнялась техника его номинации. Вплоть до 2011 года вирус не причислялся к семейству *Coronaviridae,* и его наименование носило более обобщенный характер - *вирус гепатита мышей (MHV – Murine hepatitis virus)* (Bailey et al.1949; Cheever et al. 1949; King et al. 2011).

Другая стратегия номинации – по номеру маркировки лабораторного смыва – применялась, когда началось изучение коронавирусов человека. Так, выделенный в 1965 году в Солсбери штамм первого коронавируса человека вошел в историю вирусологии как B814 (Tyrrell, Bynoe 1965). Спустя год ученые Чикагского университета изолировали новый штамм и назвали его 229E. (Hamre, Procknow 1966, 190-3).

Поиски стройной научной классификации на едином основании увенчались тем, что международная медицинская терминология выбрала как сквозной принцип сходство с солнечной короной. Эта семантическая мотивация на сегодня является ведущей в таксономии многочисленных представителей коронавирусов, сложившейся к концу 90-х годов прошлого столетия после уточнений и поправок (Pringle 1996; Francki, Fauquet et al. 1991). Она проходит красной нитью по всей классификационной таблице от наименования семейства (коронавирусы), к наименованию родов (альфакоронавирусы, бетакоронавирусы, дельтакоронавирусы и т. д.) и широко представлена в более низком таксоне многочисленных видов (коронавирус летучих мышей, длиннокрылых, китообразных, птиц, дельтакоронавирус свиней, альфакоронавирус 1-го типа, бетакоронавирус 1-го типа, коронавирус тяжелого острого респираторного синдрома и т. д.). Очевидно, что семантически прозрачная мотивация через *корону* вытесняет различные первоначальные техники номинации и становится главенствующей стратегией означивания для этого семейства вирусов.

К началу XXI века в среде вирусологов сложилось мнение о коронавирусах «как об актуальных ветеринарных патогенах, не представляющих особой опасности для человека» (Щелканов и др. 2020, 231). Жесткая закрепленность за строго очерченным кругом референтов способствовала тому, что в первые десятилетия своего существования термин *коронавирус* был моносемичным и циркулировал в узком профессиональном дискурсе биологов и медиков в значении *'вирусный возбудитель желудочно-кишечных и респираторных инфекций',* лишь спорадически появляясь на страницах научно-популярных изданий.

Ситуация кардинально изменилась, когда в конце 2019 года в Китае была зафиксирована вспышка нового, неизвестного до этого времени тяжелого респираторного заболевания, вызванного загадочным вирусом, смертельно опасным для людей. Хотя эпидемия уже была зарегистрирована в г. Ухань и данные об этом были переданы во Всемирную организацию здравоохранения, вирусологи долгое время не могли опознать ее возбудителя. В последний день 2019 г. ВОЗ публикует на разных языках, в том числе на русском, заявление о том, что

> принимает к сведению заявление для СМИ о случаях «вирусной пневмонии» в г. Ухань, Китайская Народная Республика (ВОЗ 2020)

и подчеркивает, что речь идет о «пневмонии *неизвестной этиологии*». В первые месяцы следующего года эпидемия охватила многие европейские страны, переросла в пандемию и получила широкую огласку в СМИ, вследствие чего моносемичный термин молниеносно перешагнул границы социолекта, из малочастотного превратился в высокочастотное слово, попал на страницы международной прессы и соцсетей и проник в общелитературный язык и в субстандарт и т.д. Динамичные семантические процессы, запущенные пандемией, носили международный характер и протекали параллельно в разных языках. Рассмотрим, как эти процессы развивались в русском языке.

3 Языковая динамика в период пандемии

3.1 Коронавирус: возбудитель новой смертельно опасной болезни у людей

9 января 2020 г. ВОЗ публикует сообщение о том, что «согласно выводам китайских властей, вспышка вызвана *новым коронавирусом*» (ВОЗ 2020). Агентства ТАСС и РИА Новости в тот же день транслируют эту информацию, и ее подхватывают все российские газеты. Таким образом, слово *коронавирус* в значении ‘*возбудитель новой атипичной пневмонии*’ впервые многократно фиксируется в русскоязычной прессе, адресованной широкой читательской аудитории.

В течение месяца механизмы передачи вируса, его этиология и диагностика находятся под пристальным вниманием ученых. Прогресс эпидемиологических исследований снова идет рука об руку с поисками новой номинации во всех европейских языках: мы снова видим, как фокус микроскопа сужается не только

в научных лабораториях, но и в поисках наиболее точной и однозначной номинации нового вируса. На смену весьма размытому наименованию *новый коронавирус* приходят следующие описательные обороты: *коронавирус, выявленный в Ухане*, *китайский коронавирус, новый бетакоронавирус (относящийся к тому же семейству, что и ТОРС-КоВ и БВРС-КоВ)*, *новый коронавирус 2019 г.* Из соображений политкорректности ВОЗ отказывается от номинаций, включающих прямую географическую отсылку к конкретной стране. В середине января эпидемиологи идентифицируют новый штамм, который раньше не выделялся у людей, и называют его 2019-нКоВ, калькируя английскую аббревиатуру 2019-nCoV. Это наименование оказалось неокончательным и не задержалось надолго в русском языке,[2] хотя и оставило свой короткий след в юридическом дискурсе. Наконец 11 февраля 2020 г. на глобальном форуме ВОЗ по вопросам новой коронавирусной инфекции, в котором принимало участие более 300 экспертов из 48 стран, за новым вирусом закрепилось официальное наименование SARS CoV-2 (ВОЗ 2020), сохранившееся и сегодня в русском языке в написании на латинице. SARS CoV-2 является аббревиатурой от английского[3] '*severe acute respiratory syndrome coronavirus*' (тяжелый острый респираторный синдром коронавирус 2) в официальных стилях речи. ТОРС-КоВ, её частичный аналог на кириллице, не получил широкого распространения в силу своей неточности.

В нейтральном стиле общелитературного языка для обозначения этого штамма употребляется гипероним *коронавирус* (Приемышева 2021б, 141). Как явствует из предыдущего повествования, значение этого слова в медицинском дискурсе отсылает нас не к конкретному единичному денотату, а к целому семейству вирусов. Подобная траектория семантического развития отвечает общей логике освоения терминологической лексики в языке и речи. За пределами профессиональных социолектов, укореняясь в общелитературном языке, термины обычно приобретают диффузное значение и трактуются с большой долей приблизительности, что позволяет лингвистам говорить о существовании так называемой наивной бытовой терминологии (Иомдин 2012).

2 Сходная картина наблюдается и в английском: уже в марте 2020 г. лингвисты отмечают, что словосочетание *novel coronavirus* и его аббревиатуры NCoV и 2019-NCoV отодвигаются на периферию, уступая место другим номинациям (Павлова 2020, 129).

3 Как отмечалось в предыдущих исследованиях (Павлова 2020; Янурик 2021), влияние английского языка на другие европейские языки, в том числе на русский, заметно усилилось во время пандемии коронавируса. По некоторым подсчетам, в русском языке доля английских заимствований в этом пласте лексики может доходить до 70 процентов (Редкозубова 2020, 196–197).

3.2 Коронавирус – ковид - корона: об означивании новой болезни

На февральском глобальном форуме ВОЗ рождается научное наименование не только нового вируса, но и вызываемой им болезни. 11 февраля ВОЗ публикует в Твиттере аббревиатуру COVID-19, расшифровывая ее как *Corona Virus Disease 2019*, или «заболевание, вызванное коронавирусом 2019 года». Твит, благодаря своей социальной значимости, мгновенно разлетается по всему миру, и в разных языках одновременно появляется одинаковый неологизм. По подсчетам ресурса Яндекс. Новости, новый термин встречается в этот день в 195 сообщениях русскоязычной прессы. В русском языке этот неологизм быстро проходит этап графического усвоения (Вепрева, Куприна 2021), в ходе которого аббревиатура на латинице закрепляется за официальным стилем письменной речи, а слово *ковид* входит в каждодневный нейтральный узус (Приемышева 2021б, 81-2).

11 марта 2020 г. ВОЗ объявляет, что вспышка COVID-19 может быть охарактеризована как пандемия, и эта новость приводит к взрыву частотности употребления слов *ковид, COVID и коронавирус,* а также к стремительному росту их словообразовательной продуктивности. В марте показатели частотности увеличиваются в сотни раз и на протяжении всего 2020 г. поддерживаются с небольшими колебаниями на очень высоком уровне (Приемышева 2021а, 18-19; Вепрева, Куприна 2021, 96-7). Резкий, почти безудержный рост неологической активности приводит к стихийному развитию полисемии, широкой вариативности и некоторому семантическому сумбуру. На этом фоне в живой стихии речи довольно быстро стирается строгое разграничение значений ‘вирус’ и ‘болезнь’ между двумя разными словами, предложенное ВОЗ в духе государственного дирижизма.[4]

Языковое сознание носителей русского языка с трудом различает значения ‘вирус’ и ‘болезнь’, что выражается в некорректных избыточных словоупотреблениях *коронавирус COVID-19, вирус COVID-19, французский штамм Ковид-19* и т.д. (Приемышева 2021б, 81; Вепрева, Куприна 2021, 93-4). Об объективных трудностях правильного употребления слова *коронавирус* свидетельствует поток вопросов, беспрерывно поступавших в справочную службу популярного информационного портала Грамота.ру с февраля 2020 г. по март 2022 г. Предметом сомнений стало

4 Термин, пришедший из французской лингвистики (Guespin, Marcellesi 1986, 17), где он используется для описания активного вмешательства государства в процессы нормализации языка и языковой номинации. Целью подобного вмешательства является модернизация языка и его защита от излишнего иностранного влияния (Baggioni 1974).

нормативное написание этого слова, характер заимствования и особенности его семантики и сочетаемости (Грамота 2022). Настойчивость, с которой вдумчивые читатели повторяют вопрос о правомерности использования буквы ‘а’ там, где привычнее выглядела бы соединительная гласная ‘о’, свидетельствует о том, как интенсивно происходит усвоение термина в общелитературном языке.

Как демонстрирует *Словарь русского языка коронавирусной эпохи* (Приемышева 2021б, 82-5; 141-2), в живом языковом узусе уже к маю 2021 г., слова *коронавирус* и *ковид* позиционируются как синонимы. В предельно короткий срок ключевые слова момента параллельно переживают зеркальные семантические сдвиги. Разница значений, изначально подчеркнутая в медицинском дискурсе, стирается, и оба слова начинают использоваться как для означивания болезни, так и для означивания вируса. Сейчас, когда пандемия закончилась и бурные неологические процессы в этом семантическом поле прекратились, отношения синонимии по-прежнему сохраняются. Характерно, однако, что язык как тонкий самонастраивающийся механизм отрефлексировал исконную этимологию слов. По данным гугл, на начало октября 2024 г. частотность употребления коллокации *штамм коронавируса*, непротиворечивой с точки зрения семантического согласования, в 13 раз превышает частотность коллокации *штамм ковида*. Стабилизация семантики заметна и по другим признакам: уходят на языковую периферию или совсем исчезают многие окказиональные значения, родившиеся в разгар пандемии. Например, слово *ковид* во множественном числе в значении ‘*о больном или зараженном коронавирусной инфекцией; об умершем от этой инфекции*’, зафиксированное весной 2020 г. как разговорно-сниженное (Приемышева 2021б, 85), к октябрю 2024 г. исчезло из употребления.

Наряду с нейтральными словами *коронавирус* и *ковид*, в период пандемии в русском языке активно используется их разговорный синоним *корона* (Приемышева 2021б, 136), образованный путем усечения и развивший всего за несколько месяцев полисемию со значениями ‘*вирус*’ и ‘*болезнь*’. Хотя возникновение этих значений у слова *корона* в русском языке сопоставимо с параллельными процессами в других языках мира, хронология его первой фиксации в русской прессе в 20-х числах января 2020 г., т. е. до первого появления в иноязычных ресурсах, и его вовлеченность в языковую игру и деривационные процессы позволяют утверждать, что перед нами «независимый от внешнего влияния внутриязыковой процесс […который] апеллирует к национальной культурной базе» (Громенко 2020, 49).

3.3 Номинативная избыточность и стилистическая дифференциация синонимов

Помимо номинаций *коронавирус, ковид и корона*, получивших в период пандемии международное распространение, в русском языке возникает около 130 наименований новой болезни с различной стилистической окраской. Подобная избыточность на номинативном уровне оценивается исследователями как «редкое, если не уникальное [явление] для истории русского *литературного* языка последних десятилетий» (Приемышева 2021а, 40-1).

В рекордно короткие сроки происходит дифференциация многочисленных синонимов по всем стилям современного литературного языка от научного и церковно-религиозного до разговорно-сниженного.

Из научного медицинского дискурса наименование новой болезни быстро перекочевало в дискурс юридический. Это происходит еще до объявления пандемии, сразу же после выявления нового штамма коронавирусов, и связано с необходимостью предотвратить проникновение опасной инфекции из Китая на территорию России. На высшем правительственном уровне принимаются постановления, где болезнь получает первое официальное наименование. 24 января 2020 г. Минюст России публикует за номером 57269 постановление Главного государственного санитарного врача РФ «О дополнительных мероприятиях по недопущению завоза и распространения *новой коронавирусной инфекции, вызванной 2019-nCoV*» (Консультант 2020).

Через неделю это название с небольшим сокращением вносится в официальный перечень заболеваний: постановление от 1 декабря 2004 г. № 715 «Об утверждении перечня социально значимых заболеваний и перечня заболеваний, представляющих опасность для окружающих» (Собрание законодательства Российской Федерации, 2004, № 49, ст. 4916) дополняется пунктом 16 следующего содержания: *коронавирусная инфекция (2019-nCoV*) (Правительство 2020).

Это название новой болезни фигурирует в заглавиях первых правительственных постановлений 2020 г., постепенно уступая место коллокации *новая коронавирусная инфекция (COVID-19).*[5] Авторы законодательных актов, ощущая тяжеловесность и

5 Назовем лишь некоторые документы: 19 марта 2020 г. выходит приказ Минздрава России за № 198н «О временном порядке организации работы медицинских организаций в целях реализации мер по профилактике и снижению рисков распространения *новой коронавирусной инфекции COVID-19*». 8 ноября 2021 г. обнародуют постановление Правительства Ленинградской области за № 709 «О внесении изменений в постановление Правительства Ленинградской области от 13 августа 2020 года № 573 «О мерах по предотвращению распространения *новой коронавирусной*

многосоставный характер этой коллокации, в тексте документов иногда упрощают ее до *новой коронавирусной инфекции* или используют вариант *COVID-19*, совпадающий с официальной версией ВОЗ.[6]

В отличие от юридического социолекта, стремящегося к точности и однозначности в трактовке нового явления, церковно-религиозный дискурс предпочитает иную стратегию номинации – максимально неопределенную. В марте 2020 г. Святейший Патриарх Московский и всея Руси Кирилл утвердил текст новой молитвы, дабы испросить у Господа избавления от новой коронавирусной инфекции (Патриархат 2020). По данным РИА новости, 22 марта она впервые прозвучала на патриаршей службе во время Великого поста в одном храме в Переделкино и потом долго читалась во время литургии во всех храмах Русской Православной Церкви. Текст молитвы был написан на церковно-славянском языке, болезнь в нем называется не прямо, а с помощью эвфемизма. В названии молитвы ковид обозначается как «*вредоносное поветрие*», а непосредственно в тексте – как «*губительное поветрие на ны движимое*». Использованное в эвфемизме слово *поветрие* должно было напоминать, видимо, о том, что болезнь преимущественно передается воздушно-капельным путем. Нечеткость и туманность данной номинации кажется намеренной, ибо она соответствует давней логике табу, исходя из которой прямые указания на опасные явления и предметы могут вызвать к жизни сами эти явления. Показательно, что во всех пресс-релизах, расположенных на официальном сайте Московского Патриархата, употребляется прямое наименование коронавирусной инфекции. Однако в ритуальном тексте, произносимом во время богослужения, избегают называть одиозного врага напрямую, равно как и дьявола, заклятого врага рода человеческого, в церковной традиции предпочитают именовать с помощью эвфемизма.

инфекции (COVID-19) на территории Ленинградской области и признании утратившими силу отдельных постановлений Правительства Ленинградской области». Один из последних значимых документов с данной формулировкой был широко распространен Министерством здравоохранения России 26 октября 2023 г.; это 18 версия Временных методических рекомендаций «Профилактика, диагностика и лечение *новой коронавирусной инфекции (COVID-19)*» и т. д. (Консультант 2024).

6 Приведем несколько документов, содержащих данную номинацию: постановление Главного государственного санитарного врача Российской Федерации от 18.03.2020 N 7 «Об обеспечении режима изоляции в целях предотвращения распространения *COVID-2019*»; постановление Главного государственного санитарного врача Российской Федерации от 16.10.2020 N 31 «О дополнительных мерах по снижению рисков распространения *COVID-19* в период сезонного подъема заболеваемости острыми респираторными вирусными инфекциями и гриппом» (Зарегистрировано в Минюсте России 17.11.2021 N 65867); официальное письмо Минздрава России от 18.06.2024 N 30-4/2753 «О вакцинации против *COVID-19*» и многие другие постановления и законодательные акты (Консультант 2024).

Русский язык использует разные стратегии номинации новой болезни (1) и вируса (2). В нейтральном стиле преобладают коллокации: (1) *ковидная болезнь, коронавирусная болезнь, ковидная инфекция, коронавирусная инфекция, модная болезнь, китайская чума, уханьская чума, чума XXI века, чума 2020 года, новая чума;* (2) *апокалиптический вирус, уханьский вирус, китайский вирус, уханьско-китайский вирус, вирус с короной.*

В разговорном стиле ведущими стратегиями являются аффиксальное словообразование, сложение, а также персонификация. Диапазон участвующих в деривации суффиксов охватывает наряду с регулярными и продуктивными нерегулярные и малопродуктивные единицы. В деривационные процессы одновременно вступают суффиксы различной семантики и стилистики: словообразовательный суффикс -оз, продуктивный в медицинской терминологии для обозначения заболеваний, связанных с тем, что названо мотивирующим словом (Ефремова 2005, 336), используется иронично, так как в научной терминологии он формирует название болезни, а не присоединяется к ее существующему наименованию (*ковидоз* ср. с психоз, тромбоз); -ин-, словообразовательный суффикс, образующий разговорные и просторечные синонимы мотивирующих имен существительных (Ефремова 2005, 197) (*ковидина* ср. со стилистикой слов помидорина, дурачина); очень продуктивный словообразовательный суффикс -к-, с помощью которого в разговорной речи создаются универбаты (Ефремова 2005, 237) (*ковидка, коронка, уханька* ср. с неотложка, вечерка); формообразовательный суффикс -ушк-, особенно продуктивный в разговорной речи и фольклоре (Ефремова 2005, 494) (*ковидушка, ковидушко, коронавирушка, коронушка*); суффиксы -ик с уменьшительным и -ищ- с увеличительным значением (*ковидик* и *коронавирусик, ковидище* и *коронище*); формообразовательные суффиксы -яш-, -яшк- с ласкательным и уменьшительно-ласкательным значением (Ефремова 2005, 77-8) (*короняша, короняшка*) и т. д.

Словообразование через сложение также носит ярко окрашенный разговорный, порой разговорно-сниженный характер и преимущественно исходит из четырех основ – *ковид-, корона-, коронавирус-, вирус* –, лишь изредка прибегая к игровому использованию основы *COVID* на латинице: *ковид-болячка, ковид-зараза/ COVID- зараза/ короназараза, ковид-хамелеон, короназавр, коронанапасть, коронахрень, коронасвинтус, короначума, короназябра, барановирус, зомбовирус, лоховирус, пугаловирус, фуфловирус, хароновирус, COVID-мутант.* Экспрессия подобных словообразовательных моделей усиливается за счет персонификации: *ковид-проказник, ковид-старуха, ковид-враг, корона-гитлер, корона-дьявол, корона-разрушитель, коронавирус-жнец, коронавирус-убийца/ COVID-убийца, коронавирус-злюка, коронавирус-разлучник, коронавирус-пират, коронавирус-поганец* и т. д.

4 Краткое заключение

Неологический бум периода пандемии был обусловлен уникальной внеязыковой ситуацией и активным речетворчеством носителей русского языка. Проведенный анализ свидетельствует о том, что за предельно короткий период медицинские термины были освоены общелитературным языком, стали отправной точкой лавинообразных словообразовательных процессов, развили сложные отношения синонимии и полисемии и стали предметом развернутой стилистической дифференциации. На пике пандемии отмечается множественность и избыточность номинаций, которые часто носят окказиональный характер. Язык задействовал почти весь свой творческий потенциал, чтобы не только удовлетворить объективные потребности в номинации новых явлений, но и дать им разноплановую оценку во всем разнообразии коммуникативных ситуаций.

Мониторинг описанных новообразований в 2024 году свидетельствует о том, что в бóльшей части они оказались сиюминутным, конситуативным и ярко оценочным ответом на события пандемии. Беспрецедентная асимметрия языкового знака наблюдалась недолго и ушла в прошлое. В литературный язык вошла лишь рассмотренная здесь триада *коронавирус – корона – ковид,* ключевые свойства которой были очевидны и ранее благодаря циркулированию в более широком круге текстов и вовлеченности в разнообразные словообразовательные процессы.

Литература

Вепрева, И.Т.; Куприна, Т.В. (2021). «COVID-19: ключевое слово коронавирусного лексикона на этапе графического усвоения в русском языке». Приемышева, М.Н. (отв. ред.), *Русский язык коронавирусной эпохи.* СПб: ИЛИ РАН, 91-100.

ВОЗ, Всемирная организация здравоохранения (2020). https://www.who.int/ru/news-room/detail/29-06-2020-covidtimeline.

Геккина, Е.Н. (2021). «К характеристике деривационных различий лексем ковид и коронавирус (семантико-словообразовательный аспект)». Приемышева, М.Н. (отв. ред.), *Русский язык коронавирусной эпохи.* СПб: ИЛИ РАН, 338-54.

Грамота, Справочно-информационный портал о русском языке (2022). https://gramota.ru/poisk?query=%D0%BA%D0%BE%D1%80%D0%BE%D0%BD%D0%B0%D0%B2%D0%B8%D1%80%D1%83%D1%81&mode=spravka.

Громенко, Е.С. (2020). «Коронный» потенциал русского языка начала 2020 г.». Козловская, Н.В. (отв. ред.), *Новые слова и словари новых слов. 2020.* СПб: ИЛИ РАН, 45-64.

Громенко, Е.С.; Павлова, А.С.; Приемышева, М.Н. (2020). «О “ковидно-коронавирусных” процессах в русском языке 2020 года». *Studia Slavica Hung,* 65(1), 51-70. https://doi.org/10.1556/060.2020.00005.

Ефремова, Т.Ф. (2005). *Толковый словарь словообразовательных единиц русского языка.* М.: АСТ: Астрель.

Закстельская, Л.Я.; Шеболдов А.В. (1971). «Новая группа вирусов – корона-вирусы». *Вестник Академии медицинских наук СССР,* 26(1), 64-8.

Зеленин, А.В.; Буцева, Т.Н. (2020). «От сидидомцев до коронапофигистов (Наименования лиц в период пандемии коронавируса)». *Русский язык в школе*, 81(6), 97-106. https://doi.org/10.30515/0131–6141–2020–81–6–97–106.

Иомдин, Б.Л. (2012). «О “неправильном” использовании терминов: может ли язык ошибаться?». Апресян, Ю.Д. и др. (под ред.), *Смыслы, тексты и другие захватывающие сюжеты.* М.: Языки славянской культуры, 233-51.

Консультант, законодательство РФ кодексы и законы в последней редакции (2020). https://www.consultant.ru/document/cons_doc_LAW_343902/.

Консультант, законодательство РФ кодексы и законы в последней редакции (2024). https://www.consultant.ru/law/hotdocs/t3245/?page=1.

Левина, С.Д. (2020). «Новая болезнь в глагольной лексике коронавирусной эпохи». Козловская, Н.В. (отв. ред.), *Новые слова и словари новых слов. 2020*. СПб: ИЛИ РАН, 157-66.

Минеева, З.И. (2020). «Неодериваты в русском языке эпохи пандемии». Козловская, Н.В. (отв. ред.), *Новые слова и словари новых слов. 2020*. СПб: ИЛИ РАН, 167-77.

Митурска-Бояновска, Й. (2021). «Словообразование в русском языке в эпоху коронавируса». Рацибурская, Л.В. (отв. ред.), *Активные процессы в современном русском языке: национальное и интернациональное.* Москва: Флинта, 459-69.

Павлова, А.С. (2020). «Наименования коронавирусной инфекции COVID-19 в русском, английском и немецком языках: культурно-национальная специфика». Козловская, Н.В. (отв. ред.), *Новые слова и словари новых слов. 2020*. СПб: ИЛИ РАН, 126-38.

Патриархат, официальный сайт Московского Патриархата (2020). http://www.patriarchia.ru/db/text/5610433.html.

Правительство, официальный сайт правительства России (2020). http://government.ru/docs/all/126028/.

Приемышева, М.Н. (2021а). «Ковидный лексикон русского языка: тенденции динамики лексико-семантической системы в период пандемии коронавирусной инфекции COVID-19». Приемышева, М.Н. (отв. ред.), *Русский язык коронавирусной эпохи.* СПб: ИЛИ РАН, 16-51.

Приемышева, М.Н. (отв. ред.) (2021б). *Словарь русского языка коронавирусной эпохи.* СПб: ИЛИ РАН.

Рацибурская, Л.В. (2021). «Новые словообразовательные гнезда с “коронавирусной” лексикой». Приемышева, М.Н. (отв. ред.), *Русский язык коронавирусной эпохи.* СПб: ИЛИ РАН, 355-77.

Редкозубова, Е.А. (2020). «COVID–лексика: этимологический и словообразовательный аспекты (на материале русского, английского и немецкого языков)». *Гуманитарные и социальные науки*, 4, 193-200. http://dx.doi.org/10.18522/2070–1403–2020–81–4–193–200.

Шеболдов, А.В. (1971). «Размножение вирусов в органных культурах фаллопиев труб человека». *Медицинский реферативный журнал. Раздел 3,* 14, 17-18.

Щелканов, М.Ю. и др. (2020). «История изучения и современная классификация коронавирусов (Nidovirales: Coronaviridae)». *Инфекция и иммунитет,* 10(2), 221-46. http://dx.doi.org/10.15789/2220-7619-HOI-1412.

Янурик, С. (2021). «Новые английские заимствования в русском языке коронавирусной эпохи: контаминанты и композиты». Приемышева, М.Н. (отв. ред.), *Русский язык коронавирусной эпохи.* СПб: ИЛИ РАН, 470-94.

Baggioni, D. (1974). «Dirigisme linguistique et néologie». *Langages*, 36, 53-66. https://doi.org/10.3406/lgge.1974.2274.

Bailey, O.T. et al. (1949). «A murine virus (JHM) causing disseminated encephalomyelitis with extensive destruction of myelin: II. Pathology». *Journal of Experimental Medicine,* 90(3), 195-212. https://doi.org/10.1084/jem.90.3.195.

Cheever, F.S. et al. (1949). «A murine virus (JHM) causing disseminated encephalomyelitis with extensive destruction of myelin. I. Isolation and Biological Properties of the Virus». *Journal of Experimental Medicine,* 90(3), 181-94. https://doi.org/10.1084/jem.90.3.181.

Francki, R.I.B.; Fauquet, C.M.; Knudson, D.L.; Brown, F. (eds.) (1991). *Virus Taxonomy. Classification and Nomenclature of Viruses. Fifth Report of the International Committee on Taxonomy of Viruses*. Wien; NY: Springer-Verlag. Archives of Virology. Supplementa.

Guespin, L.; Marcellesi, J.-B. (1986). «Pour la glottopolitique». *Langages*, 83, 5-34. https://doi.org/10.3406/lgge.1986.2493.

Hamre, D.; Procknow, J.J. (1966). «A new virus isolated from the human respiratory tract». *Proceedings of the Society for Experimental Biology and Medicine*, 121 (1), 190-3. https://doi.org/10.3181/00379727-121-30734.

King, A.M.Q. et al. (eds.) (2011). *Virus Taxonomy Classification and Nomenclature of Viruses. Ninth Report of the International Committee on Taxonomy of Viruses*. Amsterdam: Elsevier Academic Press.

Marazzini, C. (2020). *In margine a un'epidemia: risvolti linguistici di un virus.* https://accademiadellacrusca.it/it/contenuti/in-margine-a-unepidemia-risvolti-linguistici-di-un-virus/7895.

Pietrini, D. (2020). *Il mutamento (linguistico) del coronavirus. Parole nel turbine vasto.* https://www.treccani.it/magazine/lingua_italiana/articoli/parole/parole_nel_turbine_1.html.

Pringle, C.R. (1996). «Virus taxonomy 1996 — a bulletin from the Xth International Congress of Virology in Jerusalem». *Archives of Virology*, 141 (11), 2251-56. https://doi.org/10.1007/BF01718231.

Schalk, A.F.; Hawn, M.C. (1931). «An apparently new respiratory disease of baby chicks». *Journal of American Veterinary Medical Association,* 78, 413-22.

Tyrrell, D.A.; Bynoe, M.L. (1965). «Cultivation of a novel type of common-cold virus in organ cultures». *British Medical Journal,* 1 (5448), 1467-70. https://doi.org/10.1136%2Fbmj.1.5448.1467.

«Virology: Coronaviruses» (1968). *Nature,* 220, 650. https://doi.org/10.1038/220650b0.

Wildy, P. (1971). *Classification and Nomenclature of Viruses. First Report of the International Committee on Nomenclature of Viruses.* Basel: Karger.

e-ISSN 2785-3187

Balcania et Slavia
Vol. 4 — Num. 2 — Dicembre 2024

Модели деструктивных субстантивов русского и итальянского языков

Yuliya Rohava
Belarusian State Pedagogical University, Belarus

Abstract A significant part of the Russian and Italian vocabulary is represented by destructive nominations, the activity of which has grown due to the increase of destructive processes in the life of society. Modern journalism reflects a wide range of social relations: political, economic, cultural, sports, everyday life, etc. This article identifies groups of destructive nominations and demonstrates their realization in Russian and Italian publicism. The article also discusses forward "D (destructive lexical-semantic variant)" → "N (non-destructive lexical-semantic variant)" and inverse "N (non-destructive original lexical-semantic variant)" → "D (destructive derivative lexical-semantic variant)" models.

Keywords Destructive. Discourse. Italian. Lexical-semantic variant. Models. Noun. Russian. Vocabulary

Edizioni Ca'Foscari

Peer review

Submitted 2024-02-02
Accepted 2024-12-20
Published 2025-03-25

Open access

Citation Rohava, Yuliya (2024). "Модели деструктивных субстантивов русского и итальянского языков". *Balcania et Slavia*, 4(2), 217-236.

DOI 10.30687/BES/2785-3187/2024/02/005

1 Введение

Антропоцентрическая направленность лингвистической науки XXI века выявила большой интерес к человеческому фактору в языке и способности посредством языковых средств передавать события из окружающей реальности. В ряду таких событий и явлений бытия человека важное место занимает феномен деструктивности, активно исследующийся в современной лингвистической парадигме с разных позиций, различных методологических подходов.

Понятие *деструктивности* получило широкое распространение в философии, психологии, социологии, этике, лингвистике, в научной и публицистической литературе, средствах массовой информации, стало важной частью лексического состава многих языков. В широком смысле деструкция – это разрушение, уничтожение, нарушение обычной структуры предмета, его целостности, «это процесс, направленный в сторону превращения сложной структуры в простую, от многообразия к единообразию, от жизни к смерти, аннигиляции» (Фаткуллина 2010, 58). Наиболее важными маркерами деструкции являются такие акциональные действия и состояния, как ‘повреждение’, ‘нарушение’, ‘отклонение’, ‘расчленение’, ‘разрушение’, ‘смерть’, ‘уничтожение’.

В последнее время появилось большое число исследований, посвящённых речевой агрессии, которая напрямую связана с деструктивностью. В работах Т. А. Воронцовой, Е. В. Кокориной, Г. А. Копниной, А. П. Сковородникова внимание уделяется анализу разных форм речевой агрессии (Воронцова 2006 (а); Воронцова 2006 (б); Кокорина 1996; Копнина 2017; Сковородников 1997). Актуальность изучения феномена вербальной агрессии учёные связывают прежде всего с «неблагополучным социокультурным положением в современных логосферах, ростом асоциальности, общим снижением уровня речевой культуры, инвективизацией и вульгаризацией речи, пропагандой насилия в средствах массовой информации, существенным ослаблением коммуникативных механизмов, традиционно сдерживающих проявления агрессии слова» (Щербинина 2008, 6). По мнению К. Ф. Седова, речевой агрессией считается целенаправленное коммуникативное действие, «ориентированное на то, чтобы вызвать негативное эмоционально-психологическое состояние (страх, фрустрацию и т.п.) у объекта речевого воздействия» (Седов 2003, 200). Я. А. Волкова исследует коммуникативную категорию деструктивности и считает, что эмоции, будучи особой формой отражения и интерпретации действительности, являются пусковым механизмом деструктивности в общении, т. е. его мотивационной основой (Волкова 2014, 14).

Феномен речевой агрессии чаще всего рассматривается в дискурсах СМИ, где представлены такие тактики, как оскорбление, обвинение, угроза, злопожелание, проклятье, призыв, возмущение, колкость (Кулагина 2021, 93; Маевский 2022; Шпомер 2021; Садовников 2022; Шульгина 2021; Юрина 2021).

Е. В. Власова считает, что проявления речевой агрессии неоднородны по мотивации, ситуациям проявления, формам выражения, результатам. В своих исследованиях она рассматривает виды речевой агрессии, выражаемой в форме проклятий, упрёков и насмешек (Власова 2005).

К концепту «угрозы» обращаются К. В. Злоказов и А. Ю. Софронова, считающие, что «угроза рассматривается как осознанное речевое действие автора, нацеленное на деструктивное (манипулятивное) влияние на адресата, сопровождающееся формированием чувства тревоги и страха» (Злоказов, Софронова 2016).

Необходимо отметить, что в последнее время активизировались исследования деструктивности в сети интернет. Помимо таких субстантивов, как медианасилие, медиажестокость, медиаагрессия, группу деструктивности пополнили англицизмы селфицид ‘гибель при попытке сделать экстремальное селфи’, троллинг ‘способ проявления различных форм агрессивного, издевательского и оскорбительного поведения и речевой провокации с целью эскалации коммуникативного конфликта’, абьюзить ‘совершать насилие в любом виде с целью подчинения и подавления воли человека (жертвы абьюза)’, бодишейминг ‘публичное осуждение людей за недостатки их внешности; дискриминация тех, кто не вписывается в общепринятые стандарты красоты’, газлайтинг ‘манипуляция, выражающаяся в том, что агрессор внушает жертве искажённое видение ситуации и представление о себе самой’, сталкинг ‘поиски, подразумевающие преследование; передвижение крадучись’, хейзинг ‘одна из разновидностей психологического и физического насилия, недоброжелательность, проявляемая в виде публичных унизительных оскорблений’, виктимблейминг ‘психологическое разрушение и тирания, перекладывание вины и ответственности за насилие на жертву’, буллинг ‘агрессивное преследование одного из членов коллектива (особенно коллектива школьников и студентов) со стороны остальных членов коллектива или его части’, груминг ‘установление взрослыми дружеских отношений с несовершеннолетними через интернет для вступления с ними в интимную связь, получения фотографий различного характера, запугивания и шантажа’, моббинг ‘грубость, нападение стаей, травля’, ‘форма психологического насилия в виде травли сотрудника в коллективе с целью его последующего увольнения’.

2 Деструктивность как предмет лингвистического исследования

Понятие *деструктивности* практически во всех словарях определяется как нарушение, разрушение нормальной структуры чего-либо, изменение, деформация каких-либо объектов. В «Психологическом словаре» Р. С. Немова указывается, что деструктивные действия могут быть направлены на человека: «характеристика поведения и черта личности, которая проявляется в её стремлении к разрушению того, что создано в обществе другими людьми» (Немов 2007, 111). В «Интерактивном словаре итальянского языка» Гарзанти деструктивность определяется как способность к уничтожению, отсутствие конструктивного вклада (Garzanti 1999). В «Словаре итальянского языка» Н. Зингарелли деструкция – это характеристика того, что разрушительно; это агрессия, направленная на уничтожение чего-либо или кого-либо, которая не осуществляется для защиты себя (Zingarelli 2007). В «Итальянской энциклопедии наук, литературы и искусств» Треккани деструкция понимается как «свойство быть разрушительным, разрушительная сила» (Treccani).

Категория деструктивности привлекала и привлекает внимание многих исследователей. Сам термин *деструктивность* привнесён в лингвистику из сферы философских и естественных наук и впервые использован Ю. В. Фоменко (Фоменко 1975, 23-25).

В связи с тем, что основным и наиболее ярким репрезентатором категории деструктивности в языке является глагол, большинство научных исследований выполнено в русле вербальных номинаций.

Ф. Г. Фаткуллина в диссертации «Категория деструктивности в современном русском языке» выявляет онтологические признаки членения (изменения) действительности, описывает способы выражения концепта *деструктивность* на разных уровнях языка (лексическом, морфологическом, синтаксическом), осуществляет многомерное описание единиц с деструктивной семой. Автор отмечает, что глагол наиболее приспособлен для номинирования деструктивной ситуации и является ядром категории деструктивности, в периферийную часть которого включаются слова других частей речи (Фаткуллина 2002, 6-10, 80).

Э. Х. Суванова предпринимает попытку объединения в лексико-семантические группы разрушения 26 русских глаголов, которые выделяются на основе категориально-грамматической и категориально-лексической сем, составляющих семантическую структуру данных глаголов (Суванова 1985).

На тесную связь глаголов пейоративной семантики со сферами «уничтожение» и «повреждение» обращает внимание Т. А. Потапенко (Потапенко 1978). В одной из своих ранних статей автор

даёт определение деструктивным глаголам, понимая под последними глаголы, обозначающие такое воздействие на объект, в результате которого «нарушается структурная целостность объекта на макро- или микроуровне, и он не может выполнять ранее присущих ему функций» (Потапенко 1983, 50).

В диссертационном исследования Ф. К. Бураихи рассматриваются системные и функциональные характеристики моносемантических и полисемантических деструктивных глаголов русского языка (Бураихи 2011).

И.В. Ивлиев ставит целью диссертации «выявить описание значений лексем со значением ‘лишить жизни’ через создание единообразных толкований таким образом, чтобы каждое толкование включало в себя наряду с общей, объединяющей все эти лексемы частью, те строго индивидуальные элементы значения, которые позволили бы безошибочно отличить её от других, сходных по смыслу лексем» (Ивлиев 1998, 6).

В белорусском языкознании проблемами деструктивной семантики занимается В.Д. Стариченок. В его работах рассматриваются модели глагольной и именной семантической деривации, содержащие в исходной понятийной зоне деструктивную семантику, которая экстраполируется в реципиентную метафорическую область в виде различных аспектов экзистенциальной, социальной, межличностной, эмоционально-психологической, метеорологической, колоративной и многих других сфер (Стариченок 2022 (а)).

Работы некоторых лингвистов ориентированы на установление специфики деструктивных лексических единиц в профессиональном дискурсе. Так, Е. Г. Усик указывает на важную роль деструктивных глаголов (разрушения, разделения, отделения и т.п.), а также коррелятивных отглагольных существительных в языке экономики (Усик 2009, 25). Прагматику глаголов разрушения и созидания в дискурсе медицинской рекламы рассматривает Г.Т. Каримова (Каримова 2011, 212-216).

Таким образом, несмотря на, казалось бы, большое число исследовательских изысканий в области деструктивности, до настоящего времени всесторонне не рассмотрены способы и механизмы формирования вторичных лексико-семантических вариантов именных и глагольных лексических единиц, не выявлены реципиентные семантические сферы и специфика их соотнесённости с исходными деструктивными сферами, не рассматривался вопрос о прямых и обратимых семантических моделях семантической деривации.

3 Деструктивные субстантивные единицы

Понятие деструкции может быть эксплицировано не только как процессуальный временной предикат, но и как вневременной субстантивный предикат, представленный в семантике существительных. Тут денотативная сема ‘деструкция’ сочетается с категориально-грамматической семой ‘предметность’, в результате чего деструктивная ситуация характеризуется опредмеченным маркером (Фаткуллина 2010, 62-64).

Существительные со значением разрушения широко используются в русском и итальянском языках с инвариантным значением ‘нарушение или разрушение нормальной структуры чего-либо’, ‘расчленение, уничтожение, изменение, деформация любых объектов’, ‘изменение структуры чего-либо под действием различных сил (механических, физических и т. д.)’, ‘разрушение объекта под воздействием температуры, химических и микробиологических реакций’ и т.д.

Классификация субстантивов, выражающих различные по семантике деструктивные понятия, осуществляется по следующим семантическим показателям:

а) деструктивные и разрушительные действия, процессы;
б) участники (исполнители) таких процессов;
в) средства (приспособления, устройства), при помощи которых осуществляются деструктивные процессы;
г) результаты разрушительных действий;
д) место осуществления деструктивных действий (Старичёнок 2002 (б), 94-96).

Во всех этих группах наблюдаются процессы семантической деривации, когда основанием для метафорического переноса является первичный деструктивный лексико-семантический вариант, смысл которого трансформируется во вторичной реципиентной сфере. Развитие семантики в этом случае осуществляется по модели «D (деструктивный лексико-семантический вариант)» → «N (недеструктивный лексико-семантический вариант)». Необходимо учитывать тот факт, что при формировании вторичных лексико-семантических вариантов могут актуализироваться различные сходства, аналогии, ассоциативные связи и образные представления, позволяющие закреплять за исходной и итоговой сферами несколько различных связей, «от наименее стабильных творческих метафор до устойчивых «стертых» метафор, фиксированных в культурной традиции общества» (Баранов 2003, 77).

Тематика газетных публикаций, безусловно, зависит от экстралингвистической ситуации и связана с происходящими в обществе процессами (Рогова 2022, 79). Примеры употребления деструктивных номинаций для статьи были выявлены на

страницах Национального корпуса русского языка (НКРЯ), а также издательского дома «Беларусь сегодня» (Издательский дом «Беларусь сегодня») и итальянской газеты «La Repubblica» (La Repubblica), которые являются динамичными, современными газетами, ресурсами. (Рогова 2021, 230). Значения слов рассмотрены при помощи «Словаря русского языка» в 4 т. под редакцией А. П. Евгеньевой (Евгеньева 1999) и «Il dizionario della lingua italiana» Де Мауро (De Mauro 2000).

4 Группы деструктивных субстантивов

В русском и итальянском языках семантика действия по глаголу, процессу, состоянию выражается, как правило, отглагольными существительными, которые репрезентируют ситуацию разрушения «объектно»:

рус. *обвал* ‘падение большой оторвавшейся от чего-либо массы; груда камней, земли, лавина снега, обрушившаяся с гор’. *В американском городе Коламбус, столице штата Огайо, более 30 человек пострадали в результате* ***обвала*** *крыши во время студенческой вечеринки*. (Издательский дом «Беларусь сегодня» 30.04.2023); *В Татарстане из-за* ***обвала*** *грунта на строительстве дороги погиб гражданин Беларуси*. (Издательский дом «Беларусь сегодня» 29.08.2023). Оба примера демонстрируют деструктивный результат (*30 человек пострадали, погиб гражданин Беларуси*), который вызван *обвалом крыши/грунта* (деструктивным процессом). В публицистическом дискурсе существительное развивает новое значение → ‘экономическое ухудшение ситуации’: *В результате* ***обвала*** *экономик и краха промышленности в Союзном государстве не произошло.* (Издательский дом «Беларусь сегодня», 12.05.2023); *Итоги пяти месяцев показывают, что зарплата значительно выросла и не произошло* ***обвала*** *доходов*. (Издательский дом «Беларусь сегодня», 20.06.2023); ***Обвала*** *экономики и краха промышленности не произошло.* (Издательский дом «Беларусь сегодня» 17.08.2023). Неоднократное употребление номинатива *обвал* в сочетании с субстантивом *экономика* демонстрирует яркую реализацию вторичного значения (*обвал*) в публицистическом дискурсе.

Итал. *crollo* ‘caduta improvvisa e rovinosa; forte scossa, scrollone’ – ‘внезапное и разрушительное падение; сильное встряхивание, тряска’ → ‘crisi di stanchezza, cedimento psicologico; di ideali, sentimenti’ – ‘кризис усталости, ухудшение психологического состояния; об идеалах, чувствах’; → ‘disastro economico,

politico; di prezzi, quotazioni e sim.' – 'экономическая, политическая катастрофа (цен, котировок и подбн.)': *Stefano Massini scrive la storia della famiglia della più grande banca d'affari americana, un racconto lungo ben 160 anni, dall'arrivo negli Stati Uniti al* ***crollo*** *finanziario del 2008*. (La Repubblica 14.06.2022). – *Стефано Массини пишет историю семьи крупнейшего американского коммерческого банка, рассказ длиной в 160 лет, от прибытия в США до финансового* ***обвала*** *2008 года*. Несмотря на то, что одно из вторичных значений итальянского субстантива схоже со значением русского существительного, в итальянском словаре зафиксировано бóльшее количество вторичных лексико-семантических вариантов, при котором объём значений сравниваемых слов не совпадает.

Интересная ситуация наблюдается в номинативной паре *кораблекрушение – naufragio*. В русском языке номинатив является моносемантическим и имеет значение 'гибель корабля в море (в бурю, при сильном повреждении и так далее)'. В итальянском языке субстантив *naufragio* в первичном значении относится к крушению корабля и авариям, в результате которых самолёт падает в море, но также развивает и вторичные значения → 'totale fallimento di un'impresa, un'iniziativa' – 'полный провал предприятия, инициативы'; → 'secondo lo psicologo e filosofo K. Jaspers, l'impossibilità dell'essere umano di superare alcune situazioni limite come, ad es., la morte' – 'по мнению психолога и философа К. Ясперса, невозможность человека преодолеть некоторые пограничные ситуации, такие как, например, смерть'.

Одну из групп, в которые могут быть объединены деструктивные номинации русского и итальянского языков, образуют существительные, которые выступают в значении реальных или возможных исполнителей (участников, актантов) действий и ситуаций. В этой группе выделяются имена лиц с сильной уничижительной коннотацией:

рус. *оккупант* 'тот, кто участвует в оккупации': *Деревня Асташево была сожжена* ***оккупантами*** *вместе с жителями*. (Издательский дом «Беларусь сегодня» 12.02.2023);

итал. *occupante* 'chi occupa un luogo, un territorio, un edificio, spec. per motivi politici o sociali' – 'кто занимает место, территорию, здание, особ. по политическим или социальным причинам': *"****L'occupante*** *ha attaccato il centro regionale con dei missili*. (La Repubblica 06. 10. 2022). – ***Оккупант*** *обстрелял райцентр ракетами*. Беспредельность и агрессивность в сторону обычных людей описывают авторы статей, указывая, что жилые места были *сожжены* и *обстреляны*.

В значении ‘тот, кто охотится’ и ‘тот, кто угнетает’, схожи существительные *хищник ‘predatore’* и *угнетатель ‘oppressore’*: *Un* ***predatore*** *che si aggirava in scooter a caccia di donne sole in bicicletta, avvicinandole e toccando loro il seno.* (La Repubblica 18. 08. 2022). – ***Хищник****, который разъезжал на скутере, гонялся за одинокими женщинами на велосипеде, приближаясь к ним и касаясь их груди.* В предложении мужчина назван *хищником*, поскольку он преследует людей и позволяет себе телесный контакт, который, в данном случае, недопустим.

Автор предложения *Целью этой всенародной Отечественной войны против фашистских* ***угнетателей*** *является не только ликвидация опасности, нависшей над нашей страной, но и помощь всем народам Европы, стонущим под игом германского фашизма.* (Издательский дом «Беларусь сегодня» 22. 06. 2022) использует существительное *угнетатель* с целью передачи действий и влияния, которое оказывали фашисты на многие страны.

В «деструктивный» корпус также включается группа названий оружия и орудий, которые непосредственно связаны с деструктивными действиями и процессами. Выделяется несколько подгрупп данной тематической области.

1. Субстантивы со значением оружия и вооружения: существительные *оружие – arma* в русском и итальянском языках развиваются по аналогичной семантической модели: ‘орудие для нападения или защиты’ → ‘средство, способ для достижения, осуществления чего-либо’: *Однако теперь мы все столкнулись с условиями передела мира, где ложь ещё стала и формой психологического* ***оружия***. (Издательский дом «Беларусь сегодня» 26.09.2022); *Кроме того, атакам информационно-психологического* ***оружия*** *подвергается само население страны агрессора с целью получения общественного одобрения совершаемых действий*. (Издательский дом «Беларусь сегодня», 11.05.2023); *Un docente di Storia e filosofia insiste sul vaccino: “Questa è la nostra* ***arma*** *migliore”.* (La Repubblica 13.09.2021). – *Преподаватель истории и философии настаивает на вакцине:* «*Это наше лучшее* ***оружие***». В привычном человеку значении номинатив *оружие* употребляется в боевой, военной сферах, однако и белорусская, и итальянская газеты широко показывают возможность вторичного значения распространяться на психологическую, информационную, медицинскую и другие сферы. Образование некоторых вторичных лексико-семантических вариантов основывается на признаке скорости, стремительности. Похожий процесс семантической деривации наблюдается у субстантива *автомат* ‘ручное автоматическое скорострельное оружие’, вторичный лексико-семантический вариант

которого в русском языке характеризует человека, действующего быстро и оперативно по выработанному шаблону, в итальянском языке (*mitragliatrice*) – ‘человека, который говорит или действует без прерывания’: *Специалисты, которые давно у нас трудятся, могут на* ***автомате*** *работать быстрее, и при этом качество от этого не страдает*. (Издательский дом «Беларусь сегодня» 10.09.2022); *Canta bene, scrive bene e quando mette il turbo il suo rap è veloce come una* ***mitragliatrice***. (La Repubblica, 01.05.2021). – *Он хорошо поёт, хорошо пишет, и, когда он включает турбо, его рэп быстр как* ***пулемёт***.

2. Субстантивы со значением снарядов и взрывчатых веществ характеризуются способностью к семантической деривации и образованию вторичных недеструктивных лексико-семантических вариантов. Существительные *снаряд – proiettile* не характеризуются семантической тождественностью вторичных лексико-семантических вариантов: ‘вид боеприпасов для стрельбы из орудий, миномётов, реактивной артиллерии’ → ‘предмет, устройство для спортивных упражнений’: *Говорили о высочайших результатах в мужском метании копья, где* ***снаряд*** *улетал за 90 метров.* (Издательский дом «Беларусь сегодня» 12.09.2019). В итальянском языке ‘человек, который делает очень беглые и быстрые движения’: *Wijnaldum arriva sul pallone come un* ***proiettile*** *e sorprende Stegen.* (La Repubblica 07.05.2019) – *Вейналдум бьёт по мячу как* ***снаряд*** [*пуля*] *и удивляет Стегена.* В соотносительной паре *динамит – dinamite* русский субстантив является моносемантическим, итальянский – полисемантическим: ‘взрывчатое вещество’ → ‘то, что провоцирует сенсацию’: *Quest'ultimo episodio esplode come una* ***dinamite*** *verso il termine di quella che pare una sinfonia a più voci.* (La Repubblica 28.04.2005). – *Этот последний эпизод взрывается, как* ***динамит****, к концу того, что кажется симфонией с несколькими голосами*; → ‘очень крепкий ликёр, острая пища или специи’: *Fai attenzione, questo whisky, è* ***dinamite***! – *Будьте осторожны, это виски –* ***динамит***! (De Mauro). Как известно, динамит используется для осуществления взрыва, который характеризуется внезапностью, быстротой, мощностью, силой. За счёт этих параметров динамит сравнивается с алкогольным напитком, который обладает сильными вкусовыми качествами, и эпизодом, который появляется внезапно и быстро.

3. В группу существительных со значением холодного оружия включаются субстантивы *штык* (итал. *baionetta*) ‘холодное колющее оружие в виде большого ножа’, *сабля*

(итал. *sciabola*) ‘холодное колющее оружие с изогнутым лезвием’, *шпага* (итал. *spada*) ‘холодное колющее оружие с узким длинным лезвием’ и др. Например, в номинативных парах *штык* – *baionetta*, *сабля* – *sciabola* моносемантическими являются номинации *сабля* и *baionetta*, в то время как две других номинации развивают вторичные значения: *штык* → ‘слой земли в глубину, который можно захватить лопатой’, *sciabola* → ‘одна из трёх специальностей фехтования’.

4. Субстантивы со значением орудий труда и инструментов представлены номинациями *топор* (итал. *scure*) ‘устройство в виде клина, установленного на топоре с лезвием с одной стороны и обухом с другой’, *нож* (итал. *coltello*) ‘режущее приспособление, состоящее из ножа и ручки’, *пила* (итал. *sega*) ‘инструмент с острыми зубьями для резки дерева’, *хлыст* (итал. *scudiscio*) ‘тонкий и гибкий прут; плеть’, *плеть* (итал. *frusta*) ‘кнут из перевитых ремней или веревок; наказание ударами такого кнута’ и др. К примеру, в паре существительных *хлыст* – *scudiscio* итальянский номинатив является моносемантическим, русский развивает значение ‘цельный ствол поваленного дерева с вершиной, очищенный от сучьев’.

Немногочисленной является локативная группа со значением места, где осуществлялись деструктивные факты, события, ситуации: рус. *эшафот* (итал. *patibolo*) ‘помост для совершения смертной казни или для приведения в исполнение публичных наказаний’, *бойня* (итал. *macello*) ‘место (обычно здание) для убоя скота, скотобойня’, *Голгофа* (итал. *Golgota*) ‘холм около Иерусалима, на котором по христианскому преданию был распят Иисус Христос’, *плаха* (итал. *ceppo*) ‘обрубок дерева, на котором отсекали голову казнимого, а также помост, на котором совершалась казнь’, *ад* (итал. *inferno*) ‘место, где души умерших «грешников» подвергаются вечным мукам’.

Например, понятия *Голгофа* ‘*Golgota*’ схожи в значении ‘название места распятия Иисуса Христа, «место черепа» // место тяжких мучений, испытаний’. *Путь на* ***Голгофу*** *проходил от школы, возле церкви и сворачивал на проселочную дорогу к лесу. Там людей раздевали до нижнего белья, снимали обувь и заставляли прыгать в яму, где из винтовок убивали.* (Издательский дом «Беларусь сегодня» 02. 03. 2021). Субстантив Голгофа именует место, где издевались над людьми, которых после убивали.

В итальянском языке находим пример религиозной тематики, который затрагивает тему бытия, подчёркивает важность традиций. В Италии, начиная с декабря, люди готовятся к главным торжествам года – Непорочному зачатию Пресвятой Девы Марии

и Рождеству. В эти дни дома, улицы Италии украшает композиция Презепе (вертеп), символизирующая рождение Спасителя и состоящая из фигурок животных, людей, архитектуры и пейзажей. Сама статья посвящена тому, что несмотря на то, что весь мир объят злостью и несправедливостью, полон деструктивности, где *una ragazza viene uccisa dai propri genitori perché voleva sposare il ragazzo che amava e non chi le imponevano.* (La Repubblica 21.12.2022). – *Девушка убита своими родителями, потому что она хотела выйти замуж за парня, которого любила, а не за того, кого они навязывали ей, где milioni di profughi costretti a vivere senza elettricità, acqua e tutto il resto.* (La Repubblica 21.12.2022). – *Миллионы беженцев вынуждены жить без электричества, воды и всего остального,* человек не теряет надежды на лучшее будущее. Автор говорит о том, что Бог продолжает спускаться к людям на землю, зная, что место его распятия уже заготовлено: *Sono più di duemila anni che il nostro Dio ritorna nonostante sappia che è già è pronta per lui una croce sul* ***Golgota*** (La Repubblica 21.12.2022). – *Уже более двух тысяч лет наш Бог возвращается, несмотря на то, что знает, что для него уже готов крест на* ***Голгофе.***

В синонимической группе субстантивов со значением результатов деструктивной деятельности *руина – rovina*, *развалина – rudere* первичные лексико-семантические варианты соотносятся с остатками разрушенного строения, развалинами какого-либо сооружения. Вторичные лексико-семантические варианты указывают на то, что осталось, уцелело от чего-либо исчезнувшего, ушедшего, потерю престижа, авторитета, спад экономики и пр.: ***Развалины*** *старого света,* ***развалины*** *прежней славы, экономика превращается в* ***руины.*** (Издательский дом «Беларусь сегодня»); *Una malattia grave poteva causare la* ***rovina*** *di un'intera famiglia.* (La Repubblica 03.03.2021). – *Тяжёлая болезнь могла стать причиной* ***руины*** [*смерти*] *всей семьи*; *La sofferenza della malattia, la paura per i propri familiari e i propri amici, a cui si aggiunge in tanti casi il dolore per la perdita di persone care e la preoccupazione per la* ***rovina*** *economica di famiglie e imprese.* (La Repubblica 24.10.2020). – *Страдание от болезни, страх за членов семьи и друзей, к которому во многих случаях добавляется боль из-за потери близких и озабоченность экономической* ***руиной*** [*разорением*] *семей и предприятий.* В публицистических примерах описаны отрицательные итоги, плохие окончания, результаты того, что было раньше. В итальянском языке номинатив может употребляться с целью описания смерти, а также затрагивать материальное положение людей.

Кроме того, у существительных фиксируется персонифицированный лексико-семантический вариант 'слабый, немощный, одряхлевший от старости человек'. В данном случае

существительные описывают немощность, слабость человека, которые приходят или могут прийти к нему с возрастом: *Без движения человек – **развалина**, он жалок и несчастен*. (НКРЯ); *А потом оглянулся – и вижу, что я уже **развалина**…* (НКРЯ)*; У меня живёт теперь страшная **руина** – бывшая наша изящная и преумная Александра Петровна Тютчева*. (НКРЯ); В русском и итальянском предложениях *Она была далеко не **развалина**, а еще хоть куда*! (НКРЯ); *Però non sono un **rudere**: sto al passo coi tempi*. (La Repubblica 13.04.2023). – *Но я не **развалина**: я иду в ногу со временем* нейтрализуется вторичное значение существительных за счёт использования выражений *а ещё хоть куда* и *идти в ногу со временем,* которые показывают силы и возможности, имеющиеся у человека, а также его стремление соответствовать сегодняшнему дню.

5 Обратная семантическая модель

В корпусе деструктивных номинаций выделяются два основных разряда лексических единиц. В первом представлены глагольные и именные лексико-грамматические разряды слов с первичной деструктивной семантикой. Первичные лексико-семантические варианты в таких номинациях экстраполируют определённую часть деструктивной семантики на недеструктивные обозначения явлений и фактов окружающей действительности (сферы социальных, межличностных и экономических отношений, экзистенциальной реальности, физического и психического состояния человека и мн. др.). В таких случаях деструктивный континуум демонстрирует свой широкий семантический потенциал и возможность свободного выхода из деструктивности в другие области и сферы, где многочисленные вторичные лексико-семантические варианты образуются по прямой модели «D (деструктивный лексико-семантический вариант) → N (недеструктивный лексико-семантический вариант)». Такая модель как когнитивный аналог реальных связей сущностей объективного мира, их взаимодействий и зависимостей может считаться схемой формирования метафорического значения, которая характеризуется разной соотнесённостью первичных и вторичных лексико-семантических вариантов.

Несмотря на стабильность и постоянство деструктивного континуума, в нём наблюдаются качественные и количественные изменения, связанные с незначительным увеличением метафорических лексико-семантических вариантов, образованных от недеструктивных значений глагольных и субстантивных единиц других тематических групп. Такие вторичные деструктивные лексико-семантические варианты являются результатом

семантической деривации, представленной обратными семантическими моделями (Алешина 2003, 74).

Обратная семантическая модель «N → D» («недеструктивный исходный лексико-семантический вариант → деструктивный производный лексико-семантический вариант») в соотношении с прямой моделью «D → N» характеризуется небольшим разнообразием частных подмоделей, ограниченным числом вторичных лексико-семантических вариантов, что свидетельствует о функционально-семантической асимметрии, непропорциональном соотношении результатов прямой и обратной моделей. Поэтому приток вторичных лексико-семантических вариантов в корпус деструктивных номинаций является малопродуктивным, нерегулярным, а сами лексико-семантические варианты представлены незначительным числом образований.

Источником метафорической экспансии среди субстантивов являются наименования предметов быта. Так, вторичные лексико-семантические варианты субстантивов *мешок*, *котёл*, *клещи* указывают на полное окружение воинской группировки, захват противника: *мешок* ‘сделанное из мягкого материала вместилище для чего-нибудь сыпучего, для различных мелких предметов’ → ‘окружение вражескими войсками’, *котёл* ‘металлический сосуд округлой формы для варки пищи, нагревания воды и т.п.’ → ‘окружение больших групп войск противника’, *клещи* ‘металлический инструмент в виде щипцов’ → ‘охват противника с двух сторон’: *Из окружения демянского* ***мешка*** *немцы-таки вышли*. (НКРЯ); *Наша батарея попала в* ***мешок***, *в окружение*. (НКРЯ); *Потом, когда стали обрисовываться контуры Уманского* ***котла***, *русские войска были остановлены*. (НКРЯ); *Попал немедленно в* ***клещи*** *вместе со своим штабом*. (НКРЯ).

В итальянском языке у существительных *calderone* ‘большой котёл’, *pinza* ‘клещи’ подобные вторичные деструктивные значения не развились, однако в значении ‘пространство, в котором вражеская армия окружена обходным манёвром’ используется вторичное значение существительного *sacca* ‘сумка’: *Sembra che l’Isis voglia trasferire tutte le sue migliori forze in Siria, distruggere la* ***sacca*** *di Deir ez-Zour e trincerarsi in un triangolo fortificato fra le città di Raqqa, Palmira e Abu Kamal*. (La Repubblica, 18.01.2017). – *Похоже, ИГИЛ хочет перебросить в Сирию все свои лучшие силы, уничтожить* ***сумку*** [*котёл*] *Дейр-эз-Зор и закрепиться в укреплённом треугольнике между городами Ракка, Пальмира и Абу-Камаль*. Несколько иную смысловую соотнесённость имеет итал. *sacco* ‘мешок’, употребляющийся в значении ‘разграбление’: *Il* ***sacco*** *di Roma*. (De Mauro) – ***Мешок*** [*разграбление*] *Рима*.

Медицинский термин *операция* и его итальянский аналог *operazione* прочно вошли в лексический состав двух языков со значением ‘совокупность боевых действий, объединённых одной

целью, одним заданием': *Это была последняя наступательная* ***операция*** *фашистов на Западе*. (НКРЯ); *Высадка в Нормандии – крупнейшая десантная* ***операция*** *Второй мировой войны*. (НКРЯ); *Участие Италии в испанских событиях не ограничивалось одними военными* ***операциями***. (НКРЯ); ***Operazione*** *navale*, *aerea*, ***operazione*** *difensiva*, *offensive*. (De Mauro). – ***Операция*** *военно-морская*, *воздушная*, ***операция*** *оборонительная*, *наступательная*. Как видно, обратная модель реализована как в русском, так и итальянском языках, когда недеструктивное существительное медицинской сферы развивает вторичные значения деструктивной военной сферы.

Вторичный деструктивный лексико-семантический вариант итал. *manovra* 'тактическое или стратегическое перемещение воинской части в ходе наступательных или оборонительных действий' образовался по обратной семантической модели от первичного лексико-семантического варианта 'набор операций, необходимых для запуска машины или устройства в действие': *Taiwan, la Cina accerchia l'isola con una maxi-****manovra*** *militare*: *impegnati 71 aerei e 9 navi*. (La Repubblica 04.04.2023). – *Тайвань, Китай окружает остров максимальным военным* ***манёвром***: *задействованы 71 самолёт и 9 кораблей*.

В реципиентную сферу обратных семантических моделей часто включается понятие *смерть*, которое отождествляется с событиями и фактами завершающего характера: заключительной частью музыкальных, оперных и литературных произведений, спортивных соревнований (*эпилог*, *финал*, *финиш*): *Трагический* ***эпилог*** *жизни Пушкина – такова главная тема исторического романа.* (НКРЯ); *Даже самый печальный и неуклюжий* ***финал*** *прожитой жизни лучше, чем жизнь банально непрожитая*. (НКРЯ); *Жизнь имеет свой* ***финиш***. (НКРЯ); *Эти дни будут самыми яркими и, пожалуй, самыми счастливыми в моей сложной, теперь уже приближающейся к* ***финишу*** *жизни*? (НКРЯ). В итальянском языке существительные *epilogo* 'эпилог', *finale* 'финал' в деструктивном значении не используются, а существительное *traguardo* 'финиш' может развивать вторичный деструктивный лексико-семантический вариант 'приспособление для прицеливания огнестрельного оружия'.

В парадигму обратных семантических моделей активно включаются метеорологические термины, в семантике которых актуализируются такие маркеры, как сила, порывистость, разрушительность, стремительность, интенсивность проявления природных действий. Смысловое тождество реципиентных сфер наблюдается у номинаций *шквал – raffica*, 'сильная массированная стрельба из орудий, пулемётов, артиллерийских снарядов', *буря – bufera* 'события разрушительного характера': *Они обрушили на позиции десантников* ***шквал*** *огня из миномётов и*

гранатомётов. (НКРЯ); *Артиллерийский* ***шквал*** *уже схлынул*. (НКРЯ); *Грянула* ***буря*** *Отечественной войны*. (НКРЯ); *Si udivano raffiche di mitra, una* ***raffica*** *di fucileria*. (De Mauro). – *Слышались автоматные* ***шквалы***, *оружейный* ***шквал*** [*залп*]; *La* ***bufera*** *della guerra civile*. (De Mauro). – ***Буря*** *гражданской войны*. В русском языке, как и в итальянском, в процесс семантической деривации также включаются субстантивы *ураган*, *смерч*: *На мирно спящий дом обрушился* ***ураган*** *огня, дыма, пыли, гари, осколков*. (НКРЯ); *Раздались выстрелы, и опять разразился* ***ураган*** *курсантского огня*. (НКРЯ); *Пулемётный огонь смешивался с* ***ураганом*** *дальнобойных снарядов*. (НКРЯ); *Человек, попавший в* ***смерч*** *огня и со свистом несущихся камней, умирает мгновенно*. (НКРЯ). В итальянском языке существительное *uragano* ‘ураган’ может обозначать как ‘большой шум, гул’, так и ‘чрезмерно возбуждённого, беспокойного человека’: – *L’oratore ha appena finito il suo discorso, che viene sommerso da un* ***uragano*** *di applausi*. (La Repubblica 28.09.2011). – *Оратор только что закончил свою речь, которую охватил* ***ураган*** *аплодисментов*; *Leon ha da poco festeggiato i suoi due anni ed è un vero* ***uragano***. (La Repubblica). – *Леон недавно отпраздновал свои два года, и он настоящий* ***ураган*** [*активный ребёнок*].

6 Выводы

Таким образом, межъязыковые сходства и различия в деструктивном континууме русского и итальянского языков рельефно выделяются в субстантивных номинативных единицах и обусловлены рядом факторов лингвистического и экстралингвистического характера. Деструктивные существительные русского и итальянского языков могут быть объединены по проанализированным группам, однако количество групп может быть увеличено в связи с выявлением других номинаций пейоративного характера. Для освещения деструктивных событий 21 века, публицистические дискурсы обоих языков (русского и итальянского) широко используют пейоративные существительные, чем не только более подробно описывают происходящее, но и вызывают у читателя ещё больший интерес к освещаемым событиям. Соотношение прямой и обратной моделей носит непропорциональный, асимметрический характер, заключающийся в превалировании вторичных номинаций, образованных по прямой модели. Небольшое число обратных семантических моделей подтверждает тот факт, что деструктивная лексико-тематическая группа является практически закрытой и не допускает активного включения в свой состав лексических единиц из других тематических групп.

Литература

Алешина, О.Н. (2003). *Семантическое моделирование в лингвометафорологических исследованиях (на материале русского языка)*. Дис. … д-ра филол. наук, Новосибирск, 367 с.

Баранов, А.Н. (2003). *О типах сочетаемости метафорических моделей*. Вопросы языкознания, 2, 73-94.

Бураихи, Ф.К. (2011). *Глаголы деструктивного воздействия в современном русском языке: системные и функциональные характеристики*: автореф. дис. ... канд. филол. наук: 10.02.01. Воронеж. гос. ун-т. Воронеж, 23 с.

Власова, Е.В. (2005). *Речевая агрессия в печатных СМИ (на материале немецко- и русскоязычных газет 30-х и 90-х гг. XX в.)*: дис. … канд. филол. наук : 10.02.19. Саратов, 219 л.

Волкова, Я.А. (2014). *Деструктивное общение в когнитивно-дискурсивном аспекте* : автореф. дис. … д-ра филол. наук: 10.02.19; Волгогр. гос. пед. ун-т. Волгоград, 46 с.

Воронцова, Т.А. (2006а). *Речевая агрессия в коммуникативно-дискурсивной парадигме.* Вестн. Воронеж. гос. ун-та. Сер.: Лингвистика и межкультур. коммуникация, 1, 83-6.

Воронцова, Т.А. (2006б). *Речевая агрессия: вторжение в коммуникативное пространство.* Ижевск: Удмурт. гос. ун-т, 256 с.

Евгеньева, А.П. (1999). *Словарь русского языка*: В 4-х т., РАН, Ин-т лингвистич. исследований. Москва: Рус. яз.; Полиграфресурсы, 2985 с.

Злоказов, К.В. (2016). *К оценке риска воплощения анонимной угрозы*. Полит. лингвистика, 3. https:politlinguistika.ru/images/3-2016/12.pdf.

Ивлиев, И.В. (1998). *Лексикографическое описание глаголов со значением лишения жизни в русском языке*: дис. ... канд. филол. наук : 10.02.01. Москва,199 л.

Издательский дом «Беларусь сегодня». https://www.sb.by.

Каримова, Г.Т. (2011). *Прагматика глаголов разрушения и созидания в дискурсе медицинской рекламы*. Учен. зап. Казан. ун-та. Сер.: Гуманитар. науки. Т. 153, кн. 6, 212-218.

Кокорина, Е.В. (1996). *Стилистический облик оппозиционной прессы*. Русский язык конца XX столетия (1985–1995): коллектив. моногр. / В.Л. Воронцова [и др.]; отв. ред. Е.А. Земская; Рос. акад. наук, Ин-т рус. яз. Москва, 409-26.

Копнина, Г.А. (2017). *Речевое манипулирование:* учеб. Пособие. 6-е изд., стер., Москва: Флинта : Наука, 169 с.

Кулагина, Д.А. (2021). *Исследования вербальной репрезентации речевой агрессии*; Соц. и гуманитар. науки. Отечеств. и зарубеж. лит. Сер. 6, Языкознание, 1, 92-96. https://doi.org/10.31249/ling/2021.01.05.

Маевский, В.М. (2022). *Имплицитная речевая агрессия в текстах современных британских СМИ*. Вестн. Волгогр. гос. ун-та. Сер. 2, Языкознание. Т. 21, 4, 111-22.

Немов, Р.С. (2007). *Психологический словарь : более 11000 определений психол. терминов,* Москва : ВЛАДОС, 559 с.

НКРЯ = *Национальный корпус русского языка.* https://ruscorpora.ru.

Потапенко, Т.А. (1978). *Лексико-семантическая группа глаголов разрушительного воздействия на объект в современном русском*

литературном языке: дис. ... канд. филол. наук: 10.02.01 / Т.А. Потапенко., Москва, 208 л.

Потапенко, Т.А. (1983). *Лексико-семантическая характеристика глаголов с общим значением разрушительного воздействия на объект*. Филол. науки., 2, 50-56.

Рогова, Ю.В. (2021). *Группы деструктивных номинаций в современном публицистическом дискурсе*. Язык и межкультурные коммуникации: сб. науч. ст. / Белорус. гос. пед. ун-т, 230-233.

Рогова Ю.В. (2022). *Сложные деструктивные конструкции на страницах «СБ. Беларусь сегодня».* Образование и наука в XXI веке: ежегод. сб. науч. тр. БГПУ. Вып. 4, 79-81.

Садовников, С.А. (2022). *Приемы речевой агрессии как средство манипуляции в программе «Вести недели».* Вестн. Нижегор. ун-та им. Н.И. Лобачевского, 5, 163-67.

Седов, К.Ф. (2003). *Речевая агрессия в межличностном взаимодействии.* Прямая и непрямая коммуникация: сб. науч. ст. / Сарат. гос. ун-т; отв. ред. В.В. Дементьев. Саратов, 196-212.

Сковородников, А.П. (1997). *Теоретические и прикладные аспекты речевого общения*: науч.-метод. бюллетень. Краснояр. гос. ун-т, лаб. лингвоэкологии и речевой культуры. Красноярск; Ачинск: Краснояр. гос. ун-т, 2, 68 с.

Старычонак, В.Д. (2022а). *Семантычная прастора дзеясловаў беларускай мовы (на матэрыяле другасных намінацый)*: манаграфі. – Мінск : Беларус. дзярж. пед. ун-т, 315 с.

Старычонак, В.Д. (2022б) *Субстантыўныя дэструктыўныя найменні ў беларускай мове*. Вес. БДПУ. Сер. 1, Педагогіка. Псіхалогія. Філалогія, 3, 94-8.

Суванова, Э.Х. (1985). *Лексико-семантическая классификация непроизводных глаголов*. Актуальные проблемы русского словообразования: сб. ст. / Ташкент. гос. пед. ин-т. – Ташкент, 80-84.

Усик, Е.Г. (2009). *Прагматическая направленность метафорических средств в языке экономики (на материале подъязыка телевидения в ФРГ).* Изв. Сарат. ун-та. Н. с. Сер.: Филология. Журналистика, 9, 24-27.

Фаткуллина, Ф.Г. (2010). *Концепт «деструкция» и способы его представления в русском языке*. Вестн. РУДН. Рус. и иностр. яз. и метод. их препод, 3, 57-66.

Фаткуллина, Ф.Г. (2002). *Категория деструктивности в современном русском языке* : дис. … д-ра филол. наук : 10.02.01. Уфа, 322 л.

Фоменко, Ю.В. (1975). *Класс глаголов «удара» в русском языке*. Вопросы теории русского языка: сб. ст. / Новосиб. гос. пед. ин-т., 119, 23-26.

Шпомер, Е.А. (2021). *Проявление речевой агрессии в СМИ (на материале газеты «Комсомольская правда»)*; Вестн. Хакас. гос. ун-та им. Н.Ф. Катанова, 2, 77-83.

Шульгина, К.В. (2021). *Источники сведений о речевой ситуации оскорбления при решении экспертных зада*ч. Филология и человек, 2, 88-96.

Щербинина, Ю.В. (2008). *Вербальная агрессия* / Ю.В. Щербинина. Изд. 2-е, Москва: ЛКИ, 360 с.

Юрина, Е.; С.В. Доронина. (2021). *Речевой акт угрозы в различных коммуникативных ситуациях.* Юрислингвистика, 20, 19-23.

Digita. (1999). *Dizionario Interattivo Garzanti Della Lingua Italiana* [Risorsa elettronica]. Milano: Garzanti Linguistica.

Enciclopedia Italiana di Scienze, Lettere ed Arti. Treccani. https: www.treccani.it.

Il dizionario della lingua italiana De Mauro (2000). https://dizionario.internazionale.it

Zingarelli, N. (2007). Lo Zingarelli. *Vocabolario Della Lingua Italiana* [Risorsa elettronica]. Milano: Zanichelli.

Rivista semestrale | Semestral journal

Dipartimento di Studi Linguistici
e Culturali Comparati
Università Ca' Foscari Venezia

Università
Ca'Foscari
Venezia