

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.

Sommario 1 Introduzione. – 2 *Corpora diacronici*: caratteristiche specifiche e problematiche comuni. – 3 Rassegna dei *corpora diacronici* per la lingua russa. – 3.1 *corpora* storici inclusi nel *Nacional'niy korpus russkogo jazyka*. – 3.2 *Corpus Cyrillo-Methodianum Helsingiense*. – 3.3 I *corpora* di paleoslavo e russo antico del *Thesaurus Indogermanischer Text- und Sprachmaterialien*. – 3.4 *Trondheim-Sofia Corpus of Old Slavic*. – 3.5 «*Manuskript*». *Slavjanskoje pis'mennoe nasledie*. – 3.6 *Regensburgskij diachroničeskij korpus russkogo jazyka*. – 3.7 *Diachronen korpus na bălgarski ezik (IX-XVIII vv.)* e il sistema *Histdict*. – 4 *Corpora diacronici* specialistici. – 4.1 *Sankt-Peterburgskij korpus agiografičeskikh tekstov*. – 4.2 Il portale *Istorija russkoj leksikografii*. – 5 Conclusioni.



**Edizioni
Ca' Foscari**

Peer review

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

Open access

© 2024 Bindi | 4.0



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

111

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 korpusnuyu 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 delle 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).

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 dall'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://torotreebank.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. Pičhadze.

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 Pičhadze (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 comprensione 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 orfografiya*) 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 orfografiya*), la query viene impostata tenendo conto delle varianti ortografiche nella stessa posizione (ad esempio, alternanza tra o/e, e/ie, o/w i/i'). È infine possibile interrogare il corpus con una «ortografia moderna» (*modernizirovannaja orfografiya*), la quale non equipara soltanto le varianti ortografiche, ma anche alcune lettere, che indicano fonemi diversi (ad esempio, e/ě/ň, e tutte le loro varianti ortografiche) (Mišina, Pičhadze 2015, 102).

⁹ Si sono occupati in modo particolare dell'annotazione della risorsa G.S. Barankova, M.V. Ermolova, N.P. Iordanī, D.S. Krylov, I.I. Makeeva, E.A. Mišina, M.S. Mušinskaja, P.V. Petručin, 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 *e/ë/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. Legkikh.

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*, svolta 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 meta-testuali 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/>.

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.

A 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 sopramenzionate 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 interrograre le suddette raccolte nella loro interezza, o se invece creare ulteriori sotto-*corpora* 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 'cam eci', ne vengano rintracciate anche tutte le possibili varianti ortografiche storiche (самъ юси, са(м) 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. Zajnnullina (2017), ad esempio, hanno considerato le collocazioni verbo-nominali con la componente *životū* ('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 *čeloveči* e *čelovečiskū* ('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-academico 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 inseriscono 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. Tomšić.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. Jauhainen, U. Dieckmann, e S. Pöyhönen.

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

²⁷ V. Jagić, *Quattuor Evangeliorum versionis palaeosloveniae 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 Tomšić, 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 Pra-ga*, 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, raccolgendo 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». *Slavjanskoje 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 preeistenti.

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, Profetologgi, 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 glagoliceskich 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 glagoliceskich 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 programmati, 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 NKRJ (cf. *supra*).

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

Completano il corpus le seguenti collezioni: *kollekciya slavjanskich učitel' nich proizvedenij* ('collezione dei testi di tipo didattico') dell'XI-XIV sec., *kollekciya drevnerusskich perevodov* ('collezione delle traduzioni della Rus'), *kollekciya russkich letopisej* ('collezione delle cronache slave orientali'), *kollekciya 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 *Me-naion*, scaricabile dall'utente.

Le collezioni del corpus sono annotate a vari livelli. In primo luogo, è stata realizzata un'annotazione di tipo analitico (*analitčeskaja 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 scsvri da criticità. In primo luogo, bisogna sottolineare che il sopramenzionato 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» (*rasširennjyj 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 *Manuscript* 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 *Manuscript* 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 *Manuscript* 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 *Manuscript*, 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 proficie 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 *Manuscript* 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 *Manuscript*, 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 *Manuscript*. 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 Efremia 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 *Radzivillovskaia 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-semanticci. Nei contributi dell'autore vengono indagate le relazioni di significato esistenti tra i verbi del sapere *věděti*, *vědati*, *znati* ('sapere', 'conoscere') nella lingua slava orientale antica (Žolobov, Baranov 2021); viene offerta una descrizione linguistico-statistica delle trasformazioni storiche del gruppo lessicale *životъ* - *žiznъ* - *žtie* ('vita') nella tradizione scrittoria della Rus' (Žolobov, Baranov 2022); viene identificata la dinamica dei cambiamenti nella serie lessicale e nella struttura semantica dei lessemi *brani* - *voina* - *ratъ* ('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 Diaconica: 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 Sofja, 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-XVI* II 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» (*tärsene*). 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 soprallineari, 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 parentesi 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. бе(з)молвии - безмо(л)вии - безмолвии: безмолвии). 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. алтаря - олтаря: алтаря), o delle desinenze *-ogo/-ago* (es. студентаго - студентаго) (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'équipe 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](https://www.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](https://www.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/>.

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, corredata della numerazione delle singoli voci del dizionario, e come tabella contenente 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 interroga-zione 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 nasa-li (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'at-testazione 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 interroga-zione, predili-gendo, 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 morfolologica. 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/>.
- MANUSKRIFT: «*Manuskript*». *Slavjanskoе 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/rруди>.
- 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 Indo-germanischer 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://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 korpuze SKAT» Грамматическая разметка в корпусе SKAT (Annotazione grammaticale nel corpus SKAT). Zacharov V.P. (otv. red.), *Trudy meždunarodnoj konferencii «Korpusnaja lingvistika – 2011»* (27-29 iunija 2011 g.), Sankt-Peterburg, 69-73.
- Azarova, I.V.; Alekseeva, E.L. (2013). «Osobennosti morfo-sintaksičeskoj razmetki drevnerusskich agiografičeskikh tekstov» Особенности морфо-синтаксической разметки древнерусских агиографических текстов (Caratteristiche dell'annotazione morfo-sintattica dei testi agiografici russi antichi). Zacharov, V.P. (otv. red.), *Trudy meždunarodnoj konferencii «Korpusnaja lingvistika – 2013»* (25-27 iunija 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 statističeskiy analiz srednevekovych slavjanskikh tekstov: instrumentarij korpusa "Manuskript" i metodika ego ispol'zovaniya» Количествоный 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 čästjach, Perm, 40-9.
- Baranov, V.A. (2017b). «Statističeski značimye slova kak charakteristika srednevekovogo slavjanskogo teksta (na materiale kollekcií 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 Vserossijskoy nauchno-praktičeskoy 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 istoričeskikh korpusov slavjanskikh 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 slavjanskikh služebnyx 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](https://doi.org/10.15388/SlavViln.2019.64(2).17)
- Baranov, V.A. (2019c). «K voprosu ob ispol'zovanií statističeskikh metodov dlja poiska kollokacij i kolligacij v drevnejšich slavjanskikh tekstuach (na materiale glagoličeskix rukopisej korpusa "Manuskript")» К вопросу об использовании статистических методов для поиска коллокаций и коллизий 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*]). *Slovovo. Časopis Staroslavenskoga instituta u Zagrebu*, 69, 1-33. <https://doi.org/10.31745/s.69.1>.
- Baranov, V.A. (2020), «Korpusnye issledovaniya srednevekovych slavjanskikh rukopisej: statističeski značimye n-grammy (kollokacii) drevnerusskikh 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 soderžestvo nauk: opyt primenenija informacionnych technologij v istorii i smezhnych 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 pedagogičeskoy lingvistiki. 2. Prikladnaja lingvistika: sovremennoye resursy i perspektivy*, 94-106. <https://doi.org/10.29025/2079-6021-2022-2-94-106>.
- Baranov, V.A.; Dubovcev S.V. (2012). «Modul' statistiki informacionno-analitičeskoy sistemy "Manuskript": funkci 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 sentyabrja 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 sostavljujušie 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 oktyabrya 2010 g.), Ufa; Izhevsk, 28-31.
- Budennaia, E.V. (2020). «V poiskach triggera: knižnye i neknižnye teksty kak markery različnykh 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 piuccherperfetto nella lingua russa dei secoli XIV-XVI). *Vestnik Volgogradskogo gosudarstvennogo 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 naucnoj konferencii, posvyashenoj 100-letiju so dnja roždenija zasluzhnogo 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). «SLecl. razumět. 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). «Cerkovnoslavjanske slova imarmenja, fatunī, fortuna v proizvedenijach Maksima Greka» Церковнославянские слова имарменя, фатунъ, фортуна в произведениях Максима Грека (Le parole slavo-ecclesiastiche imarmenja, fatunī, 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». *Cyrillometodianum*, 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 Svyato-Tichonovskogo gumanitarnogo universiteta. Serija III. Filologija*, 2(47), 7-25.
- Gerd, A.S. (2004). «Elektronnyj korpus tekstov po pamjatnikam drevnerusskoj agiografičeskoy 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 agiografičeskikh tekstov SKAT: sovremennoe sostojanie i perspektivy razvitiya» Корпус древнерусских агиографических текстов SKAT: 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 naucnoj konferencii* (Iževsk, 13-17 iulija 2006 g.), Iževsk, 38-42.
- Kilina, L.F. (2020). «Problemy izuchenija ustojčivych glagol'no-imennych sočetaniij 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 sentyabrja 2020 g.), Kostroma, 85-7.
- Kilina, L.F.; Kolosova, E.I. (2020). «O grammatičeskoj semantike prilagatel'nyx čelovečči i čelovečišči v starorusskikh letopisjach» О грамматической семантике прилагательных человечь и человечьскъ в старорусских летописях (Sulla semantica grammaticale degli aggettivi čelovečči e čelovečišči 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). «Ustoččivye glagol'no-imennye sočetaniija s komponentom činiti v russkikh 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 ustoččivye sočetaniija s komponentom životi: funkcionirovanie i razvitiye (na materiale russkikh letopisej XIII-XV vv.)» Глагольно-именные устойчивые сочетания с компонентом жизнь: функционирование и развитие (на материале русских летописей XIII-XV вв.). (Collocazioni verbo-nome con la componente životi: 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 korpusnuyu lingvistiku* Введение в корпусную лингвистику (Introduzione alla linguistica dei corpora). Praha: Animedia.
- Kusennaja, T.V. (2021). «Motivirovannost' funkcionirovania iskonno russkoj i cerkovnoslavjanskoj leksiki v agiografičeskikh tekstuach 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.; Pičhadze, A.A. (2015). «Drevnerusskij podkorpus Nacional'no-go 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 russko-go 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, istoričeskikh 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 tehnologii: materiali ot Vmeždunarodna nauč. konf.* (Varna 15-20 septembri 2014 g.), Sofija; Iževsk, 214-16.
- Nikolenkova, N.V. (2017). «Sojusy v cerkonoslavjanskikh 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 moskowskogo 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 slucae pragmatikalizacii v russkom jazyke: mikrodiachronicheskoe 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 predistorii slova budto v russkom jazyke Srednevekov'ja» Кистории 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 nauchnoj konferencii* (21-23 sentjabrja 2017 g. Rjazan'), Rjazan', 69-76.
- Pen'kova, Ja.A. (2018) «Slavjanskoe predbudušee i ego grečeskie so-otvetstvija v perevodnych drevnecerkovnoslavjanskich pamjatnikach» Славянское предбудущее и его греческие соответствия в переводных древнепресторковнославянских 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, učinu, stanu, budu: korpusnoe issledovanie perifraz budušego vremeni v srednerusskoj pis'mennosti» Иму, учыну, стану, буду: корпусное исследование перифраз будущего времени в среднерусской письменности (*Imu, učinu, 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 vlijaniya 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 neopredelennykh mestoimenij: kvasirelativy 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 akademika 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 pozdrevně russkoj pis'mennosti» Инфинитивные конструкции с глаголом имѣти в поздревнерусской письменности (Costruzioni infinitive con il verbo *imeti* 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>.
- Piččadze, A.A. (2011). *Perevodčeskaja dejatel'nost' v domongol'skoj Rusi: lingvističeskiy aspekt* Переводческая деятельность в домонгольской Руси: лингвистический аспект (L'attività di traduzione nella Rus' pre-mongola: 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 orografiij i grammatiki» Корпус церковнославянских текстов, проблемы орографии 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 Ufimskogo juridičeskogo instituta MVD Rossii. Serija Obšie voprosy jazykoznanija. Jazyk i pravo. Jurislingvistika*, 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-Istoriya, 402-61.
- Romoli, F. (2018). «Sl.eccl. mudrost̄. 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 sočetaniya s komponentom dějati: semantika i funkcionirovanie» 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 lingvisticheskie issledovanija. V 2-ch čästjach, Ekaterinburg, 83-8.
- Rusanova, O.S. (2021). «Glagol'no-imennye sočetaniya s komponentom tворiti v russkich delovych tekstuach XIII-XVII vv.» Глагольно-именные сочетания с компонентом творити в русских деловых текстах XIII-XVII

- vv. (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 russkikh gramotach XIII-XVI vv.» Глагольно-именные сочетания с компонентом имѣти в русских грамотах XIII-XVI вв. (Collocazioni verbo-nome con la componente *imeti* nelle *gramoty russi* dei secoli XIII-XIV). Chramušina, Ž.A. et al. (otv. red.), *Aktual'nye voprosy perevoda, lingvistiki, istorii literatury i fol'klora* Sbornik statej X Meždunarodnoj naučnoj konferencii molodych učenych (11 fevralja 2022 g.), Ekaterinburg, 45-8.
- Ruvoletto, L. (2018). «Dvuvidovost' glagola běžat' v diachroničeskoj perspektive» Двувидовость глагола бѣжать в диахронической перспективе (Biaspettualità del verbo *bězati* in prospettiva diacronica). Patroeva, N.V. (otv. red.), *Fortunatovskie čtenija v Karelji. 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 bolěti i gorěti v drevnerusskom jazyke» Глаголы болѣти и горѣти в древнерусском языке (I verbi *bolěti* e *gorěti* nello slavo orientale antico). Gorbova E.V. (otv. red.), *Vzaimodejstvie aspeksa so smežnimi kategorijami. Materialy VII Meždunarodnoj konferencii Komissii po aspektologiji 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 starobǎlgarskija 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 zaključitel'nata 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: Gnossis.
- Uspenskij, B.A. (2002). *Istoriya russkogo literaturnogo jazyka (XI-XVII vv.)* История русского литературного языка (XI-XVII вв.) (Storia della lingua letteraria russa [XI-XVII sec.]). Izdanie 3-e, ispravленное и дополненное. Moskva: Aspent Press.
- Zacharov, V.P. (2015). «Istoricheskie korpusa i korpusnye diachronicheskie issledovaniya» 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 publčnaja naučno-tehnicheskaja 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 funkcionirovaniya» 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'klora* *Sbornik statej X Meždunarodnoj naučnoj konferencii molodych učenych* (11 fevralja 2022 g.), Ekaterinburg, 40-5.
- Zaliznjak, A.A. (2004). *Drevnenovgorodskij Dialekt* Древненовгородский диалект (L'antico dialetto di Novgorod). Vtoroe izdanie. Moskva: Jazyki slavjanskoy kul'tury.
- Žolobov, O.F. (2022). «Leksičeskij rjad bran'-voina'-rat' v staroslavjanskih i cerkovnoslavjanskikh istočnikach XI-XV vv.: Interpretacija kvantitativnykh 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 *brani'-voina'-rat'* nelle fonti paleoslave e slavo-ecclesiastiche dell'XI-XV sec.: interpretazione delle formule quantitative [sui materiali del corpus storico *Manuskript*]). *Palaeobulgaria*, 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-квантиративные 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 život’ – žizn’ – žitie: opyt lingvostatističeskogo opisanija» Трансформации лексического порядка *жизнь* – *житие*: опыт лингвостатистического 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 slavjanskikh spiskach Evangelija XII-XIII vv. (Na materiale “Pritči o bludnom syne”)» Из наблюдений над характером языковых разночтений в славянских списках Евангелия XII–XIII вв. (На материале «Причи 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 izuchenija istorii russkogo jazyka v ino jazyčnoj 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.

