

## Istoričeskie korpusa – Nacional’nyj korpus russkogo jazyka (NKRJa) Scheda di presentazione

a cura di Beatrice Bindi, Università “G. D’Annunzio” di Chieti-Pescara  
(beatrice.bindi@studenti.unich.it)

**Indirizzo web:** <https://ruscorpora.ru/> (ultimo accesso 02/2024).

**Anno di creazione:** 2004.

### Responsabile della progettazione:

- *Drevnerusskij korpus*: A.A. Pičhadze.
- *Starorusskij korpus*: O.N. Ljaševskaja.
- *Korpus berestjanych gramot*: D.V. Sičinava, A.N. Dyškant.
- *Korpus “vostočnoslavjanskaja epigrafika”*: D.V. Sičinava, D.D. Ignat’ev.
- *Cerkovnoslavjanskij korpus*: A.E. Poljakov.

**Indirizzo e-mail:** info@ruscorpora.ru

### Tipologie di contenuti:

- *Drevnerusskij korpus*: testi, originali e tradotti, redatti nella Rus’ tra l’XI e il XIV secolo.
- *Starorusskij korpus*: opere di genere vario, composte nello stato di Mosca dal XV al XVII secolo.
- *Korpus berestjanych gramot*: lettere su corteccia di betulla dei secoli XI-XV.
- *Korpus “vostočnoslavjanskaja epigrafika”*: iscrizioni (generalmente brevi didascalie) dei secoli XI-XV.
- *Cerkovnoslavjanskij korpus*: testi rappresentativi di tutti i principali generi della tradizione scritta slavo ecclesiastica, riportati, nella maggior parte dei casi, nelle loro moderne edizioni (XIX-XX secolo).

**Struttura del sito web della risorsa:** la risorsa ha un’interfaccia molto articolata, e tuttavia affatto dispersiva.

I *corpora* storici si trovano raggruppati all’interno di un apposito modulo nella pagina di accesso principale, sulla quale vengono altresì presentate tutte le differenti collezioni che compongono il Corpus nazionale nel suo complesso, assieme a una serie di informazioni riguardanti il progetto stesso, e le sue più recenti implementazioni.

I cinque *sotto-corpora* della sezione storica sono interrogabili singolarmente: ciascuno di essi dispone di una scheda informativa per l’utente, e di due distinti moduli per la ricerca “per forme esatte” (*poisk točnych form*), e “per forme lessico-grammaticali” (*poisk leksiko-grammatičeskich form*).

All’interno del modulo di ricerca per forme lessico-grammaticali si trovano ulteriori campi per

la specificazione dei parametri della *query*, in un complesso sistema ad incastri, che tuttavia si rivela di agevole utilizzo per l'utente: posizionando il cursore in corrispondenza di ciascuna casella, infatti, si apre una breve guida all'uso, che ne spiega le funzionalità, e ne suggerisce talvolta i contenuti, selezionabili da una finestra preimpostata.

**Lingua di visualizzazione:** inglese/russo.

Il *Corpus nazionale della lingua russa* (*Nacional'nij korpus russkogo jazyka*) è il più ampio e importante corpus della lingua russa. La risorsa è stata sviluppata nell'ambito del progetto *Filologia e informatica* (*Filologija i informatika*) presso l'Accademia russa delle Scienze, venendo resa 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 anche cinque *corpora* storici (*istoričeskie korpusa*), per un totale di circa quattordici milioni di parole.

Il corpus "antico russo" (*drevnerusskij*) contiene composizioni, originali o tradotte dal greco, redatte nella Rus' a partire dall'XI secolo: si tratta di cronache, testi agiografici, scritti giuridici, opere dedicate al diritto canonico, alcuni esemplari di letteratura di pellegrinaggio, e composizioni di carattere pedagogico. La selezione dei testi originali da includere nella raccolta si è basata sui più recenti studi sulle fonti storiche, mentre i monumenti letterari tradotti sono stati scelti sulla base di uno studio di A.A. Pičhadze (2011), dedicato alla considerazione dell'aspetto linguistico dell'attività traduttorica nella Rus' nel periodo pre-mongolo.

Il corpus "vecchio russo" (*starorusskij*) raccoglie opere di genere vario, composte nello stato di Mosca dal XV al XVII secolo, tra cui cronache e racconti, documenti di carattere commerciale, corrispondenze private, monumenti della letteratura religiosa, composizioni poetiche. Questa sezione mostra un carattere marcatamente ibrido: la lingua dei testi ivi contenuti, infatti, si caratterizza per una piramide funzionale piuttosto complessa di stili e registri, che rimangono peraltro in una zona d'ombra nell'ambito degli studi sull'evoluzione della lingua russa. Ai fini della compilazione della raccolta sono state utilizzate edizioni digitalizzate delle fonti, con particolare riferimento alla collana *Biblioteca della letteratura della Rus'* (*Biblioteka literatury Drevnej Rusi*, 1997-2020), e alla serie *Raccolta completa delle cronache russe* (*Polnoe sobranie russkich letopisej*, 1841-2004).

Il corpus delle "iscrizioni su corteccia di betulla" (*berestjanye gramoty*), raccoglie le lettere su corteccia di betulla (XI-XV secolo), le cui informazioni linguistiche sono state mutate dalla sezione appositamente dedicata all'interno della rivista "Voprosy jazykoznanija" (Questioni di linguistica).

Di recentissima elaborazione (2023), la raccolta "epigrafia slava orientale" (*vostočnoslavjanskaja epigrafika*) comprende iscrizioni dei secoli XI-XV, rivenute su pareti di chiese, su pietre, croci, e altri oggetti ad uso ecclesiastico o domestico.

Conclude l'elenco il corpus "slavo ecclesiastico" (*cerkovnoslavjanskij*), preposto alla raccolta di testi rappresentativi di tutti i principali generi della tradizione scritta slavo ecclesiastica (testi liturgici, testi patristici, testi di carattere ascetico-morale, sacre scritture, manuali di diritto ecclesiastico), i quali vengono riportati nelle loro moderne edizioni (XIX-XX secolo).

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 collegati da un modulo di ricerca comune, definito "corpus pancronico" (*panchroničeskij korpus*), che 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.

Tutte e cinque le raccolte della sezione storica del NKRJa sono annotate grammaticalmente e morfosintatticamente. Le fonti ivi contenute, inoltre, sono taggate con metadati, che ne rappresentano le caratteristiche testuali significative.

Ciascuna collezione è interrogabile attraverso due distinte modalità: se, da una parte, il modulo di “ricerca per forme esatte” si limita a rintracciare i contesti d’uso di una data stringa, dall’altra il modulo di “ricerca per forme lessico-grammaticali”, consente di personalizzare la *query* in base alle richieste dell’utente, che può rimuovere le condizioni non necessarie, o aggiungerne di supplementari.

I *corpora* storici del NKRJa costituiscono le risorse migliori ad oggi esistenti per una rappresentazione maggiormente rappresentativa della tradizione scritta slava orientale, in tutte le sue forme e generi, e in un arco temporale massimamente ampio (XI-XVIII). Sulla loro base è possibile realizzare studi a tutti i livelli linguistici (lessico, morfologia, sintassi), nonché estrarre utili dati statistico-distributivi.

## Bibliografia

- Dobrušina, Poljakov 2013: E.R. Dobrušina, A.E. Poljakov, *Korpus cerkovnoslavjanskogo jazyka: vozmožnosti, metody sozdanija, perspektivy*, in *Vestnik PSTGU. Serija III: Filologija*, 1 (31), 2013, pp. 32-44.
- Gavrilova, Šalganova, Ljaševkaja 2016: T.S. Gavrilova, T.A. Šalganova, O.N. Ljaševkaja, *K zadače avtomatičeskoj leksiko-grammatičeskoj razmetki staroruskogo korpusa XV-XVII vv.*, in *Vestnik PSTGU. Serija III: Filologija*, 2 (47), 2016, pp. 7-25.
- Mišina, Pičhadze 2015: E.A. Mišina, A.A. Pičhadze, *Drevnerusskij podkorpus Nacional'nogo korpusa russogo jazyka*, in A.M. Moldovan (otv. red.), *Trudy Instituta russkogo jazyka im. V.V. Vinogradova*, N. 2 (3), Moskva 2015, pp. 99-115.
- Mitrenina 2014: O. Mitrenina, *The Corpora of Old and Middle Russian texts as an advanced tool for exploring an extinguished language*, in *Scrinium*, 10 (1), 2014, pp. 455–461. <https://doi.org/10.1163/18177565-90000109>.
- Pičhadze 2011: A.A. Pičhadze, *Perevodčeskaja dejatel'nost' v domongol'skoj Rusi: lingvističeskij aspekt*, Moskva 2011.
- Poljakov 2014: A.E. Poljakov, *Korpus cerkovnoslavjanskich tekstov, problemy orfografii i grammatiki*, in

*Przegląd Wschodnioeuropejski*, V (1), 2014, pp. 245-254.

Sičinava 2016:

D.V. Sičinava, *Staroruskie/srednerusskie teksty v NKRJa: ot ekstensivnoj kollekcii k korpusu*, in V.A. Baranov, T.A. Timčenko (otv. red.), *El'Manuscript–2016. Rašytinis palikimas ir skaitmeninė technologijos. VI tarptautinė mokslinė konferencija, Vilnius, 2016 m. rugpjūčio 22–28 d.*, Vilnius 2016, pp. 208-210.

Sičinava 2022:

D.V. Sičinava, *Korpus beretjanych gramot kak parallel'nyj*, in *Trudy Instituta russkogo jazyka RAN*, 2 (32), 2022, pp. 92-106.

Sičinava, Dyškant 2021:

D.V. Sičinava, A.N. Dyškant, *Integration of the Old East Slavic Epigraphical Database. Corpora and Indices*, in *Scripta & E-Scripta. The Journal of Interdisciplinary Mediaeval Studies*, volume 21, pp. 95-108.