

Infrastructura electronică de cercetare pentru patrimoniul scris bulgar medieval: istorie și perspective

Anna-Maria Totomanova    *

Facultatea de Studii Slave, Universitatea "St. Kliment Ohridski" din Sofia, Tsar Osvoboditel Blvd. 15, 1504 Sofia, Bulgaria

Despre articol

Istoric:

Primit 21 septembrie 2021

Acceptat 30 septembrie 2021

Publicat 12 decembrie 2021

Cuvinte-cheie:

diacronie

gramatică istorică

lexicologie istorică

documente vechi

Rezumat

Prezentul articol trasează istoria sistemului *Histdict*, care a stat la baza *Infrastructurii electronice de cercetare pentru patrimoniul scris bulgar medieval* (*Electronic Research Infrastructure for Bulgarian Medieval Written Heritage*), inclusă la sfârșitul anului 2020 în baza de date națională pentru cercetare. Prin acest act statul își declară sprijinul pentru resursele noastre, care pînă acum au fost create și susținute doar pe baza unor finanțări din diverse proiecte. Acest lucru reprezintă, desigur, și o mult așteptată recunoaștere a eforturilor și realizărilor noastre. Pe de altă parte, acest act coincide cu alte două evenimente notabile: includerea RESILIENCE (Infrastructura de Cercetare pentru Studii Religioase), din care face parte și *Histdict*, pe harta europeană a infrastructurilor de cercetare și inițierea proceselor de actualizare a sistemului. Datorită acestei situații, Infrastructura se confruntă în prezent cu noi provocări—nu numai îmbunătățirea cu succes a serviciilor pe care le oferă, dar și includerea Patrimoniului Cultural Ortodox în rețeaua de cercetare europeană, ceea ce va avea drept efect promovarea și popularizarea istoriei și a culturii din sud-estul Europei.

1. Introducere

Infrastructura electronică de cercetare pentru patrimoniul scris bulgar medieval (*Electronic Research Infrastructure for Bulgarian Medieval Written Heritage*) a fost inițiată în anul 2009 sub forma proiectului de cercetare doctorală și postdoctorală BG051PO001-3.3-04-001 ICT Tools for Historical Linguistic Studies (Instrumente IT Pentru Studii de Lingvistică Istorică), finanțat prin Fondul Social European, Programul Operațional Resurse Umane, fiind continuat cu trei alte proiecte concepute pe aceeași schemă:

- BG051PO001-4.3.04-0004 *E-Medievalia* (*Electronic Resources for Distant Learning in Medieval Studies*) [Resurse electronice pentru învățarea la distanță în studii medievale] (2012–2014)
- BG051PO001-3.3.06-0024 *Informatics, Grammar, Lexicography* [Informatică, gramatică, lexicografie] (2012–2015)
- BG05M2OP001-2-009-0005 *Modern Paleoslavonic and Medieval Studies* [Studii moderne de slavă veche și medievală] (2017–2019)¹

În perioadele în care activitățile noastre nu au fost acoperite din fonduri europene, am accesat alte oportunități de proiecte precum finanțare internă în cadrul University Humanities Complex (Complexul Universitar pentru Științe Umaniste), la care am participat cu proiectul *Digital Medievalia* (2016–2018). De la finele anului 2019, Infrastructura de Cercetare a fost finanțată prin Programul Național de Cercetare „Patrimoniul Cultural, Memorie Națională și Dezvoltare Socială” finanțat de Ministerul Educației și Științei și coordonat de Universitatea din Sofia ca instituție de cercetare principală în domeniul științelor umaniste. Între timp ni s-a alocat o sumă substanțială pentru modernizarea și îmbunătățirea resurselor digitale existente prin intermediul proiectului *Heritage BG* finanțat prin Programul Operațional Știință

* Adresă de corespondență: atotomanova@abv.bg.

¹ Obiectivele și rezultatele proiectelor au fost în permanență raportate și popularizate, vezi [Ganeva \(2018\)](#), [Totomanova \(2012, 2017, 2018\)](#), [Totomanova-Paneva \(2020\)](#).

și Educație pentru Dezvoltare Inteligentă (*nasledstvo.bg*). Astăzi, după doi ani de licitații fără succes, am selectat o companie IT care are misiunea de a îndeplini această sarcină importantă.

În anul 2019 am depus aplicația pentru a deveni parte a Hărții Naționale pentru Infrastructura în Cercetare, iar la sfârșitul anului 2020 resursele noastre digitale au fost oficial acceptate ca infrastructură de cercetare.

Din martie 2018 am fost implicați în două proiecte ESFRI care s-au suprapus, al căror obiectiv a fost să devină Infrastructură Europeană de Cercetare pe domeniul studiilor religioase (ReIRes², RESILIENCE³). În consecință, pînă în prezent, nu am putut conta pe o finanțare constantă din partea statului, fiind nevoiți să ne finanțăm activitatea din diverse tipuri de proiecte și inițiative pentru a supraviețui și pentru a reuși să implementăm obiectivul de a produce un complex de resurse digitale pentru studierea și popularizarea patrimoniului scris bulgar. Resursele noastre sînt incluse pe portalul web *Cyrrillomethodiana*, unde se regăsesc informații detaliate cu privire la toate proiectele noastre legate de Infrastructura de Cercetare. Parteneriatul include trei instituții: Universitatea „St. Kliment Ohridski” din Sofia, Centrul de Cercetări Chirilo-Methodiene al Academiei de Științe din Bulgaria și Biblioteca Centrală a Academiei de Științe din Bulgaria.

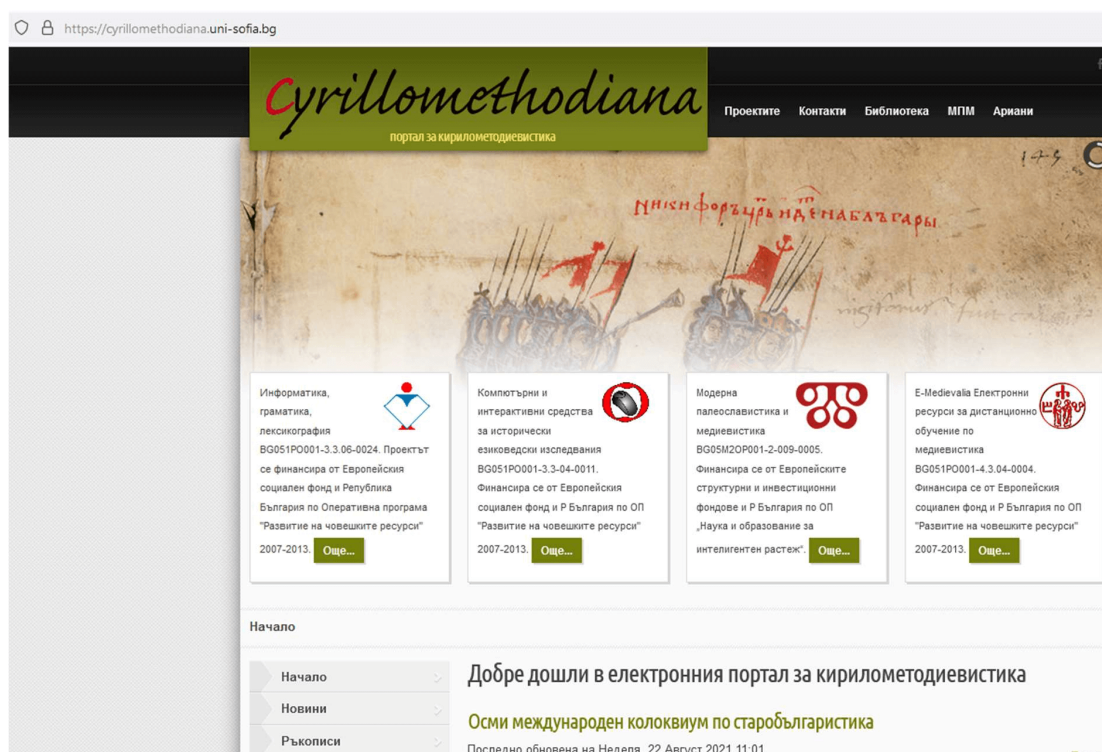


Figura 1: Portalul web *Cyrrillomethodiana*, *cyrrillomethodiana.uni-sofia.bg*

De ce patrimoniul scris?

Patrimoniul scris este sursa istorică cea mai veridică a oricărei națiuni și a oricărei țări și stă la baza identității naționale. Infrastructura noastră de cercetare furnizează resurse și instrumente digitale de cercetare pentru studiul patrimoniului scris bulgar din perioada medievală din perspectiva *lingvisticii diacronice*, a *istoriei literare*, a *istoriei bulgare și europene*, a *teologiei*, a *istoriei culturale* și a *filosofiei*. Bulgara veche (slavona bisericească), creată de sfinții frați Chiril și Methodius, a funcționat în Evul Mediu ca limbă sacră a slavilor ortodocși din sud-estul Europei, precum și în România și Moldova. Acesta este motivul pentru care cercetătorii din domeniul istoriei și culturii lumii medievale slavo-bizantine sînt interesați de utilizarea resurselor digitale disponibile.

²Vezi *reires.eu*.

³Vezi *resilience-ri.eu*.

Și de ce electronic?

- a) Utilizarea tehnologiilor digitale pentru publicarea și procesarea textelor medievale reduce timpul dedicat colectării datelor și producerii de rezultate fiabile. Altfel spus, introducerea tehnologiilor digitale într-un domeniu atât de conservator precum cel al științelor umaniste istorice nu numai că ajută la optimizarea activității cercetătorilor, dar creează și condițiile pentru noi inițiative și proiecte de cercetare.
- b) Instrumentele digitale conferă domeniului științelor umaniste istorice o mai mare notă de atractivitate pentru tinerii născuți în era digitală.

2. Resursele și instrumentele noastre digitale

Așa cum se poate observa, resursele și instrumentele digitale de cercetare pentru procesarea textelor medievale cu origine bulgară demonstrată au fost create într-o perioadă de peste zece ani, printr-o serie de proiecte științifice și educaționale. Sistemul *Histdict* include următoarele resurse și instrumente digitale:

2.1. Corpus diacronic al limbii bulgare, secolele IX–XVIII

Corpusul conține peste 150 de texte de origine bulgară certă, din secolele X–VIII, de diverse genuri, și dispune de propriul software specific, care permite adnotarea textuală și paleografică. Ținând cont de faptul că literatura bulgară a facilitat transmiterea modelului cultural și literar bizantin către alte națiuni ortodoxe din această parte a Europei, Corpusul conține atât texte medievale bulgare originale, cât și traduceri. Textele sînt redactate digital și (cu excepția lucrărilor aparținînd Sf. Kliment Ohridski) sînt reproduse în sistemul ortografic în care s-au păstrat (bulgar, sîrb, rusc). Corpusul include, de asemenea, texte bulgare din perioada modernă timpurie, precum damaskini și alte compilații, dar și unele texte neliterare, precum note ale copiștilor, inscripții și documente juridice. Dacă în 2011 aveam doar 75 de texte, în prezent numărul acestora s-a dublat, și continuăm să încărcăm materiale noi. Unele dintre acestea ne sînt furnizate de colegii noștri din străinătate, care utilizează la rîndul lor acest corpus. Unul dintre cele mai consistente texte încărcate în sistem este așa numitul *Chronograph of the Archive* [Cronograful Arhivei] sau *Jewish Chronograph* [Cronograful Evreiesc], care include cel mai vechi text al Octateuhului și al Cărților Regilor, traduse în Bulgaria în timpul domniei lui Simeon cel Mare (893–927). Fiecare text este precedat de datele arheografice corespunzătoare (titlu, sursă, datare, sistem ortografic, autor etc.). Softul care conține corpusul are nevoie permanent de reactualizări pentru a permite alte tipuri de adnotări: morfologică și de conținut.

Cyrillomethodiana

Текстов корпус

Блахо-в'български грамоти

Заглавие на латински: Юридически
 Жанр: не
 Преведен?: не
 Дата на ръкописа: XV-XVI
 Дата на преписа: -
 Православие: Среднобългарски
 Име на ръкописа: -
 Хранилище на ръкописа: -
 Сигнатура на ръкописа: -
 Издание: Ivan Bogdan - Documente și registre privitoare la relațiile Țării Rumânești cu Braşovul și Ungaria în secolul XV și XVI
 Нормализирани?: не
 Страници: -
 doc_id: doc_193

Mircea I (cel. Bătrân).
 (1386 – 31 Ian. 1418).

I.
 † Възъ жки еъ христ(а) бога възгосудини и христчанени и самодържавни Іо Мирца, великии воевода и господина, швадлацкѣ ан и господствѣцѣ ан еъси зван Ћрговадѣн и запалинскимъ при жи и кѣ татарскимъ стванам и шѣа пол по еъслибъ поддѣванѣ дѣки и до великаго море, и манастижъ воєжѣк и Дръствѣ градъ владѣцѣ, вѣлгоспронземаи господство ан своимъ вѣлгичи проимзовоинни, чистии и свѣтлыми сръдцѣи, и дарѣцѣ сии христѣбѣ господство ан и испълнѣцѣ прошини прѣлгарди Бранишевскимъ, повнѣк прѣсѣжъ господство ан, шѣко да повнѣни и да ѣтерѣдѣи законѣ, шѣо са имали шѣт ѣтерѣдѣитан господство ан за кѣширка, по тѣжрѣвоѣ ж зван господство ан и на пяти Бранишевскомъ доу до Брѣма(а) : шѣко да дѣвалѣ шѣт Ћвѣларъ ирифурѣнѣ, шѣт хѣсовѣнъ пѣр, шѣт кованѣн ѣи дѣкѣ, шѣт чѣхъ ѣ дѣкѣ, а шѣт рѣзанъ вѣларѣ шѣнѣ, и шѣт носѣт шѣнѣкѣ фѣрѣшѣкѣмѣ, шѣнѣрѣ: шѣт вѣловнѣкѣ, ан ѣи дѣкѣ, шѣт вѣно вѣловнѣкѣ ѣ дѣкѣ, шѣт конѣ кон ѣвѣнтѣ, ѣ дѣкѣ, шѣт мѣлѣкѣ вѣлѣкѣ ѣи дѣкѣ, шѣт пѣлѣр, шѣт шѣфѣнѣ шѣт вѣлѣкѣ, шѣт вѣлѣкѣ кѣманѣкѣ, шѣт ѣрѣманѣнѣ, шѣт коби и шѣт кѣшикѣ анѣ, шѣо дѣрѣдѣат шѣт анѣрѣ, шѣт ѣ шѣр. Ћ пѣр: вѣлѣжъ свѣни ѣ дѣкѣ, вѣлѣ ѣ дѣкѣ, вѣлѣ ѣ дѣкѣ, шѣт шѣмѣ дѣкѣ, шѣт кѣнѣкѣ коѣкѣ ѣ дѣкѣ; шѣо вѣлѣдѣт дѣрѣкѣ коѣкѣ с ѣмѣ, да не дѣстъ шѣнѣрѣ: шѣт анѣцъ свѣни ѣ дѣкѣ; конѣ, кон ѣвѣнтѣ анѣо Тѣрѣчѣ, ѣ вѣлѣкѣ; вѣнѣцѣ вѣнѣ; и шѣи кон ѣвѣнѣкѣт с рѣкѣ, шѣт колѣ ѣ рѣкѣ, а вѣрѣцѣ шѣо шѣт вѣнтѣ, да не дѣстъ шѣнѣрѣ: а на Брѣма(а) мѣлѣ ѣ [шѣнѣ]ѣ, а на колѣ ѣ Тѣрѣчѣ ѣ рѣкѣ; ѣ Тѣрѣчѣнѣи тѣкоѣкѣнѣ; конѣ тѣварѣн ѣкѣдѣ сѣ шѣи анѣо, ѣ дѣкѣ; и на Дѣвѣнѣцѣ, ѣ колѣ конѣ конѣ, тѣмѣконѣ и дѣкѣ, и ѣ рѣкѣ, а шѣт кон тѣварѣн ѣ вѣнѣ, а шѣт шѣнѣ кѣши шѣнѣрѣ; и конѣ свѣдовѣн ѣ дѣкѣ, и вѣнѣцѣ вѣнѣ. И шѣи жи конѣ си дѣватъ дѣвѣнтѣ жъ дѣлѣжѣ, да си шѣитъ дѣлѣжѣнѣ наи ѣнѣцѣ, шѣи ѣмѣт, а шѣт правѣа чѣвѣкѣ да са вѣрѣкѣт; а шѣнѣцѣ правѣа чѣвѣкѣ да не вѣнѣкѣт. Шѣнѣ жи кѣто са шѣт поѣкѣнтъ шѣт вѣлѣрѣ господство анѣ, вѣлѣнѣцѣ жи и мѣлѣцѣ, ѣкѣдѣ да ѣстъ кѣши тѣмѣнѣ кѣширка, ѣкѣдѣ да ѣстъ кѣши анѣо мѣнѣстѣа дѣдѣнѣ, тѣ прѣлѣкѣнтъ тѣ не стѣнтъ на шѣо сѣлѣ ѣтерѣдѣи и законѣи на прѣлѣвои законѣ, шѣо тѣковѣнѣ ѣмѣтъ прѣкѣтъ вѣлѣкѣо лѣо и шѣрѣкѣ шѣт господство анѣ. И си ѣмѣт, шѣи дѣнѣдѣ Фѣланѣнѣи и Мѣрѣнѣи и Брѣшъ Дѣлѣжѣнѣмѣнѣ. И си сѣдѣцѣнѣнѣ : Рѣдѣжѣ вѣнѣ, жѣпан ѣвѣдѣнѣшѣ, жѣпан Рѣдѣжѣ, вѣратъ жѣпан(а) Кѣзана, жѣпан Шѣрѣан ѣвѣчѣнѣ, жѣпан Огѣнка Рѣсинѣ, жѣпан Бѣдѣ Бѣлѣнѣшѣе. И азъ, мѣлѣчѣнтъ Бѣлѣдѣнѣнѣ писѣхъ, ѣвѣлѣсѣца ѣвѣстѣ(а) ѣ дѣнѣ, еъ ѣвѣтѣо дѣцѣ, нѣдѣ. Ћ

† Іо Мирца вѣнѣвѣдѣ, мѣнѣстѣа воєжѣкѣ господинѣ.

Figura 2: Corpus diacronic, histdict.uni-sofia.bg

2.2. Dicționar digital de bulgară veche (10 500 de intrări)

Obiectivul nostru principal în cadrul primului proiect a fost să elaborăm un *Dicționar istoric digital al limbii bulgare* și am decis că cea mai ușoară cale pentru a realiza acest obiectiv este să pornim de la *Dicționarul de slavonă*⁴ elaborat și publicat de Institutul de Limbă Bulgară al Academiei de Științe din Bulgaria ca bază pentru dicționarul nostru. Acesta este motivul pentru care am început digitalizarea *Dicționarului de slavonă*, însă în doi ani nu am reușit să introducem decât 10 000 de intrări, restul de 500 fiind digitalizate ulterior, în cadrul celui de-al doilea proiect, deoarece fie nu corespundeau total formatului celorlalte intrări, fie conțineau diverse tipuri de erori și au fost procesate una câte una. Versiunea electronică completă este în prezent disponibilă online și conține lexicul manuscriselor în slavonă din secolele X–XI.

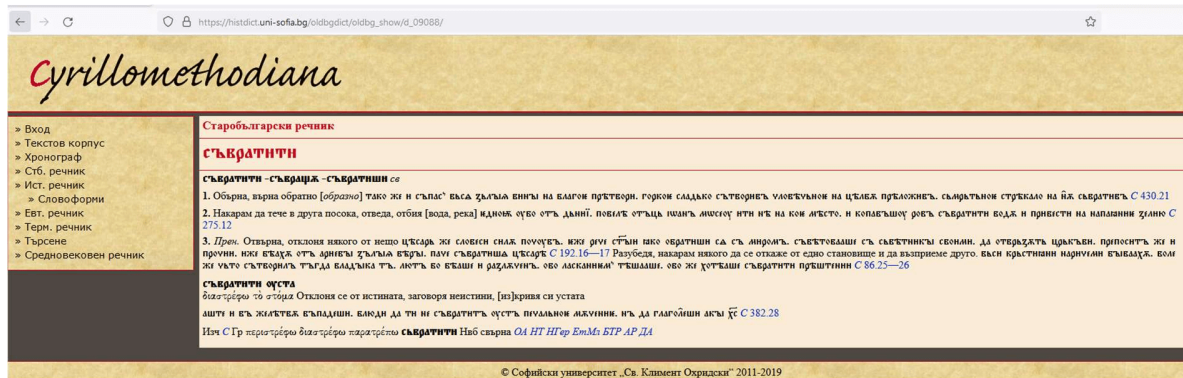


Figura 3: Dicționar online de bulgară veche, histdict.uni-sofia.bg

2.3. Dicționar invers de greacă-bulgară veche

Documentul structurat în format XML care a inclus toate intrările din Dicționarul de bulgară veche a fost utilizat pentru producerea unui dicționar invers de greacă-bulgară veche, care poate fi accesat și online (Fig. 4), fiind intens utilizat de colegii care se ocupă de practicile metafrastice ale cărturarilor care au utilizat bulgara veche. Dicționarul invers a fost un rezultat adiacent al proiectului E-Medievalia care a produs o platformă interactivă de predare-învățare în domeniul Studiilor Medievale. Platforma conține 24 de cursuri interactive pentru studenți și cursanții care frecventează cursuri post-universitare din domeniile filologie, filosofie, teologie, istorie și arte. Printre aceste cursuri se numără cursuri de slavonă, literatură bulgară veche și istoria limbii bulgare, care sînt disponibile și în limba engleză. Platforma conține și o clasă virtuală, pe care am folosit-o pentru workshop-uri și conferințe. La începutul pandemiei, această resursă digitală ne-a permis să inițiem imediat predarea online.

2.4. Dicționar istoric al limbii bulgare (o resursă de tip diacronic în dezvoltare)

Proiectarea software-ului pentru dicționarul istoric s-a dovedit a fi cea mai mare provocare pentru echipa de proiect și în special pentru specialiștii noștri în TIC, deoarece, așa cum s-a menționat mai sus, am decis să edităm și să actualizăm versiunea digitalizată a dicționarului de bulgară veche, completînd-o cu noi intrări și adăugînd o serie de sensuri noi la vechile intrări. Nici lingviștii și nici specialiștii în IT nu au conștientizat cît de dificilă avea să fie această sarcină. Abia la sfîrșitul anului 2014, cu cîteva luni înainte de finalizarea proiectului „Informatică, Gramatică, Lexicografie”, am primit un software care a permis atît editarea vechilor intrări, cît și adăugarea unora noi și astfel am reușit să începem elaborarea Dicționarului istoric al limbii bulgare. Terminologia creștină, care nu a mai fost studiată în mod corespunzător de la publicarea de către F. Miklosich a faimosului său studiu din 1876 (*Miklosich, 1876*), a fost primul grup lexical inclus în Dicționarul istoric. Versiunea în continuă dezvoltare a dicționarului este disponibilă și online (Fig. 5), intrările noi și cele editate fiind redată în culori—verde sau albastru. Dicționarul istoric a fost recent completat cu intrări sau sensuri specifice preluate din dicționare istorice sincrone care conțin lexicul extras de la singur scriitor sau de la un grup de scriitori, sau alt tip de vocabular specializat.

⁴Старобългарски речник. Т. I, 1999. София, Валентин Траянов, 1027 с. Т. II, 2009. Валентин Траянов, 1325 с.

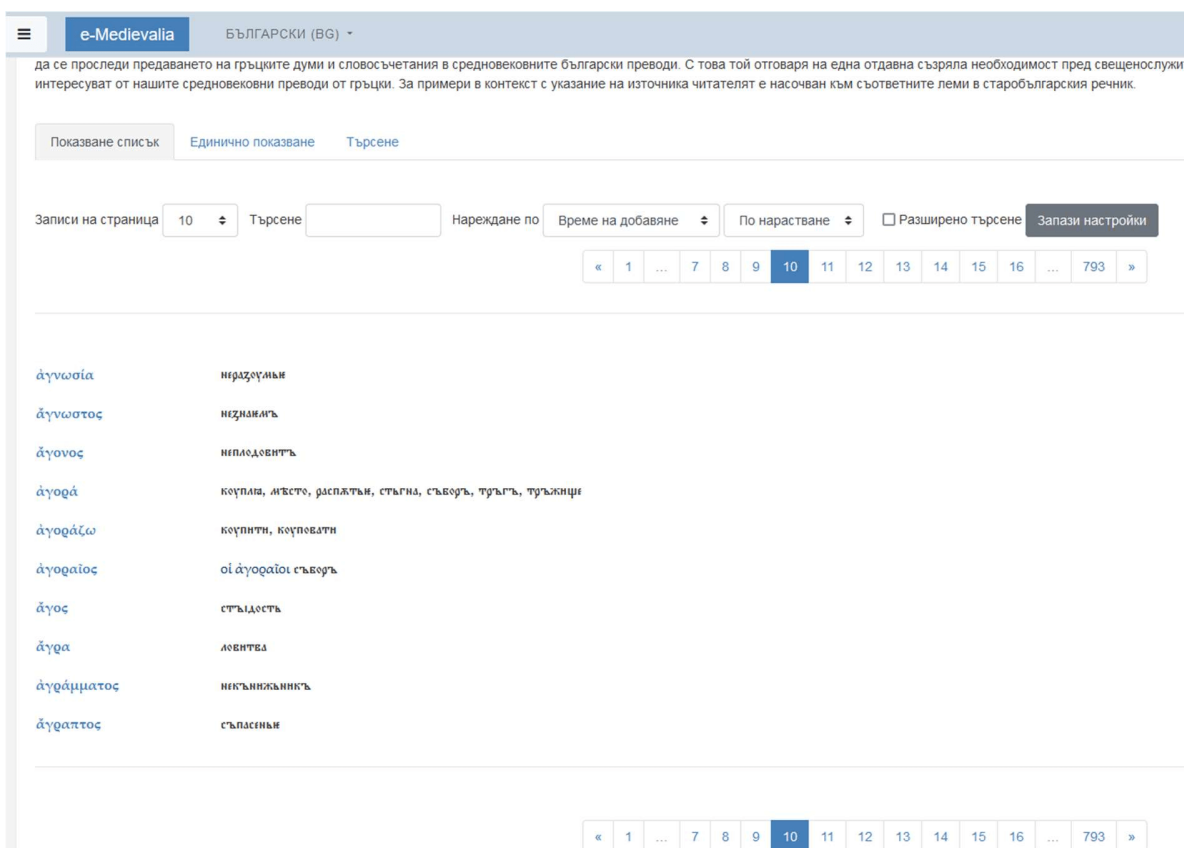


Figura 4: Dicționar online de greacă-bulgară veche: e-medievalia.uni-sofia.bg



Figura 5: Dicționar istoric, histdict.uni-sofia.bg

2.5. Dicționare istorice sincrone

În încercarea de a extinde și de a testa utilizarea de instrumente digitale în cadrul proiectului „Modern Palæoslavonic and Medieval Studies” [Studii moderne de paleoslavă și scrieri medievale], am creat două dicționare de acest tip: *Patriarch Euthymius’ Language dictionary* [Dicționar lingvistic al scrierilor Sfintului Eftimie de Târnovo] și *Terminological dictionary of John the Exarch* [Dicționar terminologic al scrierilor

lui Ioan Exarhul]. Ambele dicționare au și versiuni tipărite⁵, iar partea a doua a *Dicționarului lingvistic al scrierilor Sfântului Eftimie de Târnovo* a fost finanțată prin Programul Național de Cercetare.

Pentru elaborarea intrărilor dicționarului a fost produsă electronic o listă alfabetică a tuturor formelor cuvintelor din scrierile Sfântului Eftimie de Târnovo. Lista indică titlul opusului și locul exact în care apare forma respectivă, permițând lematizarea rapidă.

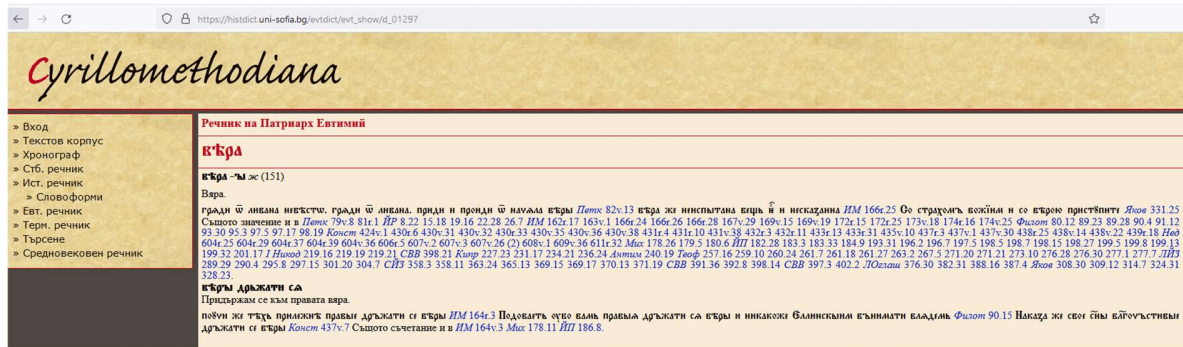


Figura 6: *Dicționar lingvistic al scrierilor Sfântului Eftimie de Târnovo*, hisdict.uni-sofia.bg

O listă similară a fost alcătuită și pentru dicționarul terminologic, însă, dată fiind complexitatea demersului, am decis să limităm materialul lingvistic și să explorăm terminologia din lucrările *Hexaemeron* și *Teologia* ale lui Ioan Exarhul. Rezultatul a fost includerea în acest dicționar specializat a aproximativ 1000 de termeni.

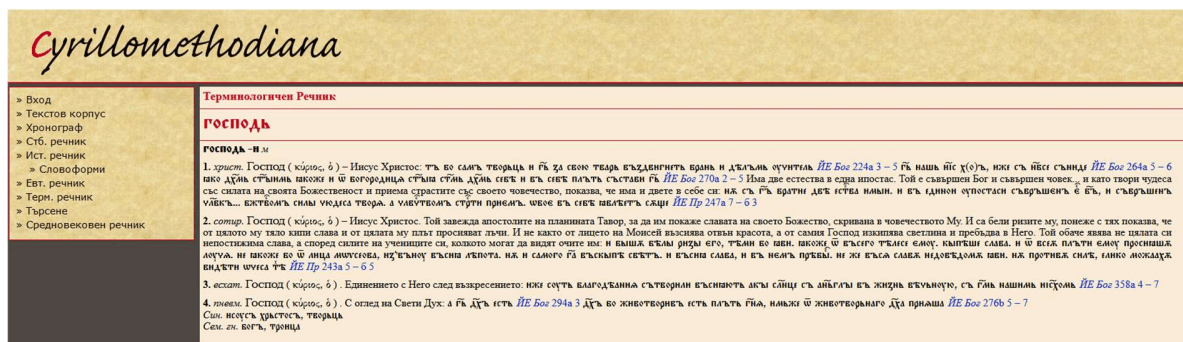


Figura 7: *Dicționar terminologic al scrierilor lui Ioan Exarhul*, hisdict.uni-sofia.bg

Participanții la proiect—în special studenți doctoranzi și postdoctoranzi care nu beneficiau de experiența proiectelor anterioare—au considerat că softul specializat pentru scrierea și editarea intrărilor de dicționar era prea complicat. De aceea am elaborat noile intrări de dicționar în format Word și le-am inclus în dicționarele specializate respective prin intermediul unui convertor. În acest scop, a fost elaborat și analizat un format special pentru intrări. Convertorul utilizat pentru dicționarul lucrărilor Sfântului Eftimie de Târnovo s-a dovedit a fi un instrument foarte util, permițându-ne să convertim 400 de intrări de dicționar în doar câteva secunde. Versiunea online a *Dicționarului lingvistic al scrierilor Sfântului Eftimie de Târnovo* este ușor diferită de versiunea tipărită, deoarece aceasta din urmă include pînă la trei exemple sub fiecare semnificație identificată. Pe de altă parte, dicționarul electronic afișează toate ocurențele lexemelor în text și permite accesarea rapidă a respectivului text în cadrul corpusului, unde utilizatorul poate găsi cuvîntul și poate vizualiza sau copia contextul respectiv. Acest lucru se poate face cu un simplu click.

Am utilizat un convertor similar și pentru *Dicționarul terminologic* și în prezent încercăm să creăm un nou convertor pentru digitalizarea dicționarului invers de greacă-slavonă elaborat de [Christov \(2019\)](#). Dicționarul conține 14 625 intrări în greacă, cărora le corespund 33 307 paralele greacă-slavonă.

⁵Речник на езика на Патриарх Евтимий. Част I. **А–Н**, Част II. **О–Ѧ**. УИ „Св. Климент Охридски“, София, 2019, 2020; Терминологичен речник на Йоан Екзарх. София, 2019.

Intrările de dicționar convertite permit redactarea online prin utilizarea unui soft specializat pentru dicționare istorice. Cu toate acestea, ținând cont de complexitatea acestui soft, intenționăm să producem un convertor și pentru încărcarea în dicționarul istoric, accelerând astfel munca pentru această resursă deosebit de importantă.

2.6. Motor de căutare și tastatură virtuală

Dificultățile în crearea unui motor de căutare adecvat provin din faptul că bulgara veche este o limbă cu flexiune morfologică complicată, în special în paradigma nominală (șase cazuri plus vocativ, trei genuri, trei numere gramaticale, forme simple și compuse ale adjectivului și participii declinabile), care a fost redusă la 2 pînă la 5 forme, în funcție de genul gramatical. Astfel, s-a dovedit că pentru a produce un motor de căutare fiabil aveam nevoie de un tagger (un instrument electronic care permite adnotarea morfologică a cuvintelor supuse inflexiunii). Pentru a produce un asemenea instrument trebuia să creăm un dicționar gramatical al slavonei bisericești, luînd în calcul toate reprezentările posibile pentru o singură formă. Cu toate acestea, aveam nevoie de un instrument de căutare rapidă care să faciliteze munca la dicționarul istoric și la utilizarea *Corpusului*. Soluția temporară a venit de la specialistul nostru în software, care a digitalizat dicționarul de slavonă bisericească și a dezvoltat softul pentru dicționarul istoric. Motorul de căutare este instalat în sistem și se pot efectua căutări pentru toate intrările (dicționare, *Corpus* și *Cronograf*) sau se poate alege una sau mai multe dintre aceste opțiuni. Dispozitivul indică lista de texte în care se află forma pe care o căutăm, iar pentru a o identifica împreună cu contextul respectiv, utilizatorul trebuie să folosească opțiunile de căutare interne din browser. Astfel, una din sarcinile principale ale companiei IT pe care am ales-o va fi să proiecteze și să instaleze un motor de căutare real care să ne permită să efectuăm căutări fără a utiliza opțiunile browser-ului și în funcție de diverși parametri: începutul sau sfîrșitul cuvîntului, grupuri de litere din interiorul cuvîntului, forme gramaticale. Tastatura virtuală, care ar avea nevoie și ea de reglaje fine, este de asemenea, parte a ecuației.

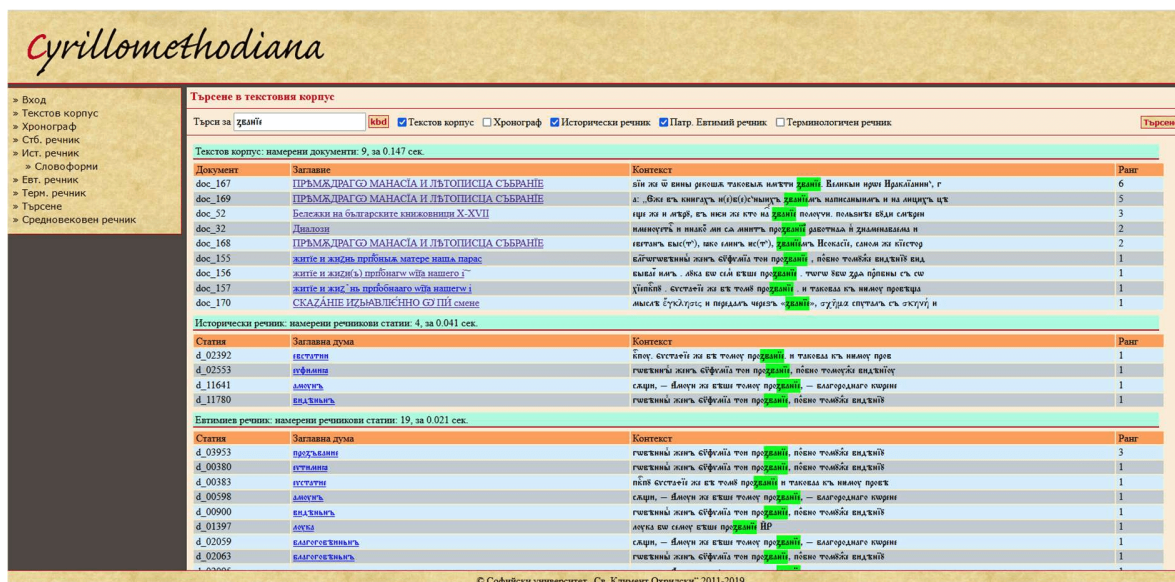


Figura 8: Motor de căutare, histdict.uni-sofia.bg

2.7. Dicționar gramatical și analizator morfologic semi-automat

Munca la dicționarul gramatical a început cu elaborarea etichetelor pentru slavonă, care includ 2200 de tag-uri descriind gramatica complicată a slavonei bisericești (Totomanova et al., 2015). Elaborarea listei de etichete a fost urmată de un dicționar gramatical, care include paradigmele cuvintelor flexibile, luînd în considerare diverse variații fonetice și ortografice (Totomanova et al., online). Paradigmele (regulile) au fost atribuite tuturor cuvintelor flexibile din dicționarul istoric, iar dicționarul final a fost instalat ca etapă finală. Acum, prin tastarea semnelui + localizat în dreptul lemei, utilizatorul poate vizualiza întreaga

paradigmă a cuvîntului respectiv. Plecînd de aici, specialiștii noștri IT au creat un prototip al programului de adnotare morfologică, care este, de asemenea, un instrument care poate fi accesat liber pe site-ul nostru.

Totuși, și aceasta este o soluție temporară, deoarece scopul nostru este să creăm un tagger automat. În acest scop actualizăm dicționarul gramatical și edităm regulile gramaticale. Introducerea în dicționarul gramatical a cuvintelor neflexibile (adverbe, prepoziții, conjuncții, particule, interjecții), precum și a formelor pronominale, s-a dovedit a fi o mare provocare. Includerea formelor de participiu a pus probleme și mai mari, datorită faptului că în slavona bisericească verbele au 5 forme de participiu, dintre care 4 pot fi declinate și au și echivalente determinate. Am dezvoltat reguli pentru generarea automată a paradigmelor participiale în funcție de tipul verbului și pînă în acest moment am produs formele de participiu din primele două conjugări din slavona bisericească (respectiv 894 de verbe cu peste 3 milioane de forme). Actualizarea dicționarului gramatical ne-a ajutat să înțelegem că trebuie să edităm lemele și definițiile gramaticale din Dicționarul istoric, ceea ce reprezintă un proces constant, fiind una din sarcinile pe care trebuie să le îndeplinească firma de IT.

The screenshot shows the web application interface for the Histdict dictionary. The browser address bar displays the URL: https://histdict.uni-sofia.bg/dictionary/resolve_forms. The page title is "Търсене на словоформа" (Search for word form). Below the title, there is a search input field containing "савесе" and a "Покажи всичко" (Show all) button. The main content area is divided into several sections:

- Исторически речник:** A table with columns "Статия", "Заглавна дума", "Словоформа", and "Спрежение". It lists two entries for "савесе" with their respective grammatical details.
- Исторически речник:** A section titled "Всички словоформи:" followed by a list of tags and their corresponding word forms.

Figura 9: Dicționar gramatical & tagger, histdict.uni-sofia.bg

2.8. Fonturi Unicode și convertor pentru bulgara veche

Pentru elaborarea Corpusului și digitalizarea Dicționarului de limbă bulgară veche am avut nevoie nu doar de soft-ul respectiv, ci și de fonturi specializate Unicode (UTF-8) pentru slavonă, care să poată fi recunoscute online. Pentru a evita dubla redactare a textelor deja disponibile în format digital, noile fonturi au fost instalate într-un convertor care a permis conversia acestora în fonturile noi. În prezent utilizăm a treia versiune a convertorului, care include trei fonturi specializate cu design diferit, *CyrillicaBulgarian10U*, *CyrillicaOchrid10U* și *CyrillicaOldStyleU*. Ultimul se folosește la redarea textelor în bulgara modernă timpurie. La început am reușit să convertim doar fonturile din familiile *Cyrillica Bulgarian*, *Cyrillica Ohrid* și *Cyrillica Shafarik*, dezvoltate de compania Synthesis Soft, utilizate la scară largă de comunitatea de cercetători în domeniul limbii paleoslave, la care au fost adăugate treptat alte fonturi pentru slavona bisericească și greaca veche: fontul italian *PopRetkov*, fontul Unicode *BukyVede*, de asemenea bazat pe designul Synthesis Soft, *TimesGreekClassic* și *TimesGreekOld*, dezvoltate de aceeași companie și convertite în prezent în *Palatino*, precum și alte varietăți ale fontului modern *TimesCyrillic* din *Times New Roman*. Fonturile noastre pentru slavonă și convertorul sînt utilizate în toată zona europeană, datorită faptului că sînt compatibile cu programele software editoriale și că sînt utilizate într-o serie de cărți și periodice din Bulgaria și din alte țări.

Sistemul open access Histdict este un complex de resurse și instrumente unice pentru publicarea și

cercetarea textelor slavone medievale care în context european și global corespunde cu *Thesaurus Linguae Graeca* și *Perseus*, reprezentînd patrimoniul scris clasic. În comparație cu acestea, infrastructura noastră de cercetare are un mare avantaj—este bazată pe legătura dintre corpusul diacronic și respectivele dicționare, care sînt elaborate prin utilizarea materialului lexical extras din corpus.

3. Concluzii: planuri de viitor și dezvoltare

Obiectivul principal al infrastructurii de cercetare este de a păstra, dezvolta și actualiza resursele și instrumentele de cercetare electronice prin crearea de noi opțiuni și capacități de cercetare. Avem în vedere următoarele activități:

- Completarea corpusului cu noi texte, cu scopul de a transpune tot patrimoniul scris bulgar în format digital
- Crearea unui corpus paralel cu traduceri în limba bulgară modernă, astfel încît lucrările autorilor care au utilizat slavona să devină accesibile și publicului larg
- Actualizarea continuă a dicționarului istoric
- Crearea de dicționare electronice sincrone pentru textele medievale
- Întreținerea și actualizarea programelor existente
- Cercetare
- Formarea studenților doctoranzi și postdoctoranzi și a tinerilor cercetători
- Publicare: cercetarea și publicarea de texte și dicționare

Pe 30 iunie 2021, Forumul ESFRI a inclus RESILIENCE pe Harta Infrastructurii de Cercetare 2021. Acest lucru înseamnă că RESILIENCE își va ocupa locul în cadrul Infrastructurilor de Cercetare strategice pentru Zona de Cercetare Europeană și că poate lucra la dezvoltarea în continuare a Infrastructurii de Cercetare pentru Studii Religioase. Astfel, situația s-a schimbat acum. Ca parte a RESILIENCE, vom putea extinde aria de potențiali utilizatori ai resurselor noastre digitale și prin includerea platformei de învățare E-Medievalia și a cursurilor acesteia de slavonă. Însă cel mai notabil rezultat al prezenței noastre în cadrul RESILIENCE este faptul că ne deschide noi orizonturi și ne oferă șansa de a include patrimoniul cultural legat de creștinism al Europei de sud-est în rețeaua de cercetare europeană. În același timp, aceasta reprezintă pentru noi și o uriașă provocare. Iar pentru a-i face față cu succes avem nevoie de noi parteneri...

Bibliografie

- Christov, I. (2019). *Гръцко-църковнославянски речник*. Съставен от Иван Христов въз основа на *Речника на църковнославянския език* от архимандрит д-р Атанасий Бончев. Редактор А. Тотоманова, Зографски манастир „Света гора”.
- Ganeva, G. (2018). *Electronic Diachronic Corpus and Dictionaries of Old Bulgarian*, în „Studia Ceranea”, 8, p. 111–119, [Crossref](#).
- Miklosich, Fr. (1876). *Die christliche Terminologie der Slavischen Sprachen: Eine sprachgeschichtliche Untersuchung von Franz Miklosich*. Denkschriften der kaiserlichen Akademie der Wissenschaften. Philosophisch-Historische Klasse. Band 24. Wien.
- Totomanova, A. (2012). *Digital Presentation of Bulgarian Lexical Heritage. Towards an Electronic Dictionary*, în „Studia Ceranea”, 2, p. 221–234, [Crossref](#).
- Totomanova, A. (2017). *Диахронный корпус болгарского языка. Состояние и перспективы*, în „Filologija”, 68, p. 223–242, [Crossref](#).
- Totomanova, A. (2018). *OCS Biblical Language in the Era of the New Technologies*, în *Дигитални и аналитични подходи към писменото наследство. Материали от 7-мата международна конференция ElManuscript „Писменото наследство и информационните технологии”*, Sofia, p. 250–265.
- Totomanova, A., Slavova, T. & Ganeva, G. (2015). *Морфосинтактичен тагсет на старобългарския книжовен език*, în *Информатика, граматика, лексикография BG051-3.3-06-0024/2012. Сборник доклади и материали от заключителната конференция*, Sofia, 29-30.06.2015, p. 17–117.
- Totomanova, A., Slavova, T. & Ganeva, G. (online). *Граматически речник на старобългарския език*, [[online](#)].
- Totomanova-Paneva, M. (2020). *Дигитални ресурси за подготовка на докторанти, постдокторанти и млади учени*, în *Девети международен есенен научно-образователен форум. Съвременният учител и предизвикателствата на информационното общество*, Sofia, p. 91–94.