

PRELUCRAREA DIGITALĂ A TEXTELOR DIALECTALE PUBLICATE LA CHIȘINĂU ÎN ANII 1969–1987

CZU: 004.912:81'28:801.82(478-25)"1969-1987"
DOI: <https://doi.org/10.52673/18570461.21.4-63.05>

Doctor în filologie **Liliana POPOVSCHI**¹

E-mail: lilianapopovschi@mail.ru

ORCID: <https://orcid.org/0000-0002-7032-142X>

Cercetător științific **Ludmila MALAHOV**²

E-mail: lmalahov@gmail.com

ORCID: <https://orcid.org/0000-0001-9846-0299>

Vlada COLESNICOVA³

E-mail: colesnicova70@gmail.com

ORCID: <https://orcid.org/0000-0001-9916-212X>

¹Institutul de Filologie Română „Bogdan Petriceicu-Hasdeu”

²Institutul de Matematică și Informatică „Vladimir Andrunachievici”

Academia de Studii Economice din Moldova³

DIGITAL PROCESSING OF DIALECTAL TEXTS PUBLISHED IN CHISINAU IN THE YEARS 1969–1987

Summary. The dialectal texts published in Chisinau in the years '60–'80 of the past century are of particular importance for research in different fields of science: linguistics (especially dialectology, the history of the Romanian language, stylistics), history, ethnography, folklore studies, sociology, psychology, ethnolinguistics, sociolinguistics, psycholinguistics. Revitalizing them through digital processing and transliteration from the Cyrillic script into the Latin script will allow their full exploration. Currently, from the six volumes of the collection *Dialectal Texts: Supplement to the Moldavian Linguistic Atlas*, the first volume has been scanned, recognized, and transliterated resulting in the Latin script text. A set of the accompanying items (recognition templates, the user alphabet, and the spelling dictionary) have been prepared to automate the process. Based on the established transliteration rules, a converter was developed to convert the recognized text of the Romanian local speech samples from the Cyrillic-based phonetic transcription into the Latin-based one.

Keywords: dialect texts, digital processing, transliteration, Cyrillic script, Latin script.

Rezumat. Textele dialectale, publicate la Chișinău în anii '60–'80 ai secolului trecut, prezintă o importanță deosebită pentru cercetările din diferite domenii ale științei: lingvistică (în special dialectologie, istoria limbii, stilistică), istorie, etnografie, folclorică, sociologie, psihologie, etnolingvistică, sociolingvistică, psiholingvistică. Actualizarea prin prelucrare digitală și transliterare din alfabet chirilic în alfabet latin va permite valorificarea lor pleneră. În prezent, din cele șase volume ale colecției *Texte dialectale. Supliment la Atlasul lingvistic moldovenesc*, primul volum a fost scanat, recunoscut și transliterat, rezultând astfel un text în grafie latină. Un set de elemente de însoțire (șabloane de recunoaștere, alfabetul utilizatorului și dicționarul) a fost pregătit pentru a automatiza procesul. Pe baza regulilor de transliterație stabilite, a fost dezvoltat un convertor pentru a converti textul recunoscut al mostrelor de grai românesc din transcrierea fonetică bazată pe grafia chirilică în cea bazată pe grafia latină.

Cuvinte-cheie: texte dialectale, prelucrare digitală, transliterare, alfabet chirilic, alfabet latin.

INTRODUCERE

Studierea limbii române vorbite la est de Prut sub aspectul variației diatopice a reprezentat o preocupare continuă a cercetătorilor Institutului de Filologie de la Chișinău încă de la înființarea lui în 1946 (ca Institut de Istorie, limbă și Literatură), când au fost inițiate primele anchete de teren [1, pp. 78-79]. După constituirea Sectorului de dialectologie și fonetică experimentală ca subdiviziune a Institutului, în 1956, membrii acestuia și-au fixat drept sarcină cercetarea graiurilor

moldovenești de pe întreg teritoriul fostei Uniuni Sovietice, obiectivul major fiind elaborarea unui atlas lingvistic care să cuprindă material comparabil cu cel din atlasele lingvistice publicate până la acea dată și să permită rezolvarea unor probleme importante de lingvistică romanică și generală.

În baza unui plan bine pus la punct și a unor principii metodologice minuțios elaborate, în perioada 1957–1965 a fost colectat material dialectal din 240 de localități din Basarabia, Transnistria, nordul Maramureșului (regiunea Transcarpatia) și al Bucovinei,

ținutul Herța, sud-estul Ucrainei, Caucaz (Abhazia și ținutul Krasnodar din Federația Rusă) și partea asiatică a ex-URSS (Kazahstan, Kârgâzstan și ținutul Primorie din Extremul Orient Rus). Ținând cont de experiența dialectologilor din România, care au dotat *Atlasul lingvistic român* (ALR) cu un supliment de texte dialectale [2] merit să permită, în mai mare măsură decât volumele atlasului, studierea fenomenelor de sintaxă, fonetică sintactică și stilistică, dialectologii chișinău-ieni au decis ca în timpul anchetei să culegă și texte. Răspunsurile obținute la întrebările din chestionarul prestabilit au servit la elaborarea celor 8 volume de atlas, dintre care 4 cu titlul *Atlasul lingvistic moldovenesc* (ALM) [3] și altele 4 cu titlul *Atlasul lingvistic român pe regiuni. Basarabia, nordul Bucovinei, Transnistria* (ALRR–Bas.), o continuare directă a ALM [4], iar aproximativ jumătate din textele înregistrate pe bandă de magnetofon au fost publicate între anii 1969–1987 în 6 volume, care poartă subtitlul *Supliment la Atlasul lingvistic moldovenesc* [5].

Textele constituie o completare prețioasă a datelor din atlasele lingvistice, deoarece „permit o mai eficiență apropiere de mecanismele vorbirii, esențială pentru studiile de limbă vorbită, abordate de lingvistica modernă” [6, p. X]. Oferind posibilitatea examinării multor fenomene fonetice, lexico-semantice, morfologice și sintactice în contexte autentice, firești de manifestare, ele sunt deosebit de importante pentru lingvistica generală, pentru lingvistica românească și cea romanică, în mod special pentru dialectologia românească, sincronică și diacronică, pentru istoria limbii, stilistică, pragmatică etc. Deși reprezintă, întâi de toate, documente de limbă română, textele dialectale furnizează totodată informații prețioase privind cultura, mentalitatea și psihologia populară, organizarea socială și evoluția comunităților din diferite zone geografice, fiind utile și pentru cercetările de etnografie, folclor, istorie, sociologie, psihologie, pentru cele de etnolingvistică, sociolingvistică, psiholingvistică ce au caracter interdisciplinar. Ele pot servi, de asemenea, drept material didactic necesar pentru ilustrarea noțiunilor teoretice studiate la diferite cursuri universitare.

Având în vedere valoarea incontestabilă a textelor dialectale publicate la Chișinău în anii '60–'80 ai secolului trecut, s-a decis readucerea lor în actualitate prin intermediul proiectului *Valorificarea științifică a patrimoniului lingvistic național în contextul integrării europene*, demarat la Institutul de Filologie Română „B. P.-Hasdeu” în 2020. Revitalizarea textelor dialectale ca parte a moștenirii culturale naționale presupune utilizarea noilor tehnologii, ceea ce a determinat cooperarea în proiect a cercetătorilor de la Institutul de Matematică și Informatică „Vladimir Andrunachievici”.

În contextul politicii europene privind conservarea digitală a patrimoniului cultural¹, prelucrarea digitală a textelor dialectale în vederea conservării, protejării și valorificării devine o prioritate. Dincolo de acest obiectiv strategic, mijloacele informatice oferă un șir de avantaje concrete.

AVANTAJE ALE PROCESĂRII DIGITALE A TEXTELOR DIALECTALE

Una dintre cele mai dificile probleme legate de reeditarea volumelor respective este cea a transliterării materialului pe care îl includ. Se știe că textele dialectale publicate la Chișinău în perioada sovietică sunt redactate cu ajutorul unui sistem de transcriere fonetică (utilizat și pentru elaborarea ALM) [7, pp. 15-17], ce are la bază grafia chirilică, oficială la acea vreme, cuprinzând, de asemenea, câteva litere latine (j, i, ŋ, h), o literă grecească (γ) și semne diacritice. Acest fapt a limitat accesul publicului la o sursă valoroasă de limbă română. Date fiind noile realități, datorate revenirii, în Republica Moldova, la grafia latină și recunoașterii unității limbii vorbite pe ambele maluri ale Prutului, se impune transliterarea acestor texte în alfabet latin, mai exact, transpunerea lor din vechiul sistem de transcriere în cel bazat pe alfabetul latin, folosit în ALR și în atlasele lingvistice regionale românești, în volumele de texte dialectale editate de dialectologii români, precum și în ALRR–Bas. Acesta este sistemul optim de transcriere a limbii române vorbite, deoarece a fost conceput pentru a reda varietatea fonetică a tuturor dialectelor și subdialectelor ei și permite notarea tuturor nuanțelor de rostire, fiind alcătuit din 94 de litere din alfabetul latin, 6 litere din alfabetul grecesc, 25 de semne diacritice și 27 de alte semne [8, p. 51]. Unele dintre aceste nuanțe nu sunt specificate în transcripția fonetică a ALM, în același timp inventarul acestui sistem include o serie de alografe consonantice cu valoare redundantă, care nici nu au fost folosite în practică [8, pp. 55-56].

Prin conjugarea eforturilor lingviștilor și informaticienilor poate fi realizată automatizarea procesului de transliterare, acțiune ce necesită, pe de o parte, stabilirea cu precizie a corespondenței dintre cele două sisteme de notație fonetică, pe de altă parte, efectuarea unor operațiuni tehnice.

Prelucrarea digitală a textelor va asigura o calitate mult mai bună a versiunii imprimată. Menționăm, în acest context, că prima ediție a fost tipărită în condiții tehnice precare, la rotaprint.

Pe viitor ar putea fi creat un corpus electronic de texte dialectale sau texte dialectale care ar putea fi incluse în corpusuri mai ample de texte de diferite tipuri

¹ https://europa.eu/european-union/topics/culture_ro

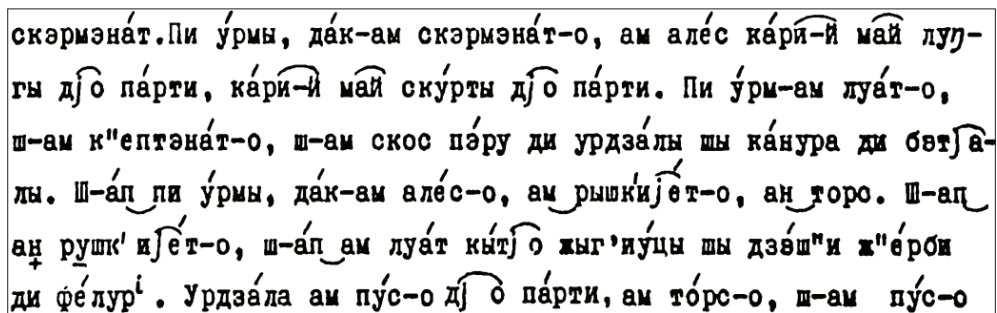


Figura 1. Imaginea paginii originale (fragment).

în limba română, ceea ce le-ar face și mai accesibile și mai ușor de consultat.

Actualizarea textelor dialectale prin prelucrare digitală și transliterare în alfabet latin va permite valorificarea lor pleneră. Noua versiune le va introduce în circuitul științific general românesc și internațional și le va pune la dispoziția unui public cât se poate de numeros, transformându-le într-un instrument de lucru util pentru cercetători, cadre didactice, studenți, masteranzi, doctoranzi și într-o importantă sursă de informații pentru toți cei interesați de limba română, de cultura și civilizația românească.

PRELUCRAREA DIGITALĂ A TEXTELOR DIALECTALE TRANSCRISE CU CARACTERE CHIRILICE

Instrumentarul pentru procesarea digitală a textelor dialectale este utilizat pentru automatizarea proceselor de conversie în format electronic și transliterare.

Conversia unui text în format electronic constă din mai multe etape: scanarea, post-procesarea resurselor scanate, pregătirea pentru OCR², OCR (recunoașterea), verificarea automată și manuală [9]. Aceste etape de prelucrare sunt descrise mai jos.

După aceasta, s-a făcut conversia (sau transliterarea) în transcrierea latinizată pe bază de IPA³, respectiv varianta românească a transcrierii. Transliterarea este descrisă în secțiunea următoare.

În prima etapă, au fost efectuate scanări la o rezoluție de până la 600 dpi. În etapa de post-procesare a fișierelor obținute s-a folosit programul Scan Tailor care efectuează operațiuni precum separarea paginilor, alinierea, adăugarea/eliminarea paginilor, conversia binară, curățarea și alte funcții⁴. Un fragment din imaginea paginii pregătite pentru recunoaștere este prezentat în figura 1.

²OCR – Optical Character Recognition (engl.), recunoașterea optică a caracterelor.

³IPA – International Phonetic Alphabet (engl.), alfabetul fonetic internațional.

⁴<https://scantailor.org/>

Recunoașterea este efectuată de programul ABBYY FineReader⁵. Inițial, este setat alfabetul utilizatorului, adică o listă completă a tuturor caracterelor care pot fi incluse în textul recunoscut.

Programul ar trebui să funcționeze de ceva timp în modul antrenament. În acest mod, sunt create șabloane pentru recunoaștere. Programul evidențiază caracter cu caracter din imaginea de text. Utilizatorul poate corecta manual limitele caracterului selectat și rezultatul recunoașterii oferit de program. Șabloanele create pot fi memorate și încărcate ulterior în program. Se poate completa setul de șabloane încărcându-l și continuând în modul antrenament.

Pe lângă alfabet și șabloane, se poate defini, de asemenea, un dicționar. Acest lucru nu este obligatoriu, dar prezența unui dicționar îmbunătățește calitatea recunoașterii. Dicționarul este utilizat pentru a clarifica cuvintele în cazul unei recunoașteri incerte și pentru a elimina cratimele la despărțirea cuvintelor. Dicționarul este format separat ca o listă de cuvinte în cod UTF-8, câte un cuvânt pe linie.

Textul nostru sursă redă o înregistrare fonetică a vorbirii dialectale și de aceea inițial nu există niciun dicționar. Dicționarul trebuie să fie completat cu cuvinte noi după ce a fost recunoscută o parte a textului. Recunoașterea unui fragment de text trebuie deseori să fie repetată după completarea dicționarului.

Un alt factor care complică utilizarea dicționarului este marea variabilitate a limbii vorbite dialectale. Chiar și același vorbitor nativ poate pronunța același cuvânt în mod diferit, ceea ce atrage variații de text.

Dicționarul a fost creat și este în curs de completare. O comparație a rezultatelor recunoașterii înainte și după adăugarea dicționarului confirmă eficiența corectorului automat, încorporat în programul FineReader. Doar 3-5 % din erorile de recunoaștere sunt lăsate pentru corecție manuală. Până în prezent a fost finalizată procesarea primului volum al cărții [5].

În figura 2 este prezentat un fragment de text recunoscut.

⁵<https://help.abbyy.com/en-us/finereader/14/>

скэрмэна'т. Пи у'рмы, да'к-ам скэрмэна'т-о, ам але'с ка'ри-й май луҥ-гы дјо парти, ка'ри-й кай ску'рты дј о па'рти. Пи у'рм-ам луа'т-о, ш-ам к"ептэна'т-о, ш-ам скос пэ'ру ди урдза'лы шы ка'нура ди бэтја-л'ы. Ш-ап пи урмы, дак-ам алес-о, ам_рышкијет-о, ан_торс. Ш-ац_аң рушк' ијет-о, ш-ап_ам луа'т кытјо жыг'иу'цы шы дзэ'ш"и ж"е'рби ди фе'лур'. Урдза'ла ам пус-о дј о па'рти, ам торс-о, ш-ам пус-о

Figura 2. Text după recunoaștere și editare.

CONVERTOR DE TRANSCRIERE FONETICĂ CU CARACTERE CHIRILICE

După recunoaștere, se efectuează o transliterare din alfabetul fonetic pe baza grafiei chirilice în cel pe baza grafiei latine. Alfabetul chirilic are aproximativ 80 de litere cu și fără diacritice, 16 diacritice, precum și ligaturi și separatori. Astfel, după recunoaștere, rezultă aproximativ 100 de caractere pentru transliterare. Poziția caracterelor într-un șir de caractere (în linie sau în suprapunct) joacă, de asemenea, un rol important. Varianta latină este configurată în mod similar. Tabelul 1 prezintă unele dintre cele 273 de reguli de transliterare elaborate și exemple ale ambelor variante de notație fonetică.

Programul de transliterare pune în aplicare aceste reguli. Programul este scris în limbajul Python, folosind un set bogat de module gata făcute. Programul este pornit din linia de comandă. Fișierele chirilice originale sunt produse de FineReader în format DOCX. Fișierele transliterate sunt în același format. Fișierele în acest format pot fi vizualizate și editate utilizând Microsoft Word, Libre Office Writer și alte programe.

Parametrii de apelare a programului sunt numele fișierelor transliterate. De asemenea, se poate seta numele și tipul fontului pentru rezultat, specifica un sufix pentru a genera numele fișierelor rezultante, seta moduri de depanare și solicita o sugestie. Pentru fiecare fișier transliterat programul afișează un mesaj cu numele fișierului sursă și al rezultatului, precum și numărul de paragrafe din text.

Programul de transliterare a trebuit să ia în considerare o serie de caracteristici, care fac mai complicată procesarea textului recunoscut.

De exemplu, atât softul OCR (FineReader), cât și persoana care verifică textul după recunoaștere înlocuiesc uneori caracterele chirilice (din setul „БКМ-Тасеорх”) cu caractere latine identice. Programul de transliterare înlocuiește toate caracterele de acest fel cu alfabetul chirilic în trecere suplimentară prin text înainte de a se efectua transliterarea. Alfabetul latin în textul fonetic chirilic este considerat o eroare care trebuie corectată.

Un fenomen mai complex este reprezentat de ambiguitatea generării caracterelor atunci când sunt prezente mai multe semne diacritice. De exemplu, un caracter cu două semne diacritice poate fi generat în trei moduri (figura 3). Standardul internațional Unicode prevede o „descompunere canonică” pentru astfel de cazuri, adică o ordine strict definită de adăugare a diacriticelor (italicize în figura 3).

$$\hat{\mathbf{a}} = \mathbf{a} + \hat{\ } + _ = \mathbf{a} + _ + \hat{\ } = \hat{\mathbf{a}} + _$$

Cod: 61 302 320 61 320 302 E2 320

Figura 3. Opțiuni de descompunere pentru un simbol compozit.

Caracterele din textul sursă sunt supuse descompunerii canonice. Numai varianta prescrisă de standard este introdusă în tabelul de transliterare. După transliterarea în alfabetul latin, se efectuează operația inversă („compoziție canonică”).

Tabelul 1

Exemplu de reguli de transliterare în alfabet latin

Regula de transliterare			Exemplu de text
Alfabet chirilic		Alfabet latin	
ý	→	ú	пу́ни → púni
ŷ	→	ц	брыŷ → brīц
ʸоá	→	ʸоá	сʸоáрили → sʸoáriili
уы	→	цâ	доуы → douâ
ʸ + vocală	→	ʸ + vocală	ʸограды → ʸográdâ

```

('y', 'u'),
('y\u0306', 'u\u032F'),
('y', [3, 'U', 'U\u032F']),
('yoa', 'U\u032Fo\u032Fa'),
('yya', 'U\u032Fu\u032Fa'),
('ы', 'u\u032Fâ'),
('э', 'ă'),
('э\u0301y', 'ă\u032F'),
('ы', [4, 'î', 'â']),

```

Figura 4. Tabel de transliterare din alfabet chirilic în alfabet latin (fragment).

Descompunerea canonică se realizează într-o trecere separată de text.

Există și alte probleme similare. Astfel, înainte de a înlocui echivalentele chirilice cu cele latine din tabel (transliterare), programul efectuează șapte treceri suplimentare de text:

- înlocuirea caracterelor care nu se află în transcrierea fonetică;
- înlocuirea caracterelor din fonturile nestandardizate folosite ocazional în transcrierea cu caractere chirilice;
- înlocuirea caracterelor latine, descompunerea canonică și conversia textului în minuscule;
- înlocuirea superscriptului cu majuscule pentru a facilita procesarea ulterioară;
- înlocuirea apostrofului după o vocală cu accentul ascuțit;
- înlocuirea, în funcție de context, a combinațiilor care conțin caractere nealfabetice;
- eliminarea spațiilor înainte de diacritice.

Figura 4 prezintă un fragment din tabelul de transliterare (limbajul Python). Opțiunile sunt afișate între paranteze drepte. De exemplu, litera *ы* este convertită fie în *î*, fie în *â*, folosind a patra variantă a condițiilor contextuale. Diacriticele sunt prezentate ca puncte de cod Unicode. De exemplu, `\u0301y` este accentul ascuțit.

Un fragment de pagină după transliterare este prezentat în figura 5.

CONCLUZII

Textele dialectale publicate la Chișinău în anii '60-'80 ai secolului trecut sunt de o valoare incontestabilă, iar în contextul politicilor privind conservarea și valorificarea patrimoniului cultural revitalizarea lor cu ajutorul noilor tehnologii devine o necesitate stringentă.

Colaborarea dintre lingviști și specialiștii în informatică a făcut posibilă prelucrarea digitală a textelor dialectale transcrise cu caractere chirilice. Automatiz-

```

scărmănat// pi urmă/ dăc-am scărmănat-o/ am alés cări-î mai lun-
gâ dîo parti/ cări-î caî scurtâ dî o pârti// pi urmă-am luât-o/
ș-am keptănăt-o/ ș-am scos păru di urdălâ și cânura di băția-
l'â// ș-ap pi urmă/ dac-am ales-o/ am_rîșkijet-o/ an_tors// ș-aț_
aņ rușk ijiet-o/ ș-ap_am luât cîțio jîgiútâ și dăși zérbi
di féluri// urdála am pus-o dî o pârti/ am tors-o/ ș-am pus-o

```

Figura 5. Fragment de text în transcriere fonetică după transliterare.

zarea procesului de transliterare a acestora în alfabet latin a necesitat, pe de o parte, stabilirea exactă a corespondenței dintre cele două sisteme de transcriere fonetică, pe de altă parte, elaborarea unui program care efectuează operațiuni tehnice specificate pentru aceasta.

Procesarea digitală a textului va oferi o calitate mult mai bună a versiunii imprimate, având în vedere că prima ediție a fost tipărită într-o specificație nesigură pe rotaprint.

Din cele 6 volume ale culegerii *Texte dialectale. Supliment la Atlasul lingvistic moldovenesc*, primul volum a fost recunoscut, a fost pregătit un set de șabloane de recunoaștere, un alfabet fonetic personalizat și un dicționar pentru a automatiza verificarea. Pe baza regulilor de transliterare stabilite, a fost dezvoltat un convertor pentru a reprezenta textul recunoscut al mostrelor de grai românesc, redată în transcrierea fonetică bazată pe alfabetul chirilic. Urmează a fi recunoscute și transliterate și celelalte volume de texte dialectale.

În viitor, ar fi posibil să se creeze un corpus electronic de texte dialectale sau să se includă textele dialectale într-un corpus mai mare de texte de diverse tipuri în limba română, ceea ce le-ar face și mai accesibile și mai ușor de vizualizat.

Actualizate prin prelucrare digitală și transliterare în alfabet latin, textele dialectale editate de dialectologii chișinăuieni vor intra în circuitul științific general românesc și internațional, devenind o importantă sursă documentară pentru cercetătorii din diferite domenii ale științei (lingvistică, istorie, etnografie, folclorică, sociologie, psihologie etc.), pentru toți cei interesați de limba română, de cultura și civilizația românească.

BIBLIOGRAFIE

1. Pavel V. Arhiva dialectologică: scurt istoric. În: *Philologia*, 2020, nr. 1-2, pp. 78-82.
2. Petrovici E. *Texte dialectale. Supliment la Atlasul lingvistic român II (ALRT II)*. Sibiu-Leipzig, 1943. 370 p.

3. Atlasul lingvistic moldovenesc, volumul I, partea 1, Fonetica, de Rubin Udler; partea 2, Fonetica, de Rubin Udler; Morfologia, de Vasile Melnic, Chișinău, 1968, hărțile 1-520; volumul II, partea 1, Lexicul, de Victor Comarnițchi, Chișinău, 1972, hărțile 521-721; partea 2, Lexicul, de Vasile Melnic și Vasile Pavel, Chișinău, 1973, hărțile 722-1034.

4. Atlasul lingvistic român pe regiuni. Basarabia, nordul Bucovinei, Transnistria, redactor coordonator Vasile Pavel, volumul I, Chișinău, 1993, 240 p.; volumul II, Chișinău, 1998, 365 p.; volumul III, Chișinău, 2002, 352 p.; volumul IV, Chișinău, 2003, 342 p.

5. Texte dialectale. Supliment la Atlasul lingvistic moldovenesc (ALM), volumul I, partea 1, culese de V. Melnic, V. Stati, R. Udler, Chișinău, 1969, 243 p.; volumul I, partea 2, culese de V. Melnic, V. Stati, R. Udler, Chișinău, 1971, 214 p.; volumul II, partea 1, publicate de A. Dumbrăveanu, E. Constantinovici, sub redacția lui R. Udler, Chișinău, 1971, 302 p.; volumul II, partea 2, publicate de A. Dumbrăveanu, sub redacția lui V. Stati, Chișinău, 1981, 307 p.; volumul III, partea 1, publicate de V. Corcimari, sub redacția lui R. Udler, Chișinău, 1981, 247 p.; volumul III, partea 2, publicate de A. Dumbrăveanu, sub redacția lui R. Udler, Chișinău, 1987, 194 p.

6. Marin M, Mărgărit Iu., Neagoe V., Pavel V. Graiuri românești din Basarabia, Transnistria, nordul Bucovinei și nordul Maramureșului. Texte dialectale și glosar. București, 2000. 531 p.

7. Atlasul lingvistic moldovenesc. Articole introductive, anexe, volumul I, partea I, redactat de R. Udler și V. Comarnițchi, cu participarea lui V. Melnic și V. Pavel. Chișinău, 1968. 176 p.

8. Pavel V., Zagaevschi V. Unitatea limbii române și sistemul unic de transcriere fonetică. În: Revistă de lingvistică și știință literară, 1990, nr. 5, pp. 49-58.

9. Cojocar S., Ciubotaru C., Colesnicov A., Malahov L., Bumbu T. Instrumentar pentru digitizarea și transliterarea textelor tipărite în limba română cu caractere chirilice. În: Revista Bibliotecii Academiei Române, nr. 2, 2017, pp. 27-38.

NOTĂ. Articolul a fost realizat în cadrul Proiectului de cercetare 20.80009.1606.01 *Valorificarea științifică a patrimoniului lingvistic național în contextul integrării europene*, Institutul de Filologie Română „B. P.-Hasdeu” al MEC.



Iraida Ciobanu. *Natură statică cu flori*, 2013, u. p., 70 × 70 cm.