Cătălina Mărănduc

PERSPECTIVE NOI DESCHISE CERCETĂRII LINGVISTICE DE INFORMATIZAREA BIBLIOGRAFIEI *DICȚIONARULUI ACADEMIC AL LIMBII ROMÂNE* (DLR)

La aproape o sută de ani de la începerea redactării marilor dicționare academice ale limbii române, înainte de a începe o reluare a primelor litere în formă actualizată, se impune un efort de modernizare a metodelor de lucru. Una dintre direcțiile de sincronizare a cercetării românești cu lexicografia internațională este informatizarea căutării de ocurențe şi prime atestări ale cuvintelor şi a extragerii de citate din sursele bibliografice.

Din păcate, datorită fondurilor insuficiente care se pot investi în această vastă întreprindere, deşi s-a început în urmă cu opt ani introducerea în calculator a textelor din bibliografia dicționarului, la ora actuală nu avem în corpusul de texte în format electronic decât aproximativ 10 procente din vasta bibliografie, cam 90 de titluri.

S-a început acum la Iaşi scanarea cărților şi periodicelor din bibliografie. O etapă ulterioară ar fi transpunerea din format fotocopiat în format text, proces numit ocerizare după numele programului utilizat, OCR (Optical Character Recognition). Se obțin texte care pot fi prelucrate, atât manual cât şi de către calculator. Această transpunere se poate însă aplica numai textelor mai recente, în alfabet latin, şi chiar şi atunci comportă numeroase erori care se corectează prin confruntarea textului ocerizat cu textul tipărit.

Echipa de la Bucureşti s-a axat în special pe transpunerea manuală în format electronic a celor mai vechi texte tipărite, scrise cu alfabet chirilic sau de tranziție (mixt), pentru care nu dispunem de programe adecvate de interpretare (OCR), combinația de litere fiind diferită de la carte la carte. Şi aceste texte, introduse manual, conțin diferite greşeli de interpretare a literei sau de dactilografie şi necesită un complicat proces de corectură (proces de rutină, care nu se poate face decât într-un ritm foarte lent).

Totuşi, investirea de timp în corectarea textelor, fie ocerizate, fie dactilografiate, se dovedeşte necesară; calitatea (acuratețea) şi conformitatea (dintre textul electronic şi textul din ediția de bază) fiind esențială pentru utilizarea lor. Actualmente, o parte din texte nefiind colaționate, citatele pe care le găsim cu ajutorul calculatorului în corpusul de texte trebuie supuse unei confruntări cu cartea, care adeseori ne duce la concluzia că sunt inutilizabile şi că nu am făcut economie, ci risipă de timp şi de efort.

LR, LVII, nr. 2, p. 157–166, Bucureşti, 2008

BDD-A631 © 2008 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-16 16:54:44 UTC)

De pildă, în *Anonimul brâncovenesc*, din *Cronicarii munteni*, vol. II, format electronic, am găsit două atestări ale cuvântului *domenii*, dar, după ce le-am găsit în volum cu destulă dificultate, am constatat că era vorba de două greşeli de dactilografie, în ambele cazuri fiind vorba despre cuvântul *domnii.* La fel, căutând atestări pentru verbul *a dezice,* am găsit una în *Cazania* lui Varlaam, de la 1642, în format electronic, dar confruntând cu textul cărții, am constatat că era vorba de o dactilografiere eronată a cuvântului *dzice,* variantă arhaică des întâlnită a lui *zice.*

În cazul în care un cercetător mai grăbit nu ar recurge la astfel de verificări ale citatelor găsite în format electronic prin confruntare cu cartea, ar putea furniza rezultate viciate, uneori ajungându-se la concluzii absolut eronate.

Deşi greoaie, datorită calității îndoielnice a dactilografiei unora dintre textele corpusului nostru electronic, pe care abia acum suntem în curs de a le colaționa, metoda este mult mai rapidă şi mai eficientă decât lectura şi fişarea clasică a cărților. La numeroase cuvinte recent publicate în volumele I–III ale literei D a DLR, prin căutarea computerizată a ocurențelor în cele 90 de titluri din bibliografia în format electronic, primele atestări au fost împinse cu un secol mai înainte sau chiar mai mult.

Am procesat aceste 90 de titluri, împreună cu alte aproximativ 200 de ediții în format electronic, ce nu aparțin bibliografiei dicționarului, dar în cazul în care găsim aici un citat interesant, îl putem transpune în edițiile din bibliografie. Procesarea s-a efectuat folosind programul Concordanțier LUCON, în varianta 5.0.[1]

Cu ajutorul acestui program se poate obține, în câteva minute, un document care să conțină toate contextele în care apare cuvântul căutat în cele 90 (sau 300) de texte din corpusul nostru. Contextele sunt segmentate automat de calculator, luându-se un număr de unități lexicale înainte şi după cuvântul căutat. Urmează ca aceste contexte să fie citite de cercetător şi, după ce sunt alese cele mai relevante, să fie decupate în conformitate cu regulile sintaxei şi logicii şi verificate cu cartea. Cu ajutorul acestei metode, am găsit prima atestare a cuvântului *domeniu* la Gh Şincai[2].

Alte citate, selectate automat din baza de texte[3], s-au dovedit a nu fi utilizabile, la o confruntare cu textul cărții, deoarece cuvântul *domeniu* apărea nu în documente, ci în comentariile editorului pe marginea documentelor. Această eroare nu este proprie metodei, nu este a calculatorului, ci ține de modul în care volumele de documente au fost introduse în baza de texte, fără a adnota în mod diferit fragmentele de document şi fragmentele de comentariu. Căutarea automată a atestărilor, în texte astfel introduse, este neproductivă.

[1] Programul, al cărui autor este Cătălin Mititelu, poate fi accesat în varianta 6.0, de pe adresa de internet: http:/sourceforge.net/projects/lucon.

[2] *Cetățile: Muncaciului, Făgăraşului şi Tocaiului cu dominiurile.* ŞINCAI, HR. III, 18/31.

[3] *Marturi: oameni din domeniul Bistriței. Iscăliți: Goian i Creciun, boiarii Jadovei* (cca 1650). HURMUZACHI, XV$_2$. *Să întîlneşte cu locurile ce se țin de Domeniul Cîmpulungului şi unde începe satul Sadova* (a. 1783). STEFANELLI, D. C. 138.

BDD-A631 © 2008 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-16 16:54:44 UTC)

Obținerea indicelui de cuvinte al unei cărți sau al mai multora, la care odinioară cercetătorii lucrau ani de zile, se poate realiza cu ajutorul calculatorului în câteva secunde. Desigur, rezultatul trebuie prelucrat de cercetător, deoarece textele noastre nefiind încă adnotate la nivel de cuvânt, calculatorul nu distinge formele flexionare ale aceluiaşi cuvânt şi nu deosebeşte numele proprii de celelalte cuvinte ale limbii.

O problemă foarte importantă care se pune este stabilirea unor principii şi norme, după care să se efectueze transpunerea în format electronic a textelor din bibliografia dicționarului, deoarece această operație este efectuată de multe persoane şi rezultă un aspect neunitar al corpusului de texte.

Până acum, s-a acordat o importanță mai mare introducerii în corpus a textelor vechi, pentru a urmări prima atestare şi frecvența arhaismelor, în vreme ce majoritatea regionalismelor nu pot fi regăsite în lucrările introduse până acum în formatul electronic, studiul lor urmând să fie urmărit într-o etapă ulterioară.

Textele la care s-a lucrat, provenind din secolele de la al XVI-lea până la prima jumătate a secolului al XIX-lea, aparțin unei perioade în care normele ortografice nu erau fixate, se foloseau diverse forme de litere chirilice, sau diverse combinații de litere chirilice şi latine, care diferă de la un volum tipărit la altul.

Transpunerea în formă scrisă a diverselor particularități fonetice ale limbii române sau transpunerea unor particularități dialectale se face de diferiți autori în diferite moduri, fie că este vorba despre literele â, î şi ă, fie că este vorba de unii diftongi pentru care alfabetele chirilice dețin o singură literă, ea, ia, iu, io etc.

Lipsa normelor morfologice şi ortografice face ca fiecare autor să aleagă formule originale de scriere sau chiar să oscileze în aceeaşi carte între mai multe modalități grafice de transpunere a aceluiaşi fenomen fonetic. Este interesant de studiat modul în care s-a petrecut acest proces laborios de cristalizare a normei ortografice, şi indicele de cuvinte realizat de calculator ne oferă o mare bogăție de informație.

După eliminarea din lista de cuvinte a tuturor formelor inexplicabile fonetic, fie că sunt greşeli de tipar, fie de dactilografie sau de ocerizare, se obține o listă a formelor flexionare, ordonate strict alfabetic şi în funcție de numărul de atestări (de frecvența lor).

De pildă, în volumul *Manual de anatomie descriptivă,* de N. A. Kretzulescu[4], al cărui format electronic l-am îngrijit, şi care foloseşte un alfabet de tranziție[5], găsim numeroase greşeli de tipar şi inconsecvențe ortografice.

După o confruntare atentă cu cartea a formelor inexplicabile din indice, am remediat greşelile de dactilografie, care nu existau în volumul tipărit, iar în cazul greşelilor de tipar, am folosit în formatul electronic al cărții următoarea notație:

[4] N. A. Kreţulescu, *Manual de anatomie descriptivă,* Bucureşti, tipografia lui Eliade, 1843, Biblioteca Academiei, cota II 3535.

[5] Se foloseşte un alfabet de tranziție, care conține litere latine (a, d, e, i, m, n, o, r, s, t, z) şi litere chirilice (**в**, **ц**, **г**, **к**, **л**, **п**, **в**, **щ**, ¶, **ꙗ**, **ѫ**, **ѳ**, ¥, ß, ä, ç, þ, x).

BDD-A631 © 2008 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-16 16:54:44 UTC)

ordin, (tip. *ardin*), *sau* (tip. *suu*), *singur* (tip. *singut*) etc. Acestea au fost eliminate din indicele de cuvinte şi forme morfologice. În unele cazuri, interpretarea unor grafii ca erori tipografice şi nu inconsecvenţe ortografice necesită mult discernământ şi este discutabilă: forme ca *idividul*, *dinaitea* ar putea fi sau nu transcrieri ale unei pronunţii simplificate.

Deci indicele de cuvinte şi forme din text, realizat de calculator, a fost folosit mai întâi la corectarea erorilor din formatul electronic al textului. După eliminarea erorilor din document şi din indice, se pot observa în materialul rămas, diverse oscilaţii ortografice. Fără să ne propunem acum aprofundarea unui asemenea studiu, am extras un număr de exemple, scriind între paranteze numărul de atestări ale fiecăruia. Am grupat exemplele în funcţie de diverse probleme care se pot pune în evidenţă.

În următoarele exemple sunetul *j* este concurat cu pronunţia arhaic-regională *gi*:

agiung (1), *ajung* (1);

diregind (1), *dirigind* (1), *dirigiate* (1), *dirijat* (1), *dirijată* (1), *dirijaţi* (1);

împregiuru (1), *împregiurate* (1), *împrejurul* (1).

După cum mărturiseşte autorul în cuvântul introductiv, lucrarea este o traducere şi o rezumare a două cărţi de anatomie, „a lui Cruveilie în patru volume şi a lui Lot într-un volum”, prin urmare, preluarea termenilor ştiinţifici este din limba franceză. Sub influenţa ortografiei acestei limbi, autorul are tendinţa de a dubla (inconsecvent) consoanele:

De pildă, găsim în document 511 atestări ale formei *află* şi 582 atestări ale formei *afflă.* Formele *ammestecătura* (1), *ammestecate* (2), *ammestecându-se* (2), *ammestecă* (13) sunt concurate de o singură atestare a ortografiei *amestică.* La fel, formele cu dublă consoană predomină şi la: *adduc* (1), *addusă* (1), *adductori* (2), *adductorii* (2), *adductoru* (2), *adducţie* (2), *aductorul* (2), *aduc* (2).

Formele *apparatu* (1), *apparatului* (1), *apparaturi* (1) alternează cu *aparatul* (11), *aparatele* (1), *aparatului* (1), de astă dată predominând numeric formele fără dublarea consoanei.

Atestările ne permit şi studiul unor probleme de morfologie, cum ar fi cele privitoare la flexiunea pronumelor şi la sudarea pronumelor compuse: *acelor-l-alte* (1), *celor-l-alte* (1), *cei-l-alte* (1), *celei-l-alte* (1).

Formele de plural în *-uri* sunt mai frecvente decât în limba actuală: *instrumenturi* (2), *licuiduri* (1), *diametruri* (1), *obiecturi* (1), *obrazuri* (1). De asemenea, apar forme cu sufix de întărire la conjugarea I şi la verbe care astăzi preferă forme fără sufix: *îndesează* (1), *îngroşază* (1).

Pluralul substantivelor feminine oscilează între forma cu desinenţa *-e* şi cea cu *-i*:

capsuli (1), *capsulile* (1), *capsule* (10), *capsulele* (4);

cameri (1), *camerile* (1), *camere* (1), *camerele* (1).

BDD-A631 © 2008 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-16 16:54:44 UTC)

Pluralul sau femininul ar trebui să antreneze alternanța între vocală şi diftong, *o/oa,* ceea ce de cele mai multe ori nu se întâmplă, nu ştim dacă e vorba de o pronunție a neologismelor sau de o transcriere ortografică prin analogie cu singularul:

arteriore (1) „artere mai mici", *interioră* (1)*, spinosă* (1)*, tendone* (1)*.*

Această carte este în mod deosebit interesantă pentru marele număr de neologisme din domeniul medical pe care le introduce în limba română[6]. Forma neologismelor este apropiată de limba din care sunt preluate şi oscilantă, în curs de fixare:

accesoar (1), *accesoor* (1), *acsesoaru* (1);

acueduc (1), *acueducul* (2), *acheduc* (1), (unde am notat grupul de litere ке = che, spre deosebire de че = ce);

cuadrilater (1), *cuadrilatiră* (1);

degetații (1), *degitații* (1), *degetațiile* (1), *degitațiile* (2), *degetală* (1), *digitală* (3), *digitale* (5);

cartelagiu (1), *cartelagiuri* (1), *cartelagiurile* (1), *cartilagiul* (23), *cartilag* (1), *cartilagiu* (12), *cartilagiuri* (12), *cartilagiurile* (20), *cartilagiului* (24);

larecsul (1), *larencsii* (1), *larencsului* (1), *laricsul* (1), *laricsului* (1), *larigea* (1), *larigs* (1), *larincsului* (1), *laringea* (1), *laringsu* (1).

Se poate observa şi o tendință spre adoptarea unor calcuri lexicale de structură, nu totale, ci parțiale, păstrând prima parte a cuvântului în forma românească:

carnevorilor (1), dar şi *carnivore* (1), *întrecostală* (1), dar şi *intercostala* (1) sau *întrecostoasa* (1), care ar putea fi considerat un calc lexical de structură total, *întreclavicular* (1), *întremetacarpiene* (1), *întreososului* (1), *întrespinoase* (1), *întremetacarpiene* (1), *întrevertebral* (1). Acestea s-au impus mai târziu în limbă cu prefixul neologic *inter-*.

Cuvintele *măduvoasă* (1), *măduvos* (1), *moşchiuloasă* (1) sunt calcuri totale, care nu s-au impus în limbă, fiind înlocuite cu *muscular, medular.*

Găsim şi câteva pitoreşti calcuri semantice, care nu s-au impus în limba română:

ligheanul (26), *lighean* (28), *ligheanului* (33) „bazin";

lighenuț (2), *lighenuțul* (2) „bazinet";

mitrii (22), *mitrei* (8), *mitră* (11) „uter";

ouşoare (1), *ouşoarele* (1), *ouşorul* (1) „ovul";

luniță (1) „semilună".

În acest text inovator există însă şi elemente care prezintă ortografie sau, poate, pronunție arhaizantă, care se pot remarca mai ales în cazul cuvintelor care oscilează între grupul inițial *îm-* sau *în-*:

[6] N. A. Ursu, *Formarea terminologiei ştiințifice româneşti*, Bucureşti, Editura Ştiințifică, 1962.

BDD-A631 © 2008 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-16 16:54:44 UTC)

îmflat (1), *îmflată* (1), *îmflate* (1), *înflate* (1), *îmflăturile* (1), *înflături* (1).
întrat (1), *întră* (1);
îmvăli (1), *învăluită* (1), dar şi *învelişuri* (1).

Nu am dat decât câteva exemple pentru a ilustra marea bogăție de fapte de limbă pe care le putem găsi în tipăriturile din această epocă şi uşurința cu care pot fi utilizate, în cazul în care textele sunt accesibile în format electronic.

În ce ne priveşte, considerăm esențială pentru studiu o cât mai fidelă redare a textului documentului în forma electronică, nefiind vorba despre o ortografie interpretativă, ca cea din edițiile critice tipărite, care se adresează unui public mai larg şi este necesar să aibă o formă cât mai accesibilă lecturii, unificând inconsecvențele ortografice şi apropiindu-le de ortografia actuală.

Ortografierea diplomatică (adică aceea care respectă cu strictețe toate particularităţile textului vechi), dacă este folosită pentru transpunerea textelor în format electronic, conservă la îndemâna specialiştilor toată bogăția de fapte de limbă, laboratorul în care s-a format vocabularul neologic românesc şi s-a cristalizat norma ortografică actuală. Desigur, specialiştii nu vor fi dezorientaţi de mulțimea variantelor şi vor şti să le clasifice în variante ortografice, morfologice, lexicale sau de pronunție, alegând din mulțimea faptelor de limbă pe acelea care prezintă interes pentru un studiu sau altul.

O altă lucrare pe care am introdus-o în formatul electronic al bibliografiei dicționarului este traducerea unei lucrări istorice de către Gh. Asachi[7]. Acest volum este tipărit mai îngrijit, erorile tipografice sunt mai rare şi o mare parte din ele sunt menționate într-o erată la final. Modificările din erată au fost operate în formatul electronic al cărţii, iar în cazul celorlalte am folosit aceeaşi notație ca mai sus.

Deşi dimensiunile celor două cărți sunt comparabile, indicele de cuvinte şi forme al cărţii lui Asachi este de dimensiuni mult mai mari, datorită faptului că lucrarea conține un foarte mare număr de substantive proprii, pentru care nu am prevăzut un marcaj recognoscibil de către calculator şi acestea apar intercalate printre cele comune.

Credem că, pe viitor, stabilirea unei modalități de adnotare a substantivelor proprii şi de excludere a lor din indicele de cuvinte format de calculator ar fi prioritară.

În ce priveşte substantivele proprii, se poate observa chiar o şi mai mare variație a modalității de ortografiere decât la substantivele comune. Iată câteva exemple (menționez că în exemplele care vor urma, pentru simplificare, lipsa menționării în paranteză a numărului de atestări semnifică faptul că este vorba despre o singură atestare): *Basarabiei, Beserabia; Magdolend, Magdonald, Macdonald; Peching, Pechingului, Pechinul; Sveziei, Sveții, Sveția.*

[7] I. Kaidanov, *Istoria Imperiii rosiene,* Tradusă de aga Gheorghe Asachi, mădular Academiii de Roma, Partea I, Ieşi, în Privileghiata Tipografie a Albinei, 1832, Partea II-a, 1833.

BDD-A631 © 2008 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-16 16:54:44 UTC)

Lucrarea este, ca şi cea discutată anterior, importantă pentru introducerea a numeroase neologisme, de astă dată din domeniul politic (*republică, canțelerie, avtocratic, conciliul, consilierul*), dar şi din cel militar (*cavalerie, comandiri, companii, admiralitate, asedie, ofensive, ofițirii*), juridic (*iurisconsult, giudicători, denonțiația*) sau social-economic (*mânufăpturi, bancorupte, banchier, monopolii, administrația*). Caracteristic stilului lui Asachi este amestecul acestora cu un mare număr de arhaisme (*agiunsese, agiun, aşezemântul, canoniri, cerşutul, chitai, condeiul, covărşitoare, crăeşti, cuminecare*, *cumpenire*), ortografii şi pronunțări regionale (*aşişdere, betejiți, carâtă, chitirea, cimilitura, cimotiile, cimpoaile, clevetiri*), preferarea unor sufixe şi prefixe arhaice, *-ui* (*arestui, comandui*)*, -arisi* (*corespondarisi*)*, -ație* în loc de *-re* pentru abstracte verbale (*colonizația, confiscații, coronația*), ortografia *go* pentru *gu* şi *av-* pentru diftongul *au-,* (*avangvardia, avtocratorică*), pronunția *țe-* a sunetului *ce-* (*canțalerii, comerțială*) ceea ce influențează forma neologismelor, mai greoaie decât la Kretzulescu; puține dintre ele s-au impus în limbă în forma preferată de Asachi.

Remarcăm importanța deosebită acordată de scriitor introducerii cuvintelor noi, ilustrată şi de prezența la finalul volumului a unui *Vocabular* în care o parte dintre neologisme sunt explicate succint:

Amuniția (pentru *muniție*) *Material de oştire: plumb, iarbă de puşcă.*

Anuscrisu (pentru *cronică*) *Descrierea întâmplărilor anilor, slavineşte Letopisițu.*

Armistare (pentru *armistițiu*) *Încetarea armilor seau a oştirilor* etc.[8]

Forma neologismelor variază, mai ales în ce priveşte interpretarea grafică a unor vocale sau diftongi: *biblioteci, bibliotice; brilanturi, brilianturi; cabenetului, cabineturile* (2), *cabinet* (3), *cabinetului* (4); *canțalerii, canțelerii; canțelarului, canțelariului, canțileriei; cetățan, cetățeneşti, cetățesc, cetăținesc, cetățănesc; curtizan, curtezenii, curtizenii; comendantul, comandiri, comandor, comanduia, comenduit* etc.

Lipsa diftongării la unele cuvinte alternează cu forma diftongată: *coroană, coronă; floatei, fometea, persoane, personă* etc.

Unele substantive sunt utilizate cu alt gen sau altă formă de plural decât cea din limba actuală: *clasului, clasurile, codruri, codrurile, exerciția, exerciție, favor, finanțului, finanțurilor, tarifă, proecturi, garnizonul, microscopilor, ghenie* (unde am notat grupul de litere ге = ghe spre deosebire de џе = ge).

Lipsa fixării normelor ortografice se remarcă şi într-un foarte mare număr de forme ale pronumelor şi adjectivelor demonstrative: *acestuie, aceştei, acialantă, acista, acee, aceea, aceaste, acestori, cialalți* etc.

Din marele număr de neologisme, vom selecta câteva dintre cele care se diferențiază de formele actuale, de obicei prin forme arhaice ale sufixelor:

[8] *Op. cit.,* p. IV.

BDD-A631 © 2008 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-16 16:54:44 UTC)

fortunoasă „norocoasă”, *gheneralnicul, gheograficesc, destronitul, clironom* (16) „moştenitor al tronului”, *clironomia* (11), *giuruință, grosietate, haractiriul, monopoliile, oaspitălime, organisirea, personalnica, practisire, reglementul, răspectui, scoposul* (57), *scoposuri* (6), *politii* (67) „oraşe”, *tirănuia, treptalnica, treptăluirea, triamvului* etc.

Alte neologisme au forma cuvintelor compuse calchiate total (acolo unde limba actuală calchiază parțial) sau parțial (acolo unde limba română utilizează împrumutul): *împreunăfăptuitori, dizcuviința, împreunălucrat* (pentru *conlucrat*)*, înaintecuvântare* (pentru *prefață*)*, armistări, vițecraiului* (pentru *vicerege*)*, bancorupte* (pentru *bancrută*)*, mânuscripturi* (pentru *manuscrise*)*, buninimoşii, mareinimie,* (cu antonimul *michinimie*)*, disțerarea, disțerații* (pentru *expatriere*, *expatriați*)*, istorioscriitor* (dar şi *istoriograful*) *locoțiitor* (dar şi *leitenant*) *manufăpturi* sau *mânufăpturi* (pentru *manufactură*) etc.

O caracteristică stilistică este formarea unui mare număr de cuvinte cu prefixul *ne-,* dintre care enumerăm pe cele mai neobişnuite: *necruțarea, necurmată* (2)*, neapărată* (21), *neapărat* (9), *necumpenită, necumpătare, necăire, nedeplin, nefățarnice, nemfrănată, nenumăroase, neominire, neorânduelile, nepractisire, nepriincioasile, nesmintită, nesălită, netemeinice, nevoință, nevărstnicul* etc.

Tot atât de numeroase (unele probabile creații personale) sunt derivatele cu prefixul *în-,* care uneori variază liber cu varianta sa *îm-: îmbunătățit* (9), *îmbunătățite* (9), *îmbărbătățit, îmfocata* (şi *în-*), *îmforma* (şi *în-*), *îmfricoşarea* (şi *în-*), *îmfrumusățarea* (şi *în-*), *îmfățoşatu* (şi *în-*), *împiedecări* „piedici”*, împizmăluire, împoporata, împregiurare* (35), *împresurarea, împustia, împăceluit, împvinovățit* (şi *în-*), *împvăpăet, înavuțit, închizeşluia, înduplicări, înşălătoriu, înțălepțind, înstrămbătățiți, înrăutățiți, înluminat, înnimici* etc.

În sfârşit, exemplificăm cu câteva arhaisme desprinse parcă din cronicarii moldoveni: *adinsu, atmanul, cerşutul, frupturi, frăține, oborirea* (3), *oborit* (3), *ocărmesc, prisosința, pregiudețe, giudicăți* (care alternează cu *judecătoare* sau *iurisconsult*), *răpegiunea, scărbirile, struncinării, suruparea, voroavă, zavistuitorii* etc.

O problemă specifică multor tipărituri cu alfabet chirilic sau mixt este ortografierea literelor ă şi î. Ambele texte discutate aici utilizează semne distincte pentru ă = ß, î = ä, â = õ. Uneori însă, litera specifică pentru â apare la început sau la final de cuvânt, cazuri în care prin transliterare s-ar obține complexe grafice neatestate în limba română, deci am putea accepta convenția de notare de a transcrie acest semn prin î la inițială, la finală de cuvânt sau chiar între prefix şi rădăcina cuvântului, context în care de obicei apare semnul ä, dar inconsecvent.

Această modalitate de scriere, explicabilă prin aceea că grafemul ä transcria în slavonă un sunet uşor nazalizat, specific inițialei de cuvânt, ar putea fi considerată un argument în favoarea actualei norme ortografice, care se bazează pe o veche tradiție a notării diferite a celor două tipuri de contexte.

BDD-A631 © 2008 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-16 16:54:44 UTC)

Deoarece textele noastre în format electronic au la bază cărţi tipărite în diverse epoci, după norme ortografice diferite, ele au caracter neunitar. Ar trebui din nou să se stabilească principial o modalitate de unificare. Pentru moment, s-a apelat la o variantă simplificată, cu â = î, a tuturor textelor, ceea ce, prin abaterea de la norma actuală, este o soluţie puţin convenabilă. Ar fi necesară elaborarea unui program pe calculator care să transpună automat textele cu semn unic pentru sunetul î în texte conforme cu ortografia actuală.

O altă convenţie de introducere a textele vechi în baza de texte, mai discutabilă, apărută prin analogia cu îngrijirea acestora în vederea publicării, ar fi unificarea variantelor ortografice cu ă şi â în interiorul cuvântului, căci de obicei aceste semne alternează şi (cu toate că unii cercetători, ca Ion Gheţie şi Andrei Avram, au scris cărţi pe această temă) nu putem şti cu certitudine care era în epocă diferenţa de pronunţie. Totuşi, eliminarea alternanţei ă/â ar duce, credem, la o prea mare sărăcire a bogăţiei variantelor grafice caracteristice acelei epoci. În aceeaşi carte cuvântul *gât* sau *lână* poate fi scris când cu *ă*, când cu *â,* deoarece nu se fixase încă o normă de scriere şi pronunţie. Deşi nu putem şti care era pronunţia în acea epocă, considerăm că, dacă nu ar fi existat diferenţe de pronunţie, limba nu ar fi folosit semne distincte şi mai ales, în cursul evoluţiei lingvistice, ele nu s-ar fi diferenţiat din ce în ce mai clar, ci una dintre acestea ar fi dispărut.

Există o diferenţă de utilitate şi obiective între îngrijirea în vederea publicării, care se adresează unui număr mai mare de cititori şi este necesar să fie accesibilă, şi îngrijirea în vederea introducerii în corpusul electronic de texte, care se adresează specialiştilor, este realizat în scopul efectuării unor căutări avansate cu calculatorul şi trebuie să conserve cât mai multe informaţii lingvistice.

Prin exemplele de mai sus, selectate din doar două texte din prima jumătate a sec. al XIX-lea, am dorit să demonstrăm vastitatea posibilităţilor pe care lucrul cu textele în format electronic îl oferă cercetării.

O altă idee care apare cu claritate este aceea că, pentru validarea rezultatelor oricărei cercetări asupra acestui material, este necesar să existe certitudinea corectitudinii şi fidelităţii transcrierii, pentru care nu trebuie precupeţit niciun efort, iar convenţiile de notare stabilite între participanţi, (de felul: nu vom transcrie decât prin *î* literele *â* sau *î* la iniţială sau la final de cuvânt), să fie cât mai puţine şi să fie urmărite cu consecvenţă de toate persoanele care prelucrează textele corpusului electronic.

În cazul în care vom unifica variantele fonetice, ortografice, morfologice etc., îi vom trimite pe viitorii cercetători înapoi la cartea veche tipărită şi mai ales, îi vom priva de posibilitatea de a efectua în text căutări avansate cu ajutorul computerului, efectuând în câteva minute o muncă pe care prin mijloace clasice ar fi posibil să o desfăşoare în luni sau ani.

BDD-A631 © 2008 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-16 16:54:44 UTC)

NOUVELLES PERSPECTIVES OUVERTES À LA RECHERCHE LINGUISTIQUE PAR L'INFORMATISATION DE LA BIBLIOGRAPHIE DU *DICTIONNAIRE ACADÉMIQUE DE LA LANGUE ROUMAINE* (DLR)

(Résumé)

On présente ici le stade de l'introduction dans un format éléctronique de la bibliographie du dictionnaire académique de la langue roumaine, à laquelle on travaille déjà depuis longtemps. Nous nous sommes axés premièrement sur les vieux textes, pour établir d'une manière plus précise les premières attestations des mots. On continuera par introduire les principales sources pour les régionalismes et les néologismes.

On présente aussi les possibilités de travail avec ce corpus restreint de textes que nous détenons et le programme de concordances pour les recherches avancées que nous utilisons.

Enfin, on donne des exemples à partir de deux livres de la première partie du XIX-ème siècle, importants pour l'aquisiton des néologismes, que nous avons récemment soignés du point de vue de la conformation avec le texte édité.

On remarque la nécéssité de la parfaite correctitude du texte éléctronique et la richesse des donnés concernant la cristalisation de la norme ortographique, de l'adoption de nouveaux mots, des changements morphologiques.

Notre opinion est qu'on a vraiment besoin d'une unité formelle de notre corpus de textes, qu'il faut établir des principes pour transposer les textes, principes respectés par tous les opérateurs, mais sans sacrifier la richesse d'informations linguistiques utiles aux spécialistes en modifiant trop les textes, conformément à ce caractère unitaire.

Institutul de Lingvistică „Iorgu Iordan – Al. Rosetti"
Bucureşti, Calea 13 Septembrie nr. 13

BDD-A631 © 2008 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-16 16:54:44 UTC)