

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 și prime atestări ale cuvintelor și a extragerii de citate din sursele bibliografice.

Din păcate, datorită fondurilor insuficiente care se pot investi în această vastă întreprindere, deși 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 Iași scanarea cărților și periodicelor din bibliografie. O etapă ulterioară ar fi transpunerea din format fotocopiă în format text, proces numit ocerizare după numele programului utilizat, OCR (Optical Character Recognition). Se obțin texte care pot fi prelucrate, atât manual cât și de către calculator. Această transpunere se poate însă aplica numai textelor mai recente, în alfabet latin, și chiar și atunci comportă numeroase erori care se corectează prin confruntarea textului ocerizat cu textul tipărit.

Echipa de la București 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. Și aceste texte, introduse manual, conțin diferite greșeli de interpretare a literei sau de dactilografie și necesită un complicat proces de corectură (proces de rutină, care nu se poate face decât într-un ritm foarte lent).

Totuși, investirea de timp în corectarea textelor, fie ocerizate, fie dactilografiate, se dovedește necesară; calitatea (acuratețea) și conformitatea (dintre textul electronic și 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 și că nu am făcut economie, ci risipă de timp și de efort.

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

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 greșeli 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.

Deși 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 și mai eficientă decât lectura și fișarea 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 și după cuvântul căutat. Urmează ca aceste contexte să fie citite de cercetător și, după ce sunt alese cele mai relevante, să fie decupate în conformitate cu regulile sintaxei și logicii și verificate cu cartea. Cu ajutorul acestei metode, am găsit prima atestare a cuvântului *domeniu* la Gh Șincai².

Alte citate, selectate automat din baza de texte³, 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 și 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, Făgărașului și Tocaiului cu dominiurile*. ȘINCAI, HR. III, 18/31.

³ *Marturi: oameni din domeniul Bistriței. Iscăliți: Goian i Creciun, boiarii Jadovei* (cca 1650). HURMUZACHI, XV₂. *Să întâlnește cu locurile ce se țin de Domeniul Cîmpulungului și unde începe satul Sadova* (a. 1783). STEFANELLI, D. C. 138.

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 aceluiași cuvânt și nu deosebește numele proprii de celelalte cuvinte ale limbii.

O problemă foarte importantă care se pune este stabilirea unor principii și norme, după care să se efectueze transpunerea în format electronic a textelor din bibliografia dicționarului, deoarece această operație este efectuată de multe persoane și 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 și 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 și 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 și ă, fie că este vorba de unii diftongi pentru care alfabetele chirilice dețin o singură literă, ea, ia, iu, io etc.

Lipsa normelor morfologice și ortografice face ca fiecare autor să aleagă formule originale de scriere sau chiar să oscileze în aceeași carte între mai multe modalități grafice de transpunere a aceluiași fenomen fonetic. Este interesant de studiat modul în care s-a petrecut acest proces laborios de cristalizare a normei ortografice, și 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 greșeli de tipar, fie de dactilografie sau de ocerizare, se obține o listă a formelor flexionare, ordonate strict alfabetic și în funcție de numărul de atestări (de frecvența lor).

De pildă, în volumul *Manual de anatomie descriptivă*, de N. A. Kretzulescu⁴, al cărui format electronic l-am îngrijit, și care folosește un alfabet de tranziție⁵, găsim numeroase greșeli de tipar și inconsecvențe ortografice.

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

⁴ N. A. Kretzulescu, *Manual de anatomie descriptivă*, București, tipografia lui Eliade, 1843, Biblioteca Academiei, cota II 3535.

⁵ Se folosește un alfabet de tranziție, care conține litere latine (a, d, e, i, m, n, o, r, s, t, z) și litere chirilice (в, ւ, г, к, л, п, в, ւ, ր, ա, ж, ф, ы, б, ä, ç, þ, x).

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

Deci indicele de cuvinte și forme din text, realizat de calculator, a fost folosit mai întâi la corectarea erorilor din formatul electronic al textului. După eliminarea erorilor din document și din indice, se pot observa în materialul rămas, diverse oscilații 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 funcție de diverse probleme care se pot pune în evidență.

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

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

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

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

După cum mărturisește autorul în cuvântul introductiv, lucrarea este o traducere și o rezumare a două cărți de anatomie, „a lui Cruveilie în patru volume și a lui Lot într-un volum”, prin urmare, preluarea termenilor științifici este din limba franceză. Sub influența ortografiei acestei limbi, autorul are tendința de a dubla (inconsecvent) consoanele:

De pildă, găsim în document 511 atestări ale formei *află* și 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 și la: *adduc* (1), *addusă* (1), *adductori* (2), *adductorii* (2), *adductoru* (2), *adducție* (2), *aductorul* (2), *aduc* (2).

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

Atestările ne permit și studiul unor probleme de morfologie, cum ar fi cele privitoare la flexiunea pronomelor și la sudarea pronomelor compuse: *acelor-l-alte* (1), *celor-l-alte* (1), *cei-l-alte* (1), *cele-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 și la verbe care astăzi preferă forme fără sufix: *îndesează* (1), *îngroșază* (1).

Pluralul substantivelor feminine oscilează între forma cu desinența *-e* și cea cu *-i*:

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

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

Pluralul sau femininul ar trebui să antreneze alternanța între vocală și diftong, *o/oa*, ceea ce de cele mai multe ori nu se întâmplă, nu știm 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 română⁶. Forma neologismelor este apropiată de limba din care sunt preluate și oscilantă, în curs de fixare:

accesoar (1), *accesoar* (1), *accesoaru* (1);

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

cuadrilater (1), *cuadrilatiră* (1);

degetații (1), *degetații* (1), *degetațiile* (1), *degetațiile* (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 și o tendință spre adoptarea unor calcuri lexicale de structură, nu totale, ci parțiale, păstrând prima parte a cuvântului în forma românească:

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

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

Găsim și câteva pitorești 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”;

oușoare (1), *oușoarele* (1), *oușorul* (1) „ovul”;

luniță (1) „semilună”.

În acest text inovator există însă și 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 științifice românești*, București, Editura Științifică, 1962.

îmflat (1), *îmflată* (1), *îmflate* (1), *înflate* (1), *îmflăturile* (1), *înflături* (1).
întrat (1), *întră* (1);
îmvăli (1), *învăluită* (1), dar și *învelișuri* (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 și ușurința cu care pot fi utilizate, în cazul în care textele sunt accesibile în format electronic.

În ce ne privește, 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 și este necesar să aibă o formă cât mai accesibilă lecturii, unificând inconsecvențele ortografice și apropiindu-le de ortografia actuală.

Ortografierea diplomatică (adică aceea care respectă cu strictețe toate particularitățile textului vechi), dacă este folosită pentru transpunerea textelor în format electronic, conservă la îndemâna specialiștilor toată bogăția de fapte de limbă, laboratorul în care s-a format vocabularul neologic românesc și s-a cristalizat norma ortografică actuală. Desigur, specialiștii nu vor fi dezorientați de mulțimea variantelor și vor ști 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. Asachi⁷. Acest volum este tipărit mai îngrijit, erorile tipografice sunt mai rare și o mare parte din ele sunt menționate într-o erată la final. Modificările din erată au fost operate în formatul electronic al cărții, iar în cazul celorlalte am folosit aceeași notație ca mai sus.

Deși dimensiunile celor două cărți sunt comparabile, indicele de cuvinte și forme al cărții 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 și acestea apar intercalate printre cele comune.

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

În ce privește substantivele proprii, se poate observa chiar o și 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*, *Pechimul*; *Sveziei*, *Sveții*, *Sveția*.

⁷ I. Kaidanov, *Istoria Imperiului rosien*, Tradusă de aga Gheorghe Asachi, mǎdular Academiei de Roma, Partea I, Ieși, în Privileghiata Tipografie a Albinei, 1832, Partea II-a, 1833.

Lucrarea este, ca și cea discutată anterior, importantă pentru introducerea a numeroase neologisme, de astă dată din domeniul politic (*republică, canțelerie, avtocratic, conciliul, consilierul*), dar și din cel militar (*cavalerie, comandiri, companii, admiralitate, asedie, ofensive, ofițirii*), juridic (*iurisconsult, judecători, denunțiația*) sau social-economic (*mănușăpturi, bancorupte, banchier, monopolii, administrația*). Caracteristic stilului lui Asachi este amestecul acestora cu un mare număr de arhaisme (*agiunse, agiun, așezământul, canoniri, cerșutul, chitai, condeiul, covârșitoare, crăești, cuminecare, cumpenire*), ortografii și pronunțări regionale (*așișdere, betejiți, carătă, chitirea, cimilitura, cimotiile, cimpoaile, clevetiri*), preferarea unor sufixe și prefixe arhaice, *-ui (arestui, comandui), -arisi (corespondarisi), -ație* în loc de *-re* pentru abstracte verbale (*colonizația, confiscății, coronația*), ortografia *go* pentru *gu* și *av-* pentru diftongul *au-*, (*avangvardia, avtocratică*), pronunția *șe-* 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 și de prezența la finalul volumului a unui *Vocabular* în care o parte dintre neologisme sunt explicate succint:

Armuniția (pentru *muniție*) *Material de oștire: plumb, iarbă de pușcă.*

Anuscrisu (pentru *cronică*) *Descrierea întâmplărilor anilor, slavinește Letopisișu.*

Armistare (pentru *armistițiu*) *Încetarea armilor seu a oștirilor etc.*⁸

Forma neologismelor variază, mai ales în ce privește interpretarea grafică a unor vocale sau diftongi: *biblioteci, bibliotice; brilanturi, brillianturi; cabenetului, cabineturile* (2), *cabinet* (3), *cabinetului* (4); *canțalerii, canțelerii; canțelarului, canțelariului, canțeleriei; cetățan, cetățenești, cetățesc, cetăținesc, cetățânesc; 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 *re = ghe* spre deosebire de *șe = ge*).

Lipsa fixării normelor ortografice se remarcă și într-un foarte mare număr de forme ale pronumelor și adjectivelor demonstrative: *acestuie, aceștei, 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.

fortunoasă „norocoasă”, *gheneralnicul*, *gheograficesc*, *destronitul*, *clironom* (16) „moștenitor al tronului”, *clironomia* (11), *giuruință*, *grosietate*, *haractiriul*, *monopoliile*, *oaspițalime*, *organisirea*, *personalnica*, *practisire*, *reglementul*, *răspectui*, *scoposul* (57), *scoposuri* (6), *politii* (67) „orașe”, *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*, *dizcuviița*, *împreunălucrat* (pentru *conlucrat*), *înaintecuvântare* (pentru *prefață*), *armistări*, *vițecraului* (pentru *vicerege*), *bancorupte* (pentru *bancrută*), *mânscripturi* (pentru *manuscrise*), *buninimoșii*, *mareinimie*, (cu antonimul *micinimie*), *disțerarea*, *disțerații* (pentru *expatriere*, *expatriați*), *istorioscriitor* (dar și *istoriograful*) *locofiitor* (dar și *leitenant*) *manușapturi* sau *mănușapturi* (pentru *manufactură*) etc.

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

Tot atât de numeroase (unele probabile creații personale) sunt derivatele cu prefixul *în-*, care uneori variaza liber cu varianta sa *îm-*: *îmbunătățit* (9), *îmbunătățite* (9), *îmbărbătațit*, *îmfocata* (și *în-*), *îmforma* (și *în-*), *îmfrișarea* (și *în-*), *îmfrumusățarea* (și *în-*), *îmfățoșatu* (și *în-*), *împiedecări* „pedici”, *împizmăluire*, *împoporata*, *împregiurare* (35), *împresurarea*, *împustia*, *împăceluit*, *îmvinovățit* (și *în-*), *îmvăpăet*, *înavuțit*, *închizeșluia*, *înduplicări*, *înșălătoriu*, *înțălepșind*, *înstrămbătații*, *înrautății*, *înluminat*, *înnimici* etc.

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

O problemă specifică multor tipărituri cu alfabet chirilic sau mixt este ortografierea literelor *ă* și *î*. Ambele texte discutate aici utilizează semne distincte pentru *ă* = *ѣ*, *î* = *ѡ*, *â* = *Ѡ*. Uneori însă, litera specifică pentru *â* apare la început sau la final de cuvânt, cazuri în care prin transliterare s-ar obține complexe grafice neatestare 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 și rădăcina cuvântului, context în care de obicei apare semnul *ă*, dar inconsecvent.

Această modalitate de scriere, explicabilă prin aceea că grafemul *ă* transcria în slavonă un sunet ușor 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.

Deoarece textele noastre în format electronic au la bază cărți 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 $\hat{a} = \hat{i}$, a tuturor textelor, ceea ce, prin abaterea de la norma actuală, este o soluție puțin convenabilă. Ar fi necesară elaborarea unui program pe calculator care să transpună automat textele cu semn unic pentru sunetul \hat{i} în texte conforme cu ortografia actuală.

O altă convenție 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 \hat{a} și \hat{i} în interiorul cuvântului, căci de obicei aceste semne alternează și (cu toate că unii cercetători, ca Ion Gheție și Andrei Avram, au scris cărți pe această temă) nu putem ști cu certitudine care era în epocă diferența de pronunție. Totuși, eliminarea alternanței \hat{a}/\hat{i} ar duce, credem, la o prea mare sărăcire a bogăției variantelor grafice caracteristice acelei epoci. În aceeași carte cuvântul *gât* sau *lână* poate fi scris când cu \hat{a} , când cu \hat{i} , deoarece nu se fixase încă o normă de scriere și pronunție. Deși nu putem ști care era pronunția în acea epocă, considerăm că, dacă nu ar fi existat diferențe de pronunție, limba nu ar fi folosit semne distincte și mai ales, în cursul evoluției lingvistice, ele nu s-ar fi diferențiat din ce în ce mai clar, ci una dintre acestea ar fi dispărut.

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

Prin exemplele de mai sus, selectate din doar două texte din prima jumătate a sec. al XIX-lea, am dorit să demonstrăm vastitatea posibilităților 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 și fidelității transcrierii, pentru care nu trebuie precupețit niciun efort, iar convențiile de notare stabilite între participanți, (de felul: nu vom transcrie decât prin \hat{i} literele \hat{a} sau \hat{i} la inițială sau la final de cuvânt), să fie cât mai puține și să fie urmărite cu consecvență 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 și 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 desfășoare în luni sau ani.

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 électronique 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'acquisition des néologismes, que nous avons récemment soignés du point de vue de la conformation avec le texte édité.

On remarque la nécessité de la parfaite correctitude du texte électronique et la richesse des données concernant la cristallisation de la norme orthographique, 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”
București, Calea 13 Septembrie nr. 13*