

Noi contribuții la informatizarea editării atlaselor lingvistice românești regionale

Florin-Teodor Olariu, Vasile Apopei, Silviu Bejinariu

1.0. Redactarea și editarea materialului dialectal cules în cadrul anchetelor efectuate pentru *Noul Atlas lingvistic român, pe regiuni* (NALR sau ALRR) constituie o prioritate a cercetătorilor dialectologi la ora actuală. Dată fiind importanța la nivel național a acestei întreprinderi științifice, responsabilitatea specialiștilor implicați în acest demers devine majoră. Nevoia urgentării acestui proces este, în aceste condiții, un imperativ al lingvisticii românești, cu atât mai mult cu cât finalizarea sa poate constitui punctul de plecare al unor noi proiecte sau sinteze la nivel național în domeniul dialectologiei, în scopul confirmării (sau, de ce nu, al redefinirii) structurii dialectale a limbii române. Acest fapt se poate realiza, în primul rând, prin posibilitatea pe care o oferă materialul NALR (prin rețeaua cartografică mult mai deasă pe care o are, în comparație cu cea prezentă în ALR) de a întreprinde o analiză mult mai detaliată a fenomenelor fonetice, morfosintactice, lexicale, semantice etc. existente la nivel dialectal.

2.0. Soluția optimă pentru rezolvarea acestei probleme stringente a venit din direcția tehnologiilor informaționale, fapt cât se poate de firesc pentru momentul actual din evoluția cercetării științifice. În acest sens, începând cu anul 2000 se derulează la Iași un proiect interdisciplinar în cadrul Academiei Române, ce are ca scop realizarea unui program de editare asistată de calculator a atlaselor lingvistice și a textelor dialectale. Proiectul respectiv, care a început sub forma unui grant academic intitulat *Cercetări în domeniul foneticii și fonologiei limbii române, cu aplicații la atlasele lingvistice românești regionale*, se desfășoară, în prezent, sub forma colaborării dintre cercetători ai Institutului de Informatică Teoretică și ai Institutului de Filologie Română „A. Philippide” (ambele făcând parte din rețeaua institutelor Academiei Române, Filiala Iași) în cadrul proiectului academic prioritar, intitulat *Proiectarea și implementarea unui sistem integrat de aplicații software pentru realizarea „Atlasului lingvistic român, pe regiuni” și editarea textelor dialectale*. Precizăm că primele rezultate al acestei colaborări au fost expuse anul trecut, într-o comunicare susținută în cadrul Simpozionului Internațional *Identitatea limbii și literaturii române în perspectiva globalizării*, simpozion organizat la Iași de Institutul de Filologie Română „A. Philippide”, împreună cu Asociația pentru Literatura Română și Cultura Poporului Român „Astra” – Despărțământul „Mihail Kogălniceanu”¹. În cele ce urmează, vom prezenta ultimele realizări ale cercetătorilor angajați în acest proiect.

¹ Lucrarea, având titlul *Editarea asistată de calculator a atlaselor lingvistice și a textelor dialectale*, a fost publicată în Ofelia Ichim, Florin-Teodor Olariu (eds.), *Identitatea limbii și literaturii române în perspectiva globalizării*, Iași, Editura „Trinitas”, p. 211-232.

Cadrul general fiind deja stabilit (prin modelarea din punct de vedere informatic a datelor grafice: conturul hărții mari, analitice, și a celei mici, sintetice, a rețelei punctelor de anchetă, stabilirea poziției răspunsurilor în raport cu punctul cartografic etc., precum și găsirea celor mai adecvate soluții pentru realizarea familiilor de fonturi necesare transcrierii fonetice în sistem NALR), s-a trecut la analiza următoarelor două puncte de pe agenda proiectului: găsirea modalităților optime pentru generarea și, ulterior, editarea hărților sintetice (interpretative) și a materialului necartografiat, în calitatea acestora de părți constitutive ale proiectului NALR, preluate din sistemul ALR, unde au constituit, la acea vreme, inovații remarcabile (și remarcate) în domeniul geografiei lingvistice². Ca punct de plecare, cele două elemente au fost concepute ca module ale aplicației inițiale, ele având ca principiu structurant, după cum vom vedea în continuare, aceeași manieră de organizare a datelor introduse în dicționarul alocat transcrierilor fonetice.

2.1. În ceea ce privește configurarea hărților interpretative, s-a avut în vedere (ca, de altfel, pe întreg parcursul desfășurării proiectului de față) păstrarea soluțiilor deja consacrate la nivel național, și anume, folosirea *hașurilor* și a *simbolurilor* (și, evident, a combinației dintre acestea) în alcătuirea acestora³. Având ca fundal această idee, următorul pas – esențial în rezolvarea generării și editării acestui tip de hartă – a fost stabilirea corespondenței punct de anchetă – tip de hașură sau simbol, precum și acordarea utilizatorului programului a posibilității ca mai multe răspunsuri identice din puncte de anchetă diferite să poată fi grupate în serii omogene, care, ulterior, să constituie sursa generării unor arii lingvistice corespunzătoare respectivelor serii de răspunsuri. Precizăm că această modalitate de generare a hărților interpretative se aplică cu aceeași ușurință atât în realizarea hărților lexicale, cât și a celor fonetice, morfosintactice, semantice etc., și aceasta datorită faptului că utilizatorului acestui program îi este acordată posibilitatea de a decide în mod *individual* și *selectiv* care puncte cartografice vor face parte din seria/aria respectivă.

În cele ce urmează vom prezenta, pe scurt, maniera în care se realizează o hartă interpretativă, având ca suport, în cazul de față, o problemă de ordin fonetic. Așadar, se pleacă, mai întâi, de la fereastra de bază a aplicației, unde se activează butonul intitulat, în mod generic, „Grupări” (Figura 1), în urma acestei operații apărând fereastra de dialog „Selecție grupare”. Aici fie se selectează o problemă

² Pentru o informare detaliată asupra inovațiilor promovate de *Atlasul lingvistic român*, a se vedea I. Mării, *Harta lexicală semantică*, Cluj-Napoca, Clusium, 2002, p. 12, 13, 109, 110, 123 etc.

³ Ținem să subliniem acest aspect deoarece, cel puțin din punctul de vedere al analistului programator, mult mai ușor ar fi fost ca diferitele arii ale unei asemenea hărți, în care se urmărește repartiția spațială a unui anumit fenomen lingvistic, să fie redată, de exemplu, prin tot atâtea culori. Delimitarea și generarea ariilor prin intermediul hașurilor a necesitat o muncă mult mai laborioasă, soluție care, totuși, a fost aleasă tocmai în ideea respectării tradiției deja existente în domeniu.

lingvistică deja prelucrată și, astfel, prezentă în respectiva fereastră, fie se trece la generarea unui nou tip de hartă interpretativă, prin activarea butonului „Adăugare”⁴.

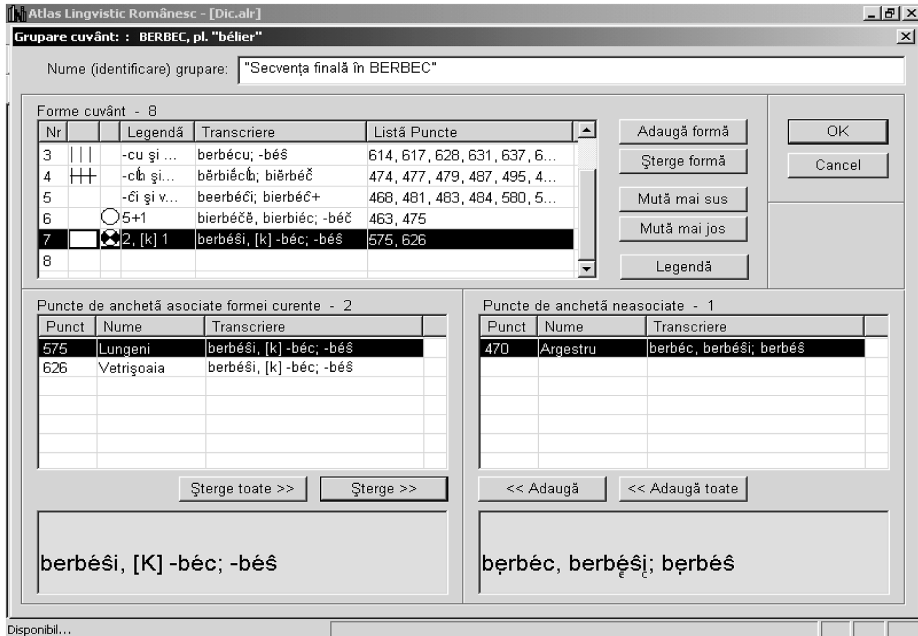


Figura 1. Fereastra de bază a aplicației (*Dicționarul „Transcrieri fonetice”*)

Odată formulat titlul grupării/hărții interpretative asupra căreia se va lucra, se va trece la selectarea și gruparea răspunsurilor identice din totalul celor deja introduse în *Dicționarul „Transcrieri fonetice”* sub cuvântul respectiv⁵, acestea din urmă fiind prezente în fereastra *Puncte de anchetă neasociate* (Figura 2). Răspunsurile care acoperă aceeași formă sau fenomen lingvistic vor fi grupate în fereastra *Puncte de anchetă asociate formei curente*, loc unde se va putea verifica, în fereastra de control din stânga-jos, identitatea formală a răspunsurilor prezente în fereastra mai sus menționată, această verificare dublând-o, de fapt, pe cea deja efectuată în fereastra de control din dreapta-jos, o dată cu realizarea selecției inițiale. Precizia selecției grupărilor atinge, în aceste condiții, un nivel ridicat, asigurându-se, astfel, omogenitatea grupării respective.

⁴ În cazul de față avem prelucrate, *demonstrativ*, două probleme lingvistice ce se bazează pe răspunsurile obținute în timpul anchetei NALR la întrebarea 899, *berbec*, pl. „[Întrebare anterioară: Cum îi zici la aceea pe care o mulgi și tunzi lâna de pe ea?] Dar dacă e parte bărbătească cum îi spui?": „Secvența finală în BERBEC” și „Grupare MN (Material Necartografiat)”.

⁵ În cazul *NALR. Moldova și Bucovina*, acest dicționar cuprinde răspunsurile înregistrate în cele 210 puncte de anchetă aferente rețelei respectivului atlas regional, puncte cuprinse în intervalul 461-670.

Figura 2. Fereastra *Grupare cuvânt*

Următoarea etapă va consta în atribuirea caracteristicilor grafice ce vor defini gruparea omogenă obținută în etapa anterioară, rezultatul acestei operațiuni găsindu-se, ulterior, în legenda hărții respective. Această operație se va desfășura (în urma activării butonului *Legendă* din fereastra *Grupare cuvânt*) în cadrul ferestrei din Figura 3, unde utilizatorul va stabili ce culoare (în cazul în care se va opta pentru editarea policromă a hărților interpretative) și ce stil va avea hașura sau conturul⁶ alocat ariei corespunzătoare formei (ca sumă de răspunsuri identice) respective, sau, în cazul în care se decide atribuirea unui simbol, acesta va alege tipul simbolului respectiv. Totodată, în această ultimă etapă se va stabili și legenda hărții, prin indicarea corespondenței *stil de hașură/simbol – formă (variantă) lingvistică acoperită (reprezentată)*. În urma stabilirii tuturor acestor detalii, se va obține o hartă interpretativă asemănătoare celei din figurile 4 și 5⁷.

⁶ Precizăm că, în forma sa finală, programul va da utilizatorului posibilitatea de a delimita, cu ajutorul unor *contururi* prestabilite, diferite arii direct pe harta mare, analitică, în stilul deja folosit de *NALR. Oltenia* sau *NALR. Muntenia și Dobrogea*.

⁷ Datorită faptului că această expunere este una pur demonstrativă, în cazul de față a fost utilizată o combinație a tuturor caracteristicilor grafice existente (hașură + contur + simbol), această combinație fiind însă mai puțin folosită în lucrările de specialitate.

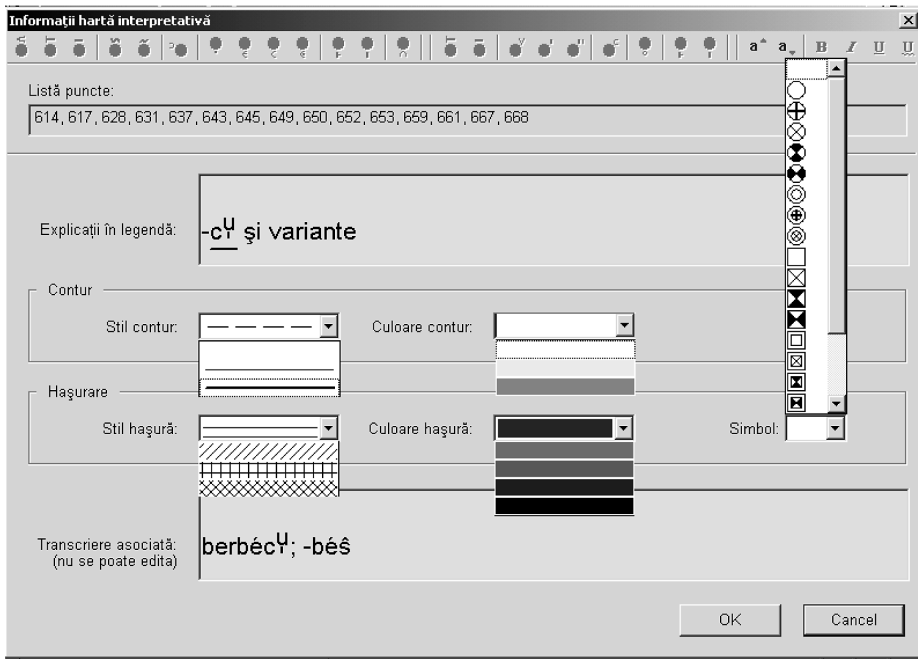
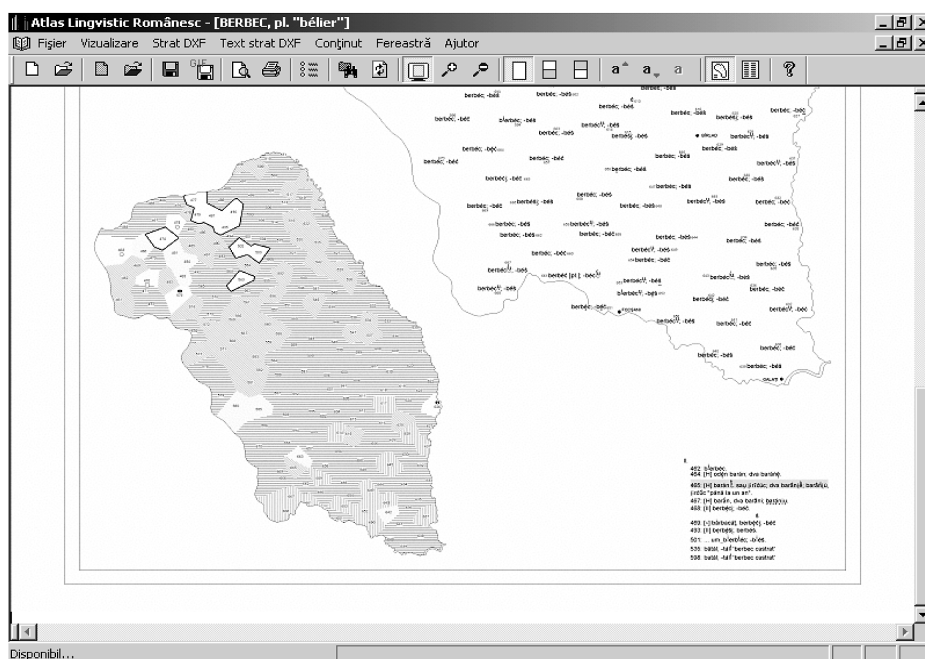


Figura 3. Fereastră *Informații hartă interpretativă*

Având în vedere faptul că *NALR. Banat* și *NALR. Moldova și Bucovina* conțin hărți interpretative grupate la sfârșitul volumului, care urmăresc un fenomen lingvistic prezent în răspunsurile aferente unui număr mai mare de întrebări din *Chestionarul NALR*, a fost necesar să se creeze posibilitatea ca hărțile interpretative să poată fi generate și editate și în mod independent față de poziția standard din harta de bază a atlasului (hartă analitică + hartă sintetică). Soluționarea acestui aspect s-a realizat prin salvarea hărții în următoarele formate: format propriu, orientat pe obiecte, format GIF sau format BMP. În acest mod, hărțile lingvistice pot deveni obiecte grafice de sine stătătoare, cărora li se pot aplica, ulterior, un set de noi operații redacționale (cum ar fi, de exemplu: introducerea datelor ce alcătuiesc legenda hărții interpretative; modificarea, la nevoie, a poziției răspunsului în raport cu punctul cartografic etc.). Această soluție are o aplicabilitate sporită din punct de vedere științific datorită faptului că, pe lângă editarea hărților interpretative de la sfârșitul atlaselor lingvistice, ea poate fi folosită și ca o aplicație separată, necesară în redactarea oricărei lucrări de specialitate care urmărește evidențierea spațială a unui anumit fenomen lingvistic.

2.2. Referitor la structurarea informației care va fi folosită la generarea planșelor ce conțin materialul necartografiat, aceasta s-a realizat, după cum menționăm și mai sus, pe baza aceluiași principiu folosit la generarea hărților

interpretative. Etapele care trebuie parcurse în vederea structurării răspunsurilor obținute la o anumită întrebare din chestionar, sub forma deja consacrată, sunt aceleași, singura deosebire fiind faptul că în cadrul acestei operațiuni se renunță la ultima etapă, și anume, aceea a atribuirii caracteristicilor grafice diferitelor serii omogene ale grupării, aceasta nemaifiind necesară în cazul de față. Așadar, operațiile necesare editării materialului necartografiat sunt cele circumscrise Figurilor 1 și 6 (aceasta din urmă fiind identică, după cum cu ușurință se poate observa, cu Figura 2), ajungându-se, în final, la rezultatele prezente în Figurile 7 și 8.



3. În continuare vom face câteva constatări în legătură cu cele prezentate mai sus, în măsură să evidențieze cât mai explicit avantajele oferite de acest sistem informațional.

3.1. Mai întâi, datorită cvasi-identității procedurale în ceea ce privește structurarea, generarea și editarea hărților interpretative și, respectiv, a materialului necartografiat, eforturile utilizatorului sunt substanțial reduse atât în ceea ce privește acomodarea sa cu acest program, cât și, mai ales, în ceea ce privește gestionarea eficientă a timpului pe întreg parcursul folosirii sistemului.

Legat de acest aspect, trebuie subliniat faptul că, prin intermediul acestui sistem, utilizatorului îi este acordată posibilitatea unui control riguros al datelor introduse. Astfel, acesta deține o evidență automată permanentă a ocurențelor (puncte cartografice) prezente în interiorul unei grupări (arii) lingvistice omogene (vezi Figura 3: *Listă puncte* sau Figura 2: *Forme cuvânt, Puncte de anchetă asociate formei curente, Puncte de anchetă neasociate*), precum și posibilitatea ca, atunci când se decide efectuarea unei modificări în *Dicționarul „Transcrieri fonetice”*, reflexul respectivei modificări în formatul grafic (hartă analitică/sintetică sau MN) să poată fi verificat imediat.

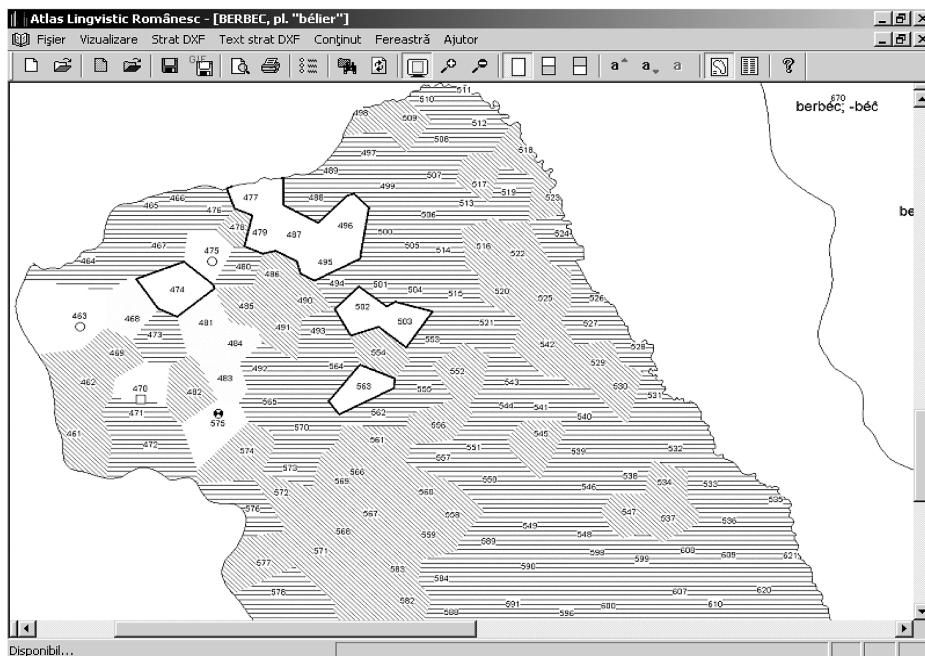


Figura 5. Hartă lingvistică interpretativă (detaliu)

3.2. Un alt argument al eficienței acestui program constă în faptul că, odată materialul dialectal introdus în *Dicționarul „Transcrieri fonetice”*, acesta devine suport al generării și editării atât a hărții analitice, cât și a celei sintetice sau,

printr-o minimă prelucrare redacțională în prealabil și în funcție de scopul urmărit, a materialului necartografiat. Posibilitatea acestei conversii își găsește aplicabilitatea cel puțin în următorul caz: în cadrul procesului de selecție și clasificare a materialului ce va fi cuprins într-un anumit volum al unui atlas lingvistic din seria NALR, autorii-redactori ai respectivului atlas decid, în urma unei discuții preliminare, efectuată pe baza fișelor de anchetă, răspunsurile căror întrebări vor fi redactate sub formă de hărți analitice și a căror sub formă de material necartografiat. Se întâmplă însă ca, odată redactat materialul sub forma unei hărți analitice, acesta să se dovedească prea omogen din punct de vedere al formelor fonetice înregistrate și, în consecință, autorii să se decidă ca redactarea materialului respectiv să se realizeze sub formă de material necartografiat. Această operațiune, destul de laborioasă în maniera clasică de redactare, va deveni una mult mai ușoară prin apelul la funcțiile sistemului de față, ea necesitând doar o minimă prelucrare a materialului *deja prezent în Dicționarul „Transcrieri fonetice”*.

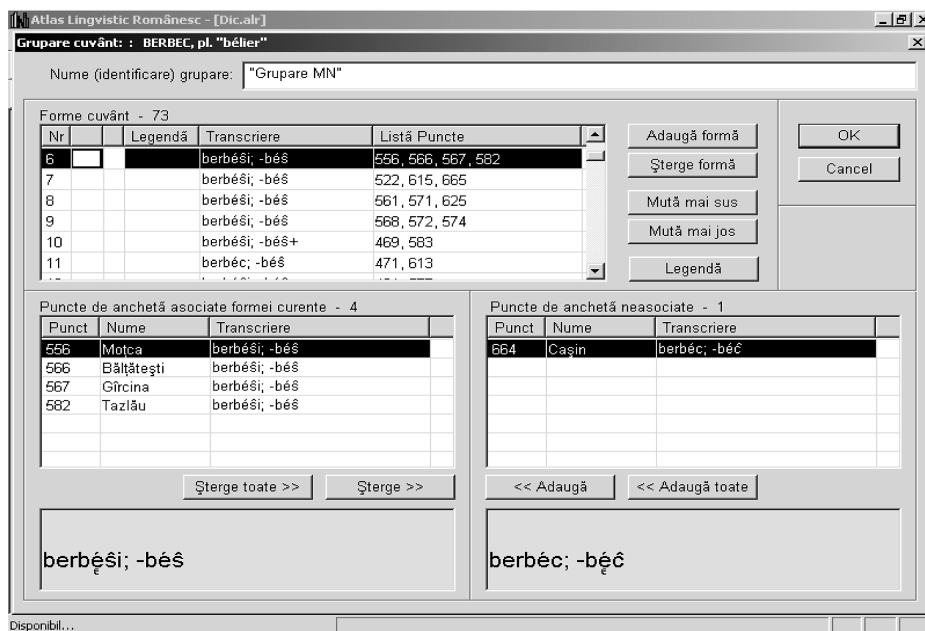


Figura 6. Fereastra *Grupare cuvânt*

3.3. Un alt avantaj important oferit de acest sistem este faptul că, în final, soluțiile de ordin informatic propuse de acesta vor putea fi „exportate” și folosite de toți specialiștii implicați în cercetări ce țin fie de domeniul geografiei lingvistice, fie de domenii adiacente, cum ar fi istoria limbii, socio- și etnolingvistica etc., dar care folosesc, în mod ocazional, instrumentarul acesteia

(ne gândim aici, în primul rând, la transcrierea fonetică și la hărțile sintetice). Elaborarea de sinteze de istoria limbii române sau de dialectologie va fi mult simplificată, prin folosirea facilităților pe care le oferă acest program. În plus, sistemul de față poate deveni o componentă de bază a instrumentarului didactic al profesorului de limba română, acesta având acum posibilitatea să ofere auditoriului, într-o manieră cât se poate de captivantă (sunet + policromie), prelegeri de dialectologie, istoria limbii române, gramatică sau semantică dublate de un material ilustrativ (în special, hărți sintetice) relevant pentru problemele prezentate.

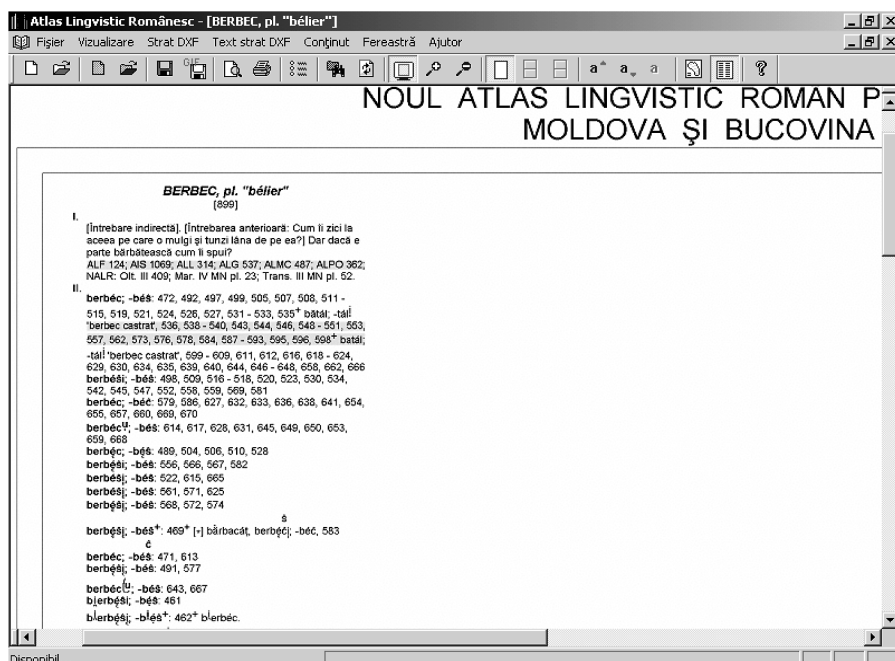


Figura 7. Planșă Material Necartografiat

Posibilitatea extinderii aplicabilității rezultatelor prezentului proiect conferă acestuia un statut privilegiat în cadrul cercetării academice interdisciplinare românești.

3.4. În forma sa finală, sistemul de față va fi compus din trei sub-sisteme (module), după cum urmează:

a) primul sub-sistem va permite stocarea informațiilor în dicționarele atlasului lingvistic electronic, precum și, pe baza materialului dialectal introdus în aceste dicționare, generarea și editarea hărților lingvistice (analitice sau sintetice) și a materialului necartografiat pentru cuvintele selectate;

b) al doilea sub-sistem va fi responsabil de editarea și consultarea, la nivel *grafic*, a hărților interpretative și a planșelor cu material necartografiat, în

calitatea acestora de părți componente ale volumului final. Acest sub-sistem va permite, pe baza informațiilor generate de primul sub-sistem, aranjări în pagină în funcție de imaginea de ansamblu a hărților și a planșelor, precum și inserări de noi elemente a căror prezență în formatul final este obligatorie (cum ar fi, de exemplu, legenda hărților interpretative);

c) al treilea sub-sistem va fi alocat editării textelor dialectale, în calitatea acestora de parte componentă a proiectului NALR.

În acest format, proiectul de față va putea fi extins la nivel național, având în vedere faptul că din cele șapte atlase românești regionale, doar două (*NALR. Oltenia* și *ALRR. Maramureș*) au fost finalizate, celelalte cinci aflându-se în diferite stadii ale redactării.

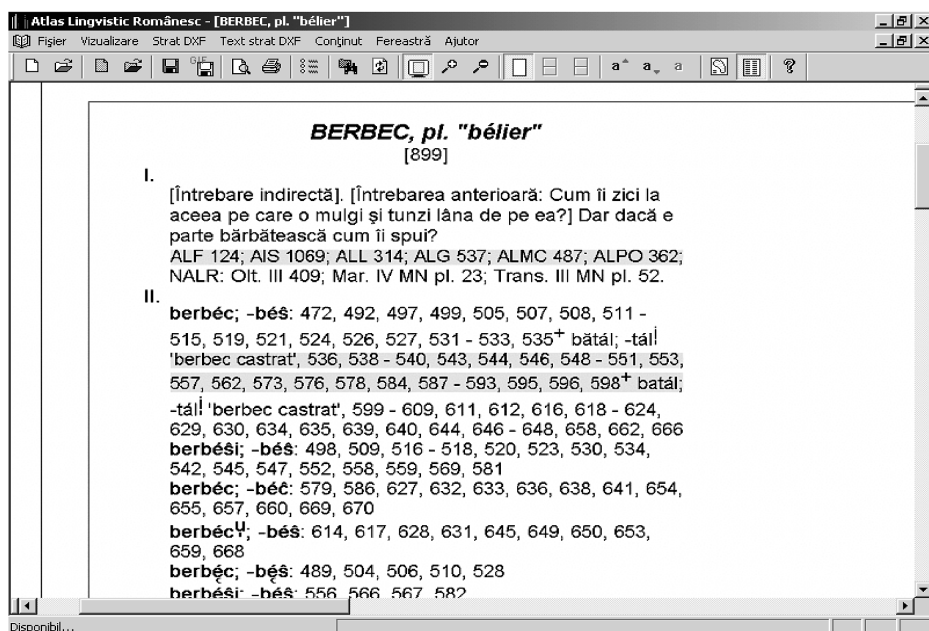


Figura 8. Planșă Material Necartografiat (detaliu)

3.5. În sfârșit, acest sistem, datorită stocării informațiilor în format electronic, oferă șansa unei promovări extrem de importante (în condițiile actuale de difuzare a informației) a atlaselor lingvistice românești regionale, și anume, prezența acestora, în format electronic, pe un sit gestionat de Academia Română, la care să aibă acces, într-o manieră elegantă și, nu în ultimul rând, avantajoasă ambelor părți, orice cercetător român sau străin interesat de probleme de dialectologie românească⁹. În acest sens, ar fi de dorit ca, imediat ce s-ar finaliza editarea ultimelor volume din seria NALR, să se înceapă introducerea în format electronic

⁹ Menționăm că deja se poate consulta o secvență dintr-o hartă lingvistică în format electronic la următoarea adresă: <http://philippide.is.edu.ro/pages/harta-dialect.html>

și a datelor cuprinse în volumele deja editate în format tradițional, numai așa reușindu-se ca întreg materialul NALR să fie disponibil pe Internet tuturor celor interesați.

Posibilitatea utilizării unei *Colecții audio*, care să dubleze, acolo unde este posibil, transcrierile fonetice din punctele de anchetă, precum și a policromiei (vezi nota 3) în scopul generării unei interfețe mult mai atrăgătoare pentru cititor constituie premise importante în direcția continuării cercetărilor în acest sens. Rezultatele obținute până la acest moment ne dau speranța că în viitorul apropiat aceste deziderate vor putea fi valorificate în primul rând în direcția unei mai bune cunoașteri a structurii dialectale a limbii române atât în rândul specialiștilor, dar și a tuturor celor interesați de astfel de probleme.

Nouvelles contributions à la publication assistée par ordinateur des atlas linguistiques roumains régionaux

La publication assistée par ordinateur des atlas linguistiques et des textes dialectales est un projet qui se déroule à l'heure actuelle dans le centre académique de Iassy. Les chercheurs de l'Institut d'Informatique Théorique et de l'Institut de Philologie Roumaine „A. Philippide” ont trouvé les meilleures solutions pour générer et éditer, à l'aide de l'ordinateur, les atlas linguistiques avec leurs composantes de base: les cartes analytiques, les cartes synthétiques (interprétatives) et les planches avec le matériel non-cartographié. Les résultats obtenus jusqu'à ce moment sont en mesure de justifier l'espoir qu'à l'avenir les atlas linguistiques roumains puissent être consultés en format électronique, sur un site de l'Académie Roumaine.