

Cercetări recente în domeniul informatizării atlaselor lingvistice românești

Florin-Teodor OLARIU, Veronica OLARIU
Marius-Radu CLIM, Ramona LUCA *

Key-words: *Romanian linguistic geography, digitized cartography, multimedia linguistic atlases, online linguistic databases*

0. Tehnologiile informatice oferă la ora actuală noi posibilități de dezvoltare/ amenajare a resurselor necesare activităților de cercetare. În aceste condiții – ale tehnologizării și informatizării cvasitotale a societății contemporane –, cercetarea filologică se află într-un plin proces de redimensionare a propriilor strategii euristice, în direcția unei compatibilizări a metodelor și a resurselor clasice de cercetare folosite în acest domeniu (dicționare, atlase, ediții de text etc.) cu noile tehnologii informatice. Crearea acestor noi instrumente de cercetare are ca prim rezultat posibilitatea promovării la o scară mult mai mare a patrimoniului cultural non-material, prin facilitarea accesului unui număr practic nelimitat de utilizatori (prin utilizarea tehnologiilor Web) la baze de date cu statut de arhive digitale interactive.

Cercetările de ultimă oră ce au ca obiect informatizarea limbii române se înscriu în direcția sincronizării metodelor și tehnicilor folosite de cercetătorii români cu cele ale omologilor lor din celelalte țări europene. Referitor la acest aspect al modernizării metodelor de analiză aplicate limbii române, precizăm faptul că, prin colaborarea excelentă care a existat în ultimul timp între lingviști și informaticieni, s-a ajuns la unele rezultate extrem de promițătoare în ceea ce privește posibilitatea prelucrării din punct de vedere informatic a diferitelor compartimente ale limbii române. Amintim, în acest sens, încercările care au fost făcute, începând cu anul 2000, de creare a unui *Consortiu de Informatizare pentru Limba Română* (<http://consilr.info.uaic.ro/>) și cercetările pe care, ulterior, membrii acestui consorțiu le-au făcut în direcția includerii limbii române în cadrul unor proiecte internaționale de anvergură, având ca scop crearea unor ontologii lexicale multilingve (EuroWordNet și BalkaNet) (Tufiș 2004); de asemenea, pornind de la această experiență în domeniul procesării electronice a limbii române, trebuie menționate recentele cercetări, vizând același nivel lexical-semantic, ce au avut ca rezultat

* Academia Română, Filiala Iași.

Această lucrare a fost realizată în cadrul proiectului TE_244/2010, finanțat de CNCS-UEFISCDI prin contractul nr. 54/28.07.2010.

crearea cadrului general al editării în format electronic a *Dicționarului limbii române* (DLR), în calitatea acestuia de dicționar tezaur, proiect la realizarea căruia au contribuit cercetători lingviști și informaticieni din centrele academice București, Iași și Cluj (Aldea et alii 2006, Dănilă 2010).

Pe lângă aceste inițiative ce au avut în vedere prelucrarea din perspectivă informatică a nivelului lexical al limbii române, există cercetări la ora actuală ce vizează și compartimentul fonetic/fonologic sau chiar nivelul suprasegmental al intonației. Amintim, cu referire la această ultimă direcție de cercetare, proiectul *Atlasul multimedia prozodic român* (AMPROM), în cadrul căruia, începând mai ales cu anul 2007, sunt documentate și analizate, folosindu-se aplicații informatice adecvate acestui tip de cercetare, principalele tipare intonative ale varietăților limbii române, atât la nivel dialectal, cât și la cel al limbii literare marcate regional, pe o rețea de puncte de anchetă în care vor fi incluse câteva zeci de localități din România și din Republica Moldova (Turculeț, Botoșineanu, Minuț 2006). Acest atlas informatizat, având structura unei baze de date, cuprinde atât înregistrările audio ale tiparelor intonative urmărite prin intermediul unui chestionar fix (enunțuri asertive, interogative, negative și interogativ-negative), cât și reprezentările grafice ale respectivelor tipare, obținute cu ajutorul unor programe speciale de prelucrare acustică folosite în acest scop (<http://amprom.uaic.ro/indexamprom.php>). Fiind conceput inițial ca o componentă regională a unui proiect similar de anvergură europeană – *Atlas multimédia prosodique de l'espace roman* (AMPER) – proiectul AMPROM s-a dezvoltat în ultima perioadă ca un proiect autonom, destinat unei configurări cât mai detaliate a specificului intonațional al limbii române.

Un al proiect destinat analizei la nivel fonetic a limbii române este cel prin care se intenționează crearea, în cadrul Institutului de Informatică Teoretică al Academiei Române, Filiala Iași, a Corpusului „Sunetele Limbii Române” (http://www.etc.tuiasi.ro/sibm/romanian_spoken_language/ro/prezentare_generala.htm), având ca principal scop constituirea unei arhive a sunetelor limbii române în format digital, pentru care, până la ora actuală au fost înregistrate peste 1500 de secvențe audio distincte. Ca scopuri derivate ale acestui proiect (și care ne interesează în mod direct din perspectiva lucrării de față), menționăm crearea unui dicționar electronic al pronunțiilor din limba română, precum și intenția autorilor de corelare a arhivei astfel constituite cu atlasele lingvistice românești, pentru realizarea unor cercetări de tip diacronic în domeniul foneticii și fonologiei limbii române. Înregistrările pentru această arhivă au fost efectuate în laborator, în condiții controlate și cu zgomot redus, pe un eșantion de vorbitori cu vârsta cuprinsă, în general, între 20 și 35 de ani, fără patologii cronice sau acute (Teodorescu 2010: 17). În plus, a fost respectată o metodologie riguroasă, ce a implicat luarea în considerare a nu mai puțin de patru criterii obligatorii în realizarea corpusului (criteriul reprezentativității statistice, criteriul reproductibilității, criteriul completitudinii și criteriul eticii științifice), acest cadru metodologic fiind completat și de realizarea, pentru fiecare caz în parte, a fișei vorbitorului, unde au fost trecute datele relevante privind caracterizarea socio-profesională și medicală a acestuia. Baza de date astfel constituită a reprezentat punctul de plecare, până la ora actuală, a unor cercetări de factură statistică privind expresivitatea emotivă în limba română, având ca scop

stabilirea unei metodologii și elaborarea unui program de analiză a intonației emotive (Teodorescu et alii 2010).

Toate aceste proiecte și inițiative științifice enumerate mai sus privind procesarea electronică a limbii române confirmă o stare de lucruri existentă la ora actuală, și anume prezența unui cadru general favorabil proiectelor de factură interdisciplinară care să vizeze realizarea unui deziderat major al cercetării academice românești: informatizarea resurselor limbii române, atât ca scop în sine, dar mai ales ca punct de plecare pentru viitoare cercetări în domeniu, care să sincronizeze cercetarea de profil românească cu cea europeană.

1. Geografia lingvistică românească, a cărui centenar a fost sărbătorit recent (1909: Gustav Weigand, *Linguistischer Atlas des dacorumänischen Sprachgebietes*, Leipzig), are o bogată tradiție în domeniul atlaselor lingvistice, nu mai puțin de trei generații de astfel de lucrări, toate cu caracter național, văzând lumina tiparului în ultima sută de ani. Pe lângă aceste trei proiecte de anvergură, unele încheiate sau altele doar parțial finalizate până la acest moment, au existat și alte inițiative de-a lungul timpului, care și-au propus fie cartografierea lingvistică a unor zone istorice ale arealului dacoromânesc (I.-A. Candrea, *Atlasul lingvistic al Banatului*), fie realizarea unor atlase ale dialectelor sud-dunărene, pentru acest din urmă tip de lucrări manifestând interes atât cercetători români, cât și specialiști din străinătate¹. Una dintre ultimile inițiative în domeniul cartografiei lingvistice românești de tip clasic o reprezintă elaborarea, începând cu anul 1998, a *ALRR. Sinteză*, al cărui prim volum, apărut în anul 2005, plasează limbă română pe o poziție privilegiată între celelalte limbi romanice, aceasta devenind „singura care, în mai puțin de un secol, dispune de trei atlase lingvistice consacrate graiurilor dacoromâne în ansamblu” (Saramandu 2005: V).

Această bogată tradiție în domeniul cartografiei lingvistice – tradiție recunoscută și apreciată ca atare la nivel european încă din prima parte a secolului al XX-lea, când, odată cu apariția primelor volume din *Atlasul lingvistic român* (1938–1942), lingviștii români introduceau o serie întregă de inovații în domeniu – a devenit în ultima perioadă o puternică motivație pentru cercetătorii din domeniu în direcția modernizării metodei cartografice, în special prin apelul la tehnologiile informatice de ultimă generație, capabile să asigure suportul logistic necesar redactării asistate de calculator a atlaselor lingvistice și, într-o etapă ulterioară, elaborării unor variante online ale acestor lucrări. În cele ce urmează vom face o prezentare a principalelor inițiative de acest gen existente la ora actuală pe teren românesc, cu scopul de a evidenția eforturile specialiștilor români în direcția sincronizării cartografiei lingvistice românești cu cea europeană.

1.1. Cercetările în direcția informatizării atlaselor lingvistice reprezintă un nou stadiu în dezvoltarea metodei geografiei lingvistice europene, tehnologia informației fiind folosită din ce în ce mai frecvent în procesul editării acestor lucrări de cartografie lingvistică². În spațiul românesc, cercetări susținute în această direcție au

¹ Pentru o prezentare a principalelor realizări ale cartografiei lingvistice românești, v. Olariu 2010.

² Maniera în care a evoluat, de-a lungul ultimelor două decenii, informatizarea cartografiei lingvistice la nivel european poate fi urmărită prin parcurgerea (și a) următoarelor lucrări: Ruffino 1992, Mouton 1994, Raimondi, Revelli 2008, Cugno, Massobrio 2010.

început în anul 2000, în cadrul Filialei Iași a Academiei Române, printr-o colaborare interdisciplinară între cercetători de la Institutul de Informatică Teoretică și de la Institutul de Filologie Română „A. Philippide”. Scopul acestui proiect a constat în realizarea unui sistem informatizat cu ajutorul căruia să se poată edita atlasele lingvistice românești regionale, precum și textele dialectale aferente respectivelor atlase. Cercetările s-au desfășurat în cadrul a două programe ale Academiei Române, intitulată *Cercetări în domeniul foneticii și fonologiei limbii române, cu aplicații la atlasele lingvistice românești regionale* (2000–2002) și, respectiv, *Proiectarea și implementarea unui sistem integrat de aplicații software pentru realizarea „Noului Atlas lingvistic român, pe regiuni” și editarea textelor dialectale* (2003–2004).

Aplicația software elaborată în cadrul acestei colaborări interdisciplinare conține module care realizează gestionarea următoarelor grupe de informații:

- simboluri pentru editarea transcrierilor fonetice;
- dicționarele atlasului lingvistic (cuvinte de bază, puncte de anchetă, transcrieri fonetice);
- informații grafice pentru descrierea hărților;
- hărțile atlasului lingvistic, care pot fi consultate și/ sau tipărite.

De cea mai mare importanță s-a dovedit a fi organizarea familiei de fonturi pentru transcrierea fonetică. În acest sens, s-a optat pentru continuitate, adică pentru acele litere, însoțite de semne grafice speciale (sedile, linii orizontale sau verticale, semicercuri, cercuri, puncte, unghiuri ascuțite, tilde ș.a.), cu care, în atlasele lingvistice românești (ALR și NALR) se marchează sunetele vorbirii, așa cum se realizează ele în graiurile populare.

Odată rezolvată problema fonturilor, a fost nevoie de o structurare a informațiilor lingvistice și grafice necesare redactării atlasului lingvistic. Soluția a fost găsită prin configurarea unor *Dicționare* în care să se realizeze colectarea informațiilor primare privind titlul hărților (cuvinte de bază), punctele de anchetă, *speech* (colecție audio), transcrierea fonetică și notele asociate transcrierii fonetice³. Fiecare dintre aceste dicționare are o funcție bine definită, după cum urmează :

a) *Dicționar „Cuvinte de bază”*: conține fondul de cuvinte (titlul hărților) care pot fi găsite în atlasul lingvistic electronic, întrebările care au fost puse la anchetă, note, observații și, eventual, imagini ce pot fi folosite în întrebările indirecte;

b) *Dicționar „Puncte de anchetă”*: conține informații (cod, nume, observații) despre punctele de anchetă prezente în cadrul atlasului lingvistic;

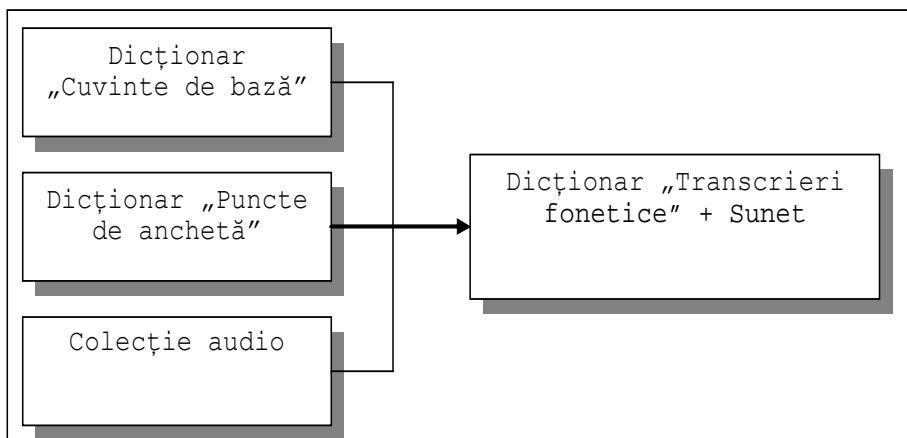
c) *Colecția audio*: include înregistrările în format WAV a răspunsurilor date în timpul anchetei, pentru fiecare cuvânt din dicționar. Această colecție se poate constitui doar dacă ancheta dialectală a fost concepută *ab initio* a se desfășura „în fața microfonului”, înregistrările astfel realizate putând fi ulterior prelucrate pentru a intra în baza de date a atlasului electronic.

d) *Dicționar „Transcrieri fonetice”*: reprezintă elementul principal al modulului și conține transcrierea fonetică a răspunsului la întrebarea pusă în anchetă

³ Rezultatele cercetărilor de până acum au făcut obiectul mai multor comunicări și articole realizate de membrii echipei, dintre care enumerăm: Florea et alii 2002, Olariu et alii 2003, Apopei et alii 2003, Bejinariu et alii 2006, Olariu et alii 2008.

pentru fiecare cuvânt din Dicționarul „Cuvinte de bază” în fiecare punct de anchetă, iar acolo unde este posibil și înregistrarea audio corespunzătoare din *Colecția Audio*.

Conexiunile dintre informațiile stocate în dicționarele ALR sunt reprezentate în figura următoare:



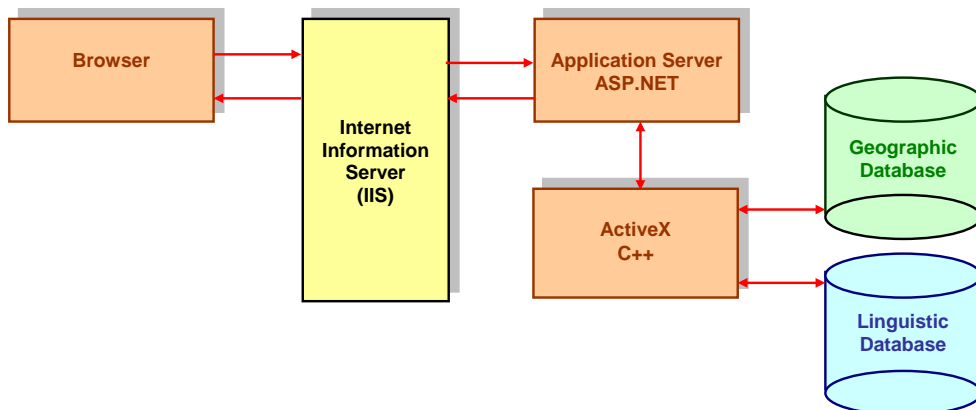
După o perioadă de testări și îmbunătățiri a variantelor succesive ale aplicației software, în anul 2005 a fost finalizat programul de editare asistată de calculator a atlasului lingvistic⁴ iar în 2007 a apărut volumul al III-lea din NALR. *Moldova și Bucovina* editat în întregime cu ajutorul calculatorului – o premieră la nivel național, primită în mod pozitiv de specialiștii din domeniu.

În ceea ce privește perspectivele acestui proiect științific, ele sunt legate de intenția de a dezvolta aplicația în direcția realizării variantei interactive care va facilita accesul on-line la baza de date a NALR. *Moldova și Bucovina*. Primele cercetări în acest sens au fost făcute deja – în aceeași formulă de colaborare interdisciplinară între dialectologi și informaticieni –, obiectivul constituindu-l de data aceasta publicarea în format Web a materialului lingvistic cuprins în primele două volume a atlasului amintit⁵. Stocarea pe un server a acestui material, la care se va putea adăuga și baza de date aferentă volumului al III-lea, deja digitalizată, precum și, în viitor, a întregii arhive din cadrul proiectului NALR. *Moldova și Bucovina*, arhivă în curs de publicare în format electronic, va permite utilizatorilor un acces direct la acest material dialectal, pe care îl vor putea consulta și utiliza în conformitate cu propriile proiecte de cercetare. Consultarea atlasului se va putea face prin intermediul unei aplicații Web, care va comunica prin intermediul unui server IIS cu aplicația ce va rula pe server. Aceasta va utiliza un control ActiveX, implementat în C++, și va genera, pe de o parte, conținutul grafic ce va fi transmis către client, iar, pe de altă parte, va asigura conexiunea cu cele două tipuri de baze de date existente în cadrul programului: bazele de date geografice și bazele de date

⁴ Aplicația informatică a fost distinsă în anul 2006 cu Premiul „Octav Mayer” al Academiei Române, Filiala Iași.

⁵ O prezentare mai detaliată a acestui nou proiect de digitalizare a cartografiei lingvistice românești se găsește în Botoșineanu, Olariu, Bejinariu.

lingvistice. Relaționarea acestor aplicații în cadrul programului Web propus de cercetătorii ieșeni poate fi urmărită în figura următoare:



Data fiind actualitatea acestui tip de cercetare la nivel european, precum și avantajele evidente pe care le poate aduce lingvisticii românești considerată în ansamblul său, credem că extinderea în viitor a acestei aplicații la nivel național și, drept urmare a acestui fapt, punerea în circuitul științific prin intermediul Internetului a întregului material lingvistic cules în cadrul proiectului NALR poate reprezenta o miză majoră pentru dialectologia românească la ora actuală.

1.2. O inițiativă similară de informatizare a atlantografiei lingvistice românești îl are ca autor pe Dorin Urișescu, dialectolog român stabilit în Canada, care, împreună cu doi specialiști informaticieni – Sheila M. Embleton și Eric S. Wheeler, au realizat cercetări în ultima perioadă în direcția digitalizării *Noului Atlas lingvistic român. Crișana*, atlas din care, până la ora actuală, au fost publicate primele două volume în format clasic. Bazându-se pe studii recente în domeniul dialectometriei (Hans Goebel) și al lingvisticii cantitative⁶, specialiștii menționați și-au propus ca, folosind tehnici statistice de genul scalării multidimensionale⁷ (al cărei principal avantaj din punct de vedere analitic este, așa cum arăta lingvistul J.K. Chambers, acela de a ignora distanțele geografice și de a reprezenta doar distanțele statistice ale fenomenelor urmărite), să realizeze o aplicație informatică aptă să ofere lingviștilor posibilitatea unor analize complexe ale variației diatopice prin utilizarea unui volum mare de material lingvistic în format digitalizat.

Sub aspect tehnic, la fel ca și în cazul proiectului de la Iași, o importanță majoră în dezvoltarea aplicației a avut-o realizarea sistemului de transcriere fonetică, sistem care, în tradiția dialectologiei românești, presupune utilizarea unui număr relativ important de caractere suplimentare în raport cu sistemul grafic de bază. Acest aspect a fost rezolvat prin fixarea caracterelor sub formă de imagini într-un editor sub formă de tastatură virtuală, cu ajutorul căruia a fost digitalizată transcrierea fonetică a întregului material lingvistic (Embleton, Urișescu, Wheeler 2007: 90).

⁶ Proiectul de față, intitulat *Romanian Online Dialect Atlas (RODA)*, se bazează pe experiența pe care cei doi informaticieni au acumulat-o odată cu informatizarea *Atlasului lingvistic al Finlandei*.

⁷ Pentru mai multe detalii, vezi Embleton, Urișescu Wheeler 2007, precum și pagina de Internet <http://www.yorku.ca/vpaweb/romanian/index.htm>.

Există și alte similitudini cu proiectul dezvoltat la Iași, care constau în: a) posibilitatea redării în cadrul aplicației a unor înregistrări audio de material dialectal; b) posibilitatea căutării unei secvențe grafice (transcriere fonetică) în întreaga bază de date a atlasului (Embleton, Uritescu, Wheeler 2008: 10), bază care se poate ridica la câteva sute de hărți lingvistice și care va cuprinde, astfel, o cantitate mare de material dialectal; c) posibilitatea adăugării manuale de simboluri aferente unor puncte cartografice (simboluri care să corespundă unui anumit fenomen urmărit), cu scopul elaborării unor hărți interpretative (Embleton, Uritescu, Wheeler 2008: 19)

O componentă majoră a programului informatic este aplicația cu ajutorul căreia se realizează hărțile interpretative, aceasta fiind poate cea mai importantă contribuție a autorilor în cadrul proiectului respectiv. Folosindu-se tehnica scalării multidimensionale⁸, utilizatorii pot realiza, cu ajutorul unui inventar predefinit de simboluri, diferite hărți interpretative de tip dialectometric, pe care vor putea fi redade „distanțele lingvistice” dintre punctele unei rețele de anchetă, în sensul diferențelor sub aspect lingvistic existente între respectivele localități anchetate (Embleton, Uritescu, Wheeler 2008: 23). Pornind de la acest aspect, una dintre facilitățile importante oferite de acest program informatic constă în faptul că, odată introdus în baza de date, materialul lingvistic va putea susține cu ușurință elaborarea unor hărți interpretative independente și conforme intereselor științifice ale fiecărui utilizator, această operațiune fiind destul de dificil de realizat (sub raport temporal și ca volum de muncă) în formatul clasic al atlaselor lingvistice:

As we continue our work, we hope to make the Romanian data available interactively, so that users can define the groups they want to see, and then see them in either a geographic or a linguistic map. Further, it should be possible to access individual items of data from the map directly. In this way, we hope that users of our online atlas will be able to define what they want to see of the data. There is still considerable work left before the interactive interface to the data is ready [Proiectul nostru continuă și sperăm să facem interactive datele atlasului, așa încât utilizatorii să aibă posibilitatea de a-și defini grupurile pe care doresc să le vadă și apoi să poată să le vizualizeze atât sub formă de hartă geografică, cât și sub formă de hartă lingvistică. În plus, s-ar putea facilita accesul la datele individuale direct de pe hartă. În acest caz, sperăm ca utilizatorii atlasului nostru online să poată să-și definească singuri ceea ce vor să vadă din baza de date disponibilă. Rămâne însă un efort considerabil de făcut până când interfața interactivă a bazei de date va fi finalizată] (Embleton, Uritescu, Wheeler 2007 : 96)⁹.

⁸ Referindu-se la această modalitate de redare a distanțelor abstracte care separă un set de puncte, autorii afirmă : „This process is analogous to making a shadow of your hand (the n-dimensional object) on a wall (the two-dimensional plane); in certain positions, the shadow will be an effective representation of your hand, and in other position the shadow is less effective” [Acest proces este similar cu umbra propriei mâini (un obiect cu „n” dimensiuni) pe un perete (un plan cu două dimensiuni); în unele poziții umbra va fi o reprezentare reală a mâinii iar în alte poziții umbra va fi una mai puțin corespunzătoare cu aceasta] (Embleton, Uritescu, Wheeler 2008: 23).

⁹ Chiar dacă informatizarea cartografiei lingvistice aduce avantaje evidente în domeniul analizei materialului dialectal, în primul rând prin posibilitatea analizei automate a unei mari cantități de date, autorii nu exclud intervenția manuală în astfel de procesări de date: „It is never certain that an automated search will return exactly what the user intended to search for. Therefore, it is important to have functions for manually reviewing and adjusting any automated search and count” [Niciodată nu este sigur că o căutare automată va oferi utilizatorului exact ceea ce acesta caută. În consecință, este

Până la această oră au fost digitalizate datele primare cuprinse în cele două volume tipărite din NALR. *Crișana*, autorii proiectului realizând, cu ajutorul aplicația pusă la punct de ei, primele cercetări asupra materialului lingvistic astfel stocat, cercetări care au vizat, în special, compartimentul fonetic al subdialectului crișan.

1.3. Ultimul proiect (în ordine cronologică) de informatizare a atlaselor lingvistice românești aparține din nou specialiștilor dialectologi și informaticieni din centrul academic ieșean, această inițiativă venind ca o continuare firească a experienței în domeniu acumulată de aceștia de-a lungul ultimilor ani. Proiectul poartă titlul *Atlasul lingvistic audiovizual al Bucovinei* (ALAB), prima fază de dezvoltare a acestuia fiind preconizată pentru anii 2010–2013, prin intermediul unui grant oferit de Consiliul Național al Cercetării Științifice (CNCS).

Un punct de plecare pentru acest proiect l-a constituit experiența celor câteva proiecte europene similare dezvoltate până la ora actuală, primul dintre acestea fiind *Sprachatlas des Dolomitenladinischen und angrenzender Dialekte* – ALD (<http://ald.sbg.ac.at/ald/ald-i/>), care a fost inițiat în anii '80 ai secolului trecut de către o echipă de dialectologi și informaticieni de la Universitatea din Salzburg. Până la ora actuală au fost editate șapte volume în format clasic din prima parte, plus trei CD-uri incluzând varianta digitalizată a bazei de date aferente ALD-I, lucrările pentru partea a doua fiind și acestea destul de avansate (<http://ald2.sbg.ac.at/a/index.php/en/>). Această lucrare de pionierat în domeniu a deschis o nouă perspectivă cercetărilor de geolingvistică (prin posibilitatea redării sunetului odată cu afișarea transcrierii fonetice a răspunsurilor), metodologia folosită în acest caz fiind urmată și de alte echipe de cercetători. O continuare directă (prin metodologie și rezultate similare) a acestei activități de cercetare îl reprezintă proiectul *Vivaio Acustico delle Lingue e dei Dialetti d'Italia* – VIVALDI (<http://www2.hu-berlin.de/Vivaldi/index.php?id=0002&lang=it>), a cărui idee a fost lansată de doi dintre autorii ALD, ce și-au propus să creeze un atlas lingvistic acustic la nivel dialectal pentru întreaga Italia (până la ora actuală aproximativ o treime din teritoriul Italia dispune deja de răspunsurile on-line la această ancheta). Un ultim exemplu pe care dorim să-l menționăm aici și care ilustrează această nouă direcție de cercetare în domeniul atlantografiei lingvistice europene este *Atlas linguistique audiovisuel du Valais romand* – ALAVAL (<http://www.unine.ch/dialectologie/AtlasPresent.html>), un atlas care, preluând metodologia ALD, a încercat să o dezvolte, încorporând posibilitatea stocării și a redării, cu ajutorul programului informatic folosit, și a imaginilor video, pe lângă utilizarea sunetului. Proiectul a început în 1994, prin redactarea chestionarului și efectuarea primelor anchete, acestea fiind finalizate în anul 2000. Și acest atlas lingvistic a fost publicat atât pe suport de hârtie, cât și on-line, fiecare din cele două versiuni având fixate scopuri diferite: versiunea imprimată cuprinde numeroase analize lingvistice suplimentare în raport cu versiunea informatizată, în timp ce aceasta din urmă are rolul de a reda materialul lingvistic sub forma unui document global, care combină aspectul verbal cu cel non-verbal al actului lingvistic. Prin caracterul complementar al celor două versiuni, ALAVAL răspunde

important să existe funcții manuale de verificare și corectare a oricărei căutări sau operații automate] (Embleton, Uritescu, Wheeler 2008: 10).

necesității unei analize polifuncționale (la nivel morfosintactic, lexical, intonațional sau chiar pragmatic) a materialului dialectal înregistrat.

În stadiul actual al cercetărilor din domeniu, proiectul ALAB încearcă să răspundă unei duble provocări: pe de o parte, intenției de a salvagarda, într-un format multimedia interactiv (posibil ca urmare a folosirii noilor tehnologii informatice), a diversității dialectale din zona etnoculturală a Bucovinei – o regiune recunoscută drept o zonă mozaic din punct de vedere dialectal, cu o lungă și bogată tradiție în sfera contactelor lingvistice și a relațiilor interculturale; pe de altă parte, hotărârii asumate cu deplină responsabilitate de echipa proiectului de a realiza, în premieră la nivel național, un atlas lingvistic audiovizual care, ținând cont de rezultatele la nivel european obținute până în acest moment, să dezvolte pe teren românesc această tehnologie, prin includerea în cadrul propriului program informatic a celor mai recente realizări în domeniu. În același timp, se are în vedere ca în elaborarea programului să se țină cont de tradiția și de specificul cartografiei lingvistice românești, ale cărei realizări s-au bucurat de-a lungul timpului de aprecierea specialiștilor străini, în principal datorită spiritului lor inovativ. Echipa proiectului, cu o expertiză științifică în domeniu deja confirmată, își propune să răspundă în condiții optime acestor provocări și să confirme, astfel, încă o dată apetența cercetătorilor români pentru soluții inovative în domeniul geografiei lingvistice.

Din punct de vedere metodologic, strategia de cercetare avută în vedere va fi tributară în cea mai mare parte experienței românești în acest domeniu, o experiență care, acumulată de-a lungul unui secol de cercetări în domeniu, constituie un punct de plecare obligatoriu și, în același timp, extrem de binevenit pentru orice nouă întreprindere de acest gen. Stabilirea rețelei de localități, realizarea unui chestionar – instrument de lucru, efectuarea unor anchete de probă sau găsirea unor informatori adecvați îndeplinirii obiectivelor propuse reprezintă tot atâtea elemente comune cu metodologia folosită în cadrul proiectelor naționale ALR/ NALR(ALRR). În plus, ca urmare a intenției de a asigura o continuitate a proiectului în raport cu atlasele lingvistice românești anterioare (aceasta fiind, de altfel, una dintre caracteristicile generale ale cartografiei lingvistice românești), se va avea în vedere ca rețeaua de puncte ALAB să includă rețeaua NALR, iar chestionarul să cuprindă, de asemenea, un număr suficient de mare de termeni și noțiuni documentate în cadrul proiectelor anterioare. Estimăm că varianta chestionarului folosit în anchete va cuprinde aproximativ 1000 de întrebări, grupate pe o serie de teme relative la viața cotidiană, stabilirea acestora făcându-se după o analiză atentă, pe baza materialului lingvistic cuprins în volumele publicate din seria NALR/ALRR, a gradului de relevanță lingvistică (la nivel fonetic, morfosintactic și lexical) a răspunsurilor obținute pe baza chestionarului NALR. Ca tipologie a informatorilor, dorim ca în fiecare localitate să anchetăm reprezentanți ai unor diferite categorii de vârstă, sex sau statut social, pentru a putea surprinde, astfel, pe lângă variația diatopică, și variațiile diastratică sau chiar diafazică. De asemenea, se va avea în vedere (în condițiile în care acest lucru mai este posibil, după o perioadă de aproximativ 40 de ani de la ancheta NALR) ca și unii dintre subiecții anchetați în cadrul proiectului NALR să fie recuperați în cadrul anchetei ALAB, în ideea decelării variației diacronice (eventual diastratic) a idiolectelor respective. Acest ultim aspect al anchetei poate deveni unul dintre pilonii principali ai proiectului de față, datorită șansei pe care

zona Bucovinei a avut-o în cadrul proiectului NALR – aceea ca răspunsurile la anchetele realizate cu chestionarul în unele localități din această regiune să fie înregistrate pe bandă de magnetofon, situație singulară pentru întreaga zonă a Moldovei. Posibilitatea de a auzi și compara on-line pronunția aceluiași idiolect pentru un același termen sau aceeași noțiune rostită la un interval de aproximativ patru decenii ar reprezenta o inovație importantă a proiectului ALAB.

În ceea ce privește partea informatică a proiectului, realizarea sistemului presupune, într-o primă fază, structurarea informațiilor în baze de date. În cadrul acestui proiect, sistemul de gestiune a bazelor de date propus, care va permite construirea, interogarea și întreținerea acestora, va fi sistemul relațional MySQL. Fiind unul dintre cele mai populare sisteme relaționale „open-source”, acest sistem poate fi achiziționat direct de pe Internet și permite administratorului crearea rapidă a unei baze de date, precum și o gestionare facilă a acesteia. Realizarea unei interfețe Web care să permită accesul utilizatorilor la informațiile stocate în baza de date va avea două componente: a) o interfață web securizată, care va fi accesibilă doar membrilor echipei proiectului ; accesul la aceasta va fi restricționat și se va face pe bază de nume utilizator și parolă, interfața urmând să asigure uneltele necesare managementului de date, cum ar fi: adăugarea de înregistrări în baza de date, modificarea (update-ul) și căutarea anumitor informații etc.; b) cea de a doua interfață va permite vizualizarea informațiilor stocate în baza de date de către toți utilizatorii interesați, prin accesarea site-ului Web, și va conține unelte de căutare care vor facilita găsirea rapidă a unor informații utile. Ca limbaj de programare necesar realizării interfeței Web, va fi folosit PHP: Hypertext Preprocessor, acesta fiind unul dintre cele mai importante limbaje de programare Web „open-source” și „server-side”, existând versiuni disponibile pentru majoritatea Web-serverelor și pentru toate sistemele de operare.

Pe lângă documentarea unor noțiuni cu ajutorul Chestionarului, este preconizată și o bază de date cu înregistrări audio-video ale unor evenimente lingvistice în situații cât mai diverse de comunicare. Înregistrarea acestor texte în format audio-video, care să fie atașate atlasului lingvistic, reprezintă un pas important în direcția realizării unui deziderat al geolingvisticii actuale – acela al atlaselor de texte, obiectiv pus în atenția specialiștilor de unul dintre cei mai avizați dialectologi contemporani, Michel Contini, pentru care aceste atlase ar reprezenta o depășire a atlaselor de cuvinte („atlanti dei testi come superamento degli atlanti di parole”). În plus, realizarea acestui nou tip de atlas lingvistic ar răspunde unui alt deziderat al geografiei lingvistice, aflat în mijlocul dezbaterilor teoretice din ultima perioadă pe marginea metodologiei anchetei dialectale, și anume, acela al trecerii de la ancheta de tip prelevare de material la ancheta dialectală de tip formă de comunicare (altfel spus, trecerea de la metoda *cuvinte și lucruri* la metoda *cuvinte și contexte comunicative*), mult mai adecvată unei redări cât mai ingenue a specificului idiomatice al unei comunități sau zone etnografice.

2. Informatizarea cartografiei lingvistice românești reprezintă la ora actuală o opțiune majoră în direcția unei dezvoltări a geografiei lingvistice în concordanță cu cerințele actuale din domeniul științelor socio-umane, cerințe care presupun abordări interdisciplinare și racordarea cercetării la contextul unei societăți informaționale. Având o tradiție de peste un secol în domeniul realizării atlaselor lingvistice,

dialectologii români au înțeles că această tradiție, pe lângă prestigiul ca atare pe care îl induce în rândul cercetărilor de profil la nivel european, presupune și asumarea unei responsabilități din partea acestora, în sensul unei adaptări continue sub aspect metodologic a cercetărilor dedicate variației diatopice a graiurilor la noile exigențe privind procesarea informatică a resurselor lingvistice. În consecință, la ora actuală există pe teren românesc mai multe proiecte privind realizarea fie a unor variante digitalizate ale atlaselor lingvistice deja existente, fie a unor atlase noi, concepute *ab initio* în format multimedia. Aceste noi proiecte de cercetare iau în considerare, de obicei, unele experiențe similare existente la nivel european, însă își dezvoltă conceptele operaționale în manieră proprie, devenind contribuții importante în domeniu sub aspect euristic și metodologic, care fac, în consecință, ca geografia lingvistică românească să reprezinte în continuare un punct de referință pentru cercetările de profil.

Bibliografie

- Aldea et alii 2006: Bogdan-Mihai Aldea, Elena Dănilă, Cristina Forăscu, Gabriela Haja, „Dicționarul limbii române” (DLR) în format electronic. Aplicații, în Elena Dănilă, Ofelia Ichim, Florin-Teodor Olariu, *Comunicare interculturală și integrare europeană*, Iași, Editura Alfa, 2006, p. 7–16.
- Cugno, Massobrio 2010: Federica Cugno, Lorenzo Massobrio, *Gli atlanti linguistici della România. Corso di Geografia linguistica*, Alessandria, Edizioni dell’Orso.
- Dănilă 2010: Elena Dănilă, *eDTLR – base de données et instrument pour la recherche pour la recherche lexicographique roumaine*, „Philologica Jassyensia”, nr. 11, 1/2010, p. 37–46.
- Apopei et alii 2003: Vasile Apopei, Silviu Bejinariu, Florin Rotaru, Florin-Teodor Olariu, *Electronic Linguistic Atlases*, în Hamid R. Arabnia (ed.), *Proceedings of the International Conference on Information and Knowledge Engineering, IKE’03*, June 23–26, 2003, Las Vegas, Nevada, USA, Volume 2, CSREA Press 2003, p. 628–633.
- Bejinariu et alii 2006: Silviu Bejinariu, Vasile Apopei, Ramona Luca, Florin-Teodor Olariu, Luminița Botoșineanu, *Electronic Linguistic Atlases. Tools for Information Analysis*, Selected papers, The 4th European Conference on Intelligent Systems and Technologies, Iași, Romania, september, 20–23, 2006, p. 107–116.
- Botoșineanu, Olariu, Bejinariu: Luminița Botoșineanu, Florin-Teodor Olariu, Silviu Bejinariu, *Un projet d’informatisation dans la cartographie linguistique roumaine: «Noul Atlas lingvistic român, pe regiuni. Moldova și Bucovina» en format électronique (e-NALR) – réalisations et perspectives*, în Actele celui de-al XXVI-lea Congres Internațional de Lingvistică și Filologie Romanică, Valencia, 6–11 septembrie 2010, în curs de apariție.
- Embleton, Uritescu, Wheeler 2007: Sheila M. Embleton, Dorin Uritescu, Eric S. Wheeler, *Romanian Online Dialect Atlas: Data Capture and Presentation*, în Peter Grzybek, Reinhard Köhler (eds), *Exact Methods in the Study of Language and Text*, Mouton de Gruyter, Berlin & New York, p. 87–96.
- Embleton, Uritescu, Wheeler 2008: *Digitalized Dialect Studies: North-Western Romanian*, Conferințele Academiei Române, Ciclul „Limba română și relațiile ei cu istoria și cultura românilor”, București, Editura Academiei Române.
- Florea et alii 2002: Ion A. Florea, Vasile Apopei, Florin-Teodor Olariu, Silviu Bejinariu, *Editarea asistată de calculator a atlaselor lingvistice și a textelor dialectale*, în Ofelia Ichim, Florin-Teodor Olariu (eds.), *Identitatea limbii și literaturii române în perspectiva globalizării*, Iași, Editura „Trinitas”, 2002, p. 211–232.
- Mouton 1994: Pilar García Mouton (ed.), *Geolingüística. Trabajos europeos*, CSIC, Madrid.

- Olariu et alii 2003: Florin-Teodor Olariu, Vasile Apopei, Silviu Bejinariu, *Noi contribuții la informatizarea editării atlaselor lingvistice românești regionale*, în Ofelia Ichim, Florin-Teodor Olariu (eds.), *Limba și literatura română în spațiul etnocultural dacoromânesc și în diaspora*, Iași, Editura „Trinitas”, 2003, p. 214–224.
- Olariu et alii 2008: Florin-Teodor Olariu, Veronica Olariu, Silviu Bejinariu, Vasile Apopei, *Los atlas lingüísticos rumanos: entre manuscrito y formato electrónico*, „Revista Española de Lingüística”, Madrid, 37 (2008), p. 215–246.
- Olariu 2010: Florin-Teodor Olariu, *Cartografia lingvistica romana: un secolo di risultati*, „Bollettino dell’Atlante linguistico italiano”, III Serie, Dispensa N. 34, 2010, p. 101–135.
- Raimondi, Revelli 2008: Gian Mario Raimondi, Luisa Revelli, *La dialectologie aujourd’hui. Acti del Convegno Internazionale „Dove va la dialettologia?”*, Saint-Vincent, Aosta, Cogne, 21–24 settembre 2006, Alessandria, Edizioni dell’Orso.
- Ruffino 1992 : Giovanni Ruffino (a cura di), *Atlanti linguistici italiani e romanzi. Esperienze a confronto*, Centro di studi filologici e linguistici siciliani, Palermo.
- Saramandu 2005: Nicolae Saramandu, *Prefață la ALRR. Sinteză*, vol. I, Editura Academiei Române, București, 2005, p. V–VII.
- Teodorescu 2010: Horia-Nicolai Teodorescu, *Metodologie pentru constituirea și analiza unui corpus adnotat de semnale vocale – cazul SROl*, în Adrian Iftene, Horia-Nicolai Teodorescu, Dan Cristea, Dan Tufiș (eds.), *Lucrările Conferinței „Resurse lingvistice și instrumente pentru prelucrarea limbii române”*, Iași, Editura Universității „Alexandru Ioan Cuza”, 2010, p. 13–22.
- Teodorescu et alii 2010: Horia-Nicolai Teodorescu, Ioan Păvăloi, Monica Feraru, *Aspecte metodologice de organizare a datelor și de analiză statistică a vocilor emoționale*, în Adrian Iftene, Horia-Nicolai Teodorescu, Dan Cristea, Dan Tufiș (eds.), *Lucrările Conferinței „Resurse lingvistice și instrumente pentru prelucrarea limbii române”*, Iași, Editura Universității „Alexandru Ioan Cuza”, 2010, p. 35–44.
- Tufiș 2004: Dan Tufiș, *BalkaNet: ontologie lexicală multilingvă*, în Costin Pribeanu, Ștefan Trășuan-Matu (eds.), *Interacțiune Om–Calculator*, București, Editura Printech, 2004, p. 9–22.
- Turculeț, Botoșineanu, Minuț 2006: Adrian Turculeț, Luminița Botoșineanu. Ana-Maria Minuț, *Un nou proiect de integrare în dialectologia romanică (AMPRom)*, în *Comunicare interculturală și integrare europeană*, volum îngrijit de Elena Dănilă, Ofelia Ichim, Florin-Teodor Olariu, Iași, Editura Alfa, 2006, p. 283–290.

Recent Researches in the Field of Computerizing Romanian Linguistic Atlases

The Romanian linguistic cartography has recently (2009) celebrated a century since its appearance, period when three series of national linguistic atlases were elaborated, doubled by other cartographic works dedicated to more restrained North or South-Danubian ethno-linguistic areas. This rich tradition in the field represents a solid argument for affirming the maturity of this research direction in Romania on one hand, but it requires a continuous modernization and its methodological update at the same time. This heuristic premise has recently been the basis of many trials of modernizing the Romanian linguistic cartography, modernization understood firstly as an involvement of the computerized methods in editing the linguistic atlases. The first results in this direction were obtained in the academic centre of Iasi where the first volume of the series NALR/ALRR (NALR. *Moldova and Bucovina*, 3rd volume) appeared in 2007, edited entirely with the help of the computer. It is also here that the first project for elaborating an audio-visual linguistic atlas in Romania is currently developed – the *Audio-Visual Linguistic Atlas of Bucovina* (ALAB).