

FLORIN-TEODOR OLARIU,
VERONICA OLARIU,
MARIUS-RADU CLIM,
RAMONA LUCA
(Iași)

***ATLASUL LINGVISTIC AUDIOVIZUAL
AL BUCOVINEI (ALAB).
PRIMA FAZĂ – UN NOU PROIECT PRIVIND
DIGITALIZAREA CARTOGRAFIEI
LINGVISTICE ROMÂNEȘTI¹***

0. Geografia lingvistică, în calitatea sa de metodă privilegiată în cercetările dialectologice, beneficiază deja de o bogată tradiție a lucrărilor în domeniu, de la apariția primelor atlase lingvistice trecând mai bine de o sută de ani. Secolul al XX-lea a constituit o perioadă de maximă efervescentă în conceperea și dezvoltarea unor proiecte de cartografie lingvistică, ce au vizat fie teritorii mai restrânse (atlase regionale), fie areale mai extinse (de la atlase naționale până la proiecte precum ALE sau ALiR). Această dinamică susținută a cercetărilor de profil a făcut ca geografia lingvistică să beneficieze în intervalul amintit de un statut bine configurat în vasta panoplie a științelor limbajului.

În ultima perioadă au fost însă voci care au încercat să avanseze ideea unei „epuizări” epistemice a acestei direcții de cercetare, argumentele în favoarea susținerii unei asemenea luări de poziție fiind structurate pe două paliere:

¹ Această lucrare, prezentată într-o primă formă la cel de-al VII-lea Congres al Societății Internaționale de Dialectologie și Geolingvistică (Viena, 23–28 iulie 2012), a fost redactată în cadrul proiectului CNCS-UEFISCDI *Atlasul lingvistic audiovizual al Bucovinei (ALAB). Prima fază*, cod TE _244 (2010–2013).

a) pe de o parte, a fost pusă în discuție însăși pertinenta gnoseologică a dialectologiei în condițiile în care graiurile, în calitatea lor de principal obiect de studiu al acestei discipline, își pierd din ce în ce mai mult statutul de unități discrete sub aspect (socio)lingvistic, variațiile inter- și intradialectale diminuându-se treptat, odată cu dispariția structurii tradiționale a comunităților rurale:

Dialectologia a suferit un proces de declin în ultima perioadă, datorită unor diverse motive: dezvoltarea unei lingvistici cu caracter universalist, constatarea dispariției universului rural tradițional – teren de cercetare privilegiat al acestei științe –, metode considerate ca învechite în comparație cu cele ale sociologiei și sociolingvisticii etc. (Le Dû 1997, p. 12).

Această linie argumentativă a fost susținută, în general, de unii sociolingviști, care vedeau în dialectologie un demers „depășit” și „arhaic” (Carton 2000, p. 129), devenit impropriu pentru studiul dinamici lingvistice din societatea contemporană;

b) pe de altă parte, semne de neîncredere în formula clasică de valorificare a cercetărilor de geolingvistică – cea a elaborării de atlase lingvistice – au venit chiar din rândul dialectologilor, care au sesizat

o restrângere [...] a câmpului de extensie epistemică a geografiei lingvistice doar la cartografia lingvistică, altfel spus [...] doar la activitatea de realizare a atlaselor, în detrimentul aspectelor explicative și interpretative care, în fond, motivase cercetarea necesară însăși realizării respectivelor atlase (Telmon 2007, p. VI).

În plus, existența unei discrepanțe majore între ritmul publicării atlaselor lingvistice și maniera în care acestea ajungeau să devină suport pentru cercetări ulterioare de mai largă anvergură dedicate configurării structurii dialectale a domeniilor astfel anchetate – fenomen ce se poate fi tradus printr-o corelare inadecvată între existența unor instrumente de cercetare (*i.e.* atlase lingvistice) relativ numeroase și slaba lor valorificare în cadrul unor cercetări științifice pentru care, practic, acestea au fost create – a dus la o diminuare a interesului pentru astfel de lucrări. Unul dintre exemplele cele mai elocvente în acest sens vine tocmai din spațiul francez – cel care a dat prima ope-

ră majoră în acest domeniu, devenită punct de reper pentru cercetările ulterioare –, unde, către sfârșitul anilor '90 ai secolului trecut, direcția de cercetare care se ocupa în cadrul CNRS cu publicarea ultimelor volume din seria atlaselor regionale și-a încetat activitatea, înainte ca respectiva serie să fie finalizată (cf. Carton 2000, p. 131).

Aceste diferite semnale care încercau să pună în discuție fie metodele, fie însuși obiectul de studiu al dialectologiei ca disciplină autonomă și, în același timp, racordată la practicile și cerințele cercetării contemporane au avut drept urmare o sensibilizare a cercetătorilor din arealul studiilor de profil înspre o problematizare de tip reflexiv asupra fundamentelor epistemologice ale respectivei discipline. Acest efort de analiză autoreflexivă a început acum mai bine de două decenii, un prim semnal în acest sens constituindu-l Colocviul intitulat „Dove va la dialettologia”, organizat în anul 1990 de renumita revistă italiană „Quaderni di Semantica”. Astfel de inițiative au continuat de-a lungul ultimelor două decenii, aici amintind doar câteva: în același an 1990 are loc o altă întrunire importantă, tot pe teren italian, dedicată, de această dată, situației cartografiei lingvistice la nivel european, și anume Congresul internațional *Atlanti linguistici italiani e romanzi. Esperienze a confronto*, Palermo; în 1991, Edgar Radtke și Harald Thun organizează simpozionul *Neue Wege der romanischen Geolinguistik*, pentru ca în 1994 să apară volumul coordonat de Pilar García Moutón intitulat *Geolinguística. Trabajos europeos*; în 1997, revista „Le Français Moderne” publică un număr intitulat *La dialectologie aujourd'hui*, iar în 2000 revista „Verbum” realizează un dosar intitulat *Géolinguistique en Europe*; în fine, în 2006 mai mulți specialiști în domeniu se reunesc în nordul Italiei (Saint-Vincent-Aosta-Cogne) în cadrul unui Colocviu care reia chiar în titlu mai vechea întrebare *Dove va la dialettologia?*, pentru ca în 2010 să apară, la Torino, o ediție mult amplificată a unui mai vechi curs de geografie lingvistică: Cugno–Massobrio 2010, o amplă dare de seamă asupra principalelor lucrări de cartografie lingvistică realizate la nivel romanic.

După cum se poate observa, această preocupare din ultima perioadă în direcția unei „défense et illustration de la géolinguistique”

(Carton 2000, p. 130) a fost una constantă și larg răspândită la nivel european, specialiști din mai multe areale culturale (italian, francez, spaniol, german) încercând să continue cercetările în acest domeniu, în principal printr-o adaptare a instrumentarului teoretico-metodologic al geolingvisticii la noul cadru epistemic al științelor limbajului.

Două au fost perspectivele adoptate de dialectologi în efortul lor de a răspunde acestei provocări:

a) mai întâi, apelul la tehnologiile informatice în vederea creării unor programe care, pe de o parte, să faciliteze redactarea atlaselor lingvistice (fie în format clasic, fie în format electronic), iar, pe de altă parte, să asigure o diseminare cât mai eficientă a acestora (în varianta lor electronică), prin posibilitatea unei consultări la scară largă (online) și interactivă de către utilizatori; astfel, începute încă din anii '60 ai secolului trecut (Vierek 1989, p. 521) cu preponderență în mediul anglo-saxon, cercetările în domeniul informatizării atlaselor lingvistice au continuat de-a lungul acestei perioade, una dintre realizările majore în această direcție constituind-o publicarea, la sfârșitul anilor '90, a primului atlas audio pe suport online: *Sprachatlas des Dolomitenladinischen und angrenzender Dialekte/ Atlas linguistique du ladin des Dolomites et des dialectes limitrophes*;

b) a doua direcție care poate fi sesizată ca urmărind o modernizare a cercetărilor de geolingvistică propune o diversificare a obiectivelor acestui domeniu, în scopul trecerii de la o geografie lingvistică *monodimensională* către o geografie lingvistică *pluridimensională* (cf. Thun 2000). Această schimbare de paradigmă a putut fi realizată prin includerea printre punctele de interes ale geolingvisticii, în afara variației diatopice, și a altor tipuri de variație, cum ar fi variația diasexuală, variația diagenerațională, variația diastratică, eventual chiar diafazică (cf. Thun 2000, p. 375–376), precum și a celei diacronice. Ca rezultate obținute deja în acest sens amintim aici publicarea, în anul 2000, a primelor două volume din *Atlas Lingüístico, Diatópico y Diastrático del Uruguay* (ADDU), sub conducerea lui Harald Thun, iar în 2002, sub conducerea aceluiași lingvist, a primului volum din

Atlas Lingüístico Guarani–Románico (ALGR), intitulat, într-o manieră sugestivă pentru perspectiva amintită de noi aici, *Sociología* (ALGR-S)².

O altă direcție de cercetare care s-a dezvoltat în ultima perioadă și care poate oferi geolingvisticii premisele unei dezvoltări solide în viitor este dată de constituirea unor baze de date dialectale complexe, în format multimedia, care să „autorizeze [...] conservarea caracterului pluridimensional al obiectului dialectal” (Dalbéra 2000, p. 139) și care să permită o exploatare din diferite perspective a materialului lingvistic astfel teaurizat. Drept exemple în acest sens amintim proiectul *Thesaurus Occitan* (THESOC), iar pentru documentarea dimensiunii diacronice a variației diatopice menționăm corpusul *Archivio Testuale del Siciliano Antico* (ARTESIA), o arhivă multimedia care își propune să teaurizeze în format interactiv o serie de documente scrise în dialectul sicilian din secolele XIV–XVI.

Informatizarea cartografiei lingvistice, precum și o diversificare tematică în sensul includerii dimensiunii sociolingvistice a variației diatopice în sfera de interes a geolingvisticii reprezintă, așadar, cele două direcții de cercetare ce conferă la ora actuală o configurație epistemică solidă și, în consecință, o perspectivă optimistă respectivului domeniu. Astfel, chiar dacă

„vremurile mărețe ale geolingvisticii clasice [...] au apus, [...] sarcina originală a geografiei lingvistice, și anume studiul sistematic al amenajării lingvistice a spațiului de către om, nu a pierdut nimic din importanța și din actualitatea sa inițială” (Bauer–Goebel 2000, p. 169–170).

Iar dezvoltarea cercetărilor de profil va depinde, în bună măsură, așa cum afirmă Fernand Carton în *Prezentarea sa a volumului 2/2000* al revistei „Verbum”, număr tematic intitulat *Géolinguistique en Europe*, de maniera în care „va favoriza interdisciplinaritatea, păstrând, în același timp, sensul terenului, menținând contactul cu faptele, care reprezintă, așa cum se știe, piatra de încercare” (Carton 2000, p. 134).

² La fel de sugestiv pentru această schimbare de optică în cadrul cercetărilor de geolingvistică este și titlul seriei editoriale în cadrul căreia au apărut aceste două lucrări: *Dialectologia pluridimensionalis romanica*.

În cele ce urmează, vom prezenta principalele realizări din ultima perioadă ale cartografiei lingvistice românești, urmărind cu preponderență proiectele ce au vizat crearea primelor atlase informatizate din acest areal lingvistic.

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 străinătate³. 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, p. V).

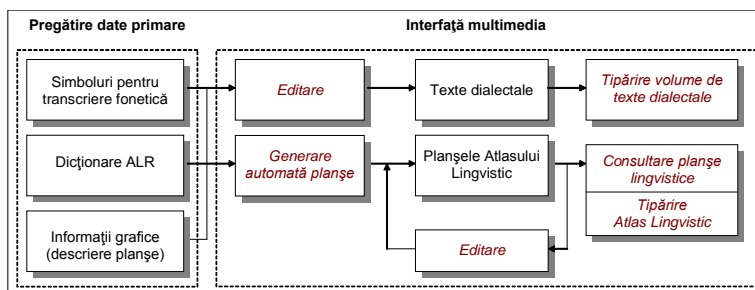
Această bogată tradiție în domeniul cartografiei lingvistice – tradiție recunoscută și apreciată ca atare la nivel european încă din prima parte a secolului al XX-lea, când, odată cu apariția primelor volume din *Atlasul lingvistic român* (1938–1942), lingviștii români introduceau o serie întregă de 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 su-

³ Pentru o prezentare a principalelor realizări ale cartografiei lingvistice românești, vezi Olariu 2010.

portul logistic necesar redactării asistate de calculator a atlaselor lingvistice și, într-o etapă ulterioară, elaborării unor variante online ale acestor lucrări.

1.1. În spațiul românesc, primele încercări de informatizare a redactării atlaselor lingvistice au fost făcute spre sfârșitul anilor '90 ai secolului trecut. Aceste cercetări au luat amploare începând cu anul 2000, când, în cadrul Filialei Iași a Academiei Române, s-au pus bazele unei colaborări interdisciplinare între cercetători de la Institutul de Informatică Teoretică și de la Institutul de Filologie Română „Al. Philippide” în scopul realizării 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, intitulate *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 a fost structurată pe două componente principale, prima vizând procedurile pentru pregătirea datelor primare, în timp ce cea de-a doua se referă la interfața multimedia:



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 adoptarea sistemului de transcriere a sunetelor folosit de *Atlasul lingvistic român* (ALR), pentru a cărui realizare informaticienii au folosit codificarea UNICODE.

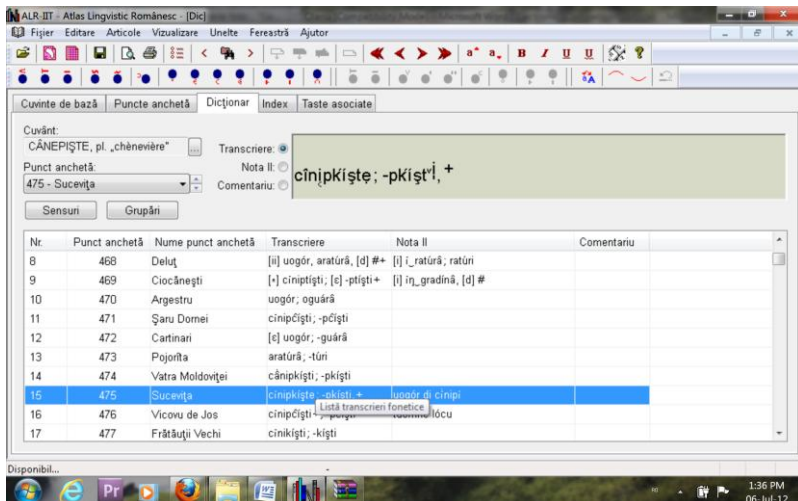
Odată rezolvată problema fonturilor, a fost nevoie de o structurare a informațiilor lingvistice și grafice necesare redactării atlasului lingvistic. Soluția a fost găsită prin configurarea unor *Dicționare* în care să se realizeze colectarea informațiilor primare privind titlul hărților (cuvinte de bază), punctele de anchetă, transcrierea fonetică și notele asociate transcrierii fonetice, precum și posibilitatea căutării de sintagme, cuvinte sau secvențe fonetice în întreaga bază de date a atlasului, prin intermediul unui *Index*⁴. 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;

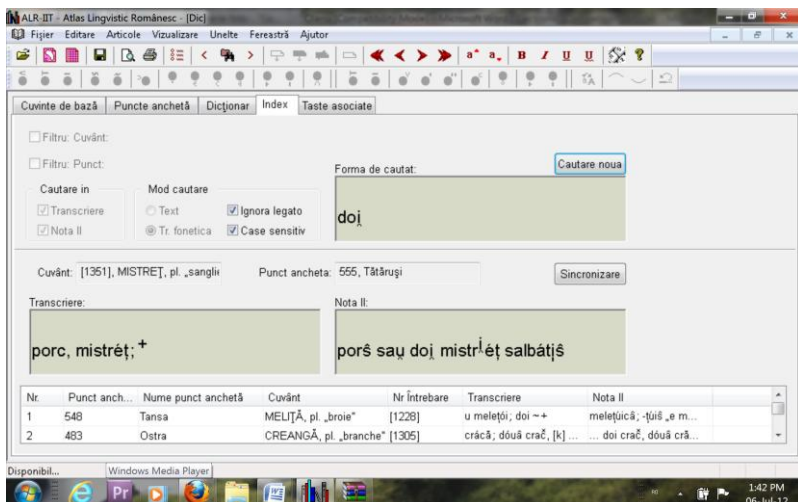
b) *Dicționar „Puncte de anchetă”*: conține informații (cod, nume, observații) despre punctele de anchetă prezente în cadrul atlasului lingvistic;

c) *Dicționar „Transcrieri fonetice”*: reprezintă elementul principal al modulului și conține transcrierea fonetică a răspunsului la întrebarea pusă în anchetă pentru fiecare cuvânt din Dicționarul „Cuvinte de bază” în fiecare punct de anchetă:

⁴ Într-o primă fază a elaborării aplicației, aceasta conținea și posibilitatea găzduirii unei *colecții audio* în format WAV aferentă răspunsurilor înregistrate. Însă, datorită faptului că ancheta cu chestionarul nu a fost programată *ab initio* pentru a răspunde unei asemenea exigențe (existând înregistrări parțiale doar pentru o mică zonă din arealul cercetat), s-a renunțat ulterior la această componentă a aplicației.



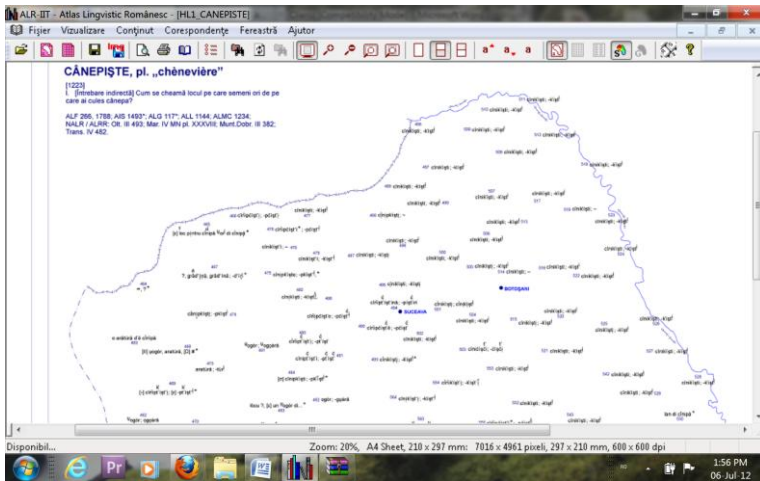
Fereastră de editare a dicționarului de transcrieri fonetice



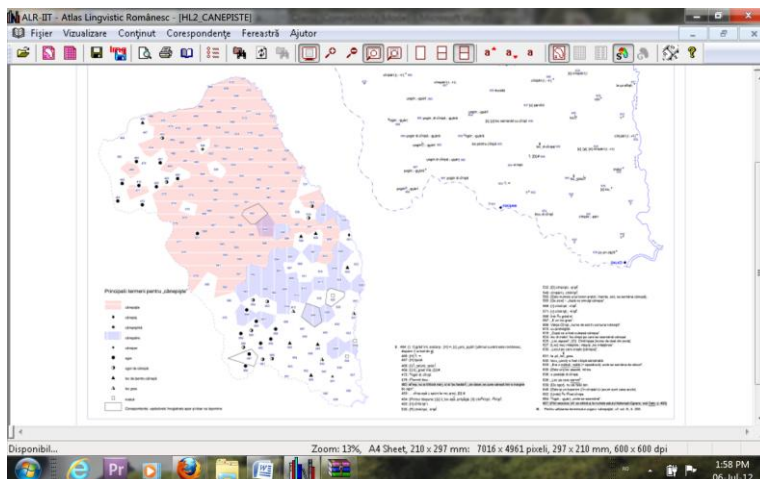
Indexul de forme, generat de aplicație

d) *Indexul*: reprezintă o aplicație care permite sortarea unor secvențe selecționate din cuprinsul întregului atlas. Această opțiune cuprinde mai multe filtre de căutare (*Cuvânt*, *Punct de anchetă*, *Transcriere fonetică*, *Notă de subsol* etc.), care facilitează atât căutarea propriu-zisă, cât și elaborarea unor indici lexicali, fonetici etc. în funcție de opțiunile utilizatorului (vezi imaginea pe pagina anterioară)

Odată introdusă informația lingvistică în dicționarele aplicației, acesta este distribuită în mod automat, fie sub formă de hartă analitică (aici existând și opțiunea de a redacta și o hartă sintetică, interpretativă, în format policrom, corespunzătoare hărții analitice), fie sub formă de listă, ca material necartografiat, această opțiune făcându-se în funcție de complexitatea variațională a respectivului material.



Secvență de hartă analitică



Secvență de hartă analitică + hartă sintetică

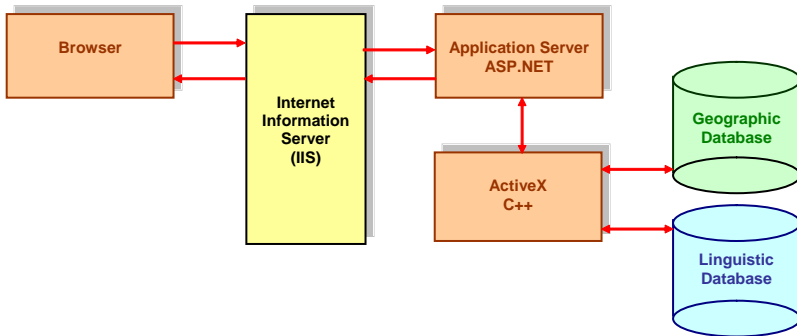
După o perioadă de testări și îmbunătățiri ale variantelor succesive ale aplicației software⁵, în anul 2005 a fost finalizat programul de editare asistată de calculator a atlasului lingvistic, iar în 2007 a apărut volumul al III-lea din *NALR. Moldova și Bucovina*, editat în întregime cu ajutorul calculatorului – 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 amintit⁶. Stocarea pe un server a acestui ma-

⁵ 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, 2008; Apopei *et alii* 2003; Bejinaru *et alii* 2006.

⁶ O prezentare mai detaliată a acestui nou proiect de digitalizare a cartografiei lingvistice românești se găsește în Botoșineanu *et alii*.

terial, la care se va putea adăuga și baza de date aferentă volumului al III-lea, deja digitalizată, precum și, în viitor, întreaga arhivă 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 lingvistice. Relaționarea acestor aplicații în cadrul programului Web propus de cercetătorii ieșeni poate fi urmărită în figura următoare:



Structura aplicației Web a atlasului lingvistic

1.2. O inițiativă similară de informatizare a atlantografiei lingvistice românești îl are ca autor pe Dorin Urițescu, 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 dome-

niul dialectometriei (Hans Goebel) și al lingvisticii cantitative⁷, specialiștii menționați și-au propus ca, folosind tehnici statistice de genul scalării multidimensionale⁸ (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 *et alii* 2007, p. 90).

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 *et alii* 2008, p. 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 (*ibidem*, p. 19)

O componentă majoră a programului informatic este aplicația cu ajutorul căreia se realizează hărțile interpretative, aceasta fiind poate

⁷ 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*.

⁸ Pentru mai multe detalii, vezi Embleton *et alii* 2007, precum și pagina de Internet <http://www.yorku.ca/vpaweb/romanian/index.htm>

cea mai importantă contribuție a autorilor în cadrul proiectului respectiv. Folosindu-se tehnica scalării multidimensionale⁹, 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 *et alii* 2008, p. 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. Astfel, sperăm ca utilizatorii atlasului nostru online să poată să-și defină singuri ceea ce

⁹ 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 et alii 2008, p. 23).

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 *et alii* 2007, p. 96)¹⁰.

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

1.3. Ultimul proiect (în ordine cronologică) de informatizare a atlaselor lingvistice românești aparține tot 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. 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 folo-

¹⁰ 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 important să existe funcții manuale de verificare și corectare a oricărei căutări sau operații automate] (Embleton *et alii* 2008, p. 10).

sită 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 Italie (până la ora actuală aproximativ o treime din teritoriul Italiei dispune deja de răspunsurile on-line la această anchetă). 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 imaginilor video, pe lângă utilizarea sunetului¹¹.

În stadiul actual al cercetărilor din domeniu, proiectul ALAB încearcă să răspundă unei duble provocări: pe de o parte, intenției de a salvarda, într-un format multimedia interactiv (posibil ca urmare a folosirii noilor tehnologii informatice), diversitatea dialectală 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 informatics*, 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 expe-

¹¹ Menționăm că cercetătorii echipei ALAB au avut, în primăvara anului 2012, schimburi de experiență atât cu echipa ALD, condusă de Hans Goebel, cât și cu echipa ALAVAL, al cărei director este Andres Kristol.

riență 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. În plus, un obiectiv important al proiectului constă în documentarea variației sociolingvistice la nivelul graiurilor din zonă, pentru a cărei realizare va fi elaborată o secțiune specială a chestionarului, ce va fi dublată și de înregistrarea unor interacțiuni verbale cotidiene din cadrul comunităților anchetate.

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 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 să permită construirea, interogarea și întreținerea acestora, este 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 ; 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 Pre-processor**, 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.

Până la această dată a fost creat site-ul proiectului : www.philippide.ro/ALAB, precum și interfața web securizată pentru administrarea proiectului.

În interfața de administrare sunt prezente uneltele de introducere, modificare, ștergere a datelor din baza de date și liste pentru vizualizarea înregistrărilor:

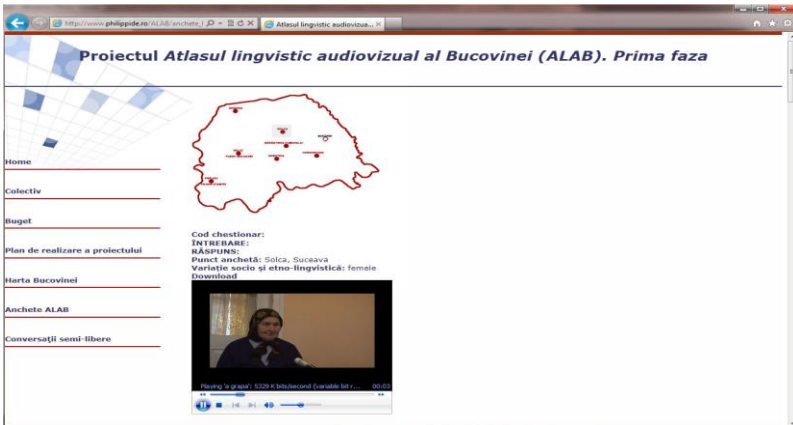
- adăugare/ modificare/ ștergere/ listă puncte de anchetă;
- adăugare/ modificare/ ștergere/ listă înregistrări în chestionar;
- adăugare/ modificare/ ștergere/ listă variație socio și etno-lin-gvistică;
- adăugare/ modificare/ ștergere/ listă subiecte de conversație semi-liberă;
- adăugare/ modificare/ ștergere/ listă anchete ALAB;
- adăugare/ modificare/ ștergere/ listă anchete NALR.



Lista de întrebări și răspunsuri din interfața de administrare

Utilizând această interfață s-a testat funcționalitatea bazei de date prin introducerea materialului audio-video înregistrat în cadrul anchetelor de probă.

În această etapă s-a realizat harta digitală a zonei anchetate care conține toate punctele de anchetă propuse în această primă fază a proiectului. Harta respectivă este accesibilă tuturor utilizatorilor de internet pe site-ul proiectului ALAB și va facilita accesul la materialul audio-video înregistrat din cadrul proiectului. Site-ul proiectului ALAB conține și liste cu înregistrările utilizate la testare din baza de date, care își păstrează însă acest caracter brut, nefinalizat încă.



Pagina accesibilă utilizatorilor Internet

Așa cum menționam mai sus, în cadrul proiectului ALAB este preconizată, pe lângă documentarea unor noțiuni cu ajutorul Chestionarului, și realizarea unei baze 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 (cf. Dumistrăcel *et alii* 1997) (altfel spus, trecerea de la metoda *cuvinte și lucruri* la metoda *cuvinte și contexte comunicative*), mult mai adecvată unei redări cât mai ingenuă a specificului idiomatic al unei comunități sau zone etnografice.

BIBLIOGRAFIE

- 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.
- Bauer–Goebel 2000 = Roland Bauer, Hans Goebel, *Utilisation nouvelle de l'informatique dans les atlas linguistiques en Europe (1980–2000)*, în „Verbum”, 2/2000, p. 169–185.
- 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 *et alii* = 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).
- Carton 2000 = Fernand Carton, *Introduction*, în „Verbum”, 2/2000 (numărul tematic intitulat *Géolinguistique en Europe*), p. 129–134.
- Cugno, Massobrio 2010 = Federica Cugno, Lorenzo Massobrio, *Gli atlanti linguistici della România. Corso di Geografia linguistica*, Alessandria, Edizioni dell'Orso, 2010.
- Dalbéra 2000 = Jean-Philippe Dalbéra, *Nouvelles technologies et perspectives nouvelles en géolinguistique*, în „Verbum”, 2/2000, p. 135–155.
- Dumistrăcel *et alii* 1997 = Stelian Dumistrăcel, Doina Hreapcă, Ion-Horia Bârleanu, *Ancheta dialectală ca formă de comunicare*, Iași, Editura Academiei Române, 1997.
- Embleton *et alii* 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, 2007, p. 87–96.
- Embleton *et alii* 2008 = Sheila M. Embleton, Dorin Uritescu, Eric S. Wheeler, *Digitalized Dialect Studies: North-Western Romanian*, București, Editura Academiei Române, 2008 (Conferințele Academiei Române. Ciclul „Limba română și relațiile ei cu istoria și cultura românilor”).
- 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 lite-*

- raturii române în perspectiva globalizării, Iași, Editura „Trinitas”, 2002, p. 211–232.
- Le Dû 1997 = Jean Le Dû, *La disparition du group des atlas et l’avenir de la géographie linguistique*, în „Le Français Moderne”, 1997, nr. 1, p. 6–12.
- Mouton 1994 = Pilar García Mouton (ed.), *Geolingüística. Trabajos europeos*, CSIC, Madrid., 1994
- 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*, în „Revista Española de Lingüística”, Madrid, 2008, nr. 37, p. 215–246.
- Olariu 2010 = Florin-Teodor Olariu, *Cartografia lingvistică română: un secol de rezultate*, „Bollettino dell’Atlante linguistico italiano”, III Serie, Dispensa N. 34, 2010, p. 101–135.
- Raimondi–Revelli (eds.) 2008 = Gian Mario Raimondi, Luisa Revelli, *La dialectologie aujourd’hui. Atti del Convegno Internazionale „Dove va la dialettologia?”*, Saint-Vincent, Aosta, Cogne, 21–24 settembre 2006, Alessandria, Edizioni dell’Orso, 2008.
- Saramandu 2005 = Nicolae Saramandu, *Prefață la ALRR. Sinteză*, vol. I, Editura Academiei Române, București, 2005, p. V–VII.
- Telmon 2007 = Tullio Telmon, *Prefazione*, în Raimondi–Revelli (eds.) 2008, p. V–IX.
- Thun 2000 = Harald Thun, *La géographie linguistique romane à la fin du XXe siècle*, în Annick Englebert et alii, *Actes du XXIIe Congrès International de Linguistique et de Philologie Romanes*, Tome III: *Vivacité et diversité de la variation linguistique*, Tübingen, Niemeyer, 2000, p. 367–388.
- Viereck 1989 = Wolfgang Viereck, *The Computerisation and Quantification of Linguistic Data: Dialectometrical Methods*, în Alan R. Thomas (ed.), *Methods in Dialectology*, Clevedon, Multilingual Matters, 1989, p. 521–547.

**ATLASUL LINGVISTIC AUDIOVIZUAL AL BUCOVINEI (ALAB).
PRIMA FAZĂ – UN NOUVEAU PROJET DE NUMÉRISATION
DE LA CARTOGRAPHIE LINGUISTIQUE ROUMAINE
(Résumé)**

La cartographie linguistique roumaine, dont le centenaire a été récemment célébré (1909: Gustav Weigand, *Linguistischer Atlas des dacorumänischen Sprachgebietes*), a une riche tradition dans le domaine des atlas linguistiques – trois générations de tels ouvrages à caractère national sont parus au cours du dernier

siècle. Cette riche tradition est la source d'une forte motivation des dialectologues roumains vers une modernisation de la méthode cartographique, principalement à travers le recours aux technologies informatiques les plus récentes, capables d'assurer le support logistique nécessaire à la rédaction des atlas tout comme à l'élaboration des premières variantes numériques de ces ouvrages.

Atlasul lingvistic audiovizual al Bucovinei (L'Atlas linguistique audiovisuel de la Bucovine – ALAB), constitue un projet de géographie linguistique multimédia qui essayera de mettre les technologies informatiques au service d'une géographie linguistique *pluridimensionnelle* (Harald Thun) et de réaliser un atlas audiovisuel avec une forte composante sociolinguistique, focalisée sur la documentation au niveau dialectal des variables diatopique et diastratique. Mettant en relation un questionnaire adéquat pour ce but avec une technique numérique syncrétique pour la collecte de données, l'équipe ALAB se propose, comme première étape du projet, de réaliser une banque de données de type 'document multimédia interactif' qui soit capable de faciliter des perspectives novatrices dans l'analyse des variations géo- et sociolinguistiques.

CUVINTE-CHEIE: *dialectologie, atlas lingvistic, lingvistică computațională, baze de date, resurse Web.*

KEYWORDS: *dialectology, linguistic atlas, computational linguistics, databases, Web resources.*

*Institutul de Filologie Română „A. Philippide”,
Institutul de Informatică Teoretică,
700481 Iași, str. Th. Codrescu, 2*