

UN INSTRUMENT MODERN DE STUDIU AL DIALECTULUI AROMÂN: CORPUS ADNOTAT MORFOSINTACTIC

VERGINICA BARBU MITITELU
MIHAELA CRISTESCU
MANUELA NEVACI

1. INTRODUCERE

În urmă cu un deceniu, rețeaua de excelență META-NET a Comisiei Europene publica rezultatele unei analize a resurselor și tehnologiilor existente la acel moment pentru limbile Europei¹. Situația era de natură să provoace îngrijorare în rândul factorilor implicați în susținerea (financiară sau de altă natură a) dezvoltării acestor resurse și tehnologii în perspectiva penetrării masive a calculatorului (și a altor dispozitive cu funcții similare, precum tabletă, telefon, ceas inteligent etc.) în viața noastră, nu doar profesională, ci și extraprofesională, și a asigurării supraviețuirii limbilor și dialectelor în noul context. După un deceniu, constatăm că perspectiva de atunci a devenit realitate curentă: comunicăm efectiv între noi prin intermediul unor astfel de dispozitive, dar comunicăm chiar și cu astfel de dispozitive: a se vedea utilizarea roboților de dialog numiți *chat bots* în interacțiunea (scrisă sau orală) cu furnizorii de diverse servicii cu largă adresabilitate, de la telefonie mobilă, la produse bancare, și o putem face în limba maternă. Aceasta reprezintă marele avantaj, aceasta este o caracteristică esențială a unor astfel de produse, care le asigură o piață de desfacere uriașă, în consecință și profituri mari. Pentru noi, lingviștii, este o nouă cerință căreia trebuie să-i facem față și o oportunitate, deopotrivă, asigurându-le astfel vorbitorilor certitudinea că își vor putea folosi limba și în această eră de digitalizare tot mai accentuată, păstrându-și astfel identitatea națională, etnică etc. Așa se explică interesul sporit acordat acum limbilor cu resurse lingvistice puține, limbilor minoritare, dialectelor și în acest efort se înscrie și demersul nostru pe care îl expunem mai jos.

Prezentăm în această lucrare crearea unei colecții de fraze din dialectul aromân, adnotate morfologic și sintactic, rezultând un corpus sub forma unei bănci de arbori sintactici (engl. *treebank*), de dimensiuni modeste, dar care este un prim pas în crearea de resurse lingvistice computaționale pentru acest dialect. Banca de arbori rezultată răspunde cerințelor actuale din comunitatea lingvisticii computaționale, atât în ceea ce privește formalismul utilizat, cât și formatul de prezentare.

Lucrarea este structurată astfel: secțiunea a doua prezintă statutul dialectului aromân, cu atenție specială asupra utilizării sale în context digital și asupra resurselor lingvistice existente. În secțiunea următoare trecem în revistă alte activități de creare a băncilor de arbori pentru limba română, pe de o parte, și pentru alte dialecte, pe de altă parte. Secțiunea a patra conține (i) descrierea corpusului creat, (ii) metoda de lucru utilizată pentru adnotarea morfologică și sintactică, (iii) descrierea succintă a principiilor de adnotare morfologică și sintactică, (iv) particularitățile dialectului aromân care au impus ajustări ale formalismelor de adnotare, (v) date statistice despre banca de arbori rezultată.

¹ Pentru situația resurselor și instrumentelor disponibile la acel moment pentru limba română, vezi Trandabăț *et al.* 2012.

2. STATUTUL DIALECTULUI AROMÂN

Dialectul aromân este unul dintre dialectele limbii române, celelalte fiind: dacoromân, meglenoromân și istroromân. Este vorbit în țări din Peninsula Balcanică (Grecia, Albania, Republica Macedonia de Nord, Bulgaria), în România, precum și în alte țări din Europa și de pe alte continente (Statele Unite ale Americii, Australia), unde au emigrat grupuri de aromâni (Saramandu 1984a: 423; Nevaci 2013b: 13). Dialectul aromân este vorbit de persoane aparținând tuturor grupelor de vârstă (generațiile vârstnică și medie, dar și tineri și copii), atât în mediul rural, cât și în mediul urban. Numărul aproximativ de vorbitori activi ai dialectului aromân este de 600.000 (Saramandu 2007a; Nevaci 2013b: 18).

2.1. Graiurile aromânești

Principalele **graiuri** ale dialectului aromân sunt: pindean (Grecia), fărșerot (Albania, Grecia și R. Macedonia de Nord), grămostean (Grecia, R. Macedonia de Nord, Bulgaria) și grabovean (Albania). La acestea se adaugă așa-numitele „graiuri insulare”, cum ar fi graiul din Muloviște și graiul din Crușova (localități din R. Macedonia de Nord) (vezi Saramandu 1984; 2007b). În descrierea particularităților fiecărui grai al dialectului aromân, s-au urmărit aspectele fonetice și fonologice, morfologice și lexicale (Nevaci 2013b). Una dintre trăsăturile diferențiatore în cadrul graiurilor aromânești o constituie sistemul vocalic: în timp ce graiurile fărșerot și grabovean conțin 6 foneme vocalice (lipsește vocala centrală închisă /i/), graiurile pindean și grămostean conțin 7 foneme vocalice (Saramandu 1984a: 424; Saramandu 2017b: 130).

Particularitățile fonetice, morfologice și lexicale ale fiecărui grai sunt exemplificate în Nevaci 2013b prin texte dialectale înregistrate în timpul anchetelor de teren. Sunt importante aceste diferențe, deoarece nu există o variantă standardizată dialectală. Prin urmare, referirea la un grai sau altul este mult mai exactă din punct de vedere științific. Anumite particularități fonetice, morfologice sau lexicale sunt specifice numai unui anumit grai, așa cum vom vedea în descrierea ce urmează, fără a fi generale în dialectul aromân.

2.2. Literatura aromână

Dialectul aromân este singurul dialect sud-dunărean al limbii române în care s-a dezvoltat o literatură cultă. În scrierea dialectului aromân s-au folosit două alfabet: grecesc (primele texte aromânești, datând din secolul al XVIII-lea) și latin (începând cu secolul al XIX-lea). Primul text redactat în dialectul aromân (cu alfabet grecesc) datează din anul 1731 și este o inscripție cu conținut religios (Saramandu 2000: 200) Menționăm lucrările semnificative dedicate literaturii dialectale aromâne, apărute în spațiul românesc.

Dr. G. Obedenaru publică la București, în 1891, *Texte macedoromâne, basme și poezii populare de la Crușova*.

Pericle Papahagi va deschide seria culegerilor de folclor și a antologiilor de literatură cultă în dialectul aromân prin lucrările: *Din literatura poporană a aromânilor*, prefață de Gr. Tocilescu, București (1900), *Basme aromâne și glosar*, București (1905), *Scriitori aromâni din sec. XVIII* (1909), *Poezia înstrăinării la aromâni*, București (1912).

Un compendiu de literatură aromână este cartea *Antologie aromânească*, a lui Tache Papahagi, publicată în 1922. Aceasta conține trei secțiuni: literatură populară (proverbe și ghicitori, poezii, cântece, balade și legende, snoave, tradiții și basme), literatură cultă și muzică populară aromânească, completată ulterior cu volumele-document *Images d'éthnographie roumaine, daco-roumaine et aroumaine*, 3 volume, 1928–1934. Tache Papahagi este primul învățat care a fixat și a impus un sistem de scriere a aromânei, folosit de eruditul lingvist și în editarea

operelor scriitorilor aromâni din colecția „Biblioteca națională a aromânilor”, inițiată de el. Au apărut din această serie șase volume, primele trei în perioada 1926–1935 (I, Nuși Tulliu, 1926; II, Zicu A. Araia, T. Caciona, 1932; III. N. Batzaria, 1935), celelalte, rămase în manuscris, fiind publicate postum (IV, M. Beza; V, C. Belimace, N. Velo; VI, N. Batzaria, T. Carafoli).

Lucrările lui Apostol N. Caciuperi (1920–1995) au fost (re)editate postum în România în 2011, într-o serie de *Opere* în patru volume: I. *Noulu Testamentu. Pri grailu armânescu*, II. *Traduțeri (dit literatura română și universală)*, III. *Gramatică armânească (proiectu di manual școlar)*, IV. *Cursu practic pi graiu armânescu* (Universalia Publishers, București, 2011). De interes pentru lingvist sunt îndeosebi cursul practic și gramatica aromână, redactate „pi graiu armânescu”, cu scopul de a folosi în învățarea dialectului.

Hristu Cândroveanu, critic literar, poet, prozator, traducător și publicist, realizează o serie de antologii în care sunt cuprinse creațiile importante în dialectul aromân: *Antologie de lirică aromână*, București, Editura Univers, 1975; *Antologie de proză aromână*, București, Editura Univers, 1977; *Un veac de poezie aromână*, 1985; *Antologia poeziei străine în grai aromân*, București, Editura Fundației Culturale Aromâne „Dimândarea Părintească”, 2000; *Antologia poeziei române clasice și moderne în grai aromân*, București, Editura Fundației Culturale Aromâne „Dimândarea Părintească”, 2003; *Caleidoscop aromân*, 5 volume, București, Editura Fundației Culturale Aromâne „Dimândarea Părintească”, 1998–2002.

În cursul anchetelor de teren la aromâni, Nicolae Saramandu a cules material etnografic și folcloric, pe care l-a valorificat în volumele: *Folclor aromân grămostean* (București, Editura Minerva, 1982, în colaborare cu N. Gh. Caraiani și Miorița la dacoromâni și aromâni (texte folclorice), publicată la Editura Minerva, în 1992 (cu Tatiana Gălușcă-Cîrșmariu).

În 2006, Matilda Caragiu-Marioțeanu și-a strâns contribuțiile într-un tom voluminos, *Aromânii și aromâna în conștiința contemporaneității*, însoțit de un volum de *Poeme aromâne*, cuprinzând creațiile poetice în dialect ale autoarei.

O antologie recentă de literatură aromână este volumul *Scriitori aromâni de ieri și de azi. Antologie de texte și studiu filologic*, de Manuela Nevaci, apărută în 2013 (Nevaci 2013c). Lucrarea reprezintă o antologie a literaturii aromâne culte, descriind operele unor autori importanți, precum Theodor Atanasie Cavallioti, Daniil Moscopoleanul, Constantin Ucuta (reprezenți ai literaturii vechi), Constantin Belimace, George Murnu, Nicolae Constantin Batzaria, Nuși Tulliu, Z.A. Araia, George Ceară, Nicolae Velo, Marcu Beza, NidaBoga, Zahu Pană, George Ceară, Nicolae Caratană, Spiru Fuchi (reprezenți ai literaturii moderne) ș.a. Operele literare ale scriitorilor aromâni culti (exemplificate prin poezii) sunt precedate de două secțiuni: *Literatura aromână veche: secolul al XVIII-lea – începutul secolului al XIX-lea* și *Sisteme de scriere utilizate în limba publicațiilor aromânești actuale*. După cum menționează autoarea, în absența unei norme supradialectale, scrierea aromânească actuală diferă de la un autor la altul.

2.3. Publicații și asociații ale aromânilor

Începând cu sfârșitul secolului al XIX-lea și continuând cu începutul secolului al XX-lea, se înființează numeroase asociații aromânești, se organizează congrese și apar publicații (periodice și volume). Apar reviste de cultură, ziare și almanahuri editate de către aromânii din țările balcanice, editate atât în dialectul aromân, cât și în româna literară. În Nevaci (2013b: 34–37) sunt enumerate titlurile publicațiilor cu articole în dialectul aromân, apărute în România, Albania, R. Macedonia de Nord, Bulgaria și Grecia. De asemenea, sunt trecute în revistă asociațiile de aromâni din diferite țări. În România, cea mai importantă societate aromânească este *Societatea de Cultură Macedo-Română*², reînființată după 1989.

² Informații pot fi găsite pe site-ul Societății, <https://scmr.ro/>.

2.4. Lucrări de popularizare a dialectului aromân publicate în România

Societatea de Cultură Macedo-Română, în parteneriat cu Institutul de Etnografie și Folclor „Constantin Brăiloiu” și cu Fundația Mușata Armână, publică la Editura Universității din București și la Editura Etnologică o serie de albume despre aromâni și graiul lor: „*Lecturi vizuale*” *etnolingvistice la aromânii din Bulgaria. Memorie, tradiție, grai, patrimoniu*, autori: Nicolae Saramandu, Emil Țîrcomnicu, Manuela Nevaci, Cătălin Alexa, București, 2019; „*Lecturi vizuale*” *etnologice la aromânii din Republica Macedonia. Regiunea Bitolia. Memorie, tradiție, grai, patrimoniu*, autori: Nistor Bardu, Emil Țîrcomnicu, Cătălin Alexa, Stoica Lascu, Ioana Iancu, București, 2018; „*Lecturi vizuale*” *etnolingvistice la aromânii din Republica Macedonia. Regiunile Ohrida, Struga, Crușova. Memorie, tradiție, grai, patrimoniu*, autori: Nicolae Saramandu, Emil Țîrcomnicu, Manuela Nevaci, Cătălin Alexa, 2018; „*Lecturi vizuale*” *etnologice la aromânii din Albania. Gramsh, Elbasan, Pogradec*, autori: Nistor Bardu, Emil Țîrcomnicu, București, 2017; „*Lecturi vizuale*” *etnologice la aromânii din Albania. Sarande și Gjirokaster*, autori: Iulia Wișoșenschi, Emil Țîrcomnicu, Cătălin Alexa, 2016; „*Lecturi vizuale*” *etnologice la aromânii din Albania. Câmpia Muzachiei*, autori: Iulia Wișoșenschi, Emil Țîrcomnicu, Cătălin Alexa, București, 2015; „*Lecturi vizuale*” *etnologice la aromânii din Albania – zona Korçe*, autori: Iulia Wișoșenschi, Emil Țîrcomnicu, Cătălin Alexa, Nistor Bardu, București, 2014.

3. STADIUL CERCETĂRILOR

3.1. Preocupări pentru informatizarea limbii române

Preocuparea de descriere a limbii române cu ajutorul formalismelor moderne a caracterizat comunitatea lingvistică românească de timpuriu, de cele mai multe ori în cadrul unor cooperări internaționale, folosind un formalism comun, lucru care permite evidențierea elementelor comune și de contrast între limbile incluse în analiză, pe de o parte, iar, pe de altă parte, ajută la ajustarea, corectarea sau extinderea formalismelor pentru a răspunde nevoilor fiecărei limbi. La nivel național, a fost constituit, în cadrul Academiei Române, Consorțiul pentru Informatizarea Limbii Române³, din păcate, nu suficient de activ de-a lungul timpului.

În cadrul acestei lucrări, perspectiva adoptată este aceea a lingvisticii corpusului. Pentru limba română, descrierea morfologică a fost făcută în proiectul MULTTEXT-East (Erjavek 2017). Paisprezece categorii (substantiv, verb, adjectiv, pronume, determinator, numeral, adverb, articol, prepoziție, conjuncție, particulă, interjecție, abreviere și reziduu) corespunzând, în mare, părților de vorbire, au fost descrise și rafinate, astfel încât o mulțime de 674 de etichete (reflectând descrieri morfosintactice ale formelor cuvintelor) s-au dovedit suficiente pentru a descrie orice formă a oricărui cuvânt din română (Tufiș *et al.* 1997). Un exemplu de formă a unui cuvânt cu analiza morfosintactică adecvată (PoStag) este următorul:

împărăteasa Ncfsry

unde interpretarea PoStagului este următoarea: substantiv (N) comun (c), genul feminin (f), numărul singular (s), cazul direct (r) (nominativ-acuzativ), formă articulată cu articol hotărât (y).

³ Domnul Nicolae Saramandu a exprimat, în ședința din 2001 a Comisiei de Informatizare pentru Limba Română, nevoia colaborării între lingviști și informaticieni pentru crearea resurselor lingvistice în format electronic, în special atlase lingvistice (<http://consilr.info.uaic.ro/consilr/ro/index.php?showpage=030101>, accesat la data de 5 mai 2021).

În ceea ce privește descrierea sintactică, aceasta s-a făcut preponderent prin gramatici de dependență (Hristea și Popescu 2003, Perez 2014, Barbu Mititelu și Irimia 2015), dar există și descrieri cu ajutorul gramaticilor de constituenți (Bick și Greavu 2010). Mai recent, o inițiativă internațională numită *Dependențe Universale* (engl. *Universal Dependencies*⁴) (UD), care are scopul de a descrie unitar gramatica limbilor lumii, cuprinde și limba română, reprezentată deja prin trei corpusuri adnotate morfologic și sintactic (bănci de arbori sintactici): RoRefTrees (RRT) (Barbu Mititelu *et al.* 2016), Nonstandard (Colhonet *al.* 2017) și SiMoNERo (Barbu Mititelu și Mitrofan 2020). Lucrarea de față descrie crearea unui nucleu al unei noi astfel de bănci, pe care am numit-o **ArT** (AromanianTreebank).

3.2. Resurse lingvistice pentru dialectul aromân

Prin resurse lingvistice înțelegem aici dicționare, gramatici și alte lucrări care descriu structura dialectului aromân. Primele lucrări cu caracter didactic destinate aromânilor au apărut în ultimele trei decenii ale secolului al XVIII-lea și aparțin cărturarilor aromâni, Theodor Anastasie Cavallioti, Daniil Moscopoleanul și Constantin Ucuta (vezi Saramandu 2013).

Theodor A. Cavallioti a publicat la Veneția, în 1770, *Πρωτοπειρία* [Prima învățătură], carte de citire pentru clasele elementare, scrisă în grecește, care are, la sfârșit, un vocabular de 1170 de cuvinte grecești, traduse în aromână și în albaneză. Lucrarea are 104 pagini și conține rugăciuni și pasaje din *Biblie*. Are și următorul text aromânesc (în original, cu alfabet grecesc): *Hristos de morți năstăsi, cu morte morteă călcându, ș-a mărmintătorlorahărzi bană harizmă*. [Hristos a înviat din morți, cu moartea pe moarte călcând și celor din mormânturi le-a dăruit viață plăcută.].

Lucrarea publicată de Daniil Moscopoleanul în 1794 la Veneția, *Εισαγωγική διδασκαλία* [Învățătură introductoare] este o carte de lectură, în grecește, care cuprinde la sfârșit *Λεχικόν Τετραγλωσσον*, un manual de conversație în care limba greacă e tradusă în aromână, bulgară și albaneză.

Constantin Ucuta este autorul lucrării apărute în 1797 la Viena, *Νέα Παιδαγωγία* [Noua Pedagogie], cu subtitlul (în traducere): „Abecedar lesnicios, pentru a-i învăța pe tineri carte romano-vlahă în uzul curent al romano-vlahilor”. Este primul abecedar aromânesc scris cu alfabet grecesc.

Din aceeași perioadă, sfârșitul secolului al XVIII-lea, datează și două lucrări cu caracter religios, rămase în manuscris și publicate ulterior: *Codex Dimonie*, publicat de Gustav Weigand în „Jahresbericht des Instituts für rumänische sprache”, Leipzig, I (1894, p. 1–78), IV (1897, p. 136–227), V (1898, p. 192–297), VI (1898–1899, p. 86–173), și *Liturghier aromânesc*, publicat în 1962 de Matilda Caragiu-Marioțeanu.

Format în ideologia Școlii Ardelene, Gheorghe Roja publică în 1809 la Buda, *Măestria ghiovăsirii românești cu litere latinești, care sânt literele românilor ceale vechi*, în care se pronunță pentru o limbă literară comună și o scriere comună, cu alfabet latin, pentru românii din nordul și din sudul Dunării. În prefața lucrării, intitulată *Cuvânt către români*, după ce vorbește despre faptul că limbile nu sunt unitare, în cadrul lor formându-se dialecte, Roja precizează că româna „în doao dialecte, adecă în cea dincoace de Dunăre și dea din colo de Dunăre se împarte, aceasta cu grecești, ceia cu schiavonești cuvinte [...] e amestecată” (p. 6). (vezi Saramandu 1984b: 55; Saramandu 2005: 752).

În 1813, Mihail. G. Boiagi publică la Viena, *Romanische oder Macedonowlachische Sprachlehre*, prima gramatică a aromânei, scrisă cu alfabet latin, prima încercare de codificare a materialului gramatical al dialectului aromân (vezi Saramandu 1984c: 605).

⁴ <https://universaldependencies.org/>

Gustav Weigand, profesor la Universitatea din Leipzig, publică în câteva monografii rezultatele cercetărilor efectuate la românii sud-dunăreni: *Die Spracheder Olympto-Walachen, nebsteiner Einleitung über Land und Leute* (1888), *Die Aromunen: Ethnographisch-philologisch-historische Untersuchungen* (I–II, 1894–1895), la care se adaugă alte lucrări, apărute în mai multe numere din „Jahresbericht”, despre aromânii din Bulgaria (vol. XIII, 1907), din Albania de Nord (vol. XVI, 1910) și despre toponime aromânești din ținutul Pindului (vol. XXI–XXV, 1919).

De asemenea, reținem lucrarea lui George G. Murnu, *Rumänische Lehnwörter im Neugriechischen mit historischen Vorbemerkungen* (München, 1902; la origine, teză de doctorat) privind prezența cuvintelor de origine aromână în dialectul grecesc din Epir, pătrunse prin împrumut, dar și prin asimilarea (grecizarea) populației aromânești, ca și contribuția lui Christea Geagea, *Elementul grec în dialectul aromân* (Cernăuți, 1932).

Primele dicționare dedicate dialectului aromân cuprind material lexical util pentru zonele pe care le ilustrează: Ștefan Mihăileanu, *Dicționar macedo-român* (1901; autorul, originar din Ohrida), I. Dalametra, *Dicționar macedo-român* (1906; autorul, originar din Veria). G. Pascu publică în două volume, în 1925, un dicționar etimologic al aromânei, *Dictionnaire étymologique macédomain*. Lucrarea prezintă însă numeroase erori în stabilirea originii cuvintelor.

Theodor Capidan publică *Aromânii. Dialectul aromân* (1932), operă fundamentală, principala sursă de informație asupra acestui dialect al limbii române. Monografia cuprinde o descriere amănunțită a caracteristicilor fonetice, morfologice, sintactice ale aromânei, cu precizarea unor particularități de grai locale.

O cercetare parțială de teren a efectuat Theodor Capidan timp de o lună, în vara anului 1928, la românii din Albania, publicând rezultatele cercetării în monografia *Fărșeroșii* (1931).

În *Romanitatea balcanică* (1936), discurs de recepție la Academia Română, și în alte două lucrări, *Macedoromânii. Etnografie, istorie, limbă* (1942) și *Limbă și cultură* (1943), Theodor Capidan lărgeste sfera informațiilor cu date de ordin istoric, etnografic și cultural. Menționăm și amplul studiu *Toponymie macédo-roumaine* (L Lit, 3, 1946), o contribuție importantă la stabilirea originii numelor topice din ținuturile locuite de aromâni în Peninsula Balcanică.

Al. Rosetti publică, în 1929–1930, *Cercetări asupra graiului românilor din Albania*, I, II, în GS, IV, 1, p. 1–83; IV, 2, p. 377–380 (și separat în 1930) rezultatele primei anchete la românii din Albania veniți în România.

Dicționarul de referință dedicat dialectului aromân este *Dicționarul dialectului aromân*, de Tache Papahagi (1963, 1974 – ediția a II-a, 2013 – ediție anastatică, ed. Nicolae Saramandu, Manuela Nevaci). În lipsa unei variante supradialectale, normate a dialectului aromân, nu putem vorbi despre existența unei gramatici propriu-zise. În categoria resurselor lingvistice menționăm lucrarea *Manual de aromână. Carți tră învițari armânești* (2005, 2007 – ediția a II-a), semnată de Matilda Caragiu-Marioțeanu și Nicolae Saramandu. Descrierea structurii dialectului aromân, la toate nivelele limbii (fonetic și fonologic, morfosintactic, lexical), a fost realizată în lucrări precum: Saramandu 2007a, Nevaci (2006, 2011, 2013a, 2013b, 2018). Toate lucrările menționate sunt disponibile în format tipărit, nu și în format electronic.

Fono-morfologie aromână. Studiu de dialectologie structurală (1968), la origine teză de doctorat, de Matilda Caragiu-Marioțeanu, este o încercare de abordare a aromânei din perspectiva lingvisticii structurale. Matilda Caragiu-Marioțeanu este autoarea unui *Compendiu de dialectologie română (nord- și sud-dunăreană)* (1975) și a capitolului „Dialectul aromân” din lucrarea colectivă *Dialectologie română* (1977), în colaborare cu Ștefan Giosu, Liliana Ionescu-Ruxândoiu, Romulus Todoran. A publicat și un *Dicționar aromân (macedo-vlah)*. *Diario*, I (A–D), 1997, în care termenii din româna literară sunt redați în aromână, dicționarul fiind destinat „celor ce

doresc ca, prin intermediul românei literare (RL), să învețe / să cunoască aromâna” (p. XIV). În colaborare cu Nicolae Saramandu a publicat un manual de aromână, *Carti tră învițari armânești* (2005), scris în dialect.

Nicolae Saramandu realizează prima cercetare de geografie lingvistică cu chestionar (*Chestionarul NALR*) la aromâni prin anchete de teren efectuate în anii 1967–1968 în Dobrogea, publicând: *Cercetări asupra aromânei vorbite în Dobrogea. Fonetică. Observații asupra sistemului fonologic* (1972, teză de doctorat), unde a delimitat pentru prima dată și a caracterizat, pe baza particularităților lingvistice, cele patru graiuri principale ale dialectului aromân: grămostean, fărșerot, pindean și moscopolean. La materialul lingvistic cules cu chestionar s-au adăugat texte înregistrate pe bandă magnetică în 19 localități dobrogene, transcrise de autor într-o riguroasă notație fonetică, ilustrând cele patru graiuri menționate mai sus. Textele, însoțite de un amplu glosar și de informații inedite privind stabilirea aromânilor în Dobrogea, au fost publicate într-un volum de peste 600 de pagini, *Aromâna vorbită în Dobrogea. Texte dialectale. Glosar* (Saramandu 2007a). Întemeiat pe bibliografia de specialitate și pe cercetările proprii, Nicolae Saramandu a realizat o descriere a dialectului aromân, publicată în *Tratat de dialectologie românească* (coord. Valeriu Rusu), sub titlul *Aromâna* (1984), urmată de *Harta graiurilor aromâne și meglenoromâne din Peninsula Balcanică* (SCL, XXXIX, 1988, 3, p. 225–245), întocmită pe baza informațiilor culese în cercetările de teren la românii din sudul Dunării. În volumul *Romanitatea orientală* (Editura Academiei Române, București, 2004; ediție în limba franceză, 2008) cuprinde un amplu studiu inedit, *Romanitatea orientală (secolele al IV-lea – al XV-lea). Surse istorice* (p. 9–80), în care se înregistrează pentru prima dată, în mod detaliat, atestările istorice din surse latine, bizantine, slave, maghiare ale romanității orientale nord- și sud-dunărene, care se află la originea limbii și a poporului român.

Rodul cercetărilor de teren efectuate timp de mai mulți ani de Nicolae Saramandu la aromânii din țările balcanice – Grecia, Albania, R. Macedonia de Nord, Bulgaria – este *Atlasul lingvistic al dialectului aromân*, prima lucrare de amploare consacrată celui mai important dialect românesc vorbit în sudul Dunării: aromâna. Din această lucrare au apărut până acum două volume: Nicolae Saramandu, *Atlasul lingvistic al dialectului aromân* (ALAR), volumul I, Editura Academiei Române, București, 2014 (editor: Manuela Nevaci) și Nicolae Saramandu, Manuela Nevaci, *Atlasul lingvistic al dialectului aromân*, volumul II, Editura Academiei Române, București, 2020. Atlasul pune în evidență graiurile românești vorbite în prezent în Grecia, Albania, R. Macedonia de Nord, Bulgaria. Atlasul este rezultatul unor îndelungate cercetări de teren efectuate de Nicolae Saramandu timp de peste patru decenii, începând cu anul 1967 și completate de Manuela Nevaci, începând cu 2007. Hărțile lingvistice cuprind întreaga arie geografică sud-dunăreană unde se vorbește în prezent dialectul aromân. Atlasul are o rețea de 61 de localități (32 în Grecia, 12 în Albania, 9 în R. Macedonia de Nord, 8 în Bulgaria). În fiecare dintre aceste localități s-a folosit *Chestionarul Noului atlas lingvistic român pe regiuni*, având 2543 de întrebări. În felul acesta, este posibilă compararea materialului dialectal aromânesc cu cel cuprins în atlasele regionale ale dialectului dacoromân, evidențiind unitatea limbii române vorbite în nordul și în sudul Dunării (țările balcanice). Hărțile atlasului au la bază un program de digitalizare, realizat de Vasile Apopei și Silviu-Ioan Bejinariu, de la Institutul de Informatică Teoretică din Iași al Academiei Române. ALAR este disponibil online (<https://www.lingv.ro/images/ALDA.pdf>).

Tot din sfera resurselor de geografie lingvistică dedicate dialectului aromân menționăm două atlase (cf. Dahmen și Kramer 1985, 1994 și Neiescu 1997), care oferă o imagine parțială asupra dialectului aromân. Rețeaua de anchetă este incompletă, fiind limitată la Grecia (la care

se adaugă trei localități din R. Macedonia) în atlasul Dahmen și Kramer, la Albania (7 localități) și R. Macedonia de Nord (5 localități), în atlasul publicat de Petru Neiescu. S-au folosit chestionare cu un număr limitat de întrebări (Chestionarul lexical al „Atlasului Limbilor Europei” în atlasul Dahmen și Kramer, care are 312 hărți și un chestionar întocmit de autor în atlasul publicat de Petru Neiescu, care are 212 hărți).

Prima monografie de sintaxă aromână, *La subordinazione circostanziale ipotattica nella frase del' dialetto aromeno (macedononeno)*, Istituto Universitario Orientale, Napoli, 1982, o datorăm lui Gheorghe Carageani.

Limba scrierilor aromânești de la sfârșitul secolului al XVIII-lea (Cavalioti, Daniil, Ucuta), partea I. *Aspecte ale grafiei. Fonetica* (Ovidius University Press, Constanța, 2004) de Nistor Bardu prezintă o analiză amănunțită a particularităților fonetice ale operelor scriitorilor aromâni din secolul al XVIII-lea.

O lucrare utilă, întemeiată pe cercetări de teren, despre *Aromânii din Rodopii Bulgariei* (Editura Beladi, Craiova, 1995) îi are ca autori pe Radu Sp. Popescu și Teodor Balkanski. Lucrarea cuprinde o descriere actualizată a zonei cercetate, care completează informațiile anterioare, fiind însoțită de o hartă, de un capitol inedit de *Onomastică* (toponimie, nume de botoz, nume de familie, nume de grup, supranume, porecle, p. 78–111) și de un corpus de texte dialectale (p. 118–139). Menționăm și o mică culegere de texte dialectale, redată într-o transcriere fonetică riguroasă: Maria Kokka, *Texte aromâne din comuna Nea Zoi, jud. Tricala* (Grecia), Universitatea din București, 1992.

Corpusurile fonetice întocmite de Adrian Turculeț în urma cercetărilor de teren pentru anchetele AMPER-ROM (partea consacrată limbii române din AMPER = *L'Atlas multimédia prosodique de l'Espace Roman*) au fost valorificate în studii precum: *Aspecte ale situației sociolingvistice a aromânilor din Albania*, (LLR – SEDD, 2003, p. 110–122), *Intonația interogativă la aromânii din R. Macedonia (cu câteva particularități segmentale ale informatorilor)* (FD, XXIX, 2010, p. 157–176), *Cercetări de (micro)fonetică aromânească (Omagiu lui Nicolae Saramandu*, București, 2011, p. 799–808).

În *Verbul în aromână. Structură și valori* (Editura Academiei Române, București, 2006) Manuela Nevaci studiază această parte de vorbire din perspectivă istorică și descriptivă, cu valorile (modale, temporale) exprimate de formule verbale, la care se adaugă o *Anexă* cu un inventar al verbelor, pe conjugări, în dialectul aromân.

Toate corpusurile românești adnotate morfologic sau chiar și sintactic, disponibile în format electronic, reflectă dialectul dacoromân. Menționăm faptul că nu există corpusuri ale dialectului aromân în format electronic. **Corpusurile scrise** sunt disponibile în format tipărit și reprezintă colecții de texte transcrise după înregistrări realizate în timpul anchetelor dialectale. Cele mai reprezentative **colecții de texte dialectale aromânești** sunt conținute în următoarele lucrări: *Aromâna vorbită în Dobrogea. Texte dialectale. Glosar* (Saramandu 2007b), *Identitate românească în context balcanic* (Nevaci 2013b), *Convergențe lingvistice balcano-romanice în dialectele românești sud-dunărene* (Nevaci 2018). În ceea ce privește **corpusurile orale**, menționăm existența unui DVD care însoțește lucrarea *Convergențe...*, din 2018, coordonată de Manuela Nevaci. DVD-ul anexat conține texte dialectale înregistrate de Manuela Nevaci în timpul anchetelor dialectale desfășurate în Albania, Grecia, R. Macedonia de Nord și Croația, în perioada 2016–2017.

O bogată bibliografie asupra dialectelor românești sud-dunărene găsim în Saramandu și Nevaci (2018), cap. III „Atlase lingvistice”; „Dialectologie română sud-dunăreană”, în Marius Sala, Nicolae Saramandu (coord.), *Lingvistica românească*, Editura Academiei Române, București, 2018.

Selectăm câteva resurse lingvistice, etnografice și muzicale online, aflate pe diverse site-uri și pagini de Facebook oficiale:

<https://www.facebook.com/Societatea-de-Cultur%C4%83-Macedo-Rom%C3%A2n%C4%83-184045761930325/>
<https://www.facebook.com/Fundatia-Nicolae-Iorga-Saranda-111315297714768/>
<https://www.facebook.com/groups/avdhela.biblioteca.di.cultura.armanasca/>
<https://www.facebook.com/arhiveleiplomaticeMAE>
https://www.facebook.com/permalink.php?story_fbid=1443077526027136&id=184045761930325
<https://www.facebook.com/groups/1428784617362384/>
https://www.revistasinteza.ro/istororomanii-zece-veacuri-de-singuratate?fbclid=iwar1ws8gbhkarx1tdtc9t_zjwebo71ns_gm8gq4uikqgsqtkwa5y9s-ooq
<https://cersipamanromanesc.wordpress.com/tag/aromanii-din-balcani-istorie/?fbclid=IwAR0CTPajrBRZN77-OuDI92nkcvbBnrX-2lWzrQdPKq-mbY7Xtiyj7ef2R-Q>
<https://www.facebook.com/photo?fbid=2784695844898612&set=pcb.10157473265616163>
<https://www.facebook.com/CIRKorce/>
<https://www.youtube.com/watch?v=JL8kPZ9cvfg>
<https://www.youtube.com/watch?v=31cnhDc2SHo>
<https://www.youtube.com/watch?v=Qx9qPNhI19g>
<https://www.youtube.com/watch?v=UaXAWSFdJJw>
 canalul “Angjel pro Albania”: www.youtube.com/watch?v=4JLvg777oXA
<https://www.youtube.com/watch?v=BbiHwkrx0jM>
<https://www.facebook.com/Societatea-de-Cultur%C4%83-Macedo-Rom%C3%A2n%C4%83-184045761930325/videos/969772260432016>
https://www.youtube.com/watch?v=ILIRt2-rcW0&fbclid=IwAR1wxw-xP7NGxMQowrvzaK_SZDTQID8zUP-2k7WaoI3Vxz8Vle7qoWF2nTI
 (921) Cantece Polifonice Palazu Mare - YouTu <https://www.youtube.com/watch?v=JL8kPZ9cvfgbe>
http://www.lingv.ro/index.php?option=com_content&view=article&id=70&Itemid=89
https://www.facebook.com/permalink.php?story_fbid=1162982010703357&id=184045761930325
https://www.facebook.com/permalink.php?story_fbid=577711199230444&id=184045761930325
 (ARHIVA SCMR)

Institute de știință și cultură:

<http://www.lingv.ro/images/ALDA.pdf>
http://www.lingv.ro/index.php?option=com_content&view=article&id=381&Itemid=270
<https://www.icr.ro/pagini/bucuresti-basme-aromane>
<https://l.facebook.com/l.php?u=https%3A%2F%2Fwww.icr.ro%2Fpagini%2Fbucuresti-basme-aromane-> (INSTITUTUL CULTURAL ROMÂN)
 Institutul de Cercetari al Romanilor din Ungaria - Maria Berényi, Familii si personalitatimacedoromane din Pesta (Secolele XVIII-XIX) (romanintezet. <https://www.facebook.com/CIRKorce/u>)

Institut de Phonétique-Musée de la Parole et du Gest:

<https://gallica.bnf.fr/ark:/12148/bpt6k10801066/f1.media?fbclid=IwAR3UVx0x8DmVyugLyDxBSLKh-fVwR7t0wKma4wwL0qX2NM9AzD7k-7s6BU#>
gallica.bnf.fr/ark:/12148/bpt6k1080127c/f1.media
gallica.bnf.fr/ark:/12148/bpt6k10801066/f1.media
<https://www.discogs.com/Various-Tumbe-Tumbe-C%C3%A2ntece-Arom%C3%A2ne%C8%99ti-Aromanian-Songs/release/13371104>
<https://www.discogs.com/label/1612653-Document:-Arhive-Folclorice-Rom%C3%A2ne%C8%99ti--Romanian-Folk-Archives>
https://www.ressources-mcm.com/s/Ibn_Battuta/item/1317
https://academiaromana.ro/ief/doc2021/Sanziene/Sarbatoarea-de-Sanziene-la-aromani.pdf?fbclid=IwAR3pK1ZsEh_yCMZcckt9f3O9Pnd56U1TOqrCpc_ZC5NZAOUJKypkhQ5jdOg

4. CORPUSUL ArT

4.1. Structura corpusului

Corpusul ArT, un nucleu deocamdată, este alcătuit din 50 de fraze extrase din lucrarea *Basme aromâne* (BA) a lui Pericle Papahagi. Ele nu ridică probleme din perspectiva dreptului de proprietate intelectuală. Frazele sunt dispartate, nu se constituie într-un text continuu, au lungimi variate, de la 4 la 30 de unități⁵, astfel încât lungimea medie a frazelor este de 11,46 de unități.

4.2. Metoda de lucru

Frazele au fost segmentate manual în unități, pentru care forma de bază (lema) a fost indicată, conform practicilor lexicografice specifice acestui dialect: pentru verbe, lema este forma de persoana întâi singular a indicativului prezent, pentru substantive este forma de nominativ-acuzativ singular nearticulat, pentru adjective este cea de masculin, nominativ-acuzativ singular nearticulat. Atunci când gramaticile sau dicționarele (vezi *supra*) indică mai multe variante pentru aceeași formă, am optat pentru consecvență în lematizare. Astfel, pentru pronumele reflexiv, lema aleasă este *si*, pentru conjuncția specifică modului conjunctiv am ales lema *să*, iar pentru cea de condițional lema este *să*. În cazul pronumelor personale, lema aleasă depinde de forma ocurentă: astfel, forme precum *î*, *u*, *lo* etc. sunt lematizate la *el*, iar forme precum *năs*, *năși*, *năsă* etc. sunt lematizate la *năs*, atât *el*, cât și *năs* fiind forme de persoana a treia, masculin, singular, nominativ-acuzativ.

Adnotarea morfologică a fost realizată manual, folosind setul de PoStaguri creat în proiectul MULTTEXT-East pentru limba română (v. *supra*), la care a fost nevoie să aducem câteva completări, dat fiind faptul că morfologia dialectului aromân prezintă anumite diferențe față de cea a dialectului dacoromân (vezi *infra*).

Pentru adnotarea la nivel sintactic s-au folosit seturile de relații sintactice definite în UD, fără vreo ajustare necesară pentru dialectul aromân. În UD se folosește un inventar de relații sintactice considerate universale, adică utilizabile la descrierea oricărei limbi, dar este permisă și formularea unor relații, subtipuri ale celor universale, necesare pentru a surprinde particularități ale unor limbi. Inventarul pentru limba română conține și câteva subtipuri de relații (Barbu Mititelu și Irimia 2015), dintre care în ArT regăsim *obl:pmod* (noua denumire a fostei relații *nmod:pmod*, folosită pentru complementul prepozițional), *ccomp:pmod* (relație folosită pentru echivalentul prepozițional al unui complement prepozițional), *expl:pv* (folosit pentru adnotarea reflexivului), *expl:poss* (folosit pentru valorile posesive ale reflexivului), *expl:impers* (folosit pentru valoarea impersonală a reflexivului).

Prezentăm în Figura 1 analiza completă a frazei din exemplul (1). Observăm că fraza are un număr de ordine în fișierul din care face parte (fișier numit „test”), și anume 28. Fiecare unitate de analiză (cuvinte și punctuație) primește un număr de ordine, corespunzător poziției în frază: *Si* are numărul de ordine 1, iar punctul final al frazei, numărul 20 (acesta fiind și numărul total de unități din frază). Aceste numere de ordine ocupă prima coloană din fișier. Formatul fișierului cuprinde zece coloane: a doua reia cuvintele în forma ocurentă în frază; coloana a treia conține lemele cuvintelor ocurente⁶; coloana a patra conține partea de vorbire în format UD⁷: PRON este pentru pronume, ADP este pentru prepoziții, NOUN este pentru

⁵ Prin unități înțelegem cuvinte și punctuație.

⁶ În cazul punctuației, coloana a treia este ocupată de același semn ca cea de-a doua.

⁷ Inventarul lor este disponibil la <https://universaldependencies.org/u/pos/index.html>. Semnele de punctuație au pe această coloană marcajul PUNCT.

substantive, DET este pentru determinatori, CCONJ este pentru conjuncții coordonatoare, PART este pentru particule (precum marca de conjunctiv aici); următoarea coloană conține adnotarea morfologetică (POS tagul) în formatul MULTEXT-East⁸, urmată, în coloana a șasea, de exprimarea aceluiași informații gramaticale în format atribut=valoare, în limba engleză; coloana a șaptea conține numărul de ordine al regentului (sau centrului) cuvântului respectiv: acesta este întotdeauna unul dintre numerele de pe prima coloană, cu excepția centrului frazei (de regulă, verbul predicativ din prima propoziție principală din frază), care are pe această coloană numărul 0, deoarece el nu depinde sintactic de niciun alt cuvânt din frază; coloana a opta conține relația sintactică stabilită între cuvântul respectiv și centrul său: de exemplu, *amirălu* are ca centru verbul *durăna* (de pe poziția 18), iar relația dintre ele este *nsubj*: *amirălu* este subiectul lexicalizat printr-un substantiv al verbului *durăna*. Ultimele două coloane rămân fără informații, deci conțin câte o linie de subliniere.

- (1) *Si mveaște tu strațiile a cafîgîluiș-acață s-ăaură ca nă cătușe pi ninga ușa iu durăna amirălu.* (Se îmbracă în straietele cafegiului și începe să miaune ca o pisică pe lângă ușa unde dormea împăratul.)

```
# sent_id = test-28
# text = Si mveaște tu strațiile a cafîgîluiș-acață s-ăaură ca nă cătușe pi ninga ușa iu durăna amirălu.
1- Si si PRON Pn3--a--y----w Case=Acc|Person=3|PronType=Prs|Reflex=Yes|Strength=Weak|Variant=Short 2- expl _ _
2- mveaște mvescu VERB Vvip3s Mood=Ind|Number=Sing|Person=3|Tense=Pres|VerbForm=Fin 0- root _ _
3- tu tu ADP Sp3s AdpType=Prep|Case=Acc 4- case _ _
4- strațiile straîli MOUN Ncf3s Case=Acc,Nom|Definite=Def|Gender=Fem|Number=Plur 2- obl _ _
5- a a DET T3 Poss=Yes|PronType=Prs 0- det _ _
6- cafîgîlui cafîgi MOUN Ncns3y Case=Dat,Gen|Definite=Def|Gender=Masc|Number=Sing 4- nmod _ _
7- și și CCONJ C3spp Polarity=Pos 8- cc SpaceAfter=No _ _
8- acață acațu VERB Vvip3s Mood=Ind|Number=Sing|Person=3|Tense=Pres|VerbForm=Fin 2- conj _ _
9- s- sã PART Os-y Mood=Sub|Variant=Short 10- mark SpaceAfter=No _ _
10- aură aură VERB Vv3ps Mood=Sub|Number=Sing|Person=3|Tense=Pres|VerbForm=Fin 8- ccomp _ _
11- ca ca ADP Sp3s AdpType=Prep|Case=Acc 13- case _ _
12- nă un DET T1fsr Case=Acc,Nom|Gender=Fem|Number=Sing|PronType=Ind 13- det _ _
13- cătușe cătușe MOUN Ncf3sm Case=Acc,Nom|Definite=Ind|Gender=Fem|Number=Sing 10- obl _ _
14- pi pi ADP Sp3s AdpType=Prep|Case=Acc 16- case _ _
15- ninga ninga ADP Sp3s AdpType=Prep|Case=Acc 16- case _ _
16- ușa ușa MOUN Ncf3s Case=Acc,Nom|Definite=Def|Gender=Fem|Number=Sing 10- obl _ _
17- iu iu ADV Rv PronType=Int,Rel 18- advmod _ _
18- durăna durã VERB Vv113 Mood=Ind|Person=3|Tense=Imp|VerbForm=Fin 16- acl _ _
19- amirălu amirã MOUN Ncns3y Case=Acc,Nom|Definite=Def|Gender=Masc|Number=Sing 18- nsubj _ SpaceAfter=No
20- . . PUNCT PERIOD 2- punct _ _
```

Figura 1. Exemplu de adnotare a unei fraze.

Celelalte relații sintactice din exemplul (1) sunt: *expl* – folosit pentru pronumele reflexive, *case* – leagă prepoziția de cuvântul pe care îl precedă⁹, *obl* – leagă un adjunct de verbul centru, *det* – leagă determinatorul de centrul său, *nmod* – leagă un modificator nominal de centrul său nominal, *cc* – leagă conjuncția coordonatoare de conjunctul care o urmează, *conj* – leagă un conjunct de primul conjunct în ordine liniară, *mark* – leagă marcatorii de centrul lor (aici, leagă conjuncția marcă a conjunctivului de verbul la acest mod), *ccomp* – leagă completiva (directă sau indirectă) de verbul regent, *advmod* – leagă adverbul de cuvântul pe care îl modifică, *acl* – leagă o atributivă de nominalul regent.). Verbul *mveaște* este fără explozia nazalității la inițială, fără vocala de sprijin *i*, ca în albaneză și în greacă. Conjunctivul *s-ăaură* este cerut de verbul de aspect *acață* ‘începe’. La verbele de conjugarea I aromâna neutralizează opoziția dintre indicativ și conjunctiv, desinența *-ă* (*ă*) de la indicativ prezent (pers. a III-a sg./pl.) extinzându-se și la conjunctiv (în loc de *-e* așteptat din conjunctivul latinesc): *s-ăaură*.

⁸ Explicațiile detaliate ale PoStagurilor pentru limba română sunt disponibile la <http://nl.ijs.si/ME/Vault/V5/msd/html/msd-ro.html>.

⁹ Prepoziția nu este centru pentru cuvintele precedate în UD, adică nu are valoare de predicat. Este doar un instrument folosit de anumite limbi pentru a exprima cazul, exprimat flexionar în alte limbi.

4.3. Particularitățile dialectului aromân reflectate parțial de ArT

Structura dialectului aromân este descrisă în Saramandu 1984a, Saramandu și Nevaci 2013a, Nevaci 2013b, avându-se în vedere nivelurile fonetic, fonologic și morfosintactic. La fel ca în celelalte dialecte ale limbii române, particularitățile morfosintactice ale dialectului aromân reprezintă dezvoltări ulterioare românei comune. Prin elementele conservatoare, în special în domeniul foneticii și fonologiei, dialectul aromân se apropie de româna comună. Fenomenele inovatoare se explică atât prin evoluția internă a dialectului, cât și prin influența exercitată de limbile balcanice.

În clasa **substantivului**, se observă următoarele trăsături care diferențiază dialectul aromân de dacoromâna vorbită astăzi: lipsa prepoziției *pe* în construcția obiectului direct realizat prin nume de persoană (ca în româna comună și ca în româna veche); lipsa prepoziției în exprimarea acuzativului numelor de locuri; lipsa morfemului de determinare definită la substantivele care exprimă obiectul posedat. Analiza exemplului (2) este redată în Figura 2: observăm că substantivul „fîcorlu” stabilește cu verbul „agudiși” relația *obj* (adică obiect direct) și că nu este precedat de o prepoziție, așa cum se întâmplă în dacoromână, deși acest substantiv are trăsătura semantică [+ persoană].

(2) *Că-țelo-agudi și fîcorlu, measă, ntribă Araplu.* (De ce îl loviși pe băiat, masă, întrebă arapul.)

```
# sent_id = test-34
# text = Că-țe lo-agudiși fîcorlu, measă, ntribă Araplu.
1. Că-țe SCONJ Ccasp Polarity=Pos 4. mark _ SpaceAfter=No
2. -țe ADV Re PronType=Int,Rel 4. advmod _
3. lo-el PRON Pp3apa-y-...-v Case=Acc|Gender=Masc|Number=Plur|Person=3|PronType=Prs|Strength=Weak|Variant=Short 4. expl _ SpaceAfter=No
4. agudiși agudescu VERB Vm1s2s Mood=Ind|Number=Sing|Person=2|Tense=Past|VerbForm=Fin 0. root _
5. fîcorlu fîcor NOUN Ncasry Case=Acc,Nom|Definite=Def|Gender=Masc|Number=Sing 4. obj _ SpaceAfter=No
6. , PUNCT COPMA 7. punct _
7. measă measă NOUN Ncfsvn Case=Voc|Definite=Ind|Gender=Fem|Number=Sing 4. vocative _ SpaceAfter=No
8. , PUNCT COPMA 7. punct _
9. ntribă ntrib VERB Vm1s3s Mood=Ind|Number=Sing|Person=3|Tense=Past|VerbForm=Fin 4. parataxis _
10. Araplu arap NOUN Ncasry Case=Acc,Nom|Definite=Def|Gender=Masc|Number=Sing 9. nsubj _ SpaceAfter=No
11. , PUNCT PERIOD 4. punct _
```

Figura 2. Exemplu cu obiect direct.

În dialectul aromân nu sunt înregistrate forme de **articol** nehotărât și de articol demonstrativ diferite de pronumele nehotărât sau de pronumele demonstrativ. În plus, pentru masculine (graiul grămostean) există o formă invariabilă *a* de articol posesiv (genitival) care însoțește pronumele posesive sau substantive în genitiv (Saramandu 1984a: 439). În Figura 1 observăm acest articol pe poziția a cincea în frază, adnotat ca determinant, conform convențiilor MULTTEXT-East, și determinând substantivul „cafișgului”. În graiul fărșerot, unele substantive masculine cunosc o dublă modalitate de articulare, atât cu articol hotărât enclitic/*a*, cât și cu articol hotărât proclitic: *al aușu* → *a aișului* (Nevaci 2011: 113)

În clasa **pronumelui**, se observă următoarele particularități: pronumele personal se realizează prin mai multe forme de nominativ-acuzativ (în exemplul (3) se observă forma „mine” pentru persoana întâi, numărul singular, cazul nominativ, iar în exemplul (5) se observă forma „io” cu aceleași caracteristici morfologice, lucru imediat dedus din faptul că cele două forme au același PoStag: Pp1-sn-----s, care codifică următoarele informații: pronume (P) personal (p), persoana întâi (1), numărul singular (s), cazul nominativ (n), formă accentuată (s)); acuzativul obiect direct nume de persoană nu este precedat de prepoziția *pe* (asemănător cu substantivul, vezi *supra*); pronumele de politețe nu există; pronumele posesiv este precedat întotdeauna de articolul posesiv (genitival) invariabil *a*; o serie de pronume și adjective nehotărâte devin pronume și adjective negative atunci când apar în contexte cu verbul negat (Cristescu 2020a, 2020b). Menționăm faptul că, în ceea ce privește uzul pronumelui personal, există o diferențiere pe graiuri: în graiul fărșerot, pentru pronumele

personal de persoana I *mine* se folosește forma de acuzativ pentru cea de nominativ, iar în graiul grămostean se folosește nominativul și pentru acuzativ: N. *mini mâc*^o ‘eu mănânc’; Ac. *Đínu la mini* ‘vino la mine’ (graiul fărșerot); N. *~o mâc*^o ‘eu mănânc’; Ac. *Đínu la ~o* ‘vino la mine’ (graiul grămostean) (Nevaci 2013b: 75; 100; Saramandu și Nevaci 2013b: 200).

În exemplul (3) este folosit pronumele „vârnă”, considerat pronume nehotărât în gramaticile dialectului aromân (Nevaci 2013b:31), ca subiect (vezi relația *nsubj*) al verbului „cârteașce”, care este însoțit de adverbul negativ „Nu”, ceea ce face ca pronumele să fie analizat ca negativ (vezi tagul Pz3msr, care are următoarea semnificație: pronume (P) negativ (z), persoana a treia (3), genul masculin (m), numărul singular (s), cazul direct (r)).

În exemplul (2) observăm că obiectul direct cu caracteristicile [+ persoana] și [+ definit] „fîcorlu” (cu tagul Ncmsry, având semnificația substantiv (N), comun (c), genul masculin (m), numărul singular (s), cazul direct (r), articulat hotărât (y)) este dublat, ca și în dacoromână, de un clitic pronominal: „lo-”. Pentru că valența de obiect a verbului este saturată de substantiv, atunci cliticului îi revine relația *expl* (expletiv). Substantivul *fîcorlu* are morfemul de determinare definită *-lu*, specific graiului grămostean; în graiul fărșerot, morfemul de determinare definită este marcat, la această categorie, prin ocurența lui *u* plenison *fîcoru*, față de forma nedeterminată definit *fîcor*^o, cu ^o final asilabic.

- (3) *Nu ti cârteașce vârnă, că va lă si pară că escu mine.* (Nu se va lega nimeni de tine, căci o să li se pară că sunt eu.)

```
# sent_id = test-4
# text = Nu ti cârteașce vârnă, că va lă si pară că escu mine.
1: Nu nu PART Qz Polarity=Neg 3 advmod
2: ti tîni PRON Pp2-sn-----w Case=Acc|Number=Sing|Person=2|PronType=Prs|Strength=Weak 3 obj
3: cârteașce cârteașcu VERB Vmip3s Mood=Ind|Number=Sing|Person=3|Tense=Pres|VerbForm=Fin 0 root
4: vârnă vârnă PRON Pz3msr Case=Acc, No|Gender=Masc|Number=Sing|Person=3|PronType=Ind 3 nsubj
5: , PUNCT COMMA 10 punct
6: că că SCOMJ Ccssp Polarity=Pos 10 mark
7: va va1 AUX Va-3p Number=Plur|Person=3 10 aux
8: lă el' PRON Pp3-pd-----w Case=Dat|Number=Plur|Person=3|PronType=Prs|Strength=Weak 10 iobj
9: si si PRON Px3--a-y-----w Case=Acc|Person=3|PronType=Prs|Reflex=Yes|Strength=Weak|Variant=Short 10 expl:impers
10: pară par VERB Vmip3 Mood=Ind|Person=3|Tense=Pres|VerbForm=Fin 3 advcl
11: că că SCOMJ Ccssp Polarity=Pos 12 mark
12: escu hîu VERB Vmip1s Mood=Ind|Number=Sing|Person=1|Tense=Pres|VerbForm=Fin 10 csbj
13: mine mini PRON Pp1-sn-----s Case=Nom|Number=Sing|Person=1|PronType=Prs|Strength=Strong 12 nsubj
14: , PUNCT PERIOD 3 punct
```

Figura 3. Pronume nehotărât cu valoare de pronume negativ.

Dintre particularitățile clasei **verbului** în dialectul aromân (care este descrisă detaliat în Nevaci 2006), le menționăm pe următoarele: o mare complexitate a flexiunii verbale, în ceea ce privește atât inventarul flectivelor, cât și structura radicalului; fluctuații de conjugare, care au condus la reducerea celor patru conjugări la trei; modul conjunctiv cunoaște patru timpuri: prezent, imperfect, perfect și mai-mult-ca-perfect (după modelul din neogreacă și albaneză, vezi și Saramandu 1969); modul condițional-optativ cunoaște un timp sintetic (prezent) (vezi exemplul (4) și Figura 4) care se construiește cu ajutorul conjuncției *să* (*s*, *s-*, *z-*) și două timpuri analitice (perfect și mai-mult-ca-perfect); modul imperativ nu are forme speciale de pozitiv (vezi exemplul (4) și Figura (4)) și negativ (vezi exemplul (5) și Figura (5)); spre deosebire de dacoromână, forma negativă este alcătuită din forma pozitivă, precedată de adverbul *nu*; infinitivul este o formă invariabilă simplă și cunoaște numai forma lungă (moștenită din latină). Forma *cârteașce* este specifică graiului crușovean, vorbit în R. Macedonia de Nord, în timp ce în celelalte graiuri este ocurentă forma *cârteaște*. Forma de indicativ prezent pers. I sg a verbului sunt, *escu*, este specifică graiurilor grămostean, pindean și fărșerot (tipul plisot), în timp ce la fărșeroți (tipul șopan) întâlnim forma *hîu* (Saramandu 2017a: 25).

- (4) *Și s-nu pistipsiri, măsură tine diznou!* (Și dacă nu crezi, măsoară tu din nou!)

```

# sent_id = test-8
# text = și s-nu pistipsiri, măsură tine diznou!
1- și- și- CCONJ- Ccssp- Polarity-Pos- 6- cc- _-
2- s- s- PART- Qc- Mood-End- 4- mark- _- SpaceAfter-No
3- nu- nu- PART- Qz- Polarity-Mag- 4- advmod- _-
4- pistipsiri- pistipsescu- VERB- Vmcp3- Mood-End|Person=3|Tense=Pres|VerbForm=Fin- 6- advcl- _- SpaceAfter-No
5- ,- ,- PUNCT- COPMA- 6- punct- _-
6- măsură- măsur- VERB- Vmm-2s- Mood-Imp|Number=Sing|Person=2|VerbForm=Fin- 0- root- _-
7- tine- tine- PRON- Pp2-sn------s- Case-Nom|Number=Sing|Person=2|PronType=Prs|Strength=Strong- 6- nsubj- _-
8- diznou- diznou- ADV- Rgp- Degree-Pos- 6- advmod- _- SpaceAfter-No
9- !- !- PUNCT- EXCL- 6- punct- _-

```

Figura 4. Condițional prezent sintetic.

În exemplul (4), verbul *s-pistipsiri* este la modul condițional-optativ, timpul prezent. Forme de condițional sintetic sunt frecvente în graiul fărșerot (vezi Nevaci 2006:144; Saramandu 1969: 160). Le întâlnim și în dacoromână, în textele din secolul al XVI-lea, și în istroromâna vorbită astăzi (cf. Densusianu 1961: 147, Neiescu 2011). PoStagul său (Vmcp3) conține aceste informații: verb (V) predicativ (m), modul condițional (c), timpul prezent (p), persoana a treia (3). Întrucât există omonimie pentru persoana a treia la singular și la plural, informația de număr nu este cuprinsă în PoStag (conform principiilor MULTTEXT-East), astfel încât acesta poate fi folosit pentru ambele forme. Verbul are și conjuncția *s-*, lematizată ca *să*, care intră în relația *mark* cu verbul. Este important de remarcat faptul că setul de PoStaguri pentru limba română a trebuit să fie modificat pentru analiza verbelor de acest fel: poziția a treia din PoStag nu permitea valoarea „c”, pentru că aceasta indică modul condițional în limbile care îl formează analitic, nu sintetic cum este cazul în dacoromână. O altă modificare a setului de PoStaguri a fost adăugarea valorii „c” pentru a indica modul condițional în cazul conjuncției specifice acestui mod: vezi în Figura (4) tagul Qc atribuit conjuncției *s-*. Până la crearea ArT, particulele (Q) puteau fi doar pentru modul conjunctiv (caz în care pe a doua poziție a PoStagului apare valoarea „s”), pentru negativ (valoarea „z”, vezi în Figura 5 PoStagul adverbului negativ *nu*), pentru infinitiv (valoare care nu se aplică dialectului).

Al doilea verb din exemplul (4), „măsură”, are PoStagul Vmm-2s, care codifică următoarele informații morfologice: verb (V) predicativ (m), modul imperativ (m), fără timp (-), persoana a doua (2), numărul singular (s). Exemplul (5) conține un verb (*bati*) la imperativ negativ. PoStagul său este același ca al verbului *misură*, dar are și adverbul negativ *nu* care intră în relație cu el.

(5) *Nu mi bati, că ț-aduc ș-alti petali io!* (Nu mă bate, că ț-aduc și-alte potcoave eu!)

```

# sent_id = test-5
# text = Nu mi bati, că ț-aduc ș-alti petali io!
1- Nu- nu- PART- Qz- Polarity-Mag- 3- advmod- _-
2- mi- mi- PRON- Pp1-sn------w- Case-Acc|Number=Sing|Person=1|PronType=Prs|Strength=Weak- 3- obj- _-
3- bati- bat- VERB- Vmm-2s- Mood-Imp|Number=Sing|Person=2|VerbForm=Fin- 0- root- _- SpaceAfter-No
4- ,- ,- PUNCT- COPMA- 7- punct- _-
5- că- că- SCONJ- Ccssp- Polarity-Pos- 7- mark- _-
6- ț- ț- PRON- Pp2-sd------w- Case-Dat|Number=Sing|Person=2|PronType=Prs|Strength=Weak- 7- iobj- _- SpaceAfter-No
7- aduc- aduc- VERB- Vwip1s- Mood-Ind|Number=Sing|Person=1|Tense=Pres|VerbForm=Fin- 3- advcl- _-
8- ș- și- ADV- Cr- 9- advmod- _- SpaceAfter-No
9- alti- alt- DET- D13fpr---s- Case-Acc- Nom|Gender=Fem|Number=Plur|Person=3|Position=Preced|PronType=Ind- 10- det- _-
10- petali- petală- NOUN- Nc-fp-n- Definite=Ind|Gender=Fem|Number=Plur- 7- obj- _-
11- io- mi- PRON- Pp1-sn------s- Case-Nom|Number=Sing|Person=1|PronType=Prs|Strength=Strong- 7- nsubj- _- SpaceAfter-No
12- !- !- PUNCT- EXCL- 3- punct- _-

```

Figura 5. Imperativ negativ.

Din punct de vedere sintactic, observăm plasarea cliticului pronominal între auxiliar și verb: în exemplul (3) pronumele personal *lă* (având PoStagul Pp3-pd-----w care codifică informațiile: pronume (P) personal (p), persoana a treia (3), numărul plural (p), cazul dativ (d), formă atonă (w)) și cel reflexiv *si* (având PoStagul Px3--a--y-----w codifică informațiile: pronume (P) reflexiv (x), persoana a treia (3), numărul plural (p), cazul acuzativ (a), formă

clitică (y), formă atonă (w)) sunt plasate între auxiliarul *va* (cu PoStagul Va--3p: verb (V) auxiliar (a), persoana a treia (3), numărul plural (p)) și verbul *pară* (cu PoStagul Vmip3: verb (V) predicativ (m), modul indicativ (i), timpul prezent (p), persoana a treia (3)).

4.4. Date statistice despre corpusul ArT

Corpusul ArT conține 50 de fraze, în care găsim 573 de unități de analiză, dintre care 484 sunt doar cuvinte, ceea ce înseamnă că punctuația reprezintă 15,5%. Frecvența părților de vorbire este redată în următorul tabel:

Tabelul 1

Frecvența părților de vorbire în corpusul ArT

PoS	Număr de ocurențe	Frecvență relativă	PoS	Număr de ocurențe	Frecvență relativă
ADJ (adjectiv)	9	1,9%	NOUN (substantiv)	79	16,3%
ADP (prepoziție)	41	8,4%	NUM (numeral)	6	1,2%
ADV (adverb)	55	11,4%	PART (particulă)	50	10,3%
AUX (verb auxiliar)	9	1,9%	PRON (pronume)	71	14,7%
CCONJ (conjunție coordonatoare)	17	3,5%	SCONJ (conjunție subordonatoare)	15	3,1%
DET (determinator)	15	3,1%	VERB	115	23,8%
INTJ (interjecție)	2	0,4%	TOTAL	484	100%

Din numărul total de cuvinte, numărul de leme unice este de 235, ceea ce înseamnă 48,6%. Diversitatea lemelor pentru fiecare parte de vorbire este redată în Tabelul 2 și se poate observa că substantivul și verbul au o diversitate destul de mare, de 76%, respectiv 59%. Niciunul dintre adjective nu se repetă în corpus.

Tabelul 2

Diversitatea vocabularului din ArT

PoS	Număr de leme unice	Procent de leme unice	PoS	Număr de leme unice	Procent de leme unice
ADJ (adjectiv)	9	100%	NOUN (substantiv)	60	76%
ADP (prepoziție)	14	34%	NUM (numeral)	4	66%
ADV (adverb)	34	62%	PART (particulă)	3	6%
AUX (verb auxiliar)	3	33%	PRON (pronume)	12	17%
CCONJ (conjunție coordonatoare)	3	18%	SCONJ (conjunție subordonatoare)	6	40%
DET (determinator)	8	53%	VERB	68	59%
INTJ (interjecție)	2	100%	TOTAL	235	48,6

În ceea ce privește relațiile sintactice dintre cuvintele din corpus, frecvența lor este prezentată în Tabelul 3.

Observăm că numărul de relații *root* este egal cu numărul de fraze, adică 50: fiecare frază este reprezentată ca un arbore, a cărui rădăcină este verbul din propoziția principală. Propozițiile subordonate sunt atributive (vezi relația *acl*), circumstanțiale realizate propozițional (vezi relația *advcl*), subiective (vezi relația *csubj*), complementive directe și indirecte (ambele adnotate cu aceeași relație *ccomp*), subordonate corespunzătoare unui complement prepozițional (vezi relația *obl:pmod*). Complexitatea frazelor se observă și din frecvența conjuncțiilor subordonatoare și a propozițiilor marcă a unor moduri verbale (vezi relația *mark*), a conjuncțiilor coordonatoare

(relația *cc*), a conjuncțiilor (relația *conj*). Numărul mai mare al conjuncțiilor decât al conjuncțiilor coordonatoare arată faptul că, uneori, coordonarea se realizează prin juxtaponere.

Tabelul 3

Relațiile sintactice din ArT

Relație sintactică	Număr de ocurențe	Relație sintactică	Număr de ocurențe	Relație sintactică	Număr de ocurențe
acl	3	det	14	nsubj	41
advcl	29	discourse	1	nummod	6
advmod	75	expl	11	obj	28
amod	7	expl:impers	1	obl	25
aux	6	expl:poss	1	obl:pmod	5
case	41	expl:pv	19	parataxis	4
cc	17	fixed	2	root	50
ccomp	9	iobj	7	vocative	2
conj	21	mark	45	xcomp	4
csubj	2	nmod	7		

Cel mai frecvent, subiectul se realizează lexical și foarte rar propozițional: sunt 41 de relații *nsubj* (cu care este adnotat subiectul lexical) și 2 relații *csubj* (pentru subiective). Cum numărul de subiecte, indiferent de felul lor, este mai mic chiar decât numărul de fraze din corpus (43 față de 50), reiese că sunt multe cazuri de subiect neexprimat: de exemplu, în Figura 4 se observă că primul verb (poziția 4) nu are un subiect exprimat, pe când al doilea verb (poziția 6) are subiectul exprimat printr-un pronume, chiar dacă este la modul imperativ.

Atributul adjectival este rar: sunt doar 7 relații *amod*. Tot atâtea sunt și atributele substantivale (vezi relația *nmod*). Obiectul direct (relația *obj*) apare de 28 de ori, iar cel indirect (relația *iobj*) de 7 ori. Realizarea lor propozițională (relația *ccomp*) se petrece de 9 ori. Complemente propoziționale (relația *obl:pmod*) sunt 5.

Observăm prezența a 10 verbe reflexive: acestea sunt centru pentru un clitic cu care intră în relația *expl:pv*. Cliticele reflexive apar sporadic și cu valoare impersonală (*expl:impers*) și posesivă (*expl:poss*). Nu există niciun verb la diateza pasivă în corpus, lucru dedus din absența relațiilor specifice pentru ilustrarea acesteia: *aux:pass* (pentru auxiliarul de pasiv), *nsubj:pass* (pentru subiectul lexicalizat), *csubj:pass* (pentru subiectiva unui verb la diateza pasivă) sau *expl:pass* (pentru un clitic reflexiv cu valoare pasivă). Acest lucru nu înseamnă că diateza pasivă nu există în dialectul aromân, ci doar că ea nu este reflectată de corpusul ArT.

În Figura 6 redăm reprezentarea arborescentă a exemplului următor:

(6) *Ľepurle ansare d-iu nu te-aștept.* (Iepurele sare de unde nu te aștepti.)

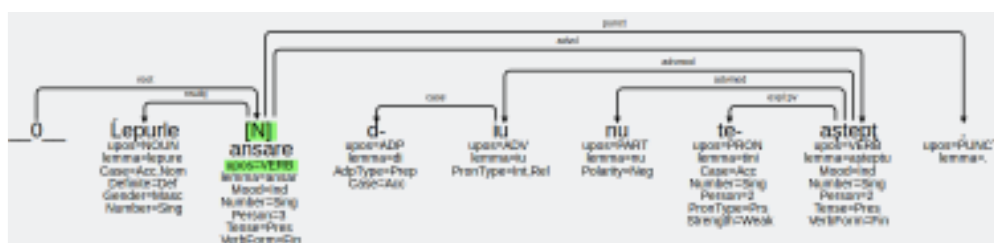


Figura 6. Exemplu de arbore sintactic din ArT.

În ceea ce privește lexicul enunțurilor analizate, observăm preponderența celui moștenit din latină și concordanța cu dacoromâna, cu excepția a patru împrumuturi:

	Dialect aromân	Coreșpondența cu dialectul dacoromân	Etimonul
1.	<i>a</i> (al)	a (al)	< Lat. <i>illum, illam</i> .
2.	<i>acață</i>	agață	< Lat. <i>*accaptiare</i>
3.	<i>aduc</i>	aduc	< Lat. <i>adducere</i>
4.	<i>agudiși</i> , pers. s.	–	< Vs. <i>goditi</i>
5.	<i>alti</i>	alte	< Lat. <i>alt[(e)rum]</i>
6.	<i>amiră</i>	–	< Tc. <i>amyr</i>
7.	<i>ansare</i>	sare, vb.	< Lat. <i>salire</i>
8.	<i>arap</i>	arap	< Bg. <i>arap</i>
9.	<i>aștepți</i> , Ind.prez. II, pl.	aștepți	< Lat. <i>*astectare (= adspectare)</i>
10.	<i>bati</i>	bate	< Lat. <i>batt(u)ere</i>
11.	<i>ca</i>	ca	< Lat. <i>quam</i>
12.	<i>cafișului</i>	cafișiu	< Tc. <i>kahveci</i>
13.	<i>cătușe</i>	cătușă (înv.)	derivat de la <i>*cată</i> (< lat. <i>catta</i> „pisică”) + <i>-ușă</i>]
14.	<i>cărteașce</i> , Indprez. 3 sg.	cărtește	< ? <i>krūtēnije</i> „bodogăneală” (Miklosich, Slaw. Elem., 27; Cihac, II, 44); cf. bg. <i>kürjã</i>
15.	<i>dî</i>	de	< Lat. <i>de</i>
16.	<i>dîznou</i>	dîn nou	<i>din + nou</i>
17.	<i>durnă</i> , imperf. 3 sg.	dormea	< Lat. <i>dormire</i>
18.	<i>escu</i>	<u>sunt</u> , cf. <i>ești</i>	< Lat. <i>sum, *fui, *fire (= fieri)</i> .
19.	<i>fîcor</i>	fêcor	< Lat. <i>*fetiolum</i> din <i>fât</i> + suf. <i>-ior</i>
20.	<i>io</i>	io (înv.)	< Lat. <i>ego</i>
21.	<i>iu</i>	unde	< Lat. <i>iuve</i>
22.	<i>lă</i> , pron. 3 pl.	le	< Lat. <i>illum, illa</i>
23.	<i>lepurle</i> , s. art.	iepurele	< Lat. <i>lepus, -oris</i>
24.	<i>measă</i>	masă	< Lat. <i>mensa</i>
25.	<i>mi</i>	mi	< Lat. <i>me</i>
26.	<i>misură</i>	măsoară	< Lat. <i>mensurare</i>
27.	<i>mveaște</i>	învește (înv. și reg.)	< Lat. <i>investire</i>
28.	<i>ňaură</i>	miaună	der. de la <i>miau</i>
29.	<i>nă</i>	ne	< Lat. <i>nos</i>
30.	<i>ninga</i>	lângă	derivat din <i>longum ad</i> , conf. Alb. <i>nëngë</i>
31.	<i>ntribă</i>	întrebă	< Lat. <i>*interrogare (= interrogare)</i> .
32.	<i>nu</i>	nu	< Lat. <i>non</i>
33.	<i>pară</i>	para	< Tc. <i>para</i>
34.	<i>petală</i>	-	< Gr. <i>πέταλον</i>
35.	<i>pi</i>	pe	< Lat. <i>super, per</i>
36.	<i>si pistipsiri</i> , condițional, 3 sg.	-	< Gr. <i>πιστεύω</i>
37.	<i>si</i>	se	< Lat. <i>se</i>
38.	<i>străhile</i>	straie	Etimologie necunoscută
39.	<i>și</i>	și	< Lat. <i>sic</i>
40.	<i>tîne</i>	tine	< Lat. <i>tu</i>
41.	<i>ușa</i>	ușa	< Lat. <i>*ustia</i> (pl., devenit sg., al lui <i>*ustium = ostium</i>).
42.	<i>va</i>	va	< Lat. <i>*volere</i>
43.	<i>vârnu</i>	vreun	< Lat. <i>*vere-unus</i>

5. CONCLUZII

Corpusul ArT este accesibil în cadrul proiectului UD, alături de celelalte corpusuri pentru limba română, fiind însă singurul care oferă analiza morfosintactică a unor fraze aparținând dialectului aromân. Crearea acestui corpus a dovedit faptul că inventarul de PoStaguri creat pentru limba română se aplică cu succes și (cel puțin) unuia dintre dialectele sale, necesitând rafinări pentru particularitățile sale în raportul cu limba. De asemenea, inventarul de relații propus în UD și folosit pentru limba română nu a avut nevoie de nicio ajustare pentru analiza frazelor din ArT. Ne exprimăm speranța că pașii de informatizare a limbii române, inclusiv a dialectelor sale, vor continua și se vor intensifica în viitorul apropiat.

SURSE

BA – Pericle Papahagi, *Basme aromâne*, București, 1905.

BIBLIOGRAFIE

- Barbu Mititelu, Irimia 2015 – Verginica Barbu Mititelu, Elena Irimia, „Description of the Romanian Syntax within Universal Dependency Project”, în: Daniela Gifu, Diana Trandabăț, Dan Cristea, Dan Tufiș (coord.), *Proceedings of the 11th International Conference „Linguistic Resources and Tools for Processing the Romanian Language”*, Iași, Editura Universității Alexandru Ioan Cuza, p. 185–194.
- Barbu Mititelu *et alii* 2016 – Verginica Barbu Mititelu, Radu Ion, Radu Simionescu, Elena Irimia, Cenel-Augusto Perez, „The Romanian Treebank Annotated According to Universal Dependencies”, în *Proceedings of HrTAL2016*, Dubrovnik, Croatia, 29 septembrie – 1 octombrie 2016.
- Barbu Mititelu, Mitrofan 2020 – Verginica Barbu Mititelu, Maria Mitrofan, „The Romanian Medical Treebank – SiMoNERo”, în: Verginica Barbu Mititelu, Elena Irimia, Dan Tufiș, Dan Cristea (coord.), *Proceedings of the 15th International Conference „Linguistic Resources and Tools for Processing the Romanian Language”*, Iași, Editura Universității Alexandru Ioan Cuza, p. 7–16.
- Bick, Greavu 2010 – Eckhard Bick, Arina Greavu, „A Grammatically Annotated Corpus of Romanian Business Texts”, în *Proceedings of Multilinguality and Interoperability in Language Processing with Emphasis on Romanian*, București, Editura Academiei Române, p. 169–183.
- Colhon *et alii* 2017 – Mihaela Colhon, Cătălina Mărănduc, Cătălin Mititelu, „A Multiform Balanced Dependency Treebank for Romanian”, în *Proceedings of Knowledge Resources for the Socio-Economic Sciences and Humanities, (Know RSH)*, Workshop at the Recent Advances in Natural Language Processing (RANLP), Varna, Bulgaria, 8 septembrie 2017, p. 9–19.
- Cristescu 2020a – Mihaela Cristescu, „Negative Concord Items in the Aromanian dialect of the Romanian Language”, în *Analele Universității din București. Limba și literatura română*, nr. LXIX, p. 69–80.
- Cristescu 2020b – Mihaela Cristescu, „Negation in the Aromanian dialect of the Romanian language: a typological analysis”, în *Fonetica și dialectologie*, XXXIX, p. 47–59.
- Dahmen, Kramer 1985, 1994 – Wolfgang Dahmen, Johannes Kramer, *Aromunischer Sprachatlas*, vol. I. (1985), vol. II (1994), Hamburg, Buske.
- DDA 1963/2013 – Tache Papahagi, *Dicționarul dialectului aromân*, București, Editura Academiei Republicii Populare Române, București, 1963; 2013 – ediție anastatică (ed. Nicolae Saramandu, Manuela Nevaci), Editura Academiei Române, București, 2013.
- Densusianu 1961 – Ovid Densusianu, *Istoria limbii române*, 2 vol., Editura Științifică.
- Erjavek 2017 – Tomaz Erjavek, „MULTEXT-East”, în Nancy Ide, James Pustejovsky (eds.), *Handbook of linguistic annotation*, Springer, p. 441–462.

- Hristea, Popescu 2003 – Florentina Hristea, Marius Popescu, „A Dependency Grammar Approach to Syntactic Analysis with Special Reference to Romanian”, în: Florentina Hristea, Marius Popescu (coord.), *Building Awareness in Language Technology*, București, Editura Universității din București, 9–16.
- Neiescu 1997 – Petru Neiescu, *Mic atlas al dialectului aromân din Albania și din Fosta Republică Iugoslavă Macedonia*, București, Editura Academiei Române, 1997.
- Neiescu 2011, 2015, 2016 – Petru Neiescu, *Dicționarul dialectului istroromân*, București, Editura Academiei Române, I (A–C), 2011; II (Č–K), 2015; III (L–Pînța2), 2016.
- Nevaci 2006 – Manuela Nevaci, *Verbul în aromână. Structură și valori*, București, Editura Academiei Române.
- Nevaci 2011 – Manuela Nevaci, *Graiul aromânilor fărșeroți din Dobrogea*, București, Editura Universitară.
- Nevaci 2013a – Manuela Nevaci, *Identitate românească în context balcanic*, București, Editura Muzeului Național al Literaturii Române.
- Nevaci 2013b – Manuela Nevaci, *Dialectele aromân și meglenoromân. Studiu sincron*, București, Editura Universității din București.
- Nevaci 2013c – Manuela Nevaci, *Scriitori aromâni de ieri și azi. Antologie de texte și studiu filologic*, București, Editura Universitară.
- Nevaci 2018 – Manuela Nevaci, „Dialectul aromân”, în: Manuela Nevaci (coord.), Carmen-Irina Floarea, Ionuț Geană, Teodora Toroipan, *Convergențe lingvistice balcano-romance în dialectele românești sud-dunărene*, București, Editura Universității din București, p. 29–67.
- Perez 2014 – Ceneal-Augusto Perez, *Linguistic Resources for Natural Language Processing* (în română), teză de doctorat, Iași, Universitatea „Alexandru Ioan Cuza”.
- Saramandu 1969 – Nicolae Saramandu, *Sistemul formelor verbale compuse în aromână*, în *FD*, VI, p. 155–162.
- Saramandu 1984a – Nicolae Saramandu, „Aromâna”, în Valeriu Rusu (coord.), *Tratat de dialectologie românească*, Craiova, Scrisul Românesc, p. 423–476.
- Saramandu 1984b – Nicolae Saramandu, *Gheorghe Constantin Rojași „Școala Normală a Nației Românești” din Pesta, Ungaria (1808–1810)*, în vol. „Contribuția românilor la îmbogățirea tezaurului cultural în Balcani”, Fundația Culturală Română, București, [1992], p. 54–60.
- Saramandu 1984c – Nicolae Saramandu, „Gramatica română sau macedoromână” de Mihai G. Boiagi. *175 de ani de la apariție*, în *LL*, nr. 4, p. 600–602.
- Saramandu 2000 – Nicolae Saramandu, *Sisteme de scriere a aromânei*, în *SCL*, LI, 2–3, p. 199–204.
- Saramandu 2004 – Nicolae Saramandu, *Romanitatea orientală*, București, Editura Academiei Române (ediția în limba franceză, Editura Academiei Române, București / Gunter Naar Verlag, Tübingen, 2008).
- Saramandu 2005 – Nicolae Saramandu, „Un structuralist avant la lettre: Gheorghe Constantin Roja”, în *Actele celui de al 4-lea Colocviu al Catedrei de limba română*, București, Editura Universității, p. 759–763.
- Saramandu 2007a – Nicolae Saramandu, „Gheorghe Constantin Roja – un precursor al dialectologiei românești”, în vol. *Lucrările primului Simpozion Internațional de Lingvistică*, București, Editura Universității din București, p. 131–138.
- Saramandu 2007b – Nicolae Saramandu, *Aromâna vorbită în Dobrogea. Texte dialectale. Glosar*, București, Editura Academiei Române.
- Saramandu 2013 – Nicolae Saramandu, „Momentul Moscopole” în istoria raporturilor lingvistice româno-albaneze, în *FD*, XXXII, p. 129–137.
- Saramandu 2014 – Nicolae Saramandu, *Atlasul lingvistic al dialectului aromân*, volumul I, București, Editura Academiei Române (editor: Manuela Nevaci).
- Saramandu 2017a – Nicolae Saramandu, *Structura dialectală a aromânei actuale. Graiul pindean*, în *SCL*, LXVIII, 1, p. 3–23.
- Saramandu 2017b – Nicolae Saramandu, *Considerații asupra graiului aromânilor fărșeroți*, în *FD*, XXXVI.

- Saramandu, Nevaci 2013a – Nicolae Saramandu, Manuela Nevaci, *Sinteze de dialectologie română*, București, Editura Universității din București.
- Saramandu, Nevaci 2013b – Nicolae Saramandu, Manuela Nevaci, „Considerații asupra aromânei actuale. Graiul grămostean”, în Daniela Răuțu, Adrian Rezeanu, Dana-Mihaela Zamfir (coord.), *„Cuvinte potrivite”. Omagiu doamnei Maria Marin la aniversare*, București, Editura Academiei Române, p. 430–440.
- Saramandu, Nevaci 2018 – Nicolae Saramandu, Manuela Nevaci, cap. III „Atlase lingvistice”; „Dialectologie română sud-dunăreană”, în: Marius Sala, Nicolae Saramandu (coord.), *Lingvistica românească*, București, Editura Academiei Române.
- Saramandu, Nevaci 2021 – Nicolae Saramandu, Manuela Nevaci, *Atlasul lingvistic al dialectului aromân*, volumul II, București, Editura Academiei Române.
- Trandabăț *et alii* 2012 – Diana Trandabăț, Elena Irimia, Verginica Barbu Mititelu, Dan Cristea, Dan Tufiș, *Limba română în era digitală – The Romanian Language in the Digital Age*, în Georg Rehm, Hans Uszkoreit (eds.), META-NET White Paper Series, Springer.
- Tufiș *et alii* 1997 – Dan Tufiș, Ana-Maria Barbu, Vasile Pătrașcu, Georgiana Rotariu, Camelia Popescu, „Corpora and Corpus-based Morpho-lexical Processing”, în: Dan Tufiș, Poul Andersen (eds.), *Recent Advances in Romanian Language Technology*, București, Editura Academiei Române, p. 35–56.

*Institutul de Cercetări pentru Inteligență Artificială
„Mihai Drăgănescu” al Academiei Române
Universitatea din București
Institutul de Lingvistică al Academiei Române
„Iorgu Iordan – Al. Rosetti”
Calea 13 Septembrie nr. 13, București*