eLEXICOGRAFIA – INTERDISCIPLINARITATEA CA PREMISĂ PENTRU CERCETAREA LIMBII ROMÂNE[1]

eLexicography – Interdisciplinarity as a Premise for the Research of Romanian Language

Scient. Researchers Dr. Elena TAMBA,
Scient. Researchers Dr. Marius CLIM,
Scient. Researchers Dr. Ana CATANĂ-SPENCHIU,
Scient. Researchers, PhD Mădălin PĂTRAŞCU
"A. Philippide" Institute of Philology of the Romanian Academy,
Iaşi Branch

Abstract

During the last years, the Romanian linguistic research in general and the lexicographic one in particular have known an intensive growth in the natural language processing area. In this direction, interdisciplinary steps have been taken to provide Romanian electronic dictionaries and text corpora.

For researchers, the Romanian Thesaurus Dictionary in electronic format (eDTLR) and the Essential Romanian Lexicographic Corpus (ERLC) will provide important and awaited tools for the Romanian language study.

Keywords: interdisciplinarity, digitalized_lexicography, lexicographic_corpus, linguistic_resources, e-lexicography

În ultimii ani, atât în domeniul lexicografiei, cât şi în cel al informaticii aplicate s-a simţit nevoia unei evoluţii interdisciplinare, ca urmare a cerinţelor de dezvoltare pe cât mai multe direcţii. Astfel, într-o primă fază la propunerea cercetătorilor informaticieni cu preocupări în domeniul prelucrării limbajului natural, au fost iniţiate unele proiecte comune, interdisciplinare, care au dus la realizarea unor corpusuri de limbă română sau a unor programe care să ajute la prelucrarea textelor în limba română, cu rezultate eficiente în domeniul cercetării acesteia.

Istoricul proiectelor de tip academic, care au presupus o interdisciplinaritate sporită, începe cu demersurile de informatizare a *Dicţionarului Tezaur al limbii române*, care reprezintă cea mai importantă lucrare lexicografică apărută sub egida Academiei Române – editarea sa a început acum 105 ani; a apărut în două serii: DA (1907-1944), DLR (1965-2010), fiind publicat în 14 tomuri / 37 volume, totalizând cca 18.000 pagini şi mai mult de 175.000 intrări, cu tot cu variante.

Astfel, începând cu anul 2005, la Iaşi, s-au desfăşurat unele proiecte interdisciplinare, în care au fost implicaţi lexicografii de la Institutul de Filologie Română „A. Philippide", Iaşi, lingvişti de la Facultatea de Litere de la Universitatea „Alexandru Ioan Cuza" din Iaşi şi informaticienii de la Facultatea de Informatică, Universitatea „Alexandru Ioan Cuza" din Iaşi şi de la Institutul de Informatică Teoretică, Academia Română, Iaşi:

[1] *Acknowledgement*: Acest articol a fost realizat în cadrul proiectului *CLRE. Corpus lexicografic românesc esenţial. 100 de dicţionare de bază din Bibliografia DLR aliniate la nivel de intrare şi la nivel de sens*, CNCS-UEFISCDI, cod TE_246/2010, 2010 – 2013TE_246 (55/2010).

a) *Dicţionarul limbii române (DLR) în format electronic. Studii privind achiziţionarea* (cod CNCSIS 1815), grant finanţat de Ministerul Educaţiei şi Cercetării (MEC) prin Consiliul Naţional al Cercetării Ştiinţifice din Învăţământul Superior (CNCSIS), desfăşurat în perioada 2003–2005. Prin acest proiect s-a verificat şi demonstrat posibilitatea transformării *Dicţionarului limbii române* din text tipărit în text electronic adnotat[2], prelucrat cu ajutorul unui program specific, DLRex – un instrument de achiziţionare, prelucrare şi consultare a DLR, bazat pe o euristică prin care sunt recunoscute diferitele câmpuri formale ale textului unui articol, putându-se identifica automat textul definiţiilor, al citatelor sau al siglelor[3].

b) *Resurse lingvistice în format electronic: Monumenta linguae Dacoromanorum. Biblia 1688. Regum I, Regum II – Ediţie critică şi corpus adnotat*. *(MLD. Biblia 1688)* (cod CNCSIS 1454), desfăşurat în perioada 2006–2007. Prin acest proiect a fost găsită o posibilă metodă de achiziţionare în format electronic a unor cărţi vechi din Bibliografia DLR, cu aplicaţie asupra a două cărţi din *Biblia* tipărită la Bucureşti în anul 1688, *A împărăţiilor cea dentâiu, A împărăţiilor a doua,* precum şi crearea unor instrumente de indexare şi adnotare automată, la nivel de cuvânt, a textelor româneşti vechi.

c) *DLRI. Bază lexicală informatizată. Derivate.* (cod CNCSIS nr. 1609), desfăşurat în perioada 2007–2008. Prin acest proiect s-a prelucrat un eşantion lexicografic format din derivatele pe terenul limbii române cu sufixul *-ime* – de origine latină, şi cele cu *-işte* – de origine veche slavă, din seria veche (DA) şi din seria nouă a dicţionarului-tezaur (DLR), şi s-a demonstrat posibilitatea unificării tehnico-lexicografice a articolelor DA – DLR, prin mijloace informatice.

Aceste trei proiecte interdisciplinare au anunţat şi pregătit realizarea unui proiect cu implicare naţională, care a vizat realizarea variantei electronice a *Dicţionarului Tezaur al limbii române* – eDTLR. Echipa proiectului a fost coordonată de Facultatea de Informatică, Universitatea „Alexandru Ioan Cuza" din Iaşi, prin prof. dr. Dan Cristea şi a inclus ca parteneri: Institutul de Lingvistică „Iorgu Iordan – Al. Rosetti", Academia Română, Bucureşti; responsabil de proiect acad. Marius Sala (prin dr. Monica Busuioc); Institutul de Filologie Română „A. Philippide", Academia Română, Iaşi; responsabil de proiect dr. Gabriela Haja; Institutul de Lingvistică şi Istorie Literară „Sextil Puşcariu", Academia Română, Cluj-Napoca; responsabil de proiect dr. Rodica Marian; Institutul de Cercetări pentru Inteligenţă Artificială, Academia Română, Bucureşti; responsabil de proiect acad. Dan Tufiş; Institutul de Informatică Teoretică, Academia Română, Iaşi; responsabil de proiect acad. Horia Neculai Teodorescu; Facultatea de Litere, Universitatea „Alexandru Ioan Cuza" din Iaşi; responsabil de proiect dr. Eugen Munteanu.

Acest proiect[4] a beneficiat iniţial de o finanţare naţională din partea CNMP, iar ulterior, după decembrie 2010, lucrându-se la finalizarea site-ului eDTLR doar cu ajutorul unor voluntari – cercetători informaticieni şi lexicografi. Proiectul menţionat, după încheierea sa, ar trebui, pe de o parte, să pună la dispoziţia tuturor cunoscătorilor sau celor interesaţi de

[2] Textul electronic adnotat este un text analizat şi marcat din punct de vedere formal astfel încât să poată fi consultat, corectat, modificat etc. de către specialiştii lexicografi, cu ajutorul calculatorului. Există posibilitatea extragerii din formatul complet a unei forme destinate numai consultării, care să se adreseze unui public mai larg decât cel al specialiştilor propriu-zişi. Pentru detalii, vezi şi Haja, Dănilă *et alii*, 2005.

[3] Pentru o prezentarea detaliată a rezultatelor acestui proiect, vezi Haja, Dănilă *et alii* 2005.

[4] Pentru o prezentare detaliată a proiectului, vezi Dănilă 2010: 37–46.

limba română formatul electronic al *Dicţionarului Academiei*, pe suport electronic – şi, poate ulterior, în funcţie de schimbarea politicii lingvistice din România, şi on-line, în acces liber sau condiţionat – şi, pe de altă parte, să pună la dispoziţia deocamdată doar a cercetătorilor o arhivă electronică care să cuprindă toate textele din Bibliografia DLR. Prin rezultatele sale, eDTLR ar pune lexicografia fundamentală pentru limba română într-o situaţie de egalitate cu cea a limbilor care au deja dezvoltate astfel de resurse: *Le Trésor de la Langue Française Informatisé* (TLFi – http://atilf.atilf.fr/); *Diccionario de la lengua espanola de la Real Academia Espagnola* (DRAE – http://buscon.rae.es/draeI/); *Tesoro della lingua italiana delle origini* (TLIO –http://tlio.ovi.cnr.it/TLIO/index2.html); *Deutsches Worterbuch der Grimm* (DWB – http://germazope.uni-trier.de/Projects/DWB); *Oxford English Dictionary* (OED – http://www.oed.com/) ş.a.

Pornind de la eDTLR a apărut necesitatea realizării unui alt proiect interdisciplinar: crearea unui corpus lexicografic românesc – *CLRE. Corpus lexicografic românesc esenţial. 100 de dicţionare din bibliografia DLR aliniate la nivel de intrare şi la nivel de sens*, care este finanţat de CNCS-UEFISCDI pentru perioada 2010 – 2013 şi care se desfăşoară tot la Institutul de Filologie Română „A. Philippide" din Iaşi.

Cu ajutorul unei echipe formate din doar 4 cercetători (3 lexicografi – Elena Tamba, Marius-Radu Clim şi Ana-Veronica Catană-Spenchiu – şi un informatician – Marius Răschip) şi cu aportul unui alt informatician voluntar – Mădălin Pătraşcu, proiectul îşi propune să realizeze o bază de date care să cuprindă dicţionarele esenţiale din Bibliografia DLR, aliniate la nivel de intrare şi parţial la nivel de sens; să construiască un mediu de programe care să permită consultarea interactivă a acestui corpus şi care să se constituie într-un cadru modern de lucru şi cercetare lexicografică, uşor adaptabil la o diversitate de obiective; să realizeze o listă de cuvinte cvasi-exhaustivă, pentru limba română, pornind de la corpusul aliniat. Astfel, prin acest demers se vizează: realizarea unui corpus scanat, format din dicţionarele de referinţă ale DLR (cu respectarea legislaţiei în vigoare în ceea ce priveşte drepturile de proprietate intelectuală); scanarea şi prelucrarea (OCR-izarea[5]; parsarea[6] textului la nivel de intrare şi, parţial, la nivel de sens) a acestor dicţionare; realizarea unei interfeţe on-line pentru validarea/corectarea parsării, precum şi validarea alinierii între textul *Dicţionarului Tezaur al limbii romane* (în format electronic, rezultat al proiectului eDTLR) şi dicţionarele de referinţă din Bibliografia DLR. Dicţionarele alese au fost incluse în trei categorii:

1. *dicţionare generale*, de tipul:

DA = *Dicţionarul limbii române*, tom I-II, Tipografia ziarului „Universul", Bucureşti, Imprimeria Naţională, 1907-1944;

DLR = *Dicţionarul limbii române*, Serie nouă, tom VI-XIV, Bucureşti, Editura Academiei, 1965-2010;

DEX = *Dicţionarul explicativ al limbii romane*. Bucureşti, Editura Academiei, 1975;

2. *dicţionare auxiliare* (care sunt strâns legate de redactarea Dicţionarului Tezaur), de tipul:

[5] Convertirea din format imagine în format text.
[6] Identificarea automată a intrărilor din dicţionarele scanate şi OCR-izate anterior.

Alexandru Ciorănescu, *Dicţionarul etimologic al limbii române*. Ediţie îngrijită şi traducere din limba spaniolă de Tudora Sandru-Mehedinţi şi Magdalena Popescu Marin. Bucureşti, Editura Saeculum I. O., 2002.

**** Dicţionarul ortografic, ortoepic şi morfologic al limbii române*. Ediţia a II-a revăzută şi adăugită, Bucureşti, Univers Enciclopedic, 2005.

Florin Marcu, *Noul dicţionar de neologisme*. Bucureşti, Editura Academiei Române, 1997.

3. *dicţionare speciale* (enciclopedice ori dicţionare speciale, alese după criteriul importanţei lor pentru perspectiva diacronică asupra limbii), de tipul:

Dicţionar enciclopedic. [Vol.] I: A–C (1993), [vol.] II: D–G (1996), [vol.] III: H–K (2000), [vol.] IV: L–N (2001), [vol. V]: O–Q (2004). [vol.] VI: R–S (2006). [vol.] VII: T–Z (2009). Bucureşti, Editura Enciclopedică;

I.-Aurel Candrea – Gh. Adamescu, *Dicţionarul enciclopedic ilustrat. Partea I: Dicţionarul limbii române din trecut şi de astăzi* de I.-Aurel Candrea. *Partea II: Dicţionarul istoric şi geografic universal* de Gh. Adamescu. Bucureşti, Editura Cartea Românească, [1926–1931];

Lexiconul tehnic român. I ş. u. Elaborare nouă. Bucureşti, Editura Tehnică, 1957 ş. u.

În proiect se utilizează, astfel, atât metode lingvistice clasice / tradiţionale (de exemplu, transliterarea intrărilor în alfabet chirilic sau de tranziţie ori studiul comparativ, la nivel semantic, al dicţionarelor), cât şi metode noi, de lexicografie computaţională.

Pe plan internaţional, există deja astfel de corpusuri, care facilitează foarte mult cercetarea atât în domeniul studiului limbii / limbilor, cât şi în cel al informaticii aplicate, al prelucrării limbajului natural:

– *Le rayon des dictionnaires* (http://www.atilf.fr/) – colecţie de dicţionare informatizate franceze, din secolul al XVI-lea pînă în secolul al XX-lea;

– *Nuevo tesoro lexicográfico de la lengua española* (http://buscon.rae.es/ntlle/SrvltGUILoginNtlle) – bază de date cuprinzând versiunile facsimilate ale tuturor dicţionarelor editate şi publicate de Real Academia Espagnola;

– *Das Wörterbuchnetz* (http://germazope.uni-trier.de/Projects/WBB/) – reţea de dicţionare de limba germană, creată la universitatea Trier din Germania.

Pentru realizarea corpusului de dicţionare (CLRE) este nevoie de parcurgerea mai multor etape care necesită o adaptare în funcţie de specificul lucrărilor lexicografice avute în vedere. În primul rând este indispensabilă transpunerea în format electronic a celor 100 de dicţionare. Acest lucru presupune scanarea şi prelucrarea textelor. Astfel se impune utilizarea atât a unor echipamente performante care să faciliteze achiziţionarea electronică a dicţionarelor, cât şi a unor programe de prelucrare a scanărilor şi de recunoaştere a caracterelor care să permită realizarea în bune condiţii a bazei de date. Pentru aceasta s-a achiziţionat un scaner special pentru cărţi, Atiz BookDrive DIY[7]. Varianta aleasă s-a dovedit a fi cea mai potrivită soluţie de digitalizare a cărţilor, din perspectiva costurilor şi a eficienţei, aceasta din urmă fiind dată de performanţele camerelor digitale SLR şi de suportul de carte în forma literei V. Scanerul Atiz este prevăzut cu două camere Canon EOS 450D cu lentile de 35 mm. Lentilele EF 35 mm permit o focalizare mai mare şi sunt folosite în special pentru cărţile

[7] Mai multe informaţii despre acest produs sunt disponibile pe site-ul http://diy.atiz.com/.

în format A3 sau A2. Scanerul poate fi folosit pentru toate tipurile de cărți, indiferent de dimensiune, grosime sau tipul de legătură. În plus, este ideal pentru scanarea şi prelucrarea cărților vechi care necesită o atenție aparte.

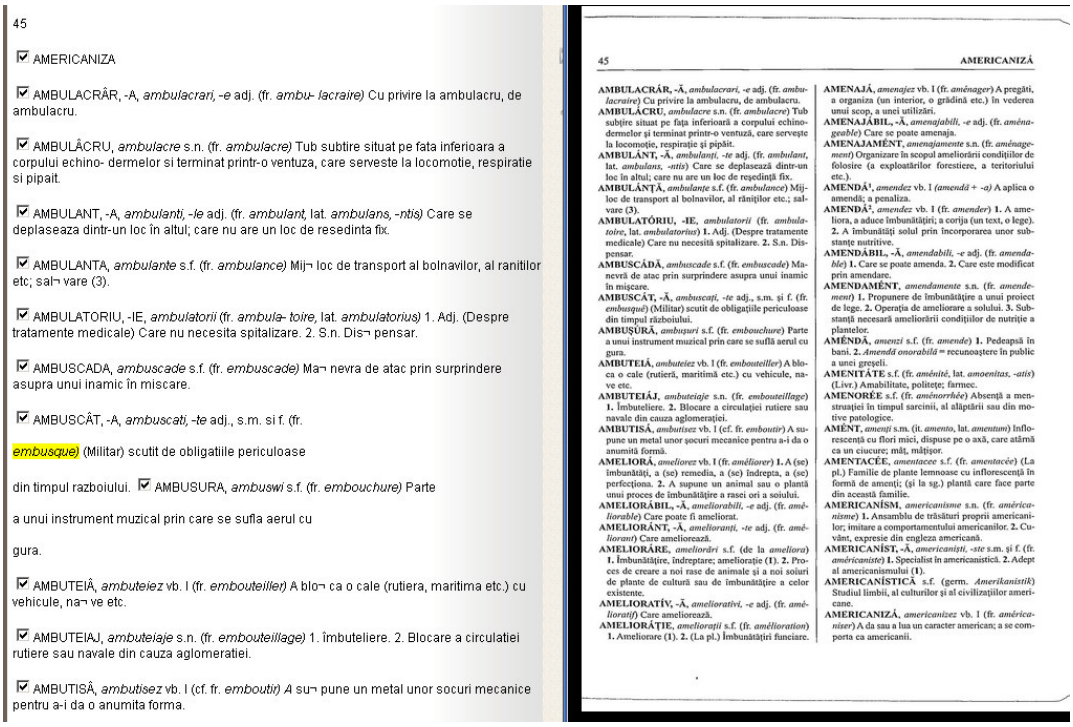
Cea de-a doua etapă presupune transformarea imaginilor scanate în text editabil. Pentru recunoaşterea optică a caracterelor din imaginile scanate şi pentru procesarea de text a fost achiziționat ABBYY FineReader Engine, care înglobează atât sistemul de recunoaştere optică a caracterelor (OCR), cât şi recunoaşterea inteligentă a caracterelor (ICR), recunoaşterea mărcii optice (OMR), recunoaşterea codurilor de bare (OBR), procesare de imagini şi conversia în format .pdf. Acest program permite convertirea rapidă şi cu acurateţe a documentelor scanate, a fişierelor în format .pdf sau a documentelor în format imagine într-o varietate de formate Office în care se pot realiza căutări şi care sunt uşor de editat.

Toate aceste echipamente şi programe facilitează o prelucrare informatică de mare acurateţe a materialului lexicografic avut în vedere.

Cel de-al treilea pas a presupus realizarea unei interfeţe de validare care să permită verificarea corectitudinii parsării şi validarea intrărilor. Aşadar, dicţionarele scanate sunt introduse într-o bază de date şi, după parsare, sunt validate de către lexicografi. Interfaţa de validare permite vizualizarea fiecărei pagini de dicţionar în parte. Pentru fiecare dicţionar este afişată pagina-titlu, iar lexicograful poate insera toate informaţiile despre dicţionar (autor, editura etc.). În baza de date dicţionarele sunt identificate după sigla din bibliografia DLR.

Lexicografii validează fiecare pagină, iar în cazurile în care parsarea s-a realizat cu erori, pagina respectivă este validată doar parţial, urmând a fi reprelucrată şi trimisă din nou spre validare.

Paginile validate corect au toate intrările recunoscute, aşa cum se poate observa şi în imaginea următoare.



45

☑ AMERICANIZA

☑ AMBULACRÂR, -A, *ambulacrari, -e* adj. (fr. *ambu- lacraire)* Cu privire la ambulacru, de ambulacru.

☑ AMBULÂCRU, *ambulacre* s.n. (fr. *ambulacre)* Tub subtire situat pe fata inferioara a corpului echino- dermelor si terminat printr-o ventuza, care serveste la locomotie, respiratie si pipait.

☑ AMBULANT, -A, *ambulanti, -le* adj. (fr. *ambulant,* lat. *ambulans, -ntis)* Care se deplaseaza dintr-un loc în altul; care nu are un loc de resedinta fix.

☑ AMBULANTA, *ambulante* s.f. (fr. *ambulance)* Mij¬ loc de transport al bolnavilor, al ranitilor etc; sal¬ vare (3).

☑ AMBULATORIU, -IE, *ambulatorii* (fr. *ambula- toire,* lat. *ambulatorius)* 1. Adj. (Despre tratamente medicale) Care nu necesita spitalizare. 2. S.n. Dis¬ pensar.

☑ AMBUSCADA, *ambuscade* s.f. (fr. *embuscade)* Ma¬ nevra de atac prin surprindere asupra unui inamic în miscare.

☑ AMBUSCÂT, -A, *ambuscati, -te* adj., s.m. si f. (fr. *embusque)* (Militar) scutit de obligatiile periculoase din timpul razboiului. ☑ AMBUSURA, *ambuswi* s.f. (fr. *embouchure)* Parte a unui instrument muzical prin care se sufla aerul cu gura.

☑ AMBUTEIÂ, *ambuteiez* vb. I (fr. *embouteiller)* A blo¬ ca o cale (rutiera, maritima etc.) cu vehicule, na¬ ve etc.

☑ AMBUTEIAJ, *ambuteiaje* s.n. (fr. *embouteillage)* 1. îmbuteliere. 2. Blocare a circulatiei rutiere sau navale din cauza aglomeratiei.

☑ AMBUTISÂ, *ambutisez* vb. I (cf. fr. *emboutir) A* su¬ pune un metal unor socuri mecanice pentru a-i da o anumita forma.

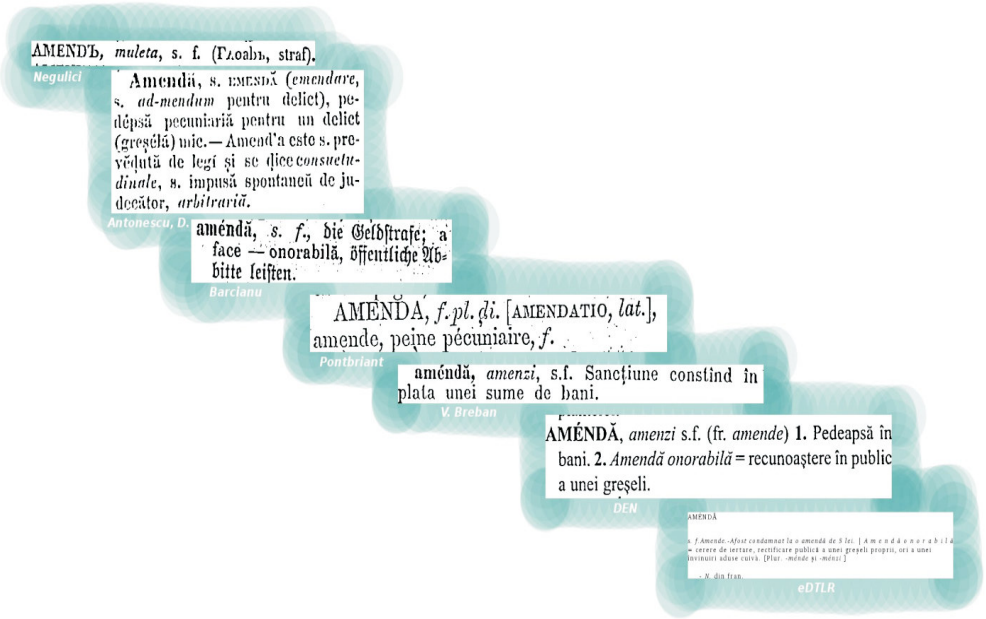
AMBULACRÁR, -Ă, *ambulacrari, -e* adj. (fr. *ambulacraire*) Cu privire la ambulacru, de ambulacru.
AMBULÁCRU, *ambulacre* s.n. (fr. *ambulacre*) Tub subțire situat pe fața inferioară a corpului echinodermelor și terminat printr-o ventuză, care servește la locomoție, respirație și pipăit.
AMBULÁNT, -Ă, *ambulanți, -te* adj. (fr. *ambulant,* lat. *ambulans, -ntis*) Care se deplasează dintr-un loc în altul; care nu are un loc de reședință fix.
AMBULÁNȚĂ, *ambulanțe* s.f. (fr. *ambulance*) Mijloc de transport al bolnavilor, al răniților etc.; salvare (3).
AMBULATÓRIU, -IE, *ambulatorii* (fr. *ambulatoire,* lat. *ambulatorius*) **1.** Adj. (Despre tratamente medicale) Care nu necesită spitalizare. **2.** S.n. Dispensar.
AMBUSCÁDĂ, *ambuscade* s.f. (fr. *embuscade*) Manevră de atac prin surprindere asupra unui inamic în mișcare.
AMBUSCÁT, -Ă, *ambuscați, -te* adj., s.m. și f. (fr. *embusqué*) (Militar) scutit de obligațiile periculoase din timpul războiului.
AMBUȘÚRĂ, *ambușuri* s.f. (fr. *embouchure*) Parte a unui instrument muzical prin care se suflă aerul cu gura.
AMBUTEIÁ, *ambuteiez* vb. I (fr. *embouteiller*) A bloca o cale (rutieră, maritimă etc.) cu vehicule, nave etc.
AMBUTEIÁJ, *ambuteiaje* s.n. (fr. *embouteillage*) **1.** Îmbuteliere. **2.** Blocare a circulației rutiere sau navale din cauza aglomerației.
AMBUTISÁ, *ambutisez* vb. I (cf. fr. *emboutir*) A supune un metal unor șocuri mecanice pentru a-i da o anumită formă.
AMELIORÁ, *ameliorez* vb. I (fr. *améliorer*) **1.** A (se) îmbunătăți, a (se) remedia, a (se) îndrepta, a (se) perfecționa. **2.** A supune un animal sau o plantă unui proces de îmbunătățire a rasei ori a soiului.
AMELIORÁBIL, -Ă, *ameliorabili, -e* adj. (fr. *améliorable*) Care poate fi ameliorat.
AMELIORÁNT, -Ă, *amelioranți, -te* adj. (fr. *améliorant*) Care ameliorează.
AMELIORÁRE, *ameliorări* s.f. (de la *ameliora*) **1.** Îmbunătățire, îndreptare; ameliorație (1). **2.** Proces de creare a noi rase de animale și a noi soiuri de plante de cultură sau de îmbunătățire a celor existente.
AMELIORATÍV, -Ă, *ameliorativi, -e* adj. (fr. *améliorati(f)*) Care ameliorează.
AMELIORÁȚIE, *ameliorații* s.f. (fr. *amélioration*) **1.** Ameliorare (1). **2.** (La pl.) Îmbunătățiri funciare.
AMENAJÁ, *amenajez* vb. I (fr. *aménager*) A pregăti, a organiza (un interior, o grădină etc.) în vederea unui scop, a unei utilizări.
AMENAJÁBIL, -Ă, *amenajabili, -e* adj. (fr. *aménageable*) Care se poate amenaja.
AMENAJAMÉNT, *amenajamente* s.n. (fr. *aménagement*) Organizare în scopul ameliorării condițiilor de folosire (a exploatărilor forestiere, a teritoriului etc.).
AMENDÁ[1], *amendez* vb. I (*amendă + -a*) A aplica o amendă; a penaliza.
AMENDÁ[2], *amendez* vb. I (fr. *amender*) **1.** A ameliora, a aduce îmbunătățiri; a corija (un text, o lege). **2.** A îmbunătăți solul prin încorporarea unor substanțe nutritive.
AMENDÁBIL, -Ă, *amendabili, -e* adj. (fr. *amendable*) **1.** Care se poate amenda. **2.** Care este modificat prin amendare.
AMENDAMÉNT, *amendamente* s.n. (fr. *amendement*) **1.** Propunere de îmbunătățire a unui proiect de lege. **2.** Operația de ameliorare a solului. **3.** Substanță necesară ameliorării condițiilor de nutriție a plantelor.
AMÉNDĂ, *amenzi* s.f. (fr. *amende*) **1.** Pedeapsă în bani. **2.** *Amendă onorabilă* = recunoaștere în public a unei greșeli.
AMENITÁTE s.f. (fr. *aménité,* lat. *amoenitas, -atis*) (Livr.) Amabilitate, politețe; fărmec.
AMENORÉE s.f. (fr. *aménorrhée*) Absență a menstruației în timpul sarcinii, al alăptării sau din motive patologice.
AMÉNT, *amenți* s.m. (it. *amento,* lat. *amentum*) Inflorescență cu flori mici, dispuse pe o axă, care atârnă ca un ciucure; mătură, mătișor.
AMENTACÉE, *amentacee* s.f. (fr. *amentacée*) (La pl.) Familie de plante lemnoase cu inflorescență în formă de amenți (și la sg.) plantă care face parte din această familie.
AMERICANÍSM, *americanisme* s.n. (fr. *américanisme*) **1.** Ansamblu de trăsături proprii americanilor; imitare a comportamentului americanilor. **2.** Cuvânt, expresie din engleza americană.
AMERICANÍST, -Ă, *americaniști, -ste* s.m. și f. (fr. *américaniste*) **I.** Specialist în americanistică. **2.** Adept al americanismului (1).
AMERICANÍSTICĂ s.f. (germ. *Amerikanistik*) Studiul limbii, al culturilor și al civilizațiilor americane.
AMERICANIZÁ, *americanizez* vb. I (fr. *américaniser*) A da sau a lua un caracter american; a se comporta ca americanii.

În această etapă fiecare dicționar trebuie tratat separat întrucât există caracteristici specifice în ceea ce privește formatul cuvântului-titlu, felul în care autorul dicționarului a delimitat fiecare intrare în parte, care este ordinea informațiilor în cadrul intrării, cum sunt prezentate definițiile etc. Dificultățile întâmpinate în această fază sunt determinate şi de cantitatea mare de material care trebuie să fie tratat din punct de vedere informatic – parsat şi aliniat (cca 100 de dicționare / aproximativ 150.000 de pagini de dicționar), dar mai ales de situația dicționarelor scrise în alfabet chirilic sau în alfabet de tranziție.

În cazul dicționarelor în alfabet chirilic şi a celor în alfabet de tranziție s-a optat pentru introducerea manuală de către lexicografi, a cuvintelor-titlu, în interfața de validare. Aceasta s-a datorat erorilor de recunoaştere automată a caracterelor, întrucât nu există încă un program de OCR-izare pentru texte vechi româneşti. Dat fiind faptul că grafia chirilică românească este destul de diferită pentru texte din epoci şi surse variate, este practic imposibil, pentru moment, să se recunoască automat un text vechi românesc. De aceea s-a renunțat la recunoaşterea automată a textului din dicționarele de acest tip, optându-se pentru o validare manuală făcută de lexicografi, care, practic, au ataşat o „etichetă" cu transcrierea în alfabet latin a cuvintelor-titlu. În plus, dicționarele vechi vor fi aliniate la nivel de imagine şi nu de text, fiind afişată în format imagine doar porțiunea aferentă cuvântului căutat.

Ultima etapă din acest proiect vizează alinierea la nivel de intrare şi, parțial, la nivel de sens a informației din dicționarele vizate.

La sfârşitul proiectului, ne dorim să oferim utilizatorului posibilitatea de a vizualiza, pentru un anumit cuvânt, toate intrările corespunzătoare din cele 100 de dicționare, într-o structură de tipul celei prezentate mai jos.



De asemenea, pentru alinierea parțială la nivel de sens, corelațiile vor fi făcute în funcție de definițiile din DLR, folosindu-se astfel rezultatele din proiectul eDTLR.

Prin acest proiect se are în vedere, aşadar, obţinerea unor rezultate care vor permite ulterior dezvoltarea de aplicaţii de anvergură privind dezambiguizarea semantică a cuvintelor, selecţii de tipuri de intrări în vederea elaborării de noi dicţionare specializate (tematice, etimologice etc.), corelarea cu alte resurse lingvistice ori multimedia, ceea ce ar aduce lexicografia românească la un nivel comparabil cu lexicografia europeană sau chiar mondială. CLRE reprezintă, astfel, şi un punct de plecare pentru cercetări viitoare, cu precădere în domeniile interdisciplinare.

În lexicografia românească actuală se pot observa câteva tendinţe care, în general, implică activităţi interdisciplinare şi care pot fi grupate astfel:

– se continuă realizarea (cu termen neprecizat) a unor ediţii „clasice" (v. continuarea DLR, reeditarea DEX etc.);

– se lucrează (deocamdată nesistematic) la realizarea unor corpusuri de texte pentru limba română;

– se realizează corpusuri lexicografice (vezi proiectul CLRE);

– a început utilizarea sistemelor de scriere de dicţionare.

Concluzii

Rezultate ale unor demersuri interdisciplinare, versiunea informatizată a *Dicţionarului Academiei* (eDTLR) şi corpusul lexicografic românesc esenţial (CLRE) vor facilita accesul specialiştilor la nişte instrumente de lucru indispensabile, mult timp aşteptate, foarte utile pentru studiul limbii române.

Bibliografie

OED = *Oxford English Dictionary* – http://www.oed.com/.

DWB = *Deutsches Wörterbuch "der Grimm"* – http://germazope.uni-trier.de/Projects/DWB.

TLFI = *Le Trésor de la Langue Française Informatisé* – http://atilf.atilf.fr/

TLIO = Tesoro della lingua italiana delle origini – http://tlio.ovi.cnr.it/TLIO/index2.html

Atkins, B.T.S.; Rundell, M. 2008: *The Oxford Guide to Practical Lexicography*. Oxford, Oxford University Press.

Clim, Dănilă *et alii* 2008: Marius Clim, Elena Dănilă, Gabriela Haja, *Premise ale informatizării cercetării lexicografice academice româneşti* în volumul *Limba română. Dinamica limbii, dinamica interpretării*, Editura Universităţii din Bucureşti, p. 585 – 591.

Cristea, Răschip *et alii* 2007: Dan Cristea, Marius Răschip, Corina Forăscu, Gabriela Haja, Cristina Florescu, Bogdan Aldea, Elena Dănilă, *The Digital Form of the Thesaurus Dictionary of the Romanian Language*, în vol. *Advances in Spoken Language Technology* (editors Corneliu Burileanu, Horia-Nicolai Teodorescu), Bucureşti, Editura Academiei Române, p. 195-206.

Dănilă 2010a: Elena Dănilă*, eDTLR – base de données et instrument pour la recherche lexicographique roumaine*, in „Philologica Jassyensia", An VI, Nr. 1 (11), 2010, p. 37–46.

Dănilă 2010b: Elena Dănilă, *Despre necesitatea realizării unui corpus lexicografic românesc esenţial*, în „Philologica Jassyensia", An VI, Nr. 2 (12), 2010, p. 41-50.

Haja, Dănilă *et alii* 2005: Gabriela Haja, Elena Dănilă, Corina Forăscu, Bogdan-Mihai Aldea, *Dicționarul limbii române (DLR) în format electronic. Studii privind achiziționarea*, Iaşi, Editura Alfa, publicat şi electronic pe www.consilr.info.uaic.ro.

Tamba Dănilă, Clim *et alii* 2012: Elena Tamba Dănilă, Marius-Radu Clim, Mădălin Pătraşcu, Ana Catană-Spenchiu, *The Evolution of the Romanian Digitalized Lexicography. The Essential Romanian Lexicographic Corpus*, in *Proceedings of the 15 th EURALEX International Congress*, 7-11 august 2012, Oslo, eds. Ruth Vatvedt Fjeld, Julie Matilde Torjusen, Press Reprosentrales, UiO, p. 225, *in extenso* (pe suport electronic) p. 1014-1017.