

Situația lexicografiei românești în context european

Elena TAMBA, Marius-Radu CLIM,
Ana CATANĂ-SPENCHIU, Mădălin PĂTRAȘCU*

Key-words: *Romanian lexicography, digitalized_lexicography, lexicographic corpus, linguistic_resources, e-lexicography*

În lexicografia din întreaga lume se observă un amplu proces de schimbare, de modernizare a mijloacelor de redactare, de consultare etc., prin intermediul unor demersuri interdisciplinare. Și lexicografia românească este marcată de această schimbare, căreia nu i se poate opune, ci se încearcă să se găsească o cale de gestionare a ei. În acest sens, principalele elemente de schimbare care se pot remarca sunt legate mai ales de crearea unor corpusuri de limbă română, adică colecții de texte de diferite tipuri – care să ajute la redactarea de dicționare – sau corpusuri lexicografice, a unor dicționare pe suport electronic, cu prezentare on-line sau nu.

Schimbarea este vizibilă în modul de lucru pentru crearea lucrărilor lexicografice. Astfel, și în lexicografia românească, într-un ritm mai mult sau mai puțin alert, se face trecerea de la dicționarele în format clasic, editate pe hârtie la dicționarele redactate cu ajutorul unor programe speciale de scriere de dicționare.

În acest scop, și în România s-au realizat în ultimii ani proiecte menite să creeze resurse similare. Astfel, pentru realizarea unor corpusuri de limba română, s-au făcut unele demersuri pentru crearea unor corpusuri lexicografice, grație unor inițiative private, care oferă accesul nelimitat la resursele dezvoltate (vezi dexonline.ro¹, care propune un corpus de cca 38 de dicționare aliniate la nivel de intrare; vocabulary.ro², corpus de 15 dicționare aliniate la nivel de intrare; site-ul editurii Litera, care oferă posibilitatea de consultare a unora dintre dicționarele pe care le-a editat³ etc.), sau grație inițiativei unor instituții academice (proiecte de informatizare a *Dicționarului-tezaur al limbii române*), însă accesul la resursele create prin aceste ultime proiecte este restricționat/limitat, în scop de cercetare (în funcție de legislația în vigoare privitoare la drepturile de autor), ceea ce conduce spre ideea necesității unor politici lingvistice coerente.

* Institutul de Filologie Română „A. Philippide”, Academia Română, Filiala Iași.

Acest articol a fost realizat în cadrul proiectului *CLRE. Corpus lexicografic românesc esențial. 100 de dicționare de bază din Bibliografia DLR aliniate la nivel de intrare și la nivel de sens*, CNCS-UEFISCDI, cod TE_246/2010, 2010 – 2013 (nr. 55/2010).

¹ <http://dexonline.ro>.

² <http://vocabulary.ro>.

³ http://nodex.litera.ro/cautare_dex/.

Dicționarul-tezaur al limbii române reprezintă cea mai importantă lucrare lexicografică apărută sub egida Academiei Române. A fost începută acum 105 ani și a apărut în două serii: DA (1907–1944) și DLR (1965–2010); în 14 tomuri/37 volume, totalizând cca 18.000 pagini și mai mult de 175.000 intrări, cu tot cu variante; elaborarea variantei electronice a început în 2007, în cadrul unui proiect, eDTLR, care a primit finanțare până în 2010, dar care este încă în lucru. Proiectul eDTLR⁴ reprezintă prima colaborarea națională între institute ale Academiei Române și facultăți de la Universitatea „Alexandru Ioan Cuza” din Iași, pentru realizarea unui demers de informatizare a unei lucrări lexicografice de asemenea dimensiuni⁵.

Acest proiect a beneficiat inițial de o finanțare națională din partea CNMP, iar ulterior, după decembrie 2010, lucrându-se la finalizarea site-ului eDTLR doar cu ajutorul unor voluntari – cercetători informaticieni și lexicografi. Proiectul menționat, după încheierea sa, ar trebui, pe de o parte, să pună la dispoziția tuturor cunoscătorilor sau celor interesați de limba română formatul electronic al *Dicționarului* Academiei, pe suport electronic – și, poate ulterior, în funcție de schimbarea politicii lingvistice din România, și on-line, în acces liber sau condiționat – și, pe de altă parte, să pună la dispoziția deocamdată doar a cercetătorilor o arhivă electronică care să cuprindă toate textele din Bibliografia DLR.

Proiectul de informatizare a *Dicționarului-tezaur* a fost pregătit de o serie de granturi anterioare⁶: *Dicționarul limbii române (DLR) în format electronic. Studii privind achiziționarea*, proiect finanțat de CNCSIS, 2003–2005, Iași; *Resurse lingvistice în format electronic: Monumenta linguae Dacoromanorum. Biblia 1688. Regum I, Regum II – Ediție critică și corpus adnotat*, proiect finanțat de CNCSIS, 2006–2007, Iași; *DLRI. Bază lexicală informatizată. Derivate*, proiect finanțat de CNCSIS, 2007–2008, Iași; *CNR – Corpus de referință al limbii române pentru constituirea de dicționare academice*, proiect finanțat de CNCSIS, 2007–2008, București.

Așa cum arătam și în alte studii⁷, varianta informatizată a *Dicționarului* academic ar pune lexicografia fundamentală pentru limba română într-o situație de egalitate cu cea a limbilor care au deja dezvoltate astfel de resurse: *Le Trésor de la Langue Française Informatisé* (tlfri – <http://atilf.atilf.fr/>); *Diccionario de la lengua espanola de la Real Academia Espanola* (drae – <http://buscon.rae.es/draeI/>); *Tesoro della lingua italiana delle origini* (tlio – <http://tlio.oiv.cnr.it/TLIO/index2>).

⁴ Pentru o prezentare detaliată a proiectului eDTLR, vezi Dănilă 2010a.

⁵ Echipa proiectului a fost coordonată de Facultatea de Informatică, Universitatea „Alexandru Ioan Cuza” din Iași, prin prof. dr. Dan Cristea, și a inclus ca parteneri: Institutul de Lingvistică „Iorgu Iordan – Al. Rosetti”, Academia Română, București; responsabil de proiect acad. Marius Sala (prin dr. Monica Busuioc); Institutul de Filologie Română „A. Philippide”, Academia Română, Iași; responsabil de proiect dr. Gabriela Haja; Institutul de Lingvistică și Istorie Literară „Sextil Pușcariu”, Academia Română, Cluj-Napoca; responsabil de proiect dr. Rodica Marian; Institutul de Cercetări pentru Inteligență Artificială, Academia Română, București; responsabil de proiect acad. Dan Tufiș; Institutul de Informatică Teoretică, Academia Română, Iași; responsabil de proiect acad. Horia Neculai Teodorescu; Facultatea de Litere, Universitatea „Alexandru Ioan Cuza” din Iași; responsabil de proiect dr. Eugen Munteanu.

⁶ Pentru o prezentare detaliată a proiectelor care au pregătit și susținut informatizarea *Dicționarului-tezaur* al limbii române, vezi Tamba Dănilă et al. 2012: 1015.

⁷ Vezi Tamba Dănilă et al. 2012: 1015.

html); *Deutsches Wörterbuch der Grimm* (dwb – <http://germazope.uni-trier.de/Projects/DWB>); *Oxford English Dictionary* (oed – <http://www.oed.com/>) ș.a. Proiectul eDLTR a condus și la ideea necesității realizării unui corpus lexicografic românesc⁸ – *CLRE. Corpus lexicografic românesc esențial. 100 de dicționare din bibliografia DLR aliniat la nivel de intrare și la nivel de sens*, care este finanțat de CNCS-UEFISCDI pentru perioada 2010–2013 și care se desfășoară tot la Institutul de Filologie Română „A. Philippide” din Iași.

Cu ajutorul unei echipe formate din doar 4 cercetători (3 lexicografi – Elena Tamba, Marius-Radu Clim și Ana-Veronica Catană-Spenchiu – și un informatician – Marius Răschip) și cu aportul unui alt informatician voluntar – Mădălin Pătrașcu –, proiectul își propune să realizeze o bază de date care să cuprindă dicționarele esențiale din Bibliografia DLR, aliniat la nivel de intrare și parțial la nivel de sens; să construiască un mediu de programe care să permită consultarea interactivă a acestui corpus și care să se constituie într-un cadru modern de lucru și cercetare lexicografică, ușor adaptabil la o diversitate de obiective; să realizeze o listă de cuvinte cvasi-exhaustivă, pentru limba română, pornind de la corpusul aliniat. Dicționarele alese au fost incluse în trei categorii: 1. *dicționare generale* (de tipul: DA = *Dicționarul limbii române*, tom I-II, Tipografia ziarului „Universul”, București, Imprimeria Națională, 1907-1944; DLR = *Dicționarul limbii române*, serie nouă, tom VI-XIV, București, Editura Academiei, 1965–2010); 2. *dicționare auxiliare*, care sunt strâns legate de redactarea *Dicționarului-tezaur* (de tipul: Alexandru Ciorănescu, *Dicționarul etimologic al limbii române*, ediție îngrijită și traducere din limba spaniolă de Tudora Sandru-Mehedinți și Magdalena Popescu Marin, București, Editura Saeculum I. O., 2002; *Dicționarul ortografic, ortoepic și morfologic al limbii române*, ediția a II-a revăzută și adăugită, București, Univers Enciclopedic, 2005 ș.a.); 3. *dicționare speciale*, enciclopedice ori dicționare speciale, alese după criteriul importanței lor pentru perspectiva diacronică asupra limbii (de tipul: *Dicționar 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), București, Editura Enciclopedică; *Lexiconul tehnic român*, I ș.u., elaborare nouă, București, Editura Tehnică, 1957 ș. u.)⁹.

Etapele de lucru în realizarea acestui corpus lexicografic presupun scanarea și OCR-izarea¹⁰ dicționarelor vizate, introducerea lor într-o bază de date, parsarea¹¹ textului la nivel de intrare și, parțial, la nivel de sens, validarea/corectarea parsării cu ajutorul unei interfețe on-line, precum și validarea alinierii între textul *Dicționarului-tezaur* al limbii române (în format electronic, rezultat al proiectului eDLTR) și dicționarele de referință din Bibliografia DLR.

În faza de validare a rezultatelor, lexicografii verifică corectitudinea decupării intrărilor și a formei cuvântului-titlu, în interfața de lucru creată special de specialiștii informaticieni:

⁸ Pentru o prezentare detaliată a acestor aspecte, vezi Dănilă 2010b.

⁹ Pentru o prezentare detaliată, vezi Dănilă 2010b: 46–47.

¹⁰ Convertirea din format imagine în format text.

¹¹ Identificarea automată a intrărilor din dicționarele scanate și OCR-izate anterior.

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.

În lexicografia internațională există deja astfel de corpusuri, care facilitează mult cercetarea atât în domeniul studiului limbii/limbilor, cât și în cel al informaticii aplicate, al prelucrării limbajului natural.

Pentru limba franceză poate fi consultată, în acces liber, o colecție de dicționare informatizate franceze, din secolul al XVI-lea până în secolul al XX-lea – *Le rayon des dictionnaires* (<http://www.atilf.fr/>), care permite căutarea unui cuvânt în patru dicționare vechi, în șase ediții ale dicționarului Academiei Franceze și în TLFi (*Trésor de la Langue Française informatisé* – <http://atilf.atilf.fr/>), cu afișarea cuvântului-titlu căutat, cu întreaga intrare în toate dicționarele care îl conțin, cronologic, și cu posibilitatea vizualizării întregii pagini de pe care a fost extras articolul respectiv, în cazul în care se optează pentru o astfel de cerință. În imaginea de mai jos se poate observa rezultatul pentru căutarea lemei *dictionnaire*.

Dictionnaires d'autrefois
Dictionnaires des 17^{ème}, 18^{ème}, 19^{ème} et 20^{ème} siècles

RECHERCHE	Accueil	Documentation	ARTFL	ATILF	Courriel
-----------	---------	---------------	-------	-------	----------

Il y a 8 entrées dans Dictionnaire de L'Académie française, 1st Edition (1694) ([Go](#)), Dictionnaire de L'Académie française, 4th Edition (1762) ([Go](#)), Dictionnaire de L'Académie française, 5th Edition (1798) ([Go](#)), Dictionnaire de L'Académie française, 6th Edition (1832-5) ([Go](#)), Dictionnaire de L'Académie française, 8th Edition (1932-5) ([Go](#))

Dictionnaire de L'Académie française, 1st Edition (1694)

Dictionnaire (Page [332](#))

Dictionnaire. s. m. Vocabulaire, recueil par ordre de tous les mots d'une langue. *Dictionnaire François. dictionnaire Latin. dictionnaire François- Latin. dictionnaire par ordre alphabétique. dictionnaire par ordre des racines. par racines. bon dictionnaire. ample dictionnaire.*

On dit aussi, *Dictionnaire Poétique-historique, historique-Geographique &c.*

[Top](#)

Dictionnaire de L'Académie française, 4th Edition (1762)

DICTIONNAIRE (Page [534](#))

DICTIONNAIRE se dit aussi de divers autres recueils faits par ordre alphabétique sur des matières de Poésie, de Géographie, d'Histoire, &c. *Dictionnaire Poétique. Dictionnaire Historique. Dictionnaire Géographique. Un Dictionnaire de Marine. Dictionnaire de rimes.*

DICTIONNAIRE. s.m. (Page [534](#))

DICTIONNAIRE s.m. Vocabulaire. Recueil de tous les mots d'une Langue, mis par ordre. *Dictionnaire François. Dictionnaire Latin. Dictionnaire Italien. Dictionnaire Espagnol. Dictionnaire par ordre alphabétique. Dictionnaire par l'ordre des racines, par racines.*

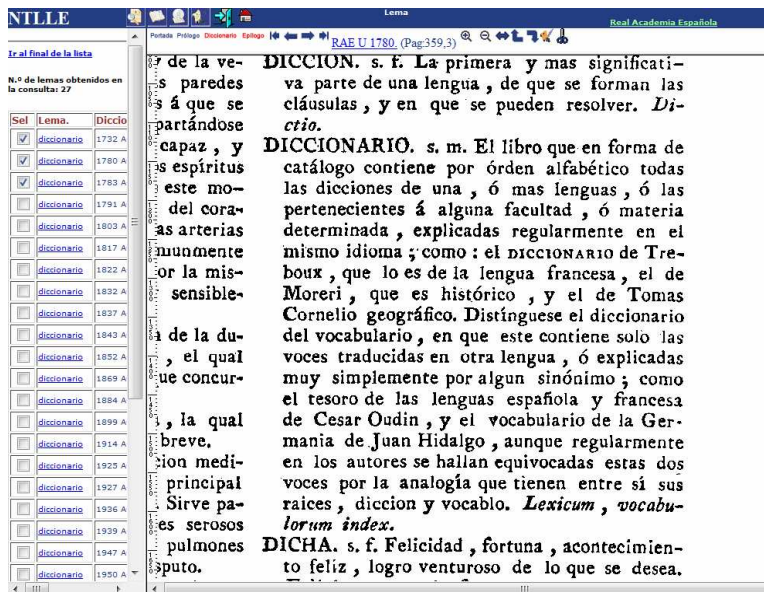
DICTIONNAIRE se dit aussi de divers autres recueils faits par ordre alphabétique sur des matières de Poésie, de Géographie, d'Histoire, &c. *Dictionnaire Poétique. Dictionnaire Historique. Dictionnaire Géographique. Un Dictionnaire de Marine. Dictionnaire de rimes.*

[Top](#)

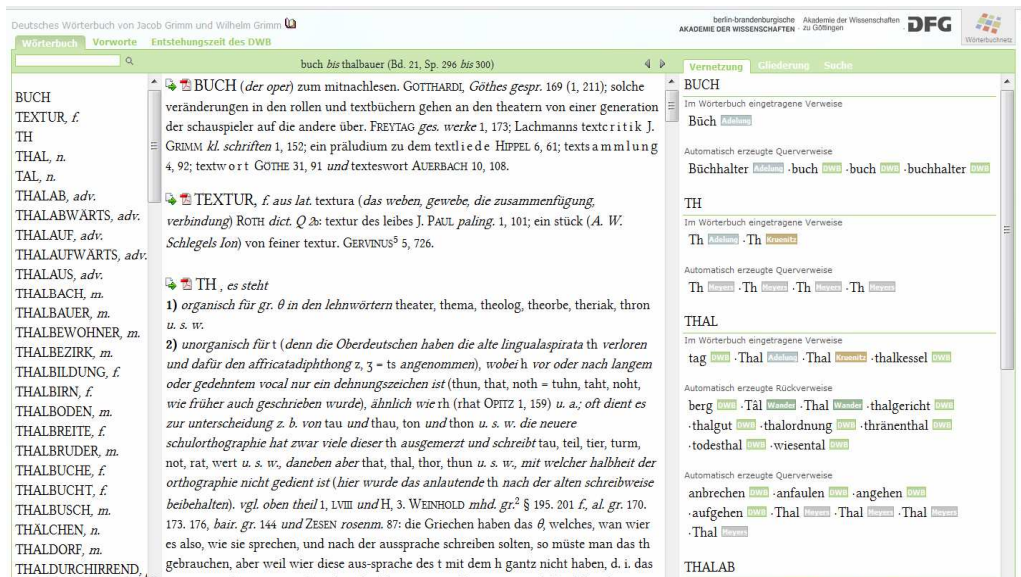
Dictionnaire de L'Académie française, 5th Edition (1798)

Pentru limba spaniolă se poate consulta o bază de date – *Nuevo tesoro lexicográfico de la lengua española* (<http://buscon.rae.es/ntlle/SrvltGUILoginNtllle>) – care cuprinde versiunile facsimilate pentru cca 70 de dicționare de limbă spaniolă, din secolul al XV-lea până în secolul al XX-lea, proiect realizat de Real Academia Espagnola.

Practic, un cuvânt poate fi căutat și va fi afișată o listă completă cu toate aparițiile sale în dicționarele aflate în baza de date, iar apoi, pentru a se vizualiza articolul căutat într-un anumit dicționar, se accesează doar acel dicționar și este afișată întreaga pagină în care este prezent articolul vizat – cuvântul *diccionario*, în cazul de față –, după cum se poate vedea și în captura de ecran următoare:



Cei interesați pot consulta o rețea de dicționare de limba germană, creată la Universitatea Trier din Germania – *Das Wörterbuchnetz* (<http://germazope.uni-trier.de/Projects/WBB/>). Interfața de consultare permite căutarea unui anumit cuvânt, care apare cu toate ocurențele sale în cele 24 de dicționare aliniat, ulterior putându-se opta pentru vizualizarea articolului dintr-un anumit dicționar, cu posibilitatea de a selecta și accesarea pdf care cuprinde articolul respectiv. Interfața de consultare permite, de asemenea, și alte opțiuni de căutare, după cum se poate observa și în printscreen-ul următor:



Practic, exemplele de corpusuri lexicografice prezentate *supra* reprezintă un argument în plus pentru realizarea *Corpusului lexicografic românesc esențial*, precum și o dovadă a posibilităților multiple de utilizare a unor astfel de resurse lingvistice electronice.

*
* *

Pentru a vedea unde se situează lexicografia românească în context european, trebuie spus că marile lucrări lexicografice se realizează:

– fie sub egida Academiiilor din anumite țări (a se vedea DWB – *Deutsches Wörterbuch “der Grimm”*, care se editează sub patronajul Academiei Germane sau DLR *Dicționarul limbii române*, redactat în institute care aparțin Academiei române);

– fie în institute special create (vezi, pentru limba franceză, L’Unité mixte de recherche ATILF – Analyse et Traitement Informatique de la Langue Française – <http://www.atilf.fr/>, organizat în 2001, sau, pentru limba slovenă – Trojina, Institute for Applied Slovene Studies – www.trojina.si/, creat în 2004).

Tendențele în lexicografia europeană actuală pot fi grupate astfel:

– continuarea editării (cu termen de finalizare de 10-15 ani) a unor ediții „clasice” (vezi DWB – *Deutsches Wörterbuch von Jacob Grimm und Wilhelm Grimm*);

– realizarea (în paralel) a unor dicționare (on-line), plecând de la corpusuri de texte din ce în ce mai mari și cu instrumente variate (cu ajutorul unor programe de extragere de exemple¹²);

– realizarea de corpusuri lexicografice (vezi *supra*, corpusurile de dicționare prezentate);

– utilizarea sistemelor de scriere de dicționar (DWS)¹³, ceea ce permite mult mai ușor alinierea ulterioară a diverselor lucrări lexicografice. În acest ultim caz este necesar și un program de colaborare între diverse instituții cu profil lexicografic, editori etc.

Prin raportare la aceste tendințe în lexicografia europeană, în lexicografia românească se observă:

– continuarea editării (cu termen neprecizat) a unor ediții „clasice” (vezi DLR, reeditare DEX etc.);

– realizarea unor corpusuri de texte pentru limba română (proces în curs de realizare, deocamdată, nesistematic);

– realizarea de corpusuri lexicografice (în lucru);

– utilizarea sistemelor de scriere de dicționar (în lucru).

Concluzii

Din informațiile prezentate în acest articol se poate observa că lexicografia românească se află în plin proces de adaptare la schimbările pe care modernizarea

¹² Pentru o prezentare a unor aspecte privitoare la Sketch Engine – sistem de căutare în corpus, utilizabil ca instrument de lucru pentru lexicografi, vezi Dănilă et al. 2011: 194–195.

¹³ Pentru o analiză a sistemelor de scriere de dicționar, vezi Dănilă et al. 2011: 195–196.

instrumentelor de lucru în lexicografia europeană/mondială le presupune. Astfel, și în lexicografia românească se observă tendința de a edita în continuare lucrări lexicografice în format clasic (pe hârtie), care este completată de realizarea unor corpusuri de texte de limbă română, a unor corpusuri lexicografice sau utilizarea (încă în fază incipientă) a unor sisteme de scriere de dicționare, pentru a eficientiza munca lexicografilor.

Astfel, prin toate aceste demersuri, lexicografia românească se aliniază la tendințele existente în lexicografia europeană.

Bibliografie

- OED = *Oxford English Dictionary* (<http://www.oed.com/>).
- DRAE = *Diccionario de la lengua espanola de la Real Academia Espanola* (<http://buscon.rae.es/draeI/>).
- 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.oivi.cnr.it/TLIO/index2.html>).
- *
* *
- Atkins, Rundell 2008: B.T.S. Atkins, M. Rundell, *The Oxford Guide to Practical Lexicography*, Oxford, Oxford University Press.
- Clim et al. 2008: Marius Clim, Elena Dănilă, Gabriela Haja, *Premise ale informatizării cercetării lexicografice academice românești*, în volumul *Limba română. Dinamica limbii, dinamica interpretării*, Editura Universității din București, p. 585–591.
- Cristea et al. 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), București, Editura Academiei Române, p. 195–206.
- Dănilă 2010a: Elena Dănilă, *eDTLR – base de données et instrument pour la recherche lexicographique roumaine*, în „*Philologica Jassyensia*”, an VI, nr. 1 (11), 2010, p. 37–46.
- Dănilă 2010b: Elena Dănilă, *Despre necesitatea realizării unui corpus lexicografic românesc esențial*, în „*Philologica Jassyensia*”, an VI, nr. 2 (12), 2010, p. 41–50.
- Dănilă et al. 2011: Elena Dănilă, Marius-Radu Clim, Ana Catană-Spenchiu, *Towards a Romanian Lexicographic Corpus*, în „*Philologica Jassyensia*”, an VII, nr. 2 (14), 2011, p. 191–198.
- Haja et al. 2005: Gabriela Haja, Elena Dănilă, Corina Forăscu, Bogdan-Mihai Aldea, *Dicționarul limbii române (DLR) în format electronic. Studii privind achiziționarea*, Iași, Editura Alfa (și în format electronic: www.consilr.info.uaic.ro).
- Tamba Dănilă et al. 2012: Elena Tamba Dănilă, Marius-Radu Clim, Mădălin Pătrașcu, 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 Representrales, UiO, p. 225, in extenso (pe suport electronic), p. 1014–1017 (on-line http://www.euralex.org/proceedings-toc/euralex_2012/).

The Status of the Romanian Lexicography in the European Context

The evolution of the new Romanian lexicography based on the web resources has changed the view concerning the dictionaries and how to create them. The Romanian lexicography is also on a high process of modernisation and computerisation. The project we presented here, called CLRE (*ERLC. Essential Romanian lexicographic corpus. 100 dictionaries from DLR bibliography aligned at entry and, partially, at meaning level*) is one of the necessary steps for Romanian lexicography to adapt itself to the digital world.

The present project has as purpose creating a Romanian lexicographic corpus, which will contain 100 dictionaries (from the 16th century to present day) aligned at entry and, partially, at meaning level. This lexicographic corpus will have the potential of being an useful database for linguistic research, in the country or abroad, becoming an indispensable instrument for future researches, for implementing and developing the Romanian lexicographical research according to international standards.