

Le Processus de Numérisation de la Lexicographie en Roumanie: Présent et Perspectives

Elena Isabelle Tamba, *Université Alexandru Ioan Cuza de Iasi, Roumanie /
Institut de Philologie Roumaine Alexandru Philippide, Académie Roumaine,
Filiale de Iasi, Roumanie (isabelle.tamba@gmail.com)*

Résumé: En Roumanie, la recherche lexicographique est en développement continu, dans un processus naturel de numérisation. Cette étape est absolument obligatoire pour la création d'instruments et de ressources électroniques, nécessaires au soutien de la langue et de la culture roumaines.

Les spécialistes roumains en linguistique et informatique appliquée, ainsi que dans les domaines de la linguistique computationnelle, ont initié des projets de recherche par lesquels ils valorisent les ressources non numérisées en les acquérant sous forme électronique ou par lesquels ils créent des dictionnaires et de nouvelles ressources et instruments directement sous forme électronique.

Le processus de numérisation place la lexicographie académique roumaine à un niveau comparable à la lexicographie internationale et permet la connexion avec des projets lexicographiques de l'étranger et l'inclusion de la Roumanie dans la sphère d'intérêt des grands réseaux lexicographiques internationaux.

Mots-clés: LEXICOGRAPHIE ROUMAINE, E-LEXICOGRAPHIE, CORPUS LEXICOGRAPHIQUE ÉLECTRONIQUE, NUMÉRISATION, RESSOURCES LEXICOGRAPHIQUES NUMÉRISÉE, ÉVOLUTION, PERSPECTIVES

Abstract: Digitalization Process of the Lexicography in Romania: Present and Perspectives. In Romania, the lexicographic research is in a continuous development, in a natural process of digitalization. This step is absolutely mandatory for creating electronic instruments and resources, which are necessary for supporting the Romanian language and culture.

The Romanian academic specialists in linguistics and applied informatics, as well as in computational linguistics fields, have initiated research projects by which they valorise the non-digitized resources by acquiring them in electronic formats or by which they create dictionaries and new resources and instruments directly in electronic format.

The digitalization process put the Romanian academic lexicography at a level comparable to the international lexicography and allows the connection with lexicographic projects from abroad and the inclusion of Romania in the sphere of interest of the great international lexicographic networks.

Keywords: ROMANIAN LEXICOGRAPHY, E-LEXICOGRAPHY, ELECTRONIC LEXICOGRAPHIC CORPUS, DIGITALIZATION, DIGITAL LEXICOGRAPHICAL RESOURCES, EVOLUTION, PERSPECTIVES

1. Introduction

La langue représente une expression de la culture, un facteur déterminant de l'identité et, en même temps, elle offre la possibilité de pouvoir communiquer. La lexicographie soutient la langue, en mettant dans les articles de dictionnaire l'histoire de chaque mot et, de cette manière, elle raconte l'histoire de la Langue. C'est pourquoi rendre plus facile l'accès à l'information est de plus en plus important. À travers l'informatisation des ressources lexicographiques, la langue roumaine s'inscrit dans cette direction de rendre l'information accessible aux personnes intéressées par la recherche de la langue roumaine / par l'étude du roumain.

La lexicographie mondiale connaît un vaste processus de changement, de modernisation des moyens de rédaction, de consultation, etc., à travers des approches qui impliquent l'interconnexion de différents domaines de recherche. Hanks (2013) fait référence aux ordinateurs et aux corpus comme le deuxième facteur le plus influent, après l'invention de l'imprimerie, en lexicographie. C'est pourquoi la nouvelle technologie du langage et du corpus a fourni aux lexicographes et aux utilisateurs des outils innovants pour compiler et consulter des dictionnaires (cf. Abdelzaher 2022).

En Roumanie, surtout au cours de la dernière décennie, la recherche lexicographique (et linguistique, d'ailleurs) connaît un développement continu, dans un processus naturel de numérisation. Cette étape est absolument obligatoire pour créer des instruments électroniques et des ressources nécessaires au soutien de la langue et de la culture roumaines.

Ce processus de numérisation de la recherche lexicographique / linguistique suppose plusieurs étapes:

- changement du format classique / en papier au format numérique, ce qui conduit à la création de corpus lexicographiques (ou linguistiques), par plusieurs approches de recherche
- numérisation des dictionnaires en papier
- création / développement des corpus lexicographiques
- création / développement des corpus linguistiques, annotés du point de vue morphologique, syntaxique et sémantique
- création des ressources lexicographiques directement en format numérique, par
- création des dictionnaires directement en format numérique
- utilisation des programmes de rédaction des dictionnaires
- utilisation des programmes pour extraire des citations, etc.

Cet article envisage la description des principaux projets et résultats dans le domaine de la lexicographie roumaine numérique, tant pour les dictionnaires contemporains que prémodernes.

2. La lexicographie numérique en Roumanie

Les professionnels roumains en lexicographie et en linguistique informatique ont initié des projets de recherche par lesquels ils veulent, d'une part, valoriser les ressources non-numérisées, en les acquérant dans des formats électroniques et, d'autre part, créer de nouvelles ressources et instruments lexicographiques / linguistiques directement en format numérique.

La plupart des efforts de numérisation des recherches lexicographiques ont été faits sous les auspices des Instituts de l'Académie Roumaine¹; mais, plus récemment, les centres de recherche de certaines universités du pays s'y sont impliqués².

À présent, en Roumanie il y a de divers projets lexicographiques (ou linguistiques) numérisés, réalisés à partir des initiatives académiques — on y inclut non seulement des instituts de langue et de littérature roumaine et les instituts d'informatique de l'Académie Roumaine, mais aussi des centres de recherche de certaines universités de Roumanie et des bibliothèques. Mais il y a aussi des initiatives privées (des projets initiés par de volontiers³; des projets de certaines maisons d'édition, etc.).

2.1 La numérisation du *Dicționarul limbii române* de l'Académie roumaine

Dicționarul limbii române [Le Dictionnaire de la langue roumaine] est l'ouvrage lexicographique le plus important pour le roumain, étant rédigé et édité par l'Académie Roumaine⁴. C'est pourquoi la création d'un format numérique, accessible aux scientifiques et à tous ceux qui s'intéressent à l'apprentissage ou à l'étude de la langue roumaine, en Roumanie ou à l'étranger, est devenue, dans la société numérique et multiculturelle, une étape absolument nécessaire à franchir.

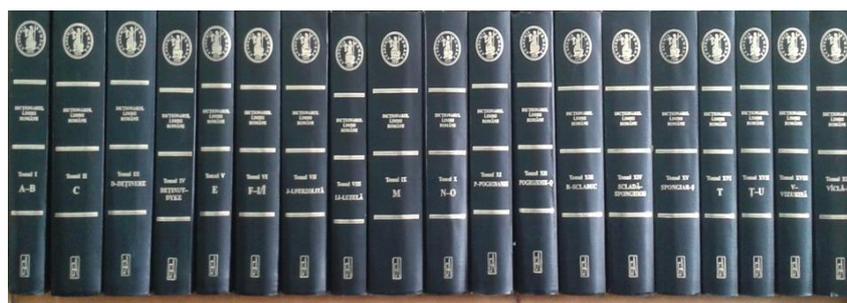


Figure 1: *Dicționarul limbii române*. Édition anastatique, 2010

Les grandes cultures disposent depuis de nombreuses années de dictionnaires et de corpus de textes au format numérique. Pour une meilleure compréhension des dimensions du *Dictionnaire-trésor de la langue roumaine*, on présente quelques-unes de ses données statistiques, comparées à d'autres grands dictionnaires européens:

- La compilation du *Dicționarul limbii române* a commencé il y a 115 ans. La première édition fut publiée en deux séries — DA (1913–1944) et DLR (1965–2010), en 14 tomes, 37 volumes, 20 000 pages de type lexique (ayant entre 7 000 et 11 000 caractères/page), plus de 175 000 entrées (avec variantes) et plus de 1 300 000 citations; l'élaboration du format numérique: première tentative, eDTLR, 2007–2010 (à présent, eDTLR est un projet repris, en plein déroulement); une autre tentative de numérisation, réussie cette fois-ci — dans le projet CLRE⁵. La deuxième édition, révisée et ajoutée, (en format numérique et en papier) est en cours de rédaction, directement en format numérique (DLR²).

Afin d'établir une comparaison entre le dictionnaire-trésor roumain et d'autres dictionnaires importants, on présente quelques informations statistiques:

- *Dictionnaire de l'Académie Française* (DAF), 1694 — la première édition imprimée; 9 éditions, consultables en ligne, 55.000 entrées;
- *Diccionario de la lengua española de la Real Academia Española* (DRAE), 1780 — la première édition imprimée; la 23e édition — 2014; 93 111 lemmes; le premier format numérique: 1992;
- *Deutsches Wörterbuch der Grimm* (DWB), 1838–1961, 32 volumes, 350.000 entrées et variantes; le premier format numérique: 1997–2004;
- *Oxford English Dictionary* (OED), la première édition — 1928, 20 volumes (la deuxième édition — 1989), 301.100 entrées, 2.412.400 exemples; le premier format numérique: 1988;
- *Trésor de la Langue Française* (TLF), XIX^e–XX^e siècle, 1971–1994 — la première édition imprimée; 16 volumes, 100.000 entrées, 270.000 définitions, 430.000 exemples; le premier format numérique: 1990–2004;
- *Tesoro della lingua italiana delle origini* (TLIO) — en ligne, 44.000 entrées (37864 publiés en ligne de 57.000).

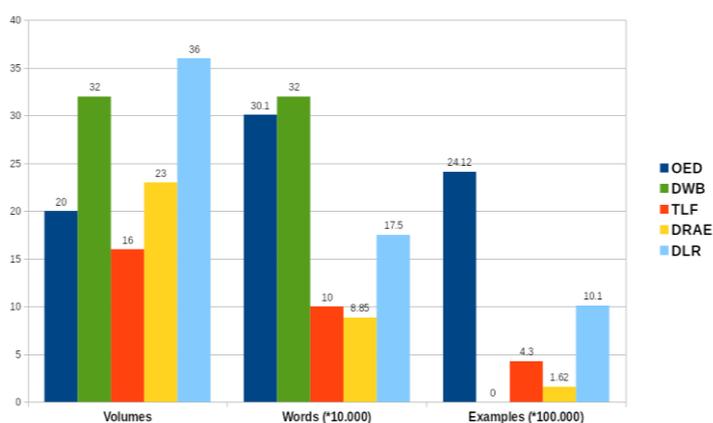


Figure 2: DLR vs. d'autres dictionnaires européens⁶

À partir des données ci-dessus, on peut observer le fait que *Dicționarul limbii române* est similaire, en ce qui concerne la conception et la réalisation, aux autres dictionnaires européens et sa numérisation représente, ainsi, une étape normale dans l'évolution de la lexicographie roumaine.

Pour mieux comprendre le parcours qui a conduit à l'étape actuelle de la lexicographie dans le cadre de l'Académie Roumaine, on va mentionner, dans ce qui suit, les projets qui ont contribué à façonner une stratégie concernant la numérisation de la recherche lexicographique académique roumaine⁷:

- 2003–2005: le projet pilote de numérisation du DLR, *Dicționarul limbii române (DLR) în format electronic. Studii privind achiziționarea* [Le Dictionnaire de la langue roumaine (DLR) en format électronique. Études sur la création du format numérique]⁸;
- 2006–2007: première édition électronique d'un ancien texte roumain de la Bibliographie du DLR, *Monumenta linguae Dacoromanorum. Bible 1688. Pars VII. Regum I, Regum II* (<https://biblia1688.solirrom.ro/7/>);
- 2007–2008: mise à jour numérisée des premiers éléments de la série DA selon les normes de rédaction DLR dans un projet nommé *DLRI. Bază lexicală informatizată. Derivate în -ime și -iște* [DLRI. Base lexicale informatisée. Dérivés en -ime et -iște]⁹;
- 2007–2010: première tentative de numérisation complète du DLR dans le projet homonyme, l'objectif principal étant l'acquisition en format numérique de la variante intégrale du *Dictionnaire-trésor de la langue roumaine*, grâce à la rétro-numérisation;
- 2010: initiation de la documentation et rédaction directement sous format électronique de la deuxième édition du *Dicționarul limbii române. DLR²*. [Le Dictionnaire de la langue roumaine. DLR²] — projet fondamental et prioritaire de l'Académie Roumaine; la deuxième édition de ce dictionnaire a été initié en 2010 et elle se rédige entièrement en format numérique (<https://dlri.ro/>); en parallèle, une version en papier est également publiée;
- 2010–2013: premier projet dont l'objectif principal est la création d'un corpus lexicographique diachronique: *CLRE. Corpus lexicografic românesc esențial* [CLRE. Corpus lexicographique roumain essentiel];
- 2014: inclusion du projet CLRE dans le plan de recherche de l'Académie Roumaine. *CLRE. Corpus lexicografic românesc electronic* [Corpus lexicographique roumain électronique] — projet fondamental de l'Académie roumaine;
- 2019–présent: le passage d'ouvrages lexicographiques numérisés, publiés isolément, à des ouvrages publiés unitairement, grâce à l'utilisation des normes et d'outils de travail communs, sur la plate-forme SOLIROM¹⁰, qui appartient aux instituts à profil philologique de l'Académie Roumaine et

qui permet la collaboration entre spécialistes, en utilisant les outils et les ressources créés, ainsi que la publication intégrée des travaux développés sous les auspices de l'Académie Roumaine.

À présent, *Dicționarul limbii române* édité sous les auspices de l'Académie Roumaine connaît trois variantes numériques, réalisées ou en train d'être réalisées dans trois projets différents, qui vont être connectées dans *CLRE. Corpus lexicografic românesc electronic*:

- (a) **DLR²** — la rédaction de la deuxième édition de *DLR. Dicționarul limbii române. Ediția a doua revăzută și adăugită* [DLR. Dictionnaire de la langue roumaine. Deuxième édition révisée et ajoutée] a commencé en 2010 et est ou sera faite entièrement directement en format numérique, à l'aide d'une interface d'écriture XML, dans le vocabulaire TEI, et publiée en ligne, au fur et à mesure que les fascicules sont finalisés — <https://dlri.ro/>. Ce projet est réalisé dans les Départements de Lexicographie des instituts de langue et de littérature roumaine de l'Académie Roumaine. La parution du premier fascicule de la deuxième édition, révisée et complétée, du *Dictionnaire de la langue roumaine*. Volume I. Lettre A. Fascicule 1 (*A–Abzițui*), rédigé en format numérique par l'équipe de lexicographes de l'Institut de Linguistique «Iorgu Iordan — Al. Rosetti», de l'Académie Roumaine, à Bucarest (<https://dlri.ro/>) a été suivi par la parution du fascicule en format classique, sur papier.
- (b) **DA, DLR** — rétro-numérisation, dans le projet CLRE (qui suppose scanner et traiter les informations dans la plate-forme, qui permet pour l'instant des consultations au niveau de l'entrée et au niveau de la catégorie lexicogrammaticale et l'affichage de l'image de la page / des pages du dictionnaire — voir les informations sur CLRE dans le sous-chapitre suivant).
- (c) **eDTLR** — rétro-numérisation dans le projet ayant le même nom. La première tentative de numérisation complète du DLR a eu lieu dans le projet *Dicționarul tezaur al limbii române în format electronic* (eDTLR) [Dictionnaire-trésor de la langue roumaine en format numérique (eDTLR)] (2007–2010), première version (partenariat entre les Instituts de l'Académie Roumaine et la Faculté d'Informatique de l'Université «Alexandru Ioan Cuza» à Iași) — l'objectif principal étant l'acquisition en format numérique de la variante intégrale du *Dictionnaire-trésor de la langue roumaine*, par des recherches qui suppose scanner, OCR-iser, corriger, analyser et introduire les données dans une plate-forme de consultation, qui permet des recherches complexes, dans tout le corps de chaque article lexicographique. À partir de 2021, le projet a été repris, dans l'Académie Roumaine — Branche Iași, en tant que projet prioritaire, afin de corriger et compléter la version numérique eDTLR et de l'inclure dans CLRE.

2.2 CLRE. *Corpus lexicografic românesc electronic* [Corpus lexicographique roumain électronique]

CLRE est un projet fondamental de l'Académie Roumaine (commencé en 2014), réalisé par les chercheurs du Département de lexicologie et lexicographie de L'Institut de Philologie Roumaine «Alexandru Philippide», L'Académie Roumaine — Filiale de Iași (<https://clre.solirom.ro/>).

Le corpus représente une collection d'éditions numériques des dictionnaires les plus représentatifs de la langue roumaine de tous les temps, alignés au niveau d'entrée et au niveau de la catégorie lexico-grammaticale.

En mai 2022, la première édition numérique anastatique au niveau d'entrée et de la catégorie lexico-grammaticale du *Dictionnaire thésaurus de la langue roumaine*, édité sous les auspices de l'Académie roumaine (1913–2010), a été publiée dans son intégralité en CLRE.



Figure 3: La première page de CLRE

Les principaux objectifs du projet CLRE sont:

- la création du plus grand corpus diachronique numérique de dictionnaires de langue roumaine, aligné, pour l'instant, au niveau de l'entrée, un corpus composé d'œuvres lexicographiques numérisées de la bibliographie DLR (transposés du format classique, sur papier, au format numérique) et à partir de dictionnaires numérisés (créés directement en format numérique éditable);
- la mise en libre accès pour le grand public des informations du CLRE. Partant du fait que la politique linguistique en Roumanie commence à prendre un contour plus ferme, les bons moyens ont été trouvés pour

pouvoir offrir un accès gratuit aux informations du corpus à tous ceux qui s'intéressent aux œuvres lexicographiques en langue roumaine;

- promotion des travaux lexicographiques réalisés sous les auspices de l'Académie Roumaine.

CLRE comprend différentes catégories de dictionnaires: généraux, explicatifs, étymologiques, encyclopédiques ou spécialisés, choisis selon le critère de leur importance pour la perspective diachronique de la langue roumaine.

À présent, le corpus CLRE, en développement continu, comprend 62 dictionnaires totalisant 124 volumes à divers stades de numérisation, dans laquelle chaque page ou image se voit attribuer des entrées annotées en lemmes ou variantes et des métadonnées sont attachées à chaque page ou image (liée à l'œuvre dont elle fait partie, numéro de page, type de page — pages précédentes, table des matières, pages arrière).

Chaque œuvre lexicographique dans CLRE est une édition anastatique, au sens de reproduction fidèle par numérisation des images des pages, la différence spécifique étant l'intégration de la collection d'images résultante dans une structure XML qui comprend également des métadonnées sur l'œuvre, des informations sur chaque mot-titre et des ajouts ou des corrections annotés (faits par les éditeurs CLRE)¹¹.

Par la conception, par la similarité en termes d'approche technique et par les dimensions, CLRE, le corpus lexicographique électronique roumain, peut être comparé à d'autres corpus lexicographiques européens:

- *Das Wörterbuchnetz* — une collection de 37 dictionnaires au format numérique, créée à l'Université de Trier en Allemagne (<https://www.woerterbuchnetz.de/>);
- *Diccionarios de la lengua española* — la base de données contenant les dictionnaires édités et publiés par la Real Academia Española (<https://www.rae.es/obras-academicas/diccionarios>);
- le corpus lexicographique pour la langue française réalisé au cadre d'*Ortolang. Outils et ressources pour un traitement optimisé de la Langue*, par le Centre National de Ressources Textuelles et Lexicales — <https://cnrtl.fr/dictionnaires/modernes/>, etc.

L'interface CLRE permet actuellement plusieurs types de recherche (par entrée, selon la première lettre ou selon la catégorie grammaticale), selon différents critères, comme on peut le voir sur les figures ci-dessous:

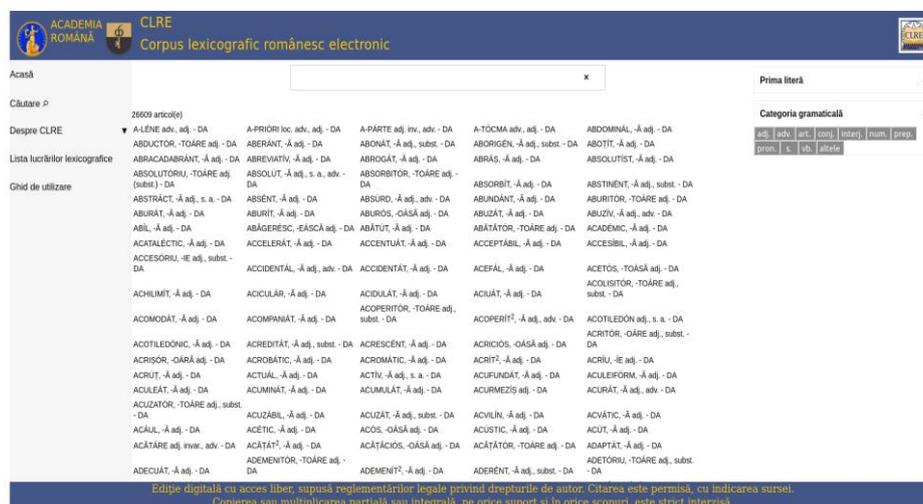


Figure 6: CLRE — le résultat de la recherche selon la catégorie grammaticale

Au fur et à mesure que le corpus se développera et que plusieurs ouvrages lexicographiques en éditions numériques, anastatiques ou créées directement en version digitale, seront accessibles au public, de nouvelles facilités de consultation seront créées pour le public-cible. Une étape essentielle à mettre en œuvre est l'alignement avec eDTLR (lorsque les résultats finaux seront accessibles) et, bien sûr, DLR².

L'utilisation du CLRE pour la rédaction du DLR² et pour d'autres projets lexicographiques et la corrélation avec d'autres ressources linguistiques ou multimédias, amènent la lexicographie roumaine à un niveau comparable à la lexicographie européenne (par exemple, avec *DÉRom. Dictionnaire Étymologique Roman* — www.atilf.fr/DERom/) ou permet l'inclusion de la lexicographie roumaine dans des réseaux internationales de lexicographie, telle, par exemple, *EneL. European Network of e-Lexicography* (www.elexicography.eu), *ELEXIS. European Lexicographic Infrastructure* (<https://elex.is/>).

2.3 D'autres projets lexicographiques numériques dans les instituts de l'Académie Roumaine

DEL.R. Dicționarul etimologic al limbii române [Le dictionnaire étymologique de la langue roumaine] est l'ouvrage lexicographique le plus complet de ce type pour la langue roumaine, comprenant à la fois les mots utilisés dans la langue standard actuelle et ceux de l'ancien roumain, qui ont disparu ou sont devenus archaïques, à côté des régionalismes et de ceux utilisés dans les langages spécialisés.

DELR est en cours de développement à l'Institut de Linguistique de l'Académie roumaine «Iorgu Iordan — Al. Rosetti» de Bucarest.

Actuellement DELR comprend les lettres A, C et la première partie de la lettre D (jusqu'au *djinn*), totalisant 30 514 mots (avec 15 251 variantes), regroupés en 8 809 nids lexicaux. DELR est publié en ligne à l'adresse: <https://delr.lingv.ro/>.

2.4 SOLIROM — instrument pour les ressources et les outils en format numérique pour la langue et la littérature roumaine

SOLIROM représente une plate-forme logicielle de l'Académie Roumaine, qui comprend des ressources linguistiques numériques et des outils linguistiques dédiés à la langue et à la littérature roumaines (DLR², CLRE, eDTLR, TDRG¹², etc.) créés dans les instituts de philologie subordonnés à la Section de philologie et littérature de l'Académie roumaine (<https://solirom.ro/>).

SOLIROM publie des résultats concernant la langue et la littérature roumaines; en outre, la plate-forme fournit aux chercheurs des outils et des ressources pour la recherche numérique, les informations étant visibles en deux sections:

- publique — on présente des ressources linguistiques numériques accessibles au public;
- privée — avec les outils numériques nécessaires à la gestion des ressources linguistiques numériques de la plate-forme, pour les chercheurs qui la développent.

ACADEMIA ROMÂNĂ SOLIROM Instrumente și resurse digitale pentru limba și literatura română

Acasă

Colecție de lucrări lexicografice

Bibliografie filologică și literară românească

Bibliografie filologică și literară străină

Publicații

Colțul cercetătorului ▼

„Spirit și limbă sunt aproape identice, iar limba și naționalitatea asemenea.” (Mihai Eminescu)

- SOLIROM este un spațiu virtual dedicat lucrărilor științifice elaborate sub coordonarea Secției de Filologie și Literatură.
- Numele ales este o abreviere a numelui Societății literare române și reprezintă o formă de omagiere a întemeietorilor Academiei Române.
- Este o platformă software care cuprinde resurse lingvistice digitalizate sau digitizate și instrumente lingvistice digitale dedicate limbii și literaturii române.
- Conține două secțiuni:
 - publică – prezintă resursele lingvistice digitale destinate accesului public,
 - privată – cu instrumentele digitale necesare pentru administrarea resurselor lingvistice digitale ale platformei, destinată cercetătorilor care dezvoltă platforma.
- Existența SOLIROM facilitează punerea în practică a unuia dintre principalele obiective ale Academiei Române: „Ocroțirea și cultivarea limbii române, a conștiinței identitare și stabilirea normelor obligatorii ale limbii române” (Statutul Academiei Române) în relație cu politicile lingvistice ale Academiei.

Noutăți

2022

- Lansare a primei versiuni a Dicționarului limbii române în cadrul CLRE
- S-a adăugat secțiunea Publicații.

2021

- Lansare CLRE. Corpus lexicografic românesc electronic
- Baza 1688, volumul 7
- ARIN. Edine de autori restrictivi
- CFLR. Colecție de texte cu adnotări morfosintactice
- Șablon pentru site-uri web pentru dicționare academice
- Mutare site-uri web pentru dicționare academice în Grifab

2020

- Interfață transcriere dicționare - acces restricționat
- Dicționarul literaturii române de la origini până la 1900
- H. Tiktin. Rumänisch-Deutsches Wörterbuch 2

2019

Figure 7: La première page du SOLIROM (<https://solirom.ro/>)

A présent l'Académie Roumaine a englobé dans sa nouvelle page électronique une zone spéciale dédiée aux ressources (numériques) en langue roumaine (des dictionnaires, la grammaire, etc.) — https://acad.ro/institutia/limba_romana.html.

2.5 CoRoLa. Corpus numérique pour la langue roumaine contemporaine

Les projets lexicographiques nécessitent aussi des corpus linguistiques à partir desquels les dictionnaires peuvent être rédigés.

Pour la langue roumaine contemporaine *CoRoLa. Corpus computațional de referință pentru limba română contemporană* [Corpus numérique de référence pour la langue roumaine contemporaine] représente un corpus linguistique, élaboré à partir de 2014 en tant que programme prioritaire de l'Académie Roumaine et qui contient divers textes, datant de 1989 à nos jours, le but de sa création étant de fournir une image objective de la langue roumaine écrite et parlée actuelle. Le corpus est ouvert au public à travers une interface de recherche de données textuelles et une de recherche de données audio.

Le corpus est consultable en ligne à l'adresse: <http://corola.racai.ro/>.



Figure 8: La première page du CoRoLa

2.6 Autres ressources lexicographiques roumaines disponibles en ligne

2.6.1 eRomLex

Primele dicționare bilingve românești (secolul al XVII-lea). Corpus digital prelucrat și aliniat

(*eRomLex*) [Les premiers dictionnaires bilingues roumains (XVIIe siècle). Corpus numérique traité et aligné (*eRomLex*)] — <http://www.scriptadacoromanica.ro/bin/view/eRomLex/>¹³ — est un projet de (retro-)numérisation, bien inédit pour le roumain par la valorisation à travers une édition comparative numérisée des premiers efforts lexicographiques roumains des lexiques bilingues roumains, partant du fait qu'ils ont la même source, afin de mettre en évidence les caractéristiques de ce réseau lexicographique et de faciliter l'accès à leur contenu. Il s'agit de six lexiques bilingues slavons-roumains du XVIIe siècle (tous manuscrits), qui représentent des éléments très importants dans la première période de développement de la lexicographie roumaine. Ces 6 lexiques sont comparés à leur source commune (le lexique slavon-ruthène publié par Pamvo Berynda à Kiev, en 1627). Ce réseau lexicographique va être complété par des études concernant leur rapport à la source, les innovations à son égard, leurs usages probables, les filiations entre lexiques, le contexte socio-culturel, le but de cette compilation, leurs utilisations potentielles, etc.

Le projet propose également une méthode pour la mise en place d'une base de données alignée qui pourrait être connectée aux dictionnaires électroniques / corpus lexicographiques électroniques roumains déjà existants (CLRE, par exemple), complétant les informations linguistiques qu'ils contiennent et lancera une approche plus étendue destinée à produire des dictionnaires traités et annotés des corpus numérisés de roumain ancien auxquels ces lexiques pourraient être reliés afin de permettre la mise en valeur des pratiques de traduction, et les usages spécifiques du matériel linguistique; ces aspects pourraient éclairer certaines questions liées à l'origine formulées jusqu'à présent et qui sont difficiles à vérifier au moyen des méthodes classiques d'interrogation des textes.

Les étapes de la réalisation de ce projet de numérisation supposent: la translittération en alphabet latin des définitions roumaines et la transcription des mots slavons; le traitement automatique du matériel (unification des graphies; élimination des diacritiques); l'alignement des entrées correspondantes; l'indexation et recherche multicritères; des études quantitatives: nombre d'entrées communes, entrées spécifiques à un certain lexique; des innovations par rapport à la source; la comparaison des inventaires lexicaux; des études quantitatives: utilité du matériel pour DLR — premières attestations, sens / formes inédit(e)s; (directions de recherches futures) alignement a d'autres ressources numériques (e.g. CLRE)¹⁴.

Le corpus réalisé par ce projet complètera l'image d'ensemble des débuts de la lexicographie roumaine.

2.6.2 Lexiconul de la Buda [Lexique de Buda]

Une autre ressource lexicographique roumaine disponible en ligne est *Lexiconul de la Buda*¹⁵, l'édition électronique du premier dictionnaire étymologique et

explicatif de la langue roumaine et c'est une référence pour la lexicographie roumaine moderne. Pour le public qui consulte le dictionnaire en ligne, l'accès est gratuit après inscription sur le site, en créant un compte et un mot de passe. L'interface est extrêmement accessible, conçue pour être fonctionnelle, utile, en fonction des différents critères de requête, permettant ainsi un survol rapide du texte (en quatre langues: roumain, latin, hongrois et allemand). Ceci est le résultat d'un projet coordonné par l'Université Babeş-Bolyai, Cluj-Napoca.

Lexiconul de la Buda est publié en ligne à l'adresse <http://bcucluj.ro/lexiconuldelabuda/site/login.php>.

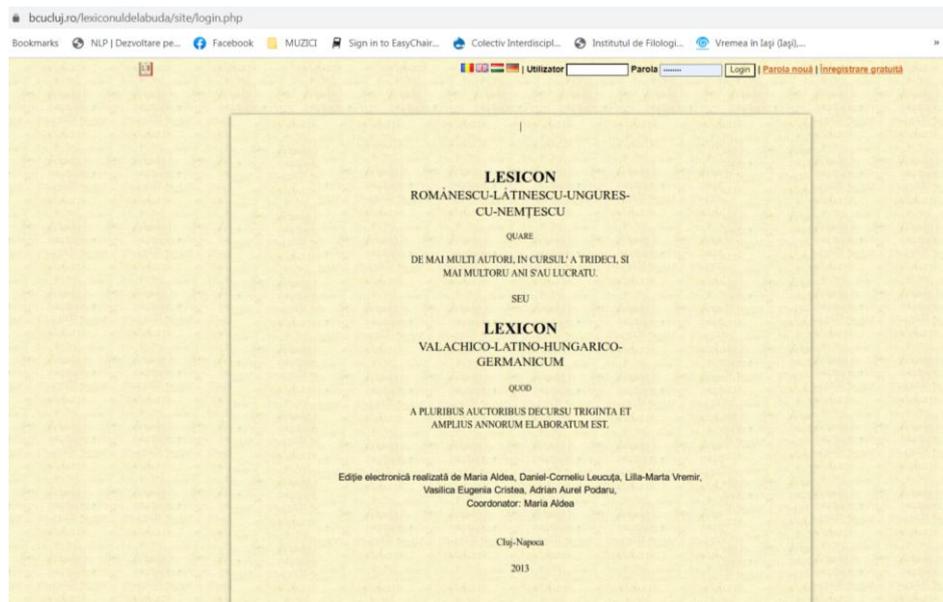


Figure 9: La première page du *Lexiconul de la Buda*

2.6.3 DIGIBUC

DIGIBUC (<http://www.digibuc.ro/>) représente la plus grande bibliothèque numérique roumaine, étant un projet réalisé par la Bibliothèque métropolitaine de Bucarest et la Bibliothèque de l'Académie roumaine, et est le partenaire officiel d'EUROPEANA, la bibliothèque numérique la plus importante d'Europe — (<http://www.europeana.eu/portal/>).

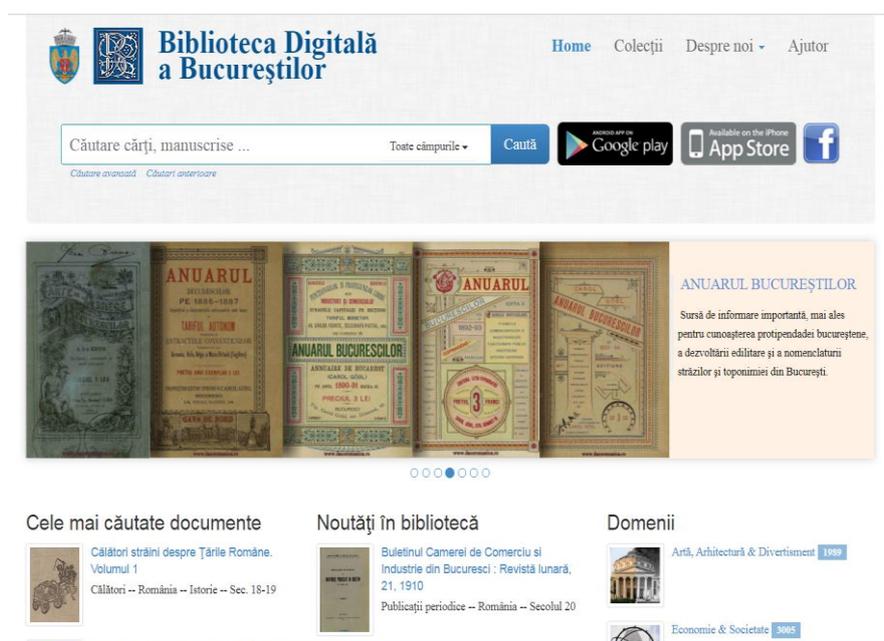


Figure 10: La première page du DIGIBUC

3. Perspectives actuelles dans la lexicographie roumaine

Les avantages de la numérisation de la recherche lexicographique roumaine résident dans le fait que les ressources créées représentent:

- des outils de travail pour les lexicographes et pour les linguistes en général;
- une source d'information facilement accessible pour ceux qui s'intéressent à la langue roumaine;
- base de création d'applications pour divers appareils nécessitant un lexique;
- des points de départ pour de futures recherches.

Ces résultats permettent également la corrélation avec d'autres ressources linguistiques ou multimédias. Ainsi, dans une prochaine étape, des connexions pourront être établies avec des informations issues de plateformes encyclopédiques telles que Wikipédia, une approche qui pourrait augmenter la visibilité des données lexicographiques incluses dans CLRE, par exemple, mais aussi enrichir les informations que le lecteur/l'utilisateur a besoin.

Donc, les perspectives de la lexicographie roumaine envisagent la création des dictionnaires numériques à partir d'un corpus textuel; le développement du corpus lexicographique CLRE; aligner les variantes numériques du *Dictionnaire-trésor de la langue roumaine* dans CLRE; rédiger / publier des dictionnaires

en format hybride (en format classique / en papier et, en même temps, en format numérique) ou exclusivement en format classique.

4. Conclusions

L'objectif de cet article est de mettre en évidence, en général, l'état actuel de la recherche lexicographique en Roumanie et ses perspectives.

Les dictionnaires électroniques et les corpus de textes, structurés en bases de données, facilitent la connaissance, la préservation et le maintien de l'identité culturelle au niveau linguistique et permettent l'inclusion d'une langue nationale dans le domaine d'intérêt de la recherche numérique des langues naturelles, au niveau mondial.

La numérisation de la recherche lexicographique roumaine est en plein développement, l'étape actuelle de l'évolution de la lexicographie roumaine impliquant, d'une part, la numérisation des ressources existantes et, d'autre part, la création de dictionnaires, de nouvelles ressources et instruments, directement en format électronique.

A travers tous les projets de numérisation de la recherche lexicographique roumaine, la langue roumaine entre plus solidement dans le circuit d'une communication plus facile.

Le processus de numérisation place la lexicographie académique roumaine à un niveau comparable à la lexicographie internationale et permet la connexion avec des projets lexicographiques de l'étranger et l'inclusion de la Roumanie dans la sphère d'intérêt des grands réseaux lexicographiques internationaux

Remerciements

Le présent article envisage aussi les résultats du projet de recherche *Primele dicționare bilingve românești (secolul al XVII-lea). Corpus digital prelucrat și aliniat* [Les premiers dictionnaires roumains bilingues (XVIIe siècle). Corpus numériques traités et alignés] <http://www.scriptadacoromanica.ro/bin/view/eRomLex/>), subventionné par le Ministère de la Recherche, de l'Innovation et de la Numérisation de Roumanie, CNCS/CCCDI — UEFISCDI, projet PNIII P11.1TE 20190517.

Endnotes

1. Pour une présentation détaillée du rôle des instituts de l'Académie Roumaine dans le processus de numérisation, voir Tamba 2022b.
2. Ainsi, dans cet article on va détailler aussi les étapes d'un projet de recherche intitulé *Primele dicționare bilingve românești (secolul al XVII-lea). Corpus digital prelucrat și aliniat (eRomLex)* [Les premiers dictionnaires bilingues roumains (XVIIe siècle). Corpus numérique traité et aligné (eRomLex)] réalisé à l'Université «Alexandru Ioan Cuza», de Iași (<http://www.scriptadacoromanica.ro/bin/view/eRomLex/>).

3. Pour exemple, <https://dexonline.ro/> — une plate-forme comprenant bien des dictionnaires de langue roumaine, projet initiée par des volontaires.
4. *Dicționarul limbii române* — appelé aussi *Dicționarul tezaur al limbii române* [Le Dictionnaire-trésor de la langue roumaine] ou *Dicționarul Academiei* [Le Dictionnaire de l'Académie] — est rédigé dans les Départements de Lexicographie des instituts de langue et de littérature roumaine de l'Académie Roumaine: Institutul de Lingvistică "Iorgu Iordan — Alexandru Rosetti" (ILB), București — <https://lingv.ro/>; Institutul de Filologie Română "A. Philippide" (IFR), Iași — <https://philippide.ro/>; Institutul de Lingvistică și Istorie Literară "Sextil Pușcariu" (ILIL), Cluj-Napoca — <http://inst-puscariu.ro/>.
5. Toutes les informations concernant le projet CLRE seront présentées plus en détail dans le sous-chapitre suivant.
6. Image extrait du Tamba (2017b: 141).
7. Pour plus de détails sur les projets réalisés au cours de la période 2003–2017, voir Haja (2017), Tamba (2017b), Haja et Tamba (2022).
8. Pour plus de détails sur ce projet, voir Haja et al. (2005).
9. Au cours des deux dernières années, au cadre de l'Académie roumaine — Filiale de Iași se déroulent un autre projet (financé au niveau national par compétition), intéressant du point de vue de la valorisation des informations de la première édition du DLR, mais aussi du point de vue de la création de ressources numérisée. Le projet TAFOC — *Terminologia astronomică românească: științific vs popular. Fenomene, obiecte cosmice și constelații* [La terminologie astronomique roumaine: scientifique vs. populaire). Des phénomènes, des objets cosmiques et des constellations] (directrice du projet: CS I dr. Cristina Michaela Florescu; <https://tafoc.sollirom.ro/>) propose une systématisation linguistique de la terminologie scientifique et populaire / commune de l'astronomie (champ lexical qui n'a pas été étudié systématiquement jusqu'à présent, tant en roumain, que dans les autres langues romanes), l'analyse ayant un caractère diachronique prononcé, en suivant l'étymologie et l'évolution historique des termes.
10. Toutes les informations concernant la plate-forme SOLIROM seront présentées plus en détail dans un autre sous-chapitre de cet article.
11. Pour une présentation détaillée sur CLRE, voir Haja et Tamba (2022), Tamba (2022a).
12. La version numérique de TDRG — H. Tiktin, *Rumänisch-Deutsches Wörterbuch* (la première édition: 1896–1926). La troisième édition de ce dictionnaire (publiée de 2003 à 2005) a été numérisée, sous le modèle de l'eDTLR, lors de la collaboration entre l'Albert-Ludwigs-Universität de Fribourg, en Allemagne, et l'Académie Roumaine. TDRG est publié en ligne à l'adresse: <https://tdrg.sollirom.ro/>.
13. Pour plus de détails sur ce projet, voir Gînsac et al. (2022).
14. Pour une présentation détaillée sur eRomLex, voir Haja et Tamba (2022), Tamba (2022a).
15. *Lesicon romanescu–latinescu–ungurescu–nemtescu quare de mai mulți autori, in cursul a trideci, si mai multor ani s'au lucrat. Seu Lexicon valachico–latino–hungarico–germanicum quod a pluribus auctoribus decursu triginta et amplius annorum elaboratum est.* Budae, Typis et Sumtibus Typographiae Regiae Universitatis Hungaricae, 1825. Ediție electronică de Maria Aldea, Daniel-Corneliu Leucuța, Lilla-Marta Vremir, Vasilica Eugenia Cristea și Adrian Aurel Podaru, Cluj-Napoca, 2013; <http://bcucluj.ro/lexiconuldelabuda/site/login.php>.

Références

Littérature secondaire

- Abdelzاهر, Esra M.** 2022. An Investigation of Corpus Contributions to Lexicographic Challenges over the Past Ten Years. *Lexikos* 32: 162-179.
- Busuioac, M.M., N. Mihai et Al. Anghelina.** 2018. Lexicografie și lexicologie. Sala, M. et N. Sarmandu. 2018. *Lingvistica românească*: 247-290. București: Editura Academiei Române.
- Clim, M.-R., E. Tamba, A. Catană-Spenchiu et M. Patrașcu.** 2016. *CLRE. Corpus lexicographique roumain essentiel*. 100 dictionnaires de la langue roumaine alignés au niveau de l'entrée et, partiellement, au niveau du sens. Éva Buchi, Jean-Paul Chauveau, Jean-Marie Pierrel (Éds.). 2016. *Actes du XXVII^e Congrès international de linguistique et de philologie romanes (Nancy, 15-20 juillet 2013)*. Vol. 2, Section 16: 1611-1622. Strasbourg: ÉLiPhi.
URL: <http://www.atilf.fr/cilpr2013/actes/section-16.html>
- Ernst, G.** 2013. Romanian. Gouws, R.H., U. Heid, W. Schweickard et H.E. Wiegand (Éds.). 2013. *Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Special Focus on Computational Lexicography*: 687-701. Berlin/Boston: De Gruyter.
- Gînsac, A.-M., M.-A. Moruz et M. Ungureanu.** 2022. The First Romanian Dictionaries (17th century). Digital Aligned Corpus. Klosa-Kückelhaus, A., S. Engelberg, C. Möhrs et P. Storzjohann (Éds.). 2022. *Dictionaries and Society. Proceedings of the XX EURALEX International Congress, 12–16 July 2022, Mannheim, Germany*: 222-229. Mannheim: IDS-Verlag.
- Gînsac, A.-M. et M. Ungureanu.** 2018. La lexicographie slavonne-roumaine au XVII^e siècle. Adaptations roumaines d'après le *Leksikon slavenorosskij* de Pamvo Berynda. *Zeitschrift für romanische Philologie* 134(3): 845-876.
- Haja, G. (Éd.).** 2017. *Lexicografia academică românească. Studii. Proiecte*. Iași: Editura Universității "Alexandru Ioan Cuza".
- Haja, G., E. Dănilă, C. Forăscu et B.-M. Aldea.** 2005. *Dicționarul limbii române (DLR) în format electronic. Studii privind achiziționarea*. Iași: Editura Alfa.
- Haja, G. et E.I. Tamba.** 2022. Publicarea online a primului volum din *CLRE. Corpus lexicografic românesc electronic — contextualizare*. Clim, M.-R. et E.I. Tamba (Coord.). 2022. *Actele Colocviului internațional "Lexicografia academică românească. Provocările informatizării", 2020–2021*: 213-218. Cluj-Napoca: Presa Universitară Clujeană.
URL: <http://www.editura.ubbcluj.ro/bd/ebooks/pdf/3188.pdf>
- Haja, G., E.I. Tamba, M.-R. Clim, C. Teodorescu et A. Anghelina.** 2020. Metode și tehnici actuale în redactarea *Dicționarului limbii române informatizat (DLRi)*. Ichim, Ofelia (Éd.). 2020. *România în spațiul euroatlantic: interferențe culturale și lingvistice*: 139-160. București: Editura Tracus Arte.
- Hanks, P.** 2013. Lexicography from Earliest Times to the Present. Allan, K. 2013. *The Oxford Handbook of the History of Linguistics*: 503-536. Oxford: Oxford University Press.
- Hartmann, R.R.K. et G. James.** 1998. *Dictionary of Lexicography*. London: Routledge.
- Kirchmeier, S.** 2020. Trends in European Language Policies with a View to Language Technology. *Standard Language/ Bendrinè Kalba* 93(2020): 1-23.
URL: <http://journals.lki.lt/bendrinekalba>

- Moruz, M.-A. et M. Ungureanu.** 2022. 17th-Century Romanian Lexical Resources and their Influence on Romanian Written Tradition. Klosa-Kückelhaus, A., S. Engelberg, C. Möhrs et P. Storjohann (Éds.). 2022. *Dictionaries and Society. Proceedings of the XX EURALEX International Congress, 12–16 July 2022, Mannheim, Germany*: 745-754. Mannheim: IDS-Verlag.
- Tamba, E.** 2014. La lexicografía Rumana. Historia y Actualidad. Córdoba Rodríguez, F., E. González Seoane et María Dolores Sánchez Palomino *Lexicografía de las lenguas románicas*: 265-282. *Perspectiva histórica. Vol. I*. Berlin/Boston: De Gruyter.
- Tamba, Elena Isabelle.** 2017a. CLRE. *Corpus lexicografic românesc esențial*. 100 de dicționare din Bibliografia DLR aliniate la nivel de intrare și la nivel de sens. Haja, Gabriela (Éd.). 2017. *Lexicografia academică românească. Studii. Proiecte*: 221-234. Iași: Editura Universității "Alexandru Ioan Cuza".
- Tamba, Elena Isabelle.** 2017b. Informatizarea lexicografiei academice românești. Haja, Gabriela (Éd.). 2017. *Lexicografia academică românească. Studii. Proiecte*, Iași: Editura Universității "Alexandru Ioan Cuza": 145-155.
- Tamba, Elena Isabelle.** 2022a. CLRE. *Corpus lexicografic românesc electronic*. Începuturi, dezvoltare și perspective. Clim, Marius-Radu et Elena Isabelle Tamba (Coord.). 2022. *Actele Colocviului internațional "Lexicografia academică românească. Provocările informatizării"*, 2020–2021: 219-232. Cluj-Napoca: Presa Universitară Clujeană.
URL: <http://www.editura.ubbcluj.ro/bd/ebooks/pdf/3188.pdf>
- Tamba, Elena Isabelle.** 2022b. The Role of the Institutes of the Romanian Academy in the Digitalization Process of the Linguistic Research. Jozić, Željko et Sabine Kirchmeier (Éds.). 2022. *The Role of National Language Institutions in the Digital Age. Contributions to the EFNIL Conference 2021 in Cavtat*: 91-100. Budapest: Nyelvtudományi Kukatóközpont.
- Tamba Dănilă, E., M.-R. Clim, M. Patrașcu et A. Catană-Spenchiu.** 2012. The Evolution of the Romanian Digitalized Lexicography. The Essential Romanian Lexicographic Corpus. Vatvedt Fjeld, R. et J.M. Torjusen (Éds.). 2012. *Proceedings of the 15th EURALEX International Congress, 7–11 August, 2012*, Oslo: 1014-1017. Oslo: Department of Linguistics and Scandinavian Studies, University of Oslo.
URL: http://www.euralex.org/proceedings-toc/euralex_2012/
- Trandabăț, D., E. Irimia, V. Barbu Mititelu, D. Cristea et D. Tufiș.** 2012. *Limba română în era digitală — The Romanian Language in the Digital Age*. META-NET White Paper Series: Europe's Languages in the Digital Age. Heidelberg/New York/Dordrecht/London: Springer.

Dictionnaires, corpus lexicographiques

- CLRE:** *Corpus lexicografic românesc electronic*, realizat de Departamentul de lexicologie și lexicografie, de la Institutul de Filologie Română "Alexandru Philippide" al Academiei Române — Filiala Iași.
<https://clre.solirom.ro/>
- DA:** Academia Română, *Dicționarul limbii române [DA]* (1913–1948), ediție digitală anastatică în *Corpus lexicografic românesc electronic*, realizat de Departamentul de lexicologie și lexicografie, de la Institutul de Filologie Română "Alexandru Philippide" al Academiei Române — Filiala Iași.
<https://dlr1.solirom.ro/>
- DAF:** *Dictionnaire de l'Académie Française*. <https://dictionnaire-academie.fr/>

DLR: Academia Română, *Dicționarul limbii române* [DLR] (1965–2010), ediție digitală anastatică în *Corpus lexicografic românesc electronic*, realizat de Departamentul de lexicologie și lexicografie, de la Institutul de Filologie Română "Alexandru Philippide" al Academiei Române — Filiala Iași.
<https://dlr1.solirom.ro/>

DLR²: *Dicționarul limbii române* [DLR] (2021–présent) Ediția a doua, revizuită și adăugită. Academia Română. București: Editura Academiei Române.
<https://dlri.ro/>

DÉRom: Buchi, Eva et Wolfgang Schweickard (Dir.). *Dictionnaire Étymologique Roman* [DERom], (2008–présent), Nancy, ATILF; publication électronique.
<http://www.atilf.fr/DERom>

DRAE: *Diccionario de la lengua española de la Real Academia Española*. <http://buscon.rae.es/drael/>

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

Lexicon romanescu–latinescu–ungurescu–nemtescu quare de mai mulți autori, in cursul a trideci, si mai multor ani s'au lucrat. Seu Lexicon valachico–latino–hungarico–germanicum quod a pluribus auctoribus decursu triginta et amplius annorum elaboratum est. Budae, Typis et Sumtibus Typografiae Regiae Universitatis Hungaricae, 1825. Ediție electronică de Maria Aldea, Daniel-Corneliu Leucuța, Lilla-Marta Vremir, Vasilica Eugenia Cristea și Adrian Aurel Podaru, Cluj-Napoca, 2013.
<http://bcucluj.ro/lexiconuldelabuda/site/login.php>

OED: *Oxford English Dictionary*. <http://www.oed.com/>

TLFi: *Le Trésor de la Langue Française Informatisé*. <http://atilf.atilf.fr>

TLIO: *Tesoro della lingua italiana delle origini*. <http://tlio.ovc.cnr.it/TLIO/index2.html>