Trotter, David/Bozzi, Andrea/Fairon, Cédrick (éd.) (2016). *Actes du XXVII^e Congrès international de linguistique et de philologie romanes (Nancy, 15-20 juillet 2013). Section 16 : Projets en cours ; ressources et outils nouveaux*. Nancy, ATILF :
http://www.atilf.fr/cilpr2013/actes/section-16.html.

CLRE. Corpus lexicographique roumain essentiel. 100 dictionnaires de la langue roumaine alignés au niveau de l'entrée

1. Introduction

Le projet *CLRE, Corpus lexicographique roumain essentiel. Les dictionnaires de la langue roumaine alignés au niveau de l'entrée* (2010–2013), continue une série d'initiatives qui visent l'informatisation de la lexicographie roumaine et qui se sont déroulées au sein de l'Institut de Philologie Roumaine « A. Philippide », de l'Académie Roumaine de Iasi. Ce dernier projet, CLRE, a été financé par le Conseil National de la Recherche Scientifique (CNCS) de Roumanie. Le collectif de recherche a été formé par des lexicographes de l'Institut de Philologie Roumaine « A. Philippide », de l'Académie Roumaine de Iasi (c'est-à-dire Elena Tamba, Ana Catană-Spenchiu, Marius Clim) et par des informaticiens de l'Université « Alexandru Ioan Cuza » de Iasi (Marius Răschip et Mădălin Ionel Patraşcu). Pendant cette période on a envisagé de créer un outil pour aider en particulier les lexicographes qui rédigent le *Dictionnaire-trésor de la langue roumaine*. CLRE représente un important outil pour les spécialistes et facilitera beaucoup le travail des lexicographes, parce que ce corpus offre surtout une perspective historique de la lexicographie roumaine et, implicitement, de la langue roumaine. Cette perspective signifie que le lexicographe aura la possibilité de voir comment un mot du vocabulaire roumain a évolué, s'il y a une variation sémantique, orthographique, ortho-épique ou morphologique. Ce corpus inclut divers types de dictionnaires : des dictionnaires généraux, des dictionnaires auxiliaires (de néologismes, d'étymologie, d'orthographie, orthoépie et de morphologie) et des dictionnaires spéciaux (c'est-à-dire des encyclopédies ou d'autres dictionnaires spécialisés, qui sont de grande importance pour la perspective diachronique sur la langue).

Mais cet outil sera également disponible au grand public qui aura accès à un grand nombre de dictionnaires, qui ont été inclus dans la base de données et alignés au niveau de l'entrée. De la Bibliographie de DLR (*Dictionnaire-trésor de la langue roumaine*) on a choisi 100 dictionnaires – 150.000 pages de dictionnaire. Cela constitue une première pour la langue roumaine, parce que ce corpus est à présent le plus grand et on peut toujours ajouter des nouveaux dictionnaires.

Ainsi, ce projet a eu pour buts :

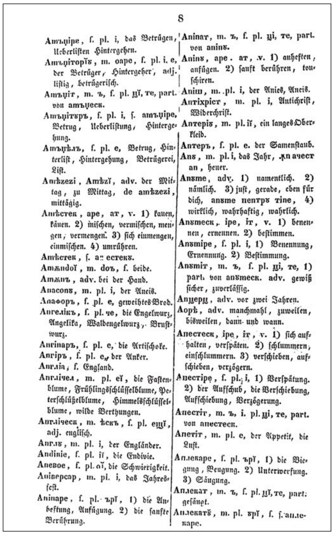
a) la réalisation d'une base de données qui comprenne les dictionnaires essentiels de la Bibliographie du DLR, alignés au niveau de l'entrée. La Bibliographie du *Dictionnaire de la langue roumaine* contient plus de 4000 titres des textes représentatifs pour la culture roumaine et elle est utilisée pour extraire des citations pour exemplifier les sens des mots-titre ;
b) l'élaboration des logiciels qui permettent la consultation interactive de ce corpus, qui peut constituer un cadre moderne de travail et de recherche lexicographique ;
c) la réalisation d'une liste de mots quasi-exhaustive, pour la langue roumaine, à partir du corpus aligné ;
d) l'augmentation de la visibilité de l'activité de recherche des professionnels linguistes et informaticiens, dans le domaine de la langue roumaine, pour promouvoir les moyens informatiques de traitement linguistique créés dans le projet.

2. La digitalisation des dictionnaires

En ce qui concerne les dictionnaires qui font partie du corpus CLRE, ils ont subi un processus de digitalisation, qui signifie qu'ils ont été numérisés à l'aide d'un scanner en V et puis traités du point de vue informatique, à l'aide d'un logiciel de reconnaissance optique des caractères (Abby Fine Reader 9) pour reconnaître les entrées et pour ensuite indexer les résultats dans une base de données XML.

Cette étape a généré de nombreux problèmes à cause du type d'alphabet utilisé, mais également à cause du format du dictionnaire.

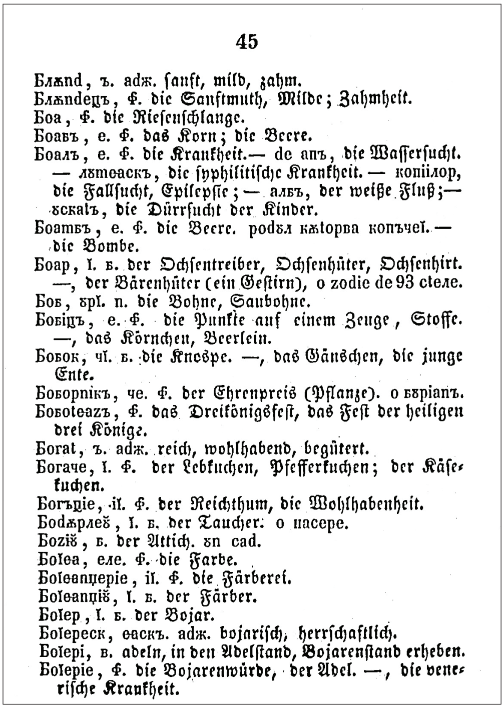
Pour la langue roumaine, les dictionnaires ont été écrits en utilisant trois types d'alphabets : l'alphabet cyrillique, l'alphabet de transition ou l'alphabet latin. Les premiers deux types ont été un véritable défi pour les informaticiens, vu que le logiciel de reconnaissance optique a été presque inefficient pour ces textes. On verra, dans les deux figures suivantes, des captures de ces dictionnaires.



8

Амъціре, f. pl. i, das Betrügen, Ueberlisten Hintergehen.
Амъціторіѕ, m. оаре, f. pl. i. e, der Betrüger, Hintergeher, adj. listig, betrügerisch.
Амъціт, m. ъ, f. pl. ції, те, part. von амъцеск.
Амъцітѕръ, f. pl. i, f. амъціре, Betrug, Ueberlistung, Hintergehung.
Амъцѣлъ, f. pl. e, Betrug, Hinterlist, Hintergehung, Betrügerei, List.
Амѣzezi, Амѣzї, adv. der Mittag, zu Mittag, de амѣzezi, mittägig.
Амѣстек, аре, ат, v. 1) kauen, käuen. 2) mischen, vermischen, mengen, vermengen. 3) sich einmengen, einmischen. 4) umrühren.
Амѣстек, f. a: естекѕ.
Амѫndoї, m. доъ, f. beide.
Амѫнъ, adv. bei der Hand.
Анаconѕ, m. pl. i, der Aneis.
Анафоръ, f. pl. e, geweihtes Brod.
Ангелікъ, f. pl. че, die Engelwurz, Angelika, Waldengelwurz, Brustwurz.
Анrinаръ, f. pl. e, die Artischoke.
Анrіръ, f. pl. e, der Anker.
Англia, f. England.
Англічел, m. pl. eї, die Fastenblume, Frühlingsschlüsselblume, Peterschlüßelblume, Himmelsschlüsselblume, wilde Werthungen.
Англіческ, m. ѣскъ, f. pl. ещї, adj. englisch.
Англѕ, m. pl. i, der Engländer.
Andinie, f. pl. iї, die Endivie.
Аневое, f. pl. ої, die Schwierigkeit.
Аніверсар, m. pl. i, das Jahresfest.
Аніnаре, f. pl. ърї, 1) die Anheftung, Anfügung. 2) die sanfte Berührung.
Аніnат, m. ъ, f. pl. ції, те, part. von аніnѕ.
Аніnѕ, аре. ат, v. 1) anheften, anfügen. 2) sanft berühren, touschiren.
Аніш, m. pl. i, der Anies, Aneis.
Антіхріст, m. pl. i, Antichrist, Widerchrist.
Антеріѕ, m. pl. iї, ein langes Oberkleid.
Антеръ, f. pl. e. der Samenstaub.
Анѕ, m. pl. i, das Jahr, ѫн ачест an, heuer.
Анѕme, adv. 1) namentlich. 2) nämlich. 3) just, gerade, eben für dich, анѕме пентрѕ tine, 4) wirklich, wahrhaftig, wahrlich.
Анѕмеск, ipe, iт, v. 1) benennen, ernennen. 2) bestimmen.
Анѕміре, f. pl. i, 1) Benennung, Ernennung. 2) Bestimmung.
Анѕміт, m. ъ, f. pl. ції, те, 1) part. von анѕмеск. adv. gewiß sicher, zuverlässig.
Анцерц, adv. vor zwei Jahren.
Аорѣ, adv. manchmahl, zuweilen, bisweilen, dann und wann.
Апестеск, ipe, iт, v. 1) sich aufhalten, verspäten. 2) schlummern, einschlummern. 3) verschieben, aufschieben, verzögern.
Апестіре, f. pl. i, 1) Verspätung. 2) der Aufschub, die Verschiebung, Aufschiebung, Verzögerung.
Апестіт, m. ъ, i. pl. ції, те, part. von апестеск.
Апетіт, m. pl. e, der Appetit, die Lust.
Аплекаре, f. pl. ърї, 1) die Biegung, Beugung. 2) Unterwerfung. 3) Säugung.
Аплекат, m. ъ, f. pl. ції, те, part. gesäugt.
Аплекатѕ, m. pl. ѕрї, f. f. аплекаре.

Figure 1 : Andrei Iser, *Vocabular românesc-nemțesc.* Brașov, I. Gött, 1850.



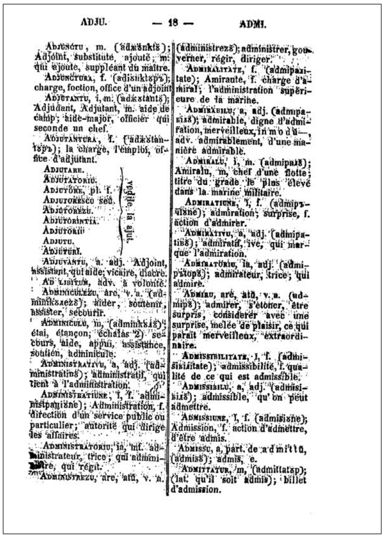
45

Блѫnd, ъ. адж. sanft, mild, zahm.
Блѫndeцъ, ф. die Sanftmuth, Milde; Zahmheit.
Боа, ф. die Riesenschlange.
Боабъ, e. ф. das Korn; die Beere.
Боалъ, e. ф. die Krankheit.— de апъ, die Wassersucht. — лѫтоаскъ, die syphilitische Krankheit. — копіілор, die Fallsucht, Epilepsie; — албъ, der weiße Fluß;— ѕскаtъ, die Dürrsucht der Kinder.
Боамбъ, e. ф. die Beere. podѕл кѫтора копъчеї. — die Bombe.
Боар, ї. б. der Ochsentreiber, Ochsenhüter, Ochsenhirt. —, der Bärenhüter (ein Gestirn), o zodie de 93 стеле.
Боб, ѕрї. н. die Bohne, Saubohne.
Бобіцъ, e. ф. die Punkte auf einem Zeuge, Stoffe. —, das Körnchen, Beerlein.
Бобок, чї. б. die Knospe. —, das Gänschen, die junge Ente.
Боборnікъ, че. ф. der Ehrenpreis (Pflanze). o ѕѕріанъ.
Боботеазъ, ф. das Dreikönigsfest, das Fest der heiligen drei Könige.
Богат, ъ. адж. reich, wohlhabend, begütert.
Богаче, ї. ф. der Lebkuchen, Pfefferkuchen; der Käsekuchen.
Богъціе, ії. ф. der Reichthum, die Wohlhabenheit.
Бодѫрлеѕ, ї. б. der Taucher. o пасере.
Бозіѕ, б. der Attich. ѕн cad.
Боіеа, еле. ф. die Farbe.
Боіеанцеріе, ії. ф. die Färberei.
Боіеанціѕ, ї. б. der Färber.
Боіер, ї. б. der Bojar.
Боіереск, еаскъ. адж. bojarisch, herrschaftlich.
Боіері, в. adeln, in den Adelstand, Bojarenstand erheben.
Боіеріе, ф. die Bojarenwürde, der Adel. —, die venerische Krankheit.

Figure 2 : S. Petri, *Vocabular portativ românesc-nemțesc și nemțesc-românesc.* Sibiu, G. de Clozius, 1861.

Le processus de reconnaissance optique des caractères n'a pas offert des résultats utilisables, parce que chaque dictionnaire en cyrillique a son propre type de caractères. Cette diversité nécessitera une adaptation du logiciel de reconnaissance optique pour chaque œuvre lexicographique. Et aussi, comme on peut y voir dans les deux captures ci-dessus, vu qu'il s'agit de dictionnaires bilingues, on a des caractères spécifiques pour chaque langue, ce qui rend la tâche du logiciel ABBYY, de reconnaissance des caractères, presque impossible. Dans le cas de dictionnaires en alphabet cyrillique et de transition, on a opté pour la translittération manuelle des entrées. De cette manière on peut aligner les entrées de ces dictionnaires avec les autres dictionnaires. Mais l'utilisateur spécialisé aura également à sa disposition la capture numérisée de l'entrée du mot choisi. Ainsi il pourra utiliser toutes les informations contenues dans ces dictionnaires.

Le format des dictionnaires a constitué tout un autre défi pour les informaticiens, qui ont dû créer un logiciel de délimitation des entrées assez flexible.


ADJU. — 18 — ADMI.

ADJUNCTU, m. (adжunktu); Adjoint, substitute, ajouté; m. qui ajoute, suppléant du maître.
ADJUNCTURA, f. (adiunktуpъ); charge, foction, office d'un adjoint.
ADJUTANTU, i, m. (adжutantu); Adjudant, Adjutant, m. aide de camp; aide-major, officier qui seconde un chef.
ADJUTANTURA, f. (adжutanturъ); la charge, l'emploi, office d'adjutant.
ADJUTARE.
ADJUTATORIU.
ADJUTÓRE, pl. f.
ADJUTORESCU seu.
ADJUTOREZU.
ADJUTORINTIA.
ADJUTORIU
ADJUTU.
ADJUTURI.
vedile la ajut.
ADJUVANTU, a. adj. Adjoint, assistant, qui aide; vicaire, diacre.
AD LIBITUM, adv. à volonté.
ADMINICULEZU, are, v. a. (adminikulezu); aider, soutenir, assister, secourir.
ADMINICULU, m, (adminkulu); étai, étançon; échalas 2) secours, aide, appui, assistance, soutien, adminicule.
ADMINISTRATIVU, a, adj. (administrativu); administratif, qui tient à l'administration.
ADMINISTRATIUNE, i, f. (administpaциune); Administration, f. direction d'un service public ou particulier; autorité qui dirige les affaires.
ADMINISTRATORIU, ia, inf. administrateur, trice; qui administ[illegible]re, qui régit.
ADMINISTREZU, are, atu, v. a. (administrezu); administrer, gouverner, régir, diriger.
ADMIRALITATE, f. (admipalitate); Amiraute, f. charge d'amiral; l'administration supérieure de la marine.
ADMIRABILU, a, adj. (admipabilu); admirable, digne d'admiration, merveilleux, in m o d u —, adv. admirablement, d'une manière admirable.
ADMIRALU, i, m. (admipalu); Amiralu, m, chef d'une flotte; titre du grade le plus élevé dans la marine militaire.
ADMIRATIUNE, i, f. (admipъциune); admiration; surprise, f. action d'admirer.
ADMIRATIVU, a, adj. (admipativu); admiratif, ive, qui marque l'admiration.
ADMIRATORIU, ia, adj. (admipъtopu); admirateur, trice; qui admire.
ADMIRU, are, atu, v. a. (admipu); admirer, s'étoner, être surpris, considerer avec une surprise, melée de plaisir, ce qui parait merveilleux, extraordinaire.
ADMISSIBILITATE, i, f. (admisibilitate); admissibilité, f. qualité de ce qui est admissible.
ADMISSIBILU, a, adj. (admisibilu); admissible, qu'on peut admettre.
ADMISSIUNE, i, f. (admisiune); Admission, f. action d'admettre, d'être admis.
ADMISSU, a, part. de a d m i t t u, (admisu); admis, e.
ADMITTATUR, m, (admittatup); (lat. qu'il soit admis); billet d'admission.

Figure 3 : Nifon Bălăşescu, *Dicionariu româno-frances*. Volum I. Fascicul I. Bucuresci, Tipografia Mitropolitului Nifon, 1859.

Dans ce cas on remarque que l'auteur du dictionnaire a soumis neuf entrées à un autre mot-titre. Mais à cause de cela, ces neuf entrées manquent d'information morphologique, à une seule exception, et on a aussi des problèmes à les délimiter.

AFLUIREA REZERVIȘTILOR 30

AFLUIREA REZERVIȘTILOR. Deplasarea rezerviștilor la unitățile la care sînt repartizați potrivit planului de mobilizare a cetățenilor cu obligații militare. Poate fi : **directă**, cînd deplasarea se face din comune sau de la organizațiile socialiste la unități ; **prin puncte de adunare**, cînd cetățenii respectivi se grupează mai întîi în anumite locuri dinainte stabilite, de unde se trimit la unități.

AGENT DE TRANSMISIUNI. Persoană prin care se transmit diferite mesaje, verbale sau scrise, între diferiți comandanți (șefi) militari, realizînd, în acest fel, legătura între aceștia. Un **A.d.T.** poate fi militar sau luptător din formațiunile de apărare și acționează pe jos sau transportat (cu motocicleta, auto, avion, elicopter).

AGENȚIE DE CAMPANIE A BĂNCII NAȚIONALE A REPUBLICII SOCIALISTE ROMÂNIA. Instituție bancară organizată de Banca Națională, cu atribuții de executare de casă pentru unități și M.U., potrivit dispozițiilor tehnice primite de la filiala de campanie și dispozițiilor organului financiar al ordonatorului de credite pe lîngă care este repartizată.

AGRAFĂ. Material genistic din sîrmă curbată și ascuțită la ambele capete, folosit la prinderea sîrmei (ghimpate) pe pari sau țăruși, pentru realizarea gardurilor și rețelelor de sîrmă.

AGRESIUNE ARMATĂ. După textul „Definiției agresiunii armate" adoptat de Comitetul special al O.N.U. la 15 aprilie 1974 și însușit de Adunarea Generală în același an agresiunea este „folosirea forței armate de către un stat împotriva suveranității, integrității teritoriale sau independenței politice a unui stat sau în orice mod incompatibil cu Carta Națiunilor Unite". Statul care a recurs primul la forța armată, în contradicție cu Carta O.N.U., este calificat agresor. Se consideră **A.A.** acel act comis cu intenție, de o anumită gravitate, care se deosebește de așa-numitele „incidente minore" (de pildă, incidente sporadice de frontieră). Cazurile tipice de **A.A.** sînt : invadarea sau atacarea teritoriului unui stat, bombardarea, blocarea porturilor sau coastelor sale de către

Figure 4 : *Lexicon militar*. Bucureşti, Editura Militară, 1980.

Dans cette figure on remarque une autre spécificité lexicographique, c'est-à-dire qu'il n'y a pas un seul mot-titre, mais un syntagme, une construction fixe ou une expression. Dans ce cas il est impossible d'englober tous les mots dans une entrée et le lexicographe choisira le mot-titre pour pouvoir aligner ce dictionnaire de cette façon.

17 ALGOL

canonice sînt unice, ceea ce rezultă simplu din modul de construire al acestora. Formele canonice sînt importante pentru că permit să se precizeze, fără a fi necesară determinarea valorilor pentru toate combinațiile, dacă două expresii sînt echivalente (reprezintă aceeași funcție). Astfel, considerînd două expresii scrise în forma disjunctivă, pentru a putea vedea că ele sînt echivalente, se deduce mai întîi forma canonică disjunctivă cu ajutorul proprietăților expresiilor booleene. Dacă expresiile obținute conțin aceiași mintermeni, expresiile date sînt echivalente. Să utilizăm acest procedeu pentru a demonstra că expresiile $f(x, y, z)$, $E(x, y, z)$ sînt echivalente. Întrucît $E(x, y, z)$ constituie o formă canonică disjunctivă, este necesar să deducem forma canonică pentru $f(x, y, z)$; se obține expresia: $f(x, y, z) = xy + y\bar{z} + \bar{x}yz = xy \cdot 1 + y\bar{z} \cdot 1 + \bar{x}yz = xy(z + \bar{z}) + y\bar{z}(x + \bar{x}) + \bar{x}yz = xyz + xy\bar{z} + \bar{x}y\bar{z} + xy\bar{z} + \bar{x}yz = \bar{x}yz + \bar{x}y\bar{z} + xy\bar{z} + xyz$, care este echivalentă cu $E(x, y, z)$. Cîteva exemple de funcții de două variabile întîlnite frecvent în informatică sînt: $f_1 = x + y$, $f_2 = x \cdot y$, $f_3 = x \cdot y$, $f_4 = x + y$, $f_5 = x\bar{y} + \bar{x}y = x \oplus y$, $f_6 = xy + \bar{x}\bar{y}$. Funcțiile f_1 și f_2 reprezintă operațiile SAU, respectiv ȘI definite anterior. Funcția f_3 constituie operația ȘI-NU, iar f_4 operația SAU-NU (NICI). Funcțiile f_5, f_6 reprezintă respectiv operațiile SAU-EXCLUSIV (sumă modulo doi) și COINCIDENȚĂ. Din expresiile lui f_5 și f_6 rezultă că funcția f_5 este 1 dacă $x \neq y$, iar f_6 este 1 dacă $x = y$ (deci cînd valorile lui x și y coincid) și că f_5 și f_6 sînt complementare. Deoarece orice funcție booleană admite două forme canonice, exprimate cu ajutorul operațiilor ȘI, SAU, NU, rezultă că acestea pot fi utilizate pentru reprezentarea tuturor funcțiilor. Se numește set complet, sau funcțional complet, de operații mulțimea operațiilor ce permit reprezentarea oricărei funcții booleene; deci operațiile ȘI, SAU, NU formează un set complet. Dacă ținem seama de relațiile lui de Morgan, care permit obținerea operației SAU (ȘI) cu ajutorul operației ȘI(SAU) și al negației, rezultă că perechile SAU, NU și, respectiv ȘI, NU constituie, de asemenea, seturi complete. Întrucît funcțiile logice sînt generate cu ajutorul unor circuite al căror cost depinde de expresiile acestora, se urmărește obținerea unei expresii cît mai simple. Obținerea expresiei cu forma cea mai simplă este numită minimizare, existînd metode eficiente pentru minimizarea funcțiilor booleene. *(P.D.).*

algebră logică (engl.: boolean algebra), **algebră booleană.**

ALGOL (**ALGO**rithmic **L**anguage), limbaj de programare de nivel înalt destinat, în primul rînd, aplicațiilor cu caracter științific. Este primul limbaj de programare elaborat prin cooperare internațională și definitivat, în varianta **ALGOL**-60 în 1960, odată cu publicarea binecunoscutului document „Raport Revizuit asupra Limbajului Algoritmic **ALGOL**-60". **ALGOL**-60 este cunoscut sub trei forme: limbajul de referință, limbajul de lucru și limbajul de publicare. Limbajul de referință corespunde definiției din Raportul Revizuit. În această formă, limbajul

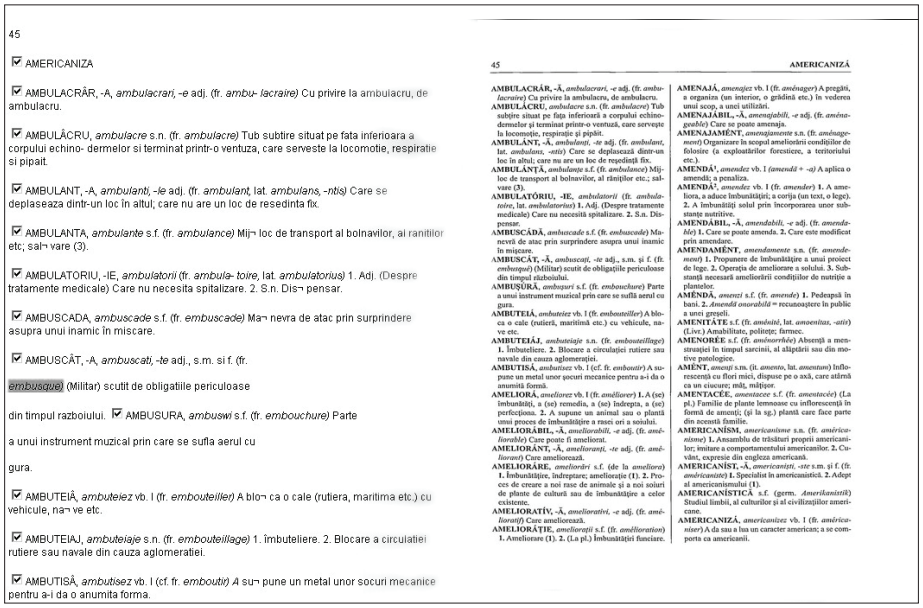
2

Figure 5: *Dicționar de informatică*. București, Editura Științifică și Enciclopedică, 1981.

Dans ce dictionnaire il y a une différenciation entre deux types d'entrées. La première est écrite avec des minuscules, ce qui dénote qu'on a une expression ou un syntagme qui utilise un mots-titre déjà défini au-dessus. Dans la deuxième situation, le mot-titre est utilisé aussi dans le texte de la définition de la même manière formelle. Vu que le mot-titre est écrit plusieurs fois, cela pose des problèmes lors de la délimitation des entrées, le logiciel pouvant les considérer comme des entrées distinctes.

3. La validation de la segmentation des dictionnaires

Ce processus a été réalisé par les lexicographes, mais également avec l'aide de quelques volontaires qui possédaient des connaissances philologiques assez solides.



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.

embusqué) (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.
AMBULĂCRU, *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, -antis*) 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.** Procedeu 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. *amélioratif*) 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.
AMENAJÁMÂNT, *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.
AMENDAMÊNT, *amendamente* s.n. (fr. *amendement*) **1.** Propunere de îmbunătățire a unui proiect de lege. **2.** Operație de ameliorare a solului. **3.** Substanță necesară ameliorării condițiilor de nutriție a plantelor.
AMÊNDĂ, *amenzi* s.f. (fr. *amende*) **1.** Pedeapsă în bani. **2.** *Amendă onorabilă* = recunoaștere în public a unei greșeli.
AMENITÁTE s.f. (fr. *amènité,* lat. *amoenitas, -atis*) (Liter.) Amabilitate, politețe; farmec.
AMENOREÉ s.f. (fr. *aménorrhée*) Absență a menstruației în timpul sarcinii, al alăptării sau din motive patologice.
AMÊNT, *amenți* s.m. (lt. *amento,* lat. *amentum*) Inflorescență cu flori mici, dispuse pe o axă, care atârnă ca un ciucure; mâț, mițișor.
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.
AMERICANISM, *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*) **1.** 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.

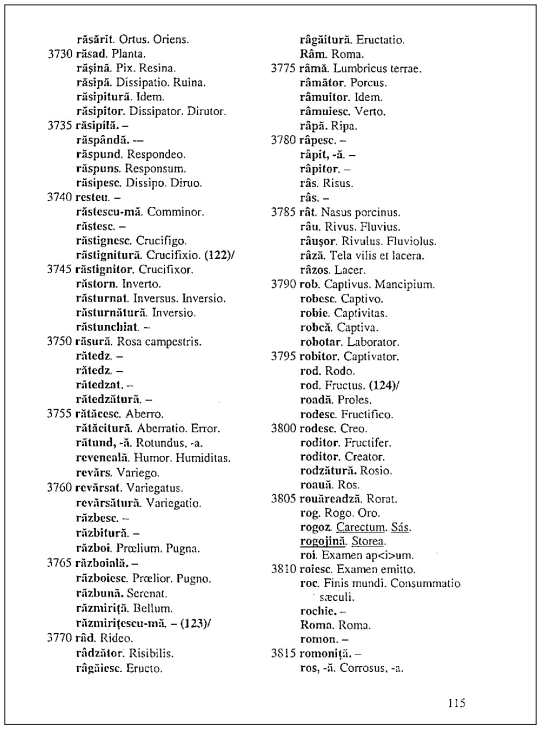
Figure 6 : La validation de la segmentation en entrées d'une page d'un dictionnaire

Dans cette étape, les informaticiens ont créé un logiciel qui essaie de reconnaître chaque entrée d'après ses caractéristiques formelles. Ce logiciel signale les entrées qui sont considérées comme véritables, mais il souligne également les autres mots qui peuvent être envisagés comme des entrées. Le lexicographe doit valider chaque entrée à son tour. Par cette démarche on valide les entrées correctes, mais, en même temps, chaque entrée est corrigée pour pouvoir les aligner facilement. Pour chaque page validée, le lexicographe a accès également à la page numérisée, pour pouvoir corriger les erreurs. S'il y a des mots qui ne peuvent pas être validés comme des entrées, alors le lexicographe peut choisir de valider partiellement la page respective. Après ça, l'informaticien traitera de nouveau cette page pour reconnaître les entrées restantes.

4. La segmentation des entrées de dictionnaire

Un autre élément distinct et définitoire pour le projet CLRE c'est le fait qu'il met à disposition des utilisateurs la variante numérisée du dictionnaire, pour pouvoir vérifier chaque entrée. Ainsi, chaque entrée est segmentée de façon automatique après l'opération de validation réalisée par les lexicographes.

Mais l'exactitude de la segmentation est liée, bien sûr, au format des dictionnaires. Et pour cela les informaticiens ont dû adapter leur logiciel pour toute une variété des formats. En général, le validateur statistique utilise les suivants éléments pour faire la segmentation : l'alignement à gauche et à droite du texte et la police (le style de la police– italique, souligné, gras –) et les types des caractères (majuscules, minuscules).

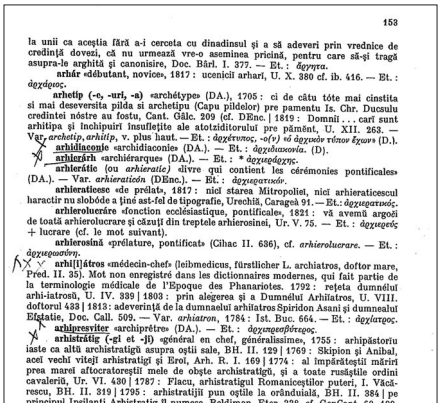


răsărit. Ortus. Oriens.
3730 **răsad**. Planta.
răşină. Pix. Resina.
răsipă. Dissipatio. Ruina.
răsipitură. Idem.
răsipitor. Dissipator. Dirutor.
3735 **răsipită**. –
răspândă. –
răspund. Respondeo.
răspuns. Responsum.
răsipesc. Dissipo. Diruo.
3740 **resteu**. –
răstescu-mă. Comminor.
răstesc. –
răstignesc. Crucifigo.
răstignitură. Crucifixio. **(122)**/
3745 **răstignitor**. Crucifixor.
răstorn. Inverto.
răsturnat. Inversus. Inversio.
răsturnătură. Inversio.
răstunchiat. –
3750 **răsură**. Rosa campestris.
rătedz. –
rătedz. –
rătedzat. –
rătedzătură. –
3755 **rătăcesc**. Aberro.
rătăciturâ. Aberratio. Error.
rătund, -ă. Rotundus, -a.
revenealâ. Humor. Humiditas.
revârs. Variego.
3760 **revârsat**. Variegatus.
revârsătură. Variegatio.
răzbesc. –
răzbiturâ. –
război. Prœlium. Pugna.
3765 **războială**. –
războiesc. Prœlior. Pugno.
răzbună. Serenat.
răzmiriţă. Bellum.
răzmiriţescu-mă. – **(123)**/
3770 **râd**. Rideo.
rădzător. Risibilis.
răgâiesc. Eructo.
râgâiturâ. Eructatio.
Râm. Roma.
3775 **râmă**. Lumbricus terrae.
râmător. Porcus.
râmuitor. Idem.
râmuiesc. Verto.
râpă. Ripa.
3780 **râpesc**. –
râpit, -ă. –
râpitor. –
râs. Risus.
râs. –
3785 **rât**. Nasus porcinus.
râu. Rivus. Fluvius.
râuşor. Rivulus. Fluviolus.
râză. Tela vilis et lacera.
râzos. Lacer.
3790 **rob**. Captivus. Mancipium.
robesc. Captivo.
robie. Captivitas.
robcă. Captiva.
robotar. Laborator.
3795 **robitor**. Captivator.
rod. Rodo.
rod. Fructus. **(124)**/
roadă. Proles.
rodesc. Fructifico.
3800 **rodesc**. Creo.
roditor. Fructifer.
roditor. Creator.
rodzătură. Rosio.
roauă. Ros.
3805 **rouăreadză**. Rorat.
rog. Rogo. Oro.
rogoz. Carectum. Sás.
rogoĭină. Storea.
roi. Examen ap<i>um.
3810 **roiesc**. Examen emitto.
roc. Finis mundi. Consummatio sæculi.
rochie. –
Roma. Roma.
romon. –
3815 **romoniţă**. –
ros, -ă. Corrosus, -a.

115

Figure 7 : *Dictionarium Valachico-Latinum. Primul dicţionar al limbii române*. Studiu introductiv, ediţie, indici şi glosar de Gh. Chivu. Bucureşti, Editura Academiei Române, 2008 [cca 1650]

Dans ce dictionnaire, considéré comme le premier dictionnaire de la langue roumaine, l'auteur a réalisé aussi une numérotation des mots introduits et a marqué cela en faisant une note tous les cinq mots et en inscrivant le nombre correspondant. Le résultat est que le logiciel de segmentation reconnaît les chiffres, et non pas les mots, comme étant les entrées.



153

la unii ca aceştia fără a-i cerceta cu dinadinsul şi a să adeveri prin vrednice de credinţă dovezi, că nu urmează vre-o aseminea pricină, pentru care să-şi tragă asupra-le arghită şi canonisire, Doc. Bâri. I. 377. — Et. : ἄρχυρα.

arhár «débutant, novice», 1817 : ucenicii arhari, U. X. 380 cf. ib. 416. — Et. : ἀρχάριος.

arhetip (-o, -uri, -a) «archétype» (DA.), 1705 : ci de căţu tóte mai cinstita si mai deseversita pilda si archetipu (Capu pildelor) pre pamentu Is. Chr. Ducsulu credintei nóstre au fostu, Cant. Gâlc. 209 (cf. DEnc. | 1819 : Domnii . . . carî sunt arhitipa şi închipuiri însufleţite ale atotziditoruluí pre pământ, U. XII. 263. — Var. *archetip, arhiti-p*, v. plus haut. — Et. : *ἀρχέτυπος, -ον(ν) τό ἀρχικόν τίπτον ἔχων* (D.).

arhidiaconíe «archidiaconie» (DA.). — Et. : *ἀρχιδιακονία*. (D).

arhierárh «archiérarque» (DA.). — Et. : * *ἀρχιερέαρχης*.

arhierátic (ou *arhieratic*) «livre qui contient les cérémonies pontificales» (DA.). — Var. *arhieraticón* (DEnc.). — Et. : *ἀρχιερατικόν*.

arhieraticéscе «de prêtre», 1817 : nici starea Mitropoliei, nici arhieraticescul haractir nu slobóde a ţine ast-fel de tipografie, Urechiă, Caragea 91. — Et.: *ἀρχιερατικός*.

arhieroluerâre «fonction ecclésiastique, pontificale», 1821 : vă avemŭ argoši de toată arhieroluerare şi căzuţi din treptele arhierosinei, Ur. V. 75. — Et. : *ἀρχιερεύς* + lucrare (cf. le mot suivant).

arhierosînă «prêlature, pontificat» (Cihac II. 636), cf. *arhierolucrare*. — Et. : *ἀρχιερωσύνη*.

arhi[i]átros «médecin-chef» (leibmedicus, fürstlicher L. archiatros, doftor mare, Pred. II. 35). Mot non enregistré dans les dictionnaires modernes, qui fait partie de la terminologie médicale de l'Époque des Phanariotes. 1792 : reţeta dumnélui arhi-iatrosŭ, U. IV. 339 | 1803 : prin alegerea şi a Dumnélui Arhiiatros, U. VIII. doftorul 433 | 1813 : adeverinţă de la dumnealui arhiiatros Spiridon Asani şi dumnealui Efstatie, Doc. Call. 509. — Var. *arhiatron*, 1784 : Ist. Buc. 664. — Et. : *ἀρχίατρος*.

archiereíter «archiprêtre» (DA.). — Et. : *ἀρχιερατεῖτερος*.

arhistrátig (-gi et -ji) «général en chef, généralissime», 1755 : arhipăstoriu iaste ca altŭ archistratigŭ asupra oştii sale, BH. II. 129 | 1769 : Skipion şi Aníbal, acei vechi viteji arhistratigi şi Eroi, Arh. R. I. 169 | 1774 : al împărăteştii măriri prea marei aflocratoreştii mele de obşte archistratigŭ, şi a toate rusăştile ordini cavalerieŭ, Ur. VI. 430 | 1787 : Flacu, arhistratigul Romanicestilor puteri, I. Văcărescu, BH. II. 319 | 1795 : arhistratiji pun oştile la crănduială, BH. II. 384 | pe principul Ipsilanti Arhistratig II numesc, Beldiman, Eter. 338, cf. GenCant. 60, 199, 273, etc. Pour un ex. dans Asachi cf. Hanеş, L. lit. 46. — Et. : *ἀρχιστράτηγος*. Au XVII[e] s. le même mot, venu par la filière du slave ecclésiastique, a désigné les chefs des anges, les archanges (cf. DA.). La var. *arhistratég* est un néologisme d'origine française (= archistratège). Cf. aroum. *arhistatigó (a. Mihail*, Geagea 314).

arhitécton «architecte», 1828 : făcăndusе un tahmin de către un architecton, Ur. VII. 171 | 1837 : şi spre siguranţie am dat chizăş pentru noi pe dumnealui architectonul, Iorga, Scris. Meşt. 94. A la même époque on a aussi *arhitector*, 1813 : va fi îngrijirea acésta asupra Domniei Sale vel Vornic de politie alcătuinduse şi arhitectorŭ al oraşului, Ur. I. 198, cf. 1824 : Bibl. R. IV. 97. Cette dernière forme est la relatinisation du terme italien vieilli *architettore* (cf. DA.). Déjà en 1793 on a *arhitect* (= all. Architekt, fr. *architecte*, cf. Pred. I. 206). — Et. : *ἀρχιτέκτων*.

arhitectonie «architecture» (DA.). — Et. : *ἀρχιτεκτονία*.

arhizugráv chef des peintres», 1797 : să mai adoge obligaţiunea, ca atunci căndŭ vre-unulŭ din zugravii cei învăţaţi deseverşiţi vor locui unŭ lucru din cele mai mari, archi-zugravulŭ să aibă a răndui lângă acelŭ zugrav învăţatŭ şi calfă, séŭ ucenici din cei mai proşti. U. III. 601. — Et. : *ἀρχιζωγράφος*. L'alternance *zugrav* (Murnu) ~ *zograv* montre nettement qu'il s'agit d'un emprunt récent où la voyelle radicale s'est mieux conservée, cf. aroum, *zuyráf* (-*hi*, Dal.).

Figure 8 : *Gáldi Ladislas, Les mots d'origine néogrecque en roumain à l'époque des Phanariotes*. Budapest, 1939.

Il y a aussi des dictionnaires, en particulier des dictionnaires anciens, qui ne sont pas faciles à trouver et qui en plus présentent de « traces » d'utilisation. Dans ce cas, c'est presque impossible, pour le logiciel, de reconnaître une entrée, si devant l'entrée il y a toute sorte de notations faites par de divers utilisateurs. Dans ce cas, c'est le validateur humain qui décide où commence l'entrée et qui corrige les éventuelles erreurs.

Les dictionnaires spécialisés posent de divers problèmes ayant trait à leur spécificité. Il y a toute sorte de marques d'orthographe, de morphologie et bien sûr, l'accent.

Après la validation, on essaye d'améliorer les résultats du logiciel de reconnaissance optique des caractères en passant par un nouveau processus de reconnaissance des caractères des dictionnaires analysés.

Les deux dernières étapes du projet ont envisagé l'alignement des entrées des dictionnaires et les possibilités d'interroger cette base de données.

En ce qui concerne l'alignement des dictionnaires, après avoir fini le processus de segmentation des entrées à l'aide de l'interface spécialement créée, on a réalisé un alignement primaire de tous les mots des dictionnaires et ensuite fait un alignement avec l'eDTLR[1]. Les entrées alignées sont incluses dans la base de données qui

[1] Le projet national *eDTLR. Le Dictionnaire (Trésor) de la Langue Roumaine en format électronique* a été financé par le Conseil National du Management des Projets (CNMP), pour la période 2007-2010 et il représente l'un de plus importants projets de type collaboratif des

permettra faire des recherches ultérieures, en utilisant des divers critères. Ainsi, on peut chercher dans un seul dictionnaire ou dans tous les dictionnaires introduites en CLRE, mais on peut effectuer également une recherche selon d'autres critères : on peut chercher les mots commençant avec la même lettre ou qui incluent un certain groupe de lettres. Le résultat de la recherche permet la visualisation de l'entrée du dictionnaire, tout d'abord en format image. Sur la même page, on peut voir aussi la variante reconnue par le logiciel ABBYY, mais elle ne peut pas être utilisée dans le cas des dictionnaires en alphabet cyrillique ou de transition. L'utilisateur peut voir également les informations concernant le dictionnaire qui a servi de source pour le mot en question.

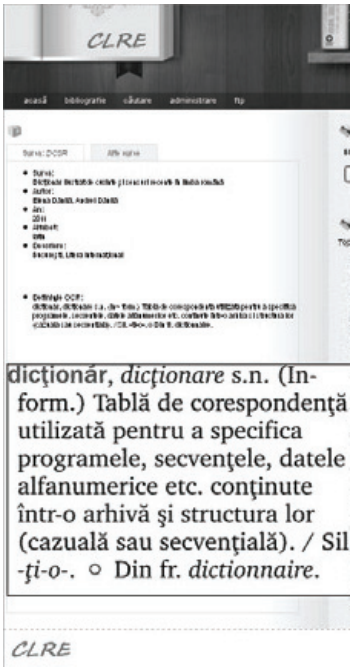


Figure 9 : Une capture d'écran de l'article « dictionnaire » d'un lexique écrit en alphabet latin.

dernières années, qui entremêle l'expérience des lexicographes et des informaticiens pour créer des outils électroniques. Le projet a été coordonné par dr. Dan Cristea de la Faculté d'Informatique de l'Université « Alexandru Ioan Cuza » de Iasi.

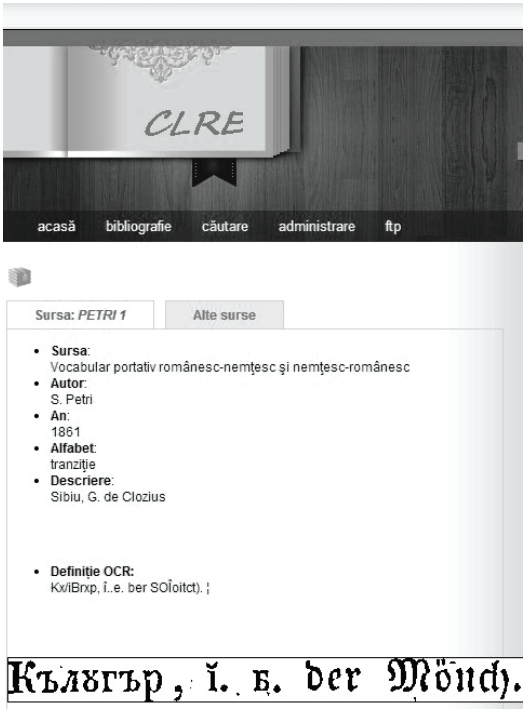


Figure 10 : Une capture d'écran de l'article « moine » d'un dictionnaire écrit en alphabet cyrillique.

5. Conclusions

Tant le spécialiste que l'utilisateur non initié peuvent consulter virtuellement, à l'aide du CLRE, une multitude de dictionnaires qui, autrement, ne seraient disponibles que dans les bibliothèques.

Ce projet est lié avec d'autres projets en ligne réalisés en Roumanie (eDTLR et la troisième édition du dictionnaire RDW[3], version informatisée) mais également avec le processus de réalisation du *Dictionnaire de la Langue Roumaine*, soutenu par l'Académie Roumaine. Il y aussi une liaison entre CLRE et d'autres projets internationaux : par exemple avec DÉRom[2] (le CLRE étant considéré comme source de documentation) ou avec *ENeL : European Network of e-Lexicography* qui est une action de la ligne de financement COST et qui cherche à créer un réseau paneuropéen des ressources lexicographiques pour l'usage académique et pour le public. Ces collaborations servent à confirmer l'importance de cette ressource électronique pour la langue roumaine.

[2] *Dictionnaire Étymologique Roman* (DÉRom) (www.atilf.fr/DERom/).

L'information accessible par l'intermédiaire du CLRE peut être comparée avec celle offerte du modèle des corpus lexicographiques crées pour d'autres langues : *Le rayon des dictionnaires*, http : //www.atilf.fr/ – collection de dictionnaires informatisés français, du XVI[e] jusqu'au XX[e] siècle ou *Das Wörterbuchnetz* ; *Nuevo tesoro lexicográfico de la lengua española*, http : //buscon.rae.es/ntlle/SrvltGUILoginNtlle – base de données qui inclut les versions facsimilées de tous les dictionnaires édités et publiés par la Real Academia Española ; http : //germazope.uni-trier.de/Projects/WBB/_– réseau de dictionnaires de langue allemande, créée à l'Université Trier d'Allemagne, etc.

De cette façon, CLRE offre aux spécialistes intéressés par l'étude de la langue roumaine, aux romanistes, mais aussi aux spécialistes informaticiens qui désirent analyser le traitement du langage naturel à partir de la langue roumaine, la désambiguïsation sémantique, etc. – un instrument de travail extrêmement utile.

L'Académie Roumaine, Filiale de Iasi — Marius-Radu CLIM
L'Académie Roumaine, Filiale de Iasi — Elena TAMBA,
L'Institut de Philologie Roumaine « A. Philippide » — Ana CATANĂ-SPENCHIU
L'Institut de Philologie Roumaine « A. Philippide » — Mădălin PĂTRAŞCU

Références bibliographiques

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

Clim, Marius / Dănilă, Elena / Haja, Gabriela, 2008. « Premise ale informatizării cercetării lexicografice academice româneşti », in : *Limba română. Dinamica limbii, dinamica interpretării*, Bucureşti, Editura Universităţii din Bucureşti, 585-591.

Dănilă, Elena, 2013. «Corpus lexicographique roumain essentiel. Les dictionnaires de la langue roumaine alignés au niveau de l'entrée», in : Herrero, Casanova / Calvo Rigual, Cesareo (ed.), 2013. *Actas del XXVI Congreso Internacional de Lingüística y de Filología Románicas 6-11 septiembre 2010, Valencia* (6-11 septembre 2010), Berlin / New York, La maison d'édition Walter de Gruyter, volum VIII, 125-134.

Dănilă, Elena / Clim, Marius-Radu / Catană-Spenchiu, Ana, 2011. «Towards a Romanian Lexicographic Corpus», *Philologica Jassyensia*, An VII, Nr. 2 (14), 191-198.

Tamba, Dănilă Elena / Clim, Marius-Radu / Pătraşcu, Mădălin / Catană-Spenchiu, Ana, 2012. «The Evolution of the Romanian Digitalized Lexicography. The Essential Romanian Lexicographic Corpus», in : Fjeld, Ruth Vatvedt, Torjusen, Julie Matilde, (ed.), 2012. *Proceedings of the 15 th EURALEX International Congress*, 7-11 august 2012, Oslo, Press Reprosentrales, UiO, 1014-1017 ; on-line : http : //www.euralex.org/proceedings-toc/euralex_2012/ .