

---

# Un Projet Accompli: le *Lexicon de Buda (1825)* en Édition Électronique

(Software Demonstration)

**Maria Aldea**

Université Babeş-Bolyai, Roumanie

e-mail: aldea\_maria@yahoo.com

## Abstract

Le but de notre ouvrage est de présenter le résultat du projet lexicographique accompli au sein de l'Université Babeş-Bolyai de Cluj-Napoca, intitulé *Le Lexicon de Buda (1825). Édition informatisée en vue d'une consultation en ligne*, id est l'édition électronique de ce dictionnaire quadrilingue ancien ayant pour langue de base le roumain (<http://dspace.bcuccluj.ro/handle/123456789/48723>).

**keywords** : le Lexicon de Buda; lexicographie roumaine numérisée; e-dictionnaires

## 1 La Lexicographie Roumaine Digitale : un État des Lieux

Depuis quelques années, suite à des initiatives institutionnelles et privées, la lexicographie roumaine se trouve dans un processus toujours plus avancé de numérisation. On y vise à la fois des ressources lexicographiques anciennes et contemporaines et la création de nouveaux dictionnaires électroniques, ce qui place la lexicographie roumaine parmi d'autres tentatives similaires internationales.

On retient dans ce sens l'*eDTLR – le Dictionnaire trésor de la langue roumaine en format électronique*, projet déroulé entre 2003 et 2005 sous la direction du professeur Dan Cristea de l'Université « Al. I. Cuza » de Iaşi, et qui deviendra par la suite la base du projet de digitalisation du *Dictionnaire trésor de la langue roumaine*, sous l'égide de l'Académie roumaine (en cours de réalisation) ; le *Corpus de référence de la langue roumaine pour la création des dictionnaires académiques (CNR)*, projet déroulé entre 2007 et 2008 sous la direction de Monica Busuioc de l'Institut de Linguistique « I. Jordan – Al. Rosetti » de Bucarest ; le *DLRI. Base lexicale informatisée. Dérivés*, programme déroulé entre 2007 et 2008 sous la direction de Cristina Florescu, chercheuse principale à l'Institut de Philologie roumaine « Al. Philippide » de Iaşi ; le *Corpus lexicographique roumain essentiel (CLRE)*, projet déroulé entre 2010 et 2013 sous la direction de Elena Tamba, auquel a participé une équipe de chercheurs provenant de l'Institut de Philologie roumaine « Al. Philippide » de Iaşi ; le *Dictionnaire explicatif de la langue roumaine en format électronique* ([www.dexonline.ro](http://www.dexonline.ro)), connu surtout comme le *DEX on-line* (pour une description plus détaillée de ces projets, voir Tamba 2014 : 273-276 ; Clim 2015 : 97-101) ; *La terminologie roumaine météorologique (scientifique vs populaire) des états atmosphériques. Étude linguistique*, projet développé entre 2011 et 2015 sous la direction de Cristina Florescu de l'Institut de Philologie roumaine « Al. Philippide » de Iaşi (<http://dfa.philippide.ro/ro/index.php>). La liste pourrait bel et bien se prolonger.

Parmi ces réalisations s'inscrit également notre projet de recherche consacré à un dictionnaire en quatre langues (roumain, latin, hongrois et allemand), qui a le mérite d'être le premier ouvrage lexicographique ancien qui soit numérisé dans l'espace public roumain.

## 2 Quelques Considérations sur le Projet

Tenu pour un repère fondamental dans la lexicographie roumaine moderne, *le Lexicon de Buda (1825)* est un ouvrage lexicographique tout à fait singulier, étant le fruit des efforts intellectuels conjugués de plusieurs érudits provenant de la région de Transylvanie – des personnalités marquantes de l'École latiniste de Transylvanie<sup>1</sup>. La rédaction du dictionnaire s'est étendue sur une période de plus de 30 ans. Rédigé en quatre langues (roumain, latin, hongrois et allemand) avec de nombreux renvois au français, à l'italien ou à l'espagnol, *le Lexicon de Buda (1825)* attire l'attention des spécialistes en étant le premier dictionnaire explicatif et étymologique de la langue roumaine à caractère scientifique, posant ainsi les assises de la lexicographie roumaine académique (Seche 1966 / 1969). Et pourtant, malgré son importance et étant données les nombreuses difficultés d'ordre technique et linguistique qu'il soulève, *le Lexicon de Buda (1825)* n'a jamais fait l'objet d'une réédition.

Pour relancer dans le circuit académique roumain de même qu'étranger cet ouvrage lexicographique exceptionnel, nous avons conçu l'idée de mettre en place une édition numérisée de celui-ci. Réunissant quatre linguistes (Maria Aldea, Lilla-Marta Vremir, Vasilica Eugenia Cristea et Adrian Aurel Podaru) et un informaticien (Daniel-Corneliu Leucuța) et bénéficiant du soutien financier de l'Autorité Nationale Roumaine pour la Recherche Scientifique (UEFISCDI-CNCS), notre projet de recherche, intitulé *Le Lexicon de Buda (1825). Édition informatisée en vue d'une consultation en ligne* (code du projet : PN-II-RU-TE-2011-3-0170), s'est déroulé sous l'égide de l'Université Babeș-Bolyai de Cluj-Napoca, pour une durée de 24 mois, entre 2011 et 2013 (voir [http://lexicon.dyndns.info/prezentare/index\\_files/Page558.htm](http://lexicon.dyndns.info/prezentare/index_files/Page558.htm)). Ce projet a bénéficié également du soutien de la Bibliothèque Universitaire Centrale « Lucian Blaga » de Cluj-Napoca qui a mis à notre disposition, à titre gratuit, un exemplaire scanné du *Lexicon* afin qu'on puisse découper et afficher les images, tout en abritant sur son site l'édition numérisée de l'ouvrage (<http://dspace.bcucuj.ro/handle/123456789/48723>).

### 2.1 Le Travail de Numérisation

L'édition électronique reproduit le texte de l'ouvrage publié en 1825 à l'Imprimerie de Buda, qui compte environ 900 pages et qui est divisé en plusieurs volets répartis comme suit : *Cuvântul către cititori* [*Avant-propos*] (rédigé en latin et en roumain, la dernière version étant rédigée à son tour en caractères latins et cyrilliques ; 6 pages), *Prefața la Ortografie* [*Préface au volet L'Orthographe*] (6 pages) et *Ortografia* [*L'Orthographe*] (53 pages), tous les deux rédigés en latin, *Dialogul pentru începutul limbii române între nepotu și unchiu* [*Dialogue sur les origines du roumain entre un neveu et son oncle*] (rédigé en roumain avec des caractères latins et cyrilliques ; 49 pages), *Extrasul cu regulile de pronunțare* [*Extrait avec des règles de prononciation*] (une page) et *Dicționarul* [*le Dictionnaire*] en quatre langues (771 pages).

L'examen du texte retenu pour la numérisation a mis en évidence plusieurs aspects dont le plus important a été la diversité des systèmes orthographiques présents dans le texte, à savoir le système roumain à caractères cyrilliques et étymologiques, les caractères gothiques spécifiques aux séquences allemandes et les signes diacritiques employés en hongrois. Comme les autres versions scannées du *Lexicon* déjà existantes sur le web ont affiché « une qualité insuffisante » (Leucuța *et al.* 2012 : 908), l'informaticien du projet a créé « une nouvelle plateforme sur Internet dans le but de

<sup>1</sup> L'école latiniste de Transylvanie était un mouvement idéologique et culturel qui a milité pour l'émancipation des Roumains, pour l'égalité en droits à ceux d'autres ethnies vivant sur le territoire transylvanien.

rendre ce lexicon accessible en ligne » [nous traduisons] (Leucuța *et al.* 2012 : 908). Cette nouvelle plateforme a été réalisée à l'aide d'outils informatiques libres et gratuits :

le langage de script (côté serveur) PHP, le système de gestion de bases de données MySQL (About MySQL), le langage de base XHTML (XHTML 1), le langage de programmation de scripts JavaScript (JavaScript), d'autres outils informatiques comme les CSS (feuilles de style en cascade), le DOM (Document Object Model), AJAX (AJAX : A New Approach to Web Applications), la bibliothèque jQuery (jQuery), des pages web hébergées par un serveur HTTP Apache (Apache HTTP Server). [nous traduisons] (cf. Leucuța *et al.* 2012 : 903 ; voir également Aldea *et al.* 2013b : 13-15)

Les membres de l'équipe ont transcrit à la main le texte entier du *Lexicon*, tout en marquant les champs informationnels correspondants aux étiquettes métalinguistiques établies par les rédacteurs du texte original et qui ont été conservées (pour une description plus détaillée des tâches qui ont incombé à chaque membre de l'équipe, voir Aldea 2013a).

Pour rendre visible notre version électronique sur divers navigateurs Internet, on a procédé à la création de nouvelles polices de caractères, plus proches de la réalité du texte :

la police gratuite *Gentium Plus* (*Gentium – une police de caractères pour les nations*) nous a servi à rendre les lettres et les signes diacritiques pour l'écriture étymologique en roumain, de même que les caractères latins et hongrois. De même, la police gratuite *Leipzig Fraktur* nous a aidés à transcrire les caractères de l'écriture gothique allemande. La tâche la plus ardue a été de rendre les glyphes cyrilliques. Aucune série de glyphes pour cette époque-là n'était malheureusement pas disponible et, malgré l'existence de plusieurs polices de caractères cyrilliques pour différentes langues, il n'y avait aucune à réunir à la fois toutes les lettres, les signes diacritiques et la forme correcte de ces glyphes. La solution a été de créer de nouveaux caractères cyrilliques à l'aide du logiciel FontForge. [nous traduisons] (voir Leucuța *et al.* 2012 : 904 ; Aldea *et al.* 2013b : 13-15)

Malgré les difficultés rencontrées, parmi lesquelles on compte aussi le manque d'une organisation cohérente et unitaire des champs informationnels, de même que la nécessité de prévoir des services de sécurité du système contre les risques d'attaque informatique, nous avons toutefois réussi à placer en libre accès cette édition électronique. Le public intéressé pourra non seulement accéder à une consultation plus rapide du texte, mais aussi en disposer de plusieurs manières différentes (cf. Aldea 2013a).

### 3 La Consultation de l'Édition Électronique du Lexicon de Buda

Pour avoir accès au contenu du *Lexicon*, une exigence de base est de s'inscrire sur la page web de celui-ci : <http://dspace.bcucuj.ro/handle/123456789/48723>. L'inscription est entièrement gratuite. Après avoir rempli les champs concernant le nom et le prénom, le nom d'utilisateur, l'adresse électronique et le mot de passe, un lien sera envoyé dans votre messagerie électronique. En cliquant sur ce lien, votre compte sera activé et vous pourrez accéder au contenu par une simple connexion.

Il convient de souligner que l'interface de la page web du projet est rédigée en quatre langues : roumain, anglais, hongrois et allemand.

Les textes précédant le dictionnaire proprement dit peuvent être visualisés par l'affichage en miroir du texte originel dans la partie gauche de l'écran et de sa transcription dans la partie droite. Il convient de mentionner que l'image scannée du texte originel peut subir une commutation avec le texte cueilli par les éditeurs ; seule la transcription interprétative reste inchangeable. De plus, il y a la possibilité de faire des recherches au sein de tous ces textes introductifs, de manière soit globale, soit restrictive à l'intérieur d'un certain texte.

En ce qui concerne le dictionnaire proprement dit, nous offrons trois options de recherche au sein de

la base des données.

La première vise la consultation *par feuilletage*. Cette option vous permet d'accéder à une liste par ordre alphabétique qui réunit des mots-vedettes, des co-vedettes (des doublets graphiques et lexicaux) et des sous-vedettes (ou des entrées secondaires). Un simple clic sur un mot permet l'affichage de l'entrée, sans tenir compte des signes diacritiques. Prenons l'exemple du mot *familie* « famille ». Il suffit de cliquer sur le mot pour que la barre d'outils de l'entrée s'affiche (voir Figure 1 : Entrée *familie* « famille »).



Figure 1 : Entrée *familie* « famille ».

On y retrouvera ainsi le texte cueilli par les éditeurs, la transcription du contenu et l'image découpée du texte scanné. Il y a également la possibilité d'afficher la page intégrale du texte originel où se trouve le mot recherché, de colorier la séquence textuelle qui vous intéresse et de bénéficier de renvois au site officiel du *DEX. Dictionnaire explicatif de la langue roumaine* ([www.dexonline.ro](http://www.dexonline.ro)), afin d'accéder à des explications des mots, de leurs valeurs sémantiques et de leur évolution. En ce qui concerne la transcription du contenu, nous avons choisi de l'afficher pour permettre au public non averti de découvrir et de comprendre le contenu de l'article de dictionnaire. Cette version mise à jour est donc une version amendée du texte, épurée des fautes typographiques évidentes. Ainsi, afin de faciliter la compréhension du texte en accord avec le contenu du dictionnaire, nous avons procédé à une mise à jour orthographique des contenus et à la mise en place d'une ponctuation spécifique à chaque langue présente dans le dictionnaire. On a abrégé d'une manière uniforme les classes et les catégories grammaticales, de même que les renvois, et on a uniformisé la ponctuation en fonction des emplois spécifiques (point, virgule, point-virgule), etc. (cf. Aldea 2013a).

Une seconde option d'interrogation concerne la *recherche simple*. Ce type de recherche se fait par le mot-vedette. Toutes les occurrences du mot recherché sont affichées sur la page. En cliquant sur le mot concerné, l'entrée apparaît de la même façon que dans la version *feuilletage*.

La dernière option met en évidence un moteur de *recherche avancée* qui permet d'effectuer des recherches avec croisement des informations. Ce type de recherche permet d'entreprendre une interrogation plus ample dans la base des données, c'est-à-dire à l'intérieur de l'information présentée dans le cadre de l'article ou, d'une manière restrictive, dans le cadre d'une seule langue (fût-elle le roumain, le latin, le hongrois ou l'allemand) (cf. Aldea 2013a) (voir Figure 2 : Composantes du menu *Avancé*).

Figure 2 : Composantes du menu *Avancé*.

De même, les recherches peuvent s'effectuer selon

(a) un champ informationnel grammatical : des classes grammaticales (noms, pronoms, numéraux, adjectifs, adverbes, interjections, etc.), des catégories grammaticales (masculin, féminin, singulier, pluriel, *singulare tantum*, *plurale tantum*), des degrés de comparaison, des expressions idiomatiques et des locutions, des suffixes ;

(b) un champ concernant l'information étymologique et les correspondances : des éléments lexicaux appartenant aux dialectes sud-danubiens, à d'autres langues romanes telles que le français, l'italien ou l'espagnol et à des langues comme le latin, le grec ou le turc ;

(c) d'autres champs informationnels concernant le domaine, les valeurs sémantiques et stylistiques, les marques d'usage, les proverbes, les citations avec des renvois aux œuvres ou aux auteurs cités. Prenons l'exemple du mot *mătușă* « tante » (voir Figure 3: Renvoi à l'entrée *mătușă* « tante »). Le mot-titre et son correspondant graphique en caractères cyrilliques mis à part, l'entrée n'enregistre qu'une seule information, c'est-à-dire un seul renvoi qui est colorié. Un clic sur celui-ci fait ouvrir une autre page tout en affichant le contenu de la nouvelle entrée (voir Figure 4: Entrée *mătușă* « tante »).

Figure 3 : Renvoi à l'entrée *mătușă* « tante ».



Figure 4 : Entrée mătușă « tante ».

À toutes ces options de visualisation et d'interrogation du contenu du *Lexicon de Buda (1825)* s'ajoutent un index des mots (provenant des dialectes sud-danubiens, de l'italien, du français, de l'espagnol et du grec), un index des proverbes et un index des œuvres et des auteurs cités (cf. Aldea 2013a).

#### 4 En Guise de Conclusion

Par la mise en format électronique du *Lexicon de Buda (1825)*, nous avons voulu, d'une part, placer cet ouvrage lexicographique en libre accès pour tout public intéressé, spécialiste ou non averti, afin d'une part, lui faciliter la consultation plus rapide du texte selon différents champs informationnels et, de l'autre, mettre à la portée des romanistes une information précieuse vu que le *Lexicon de Buda* conserve « la mémoire sémantique » de la Transylvanie de la fin du XVIII<sup>e</sup> siècle et du début du XIX<sup>e</sup> siècle.

#### 5 Références bibliographiques

- Aldea, M. (2013a). Avant-propos. In *Le Lexicon de la Buda (1825)*, <http://dspace.bcucluj.ro/jspui/handle/123456789/48723> [10/03/2016].
- Aldea, M., Harhătă, B., Leucuța, D.C., Vremir, L.M. (2013b). Primii pași în realizarea ediției electronice a Lexiconului de la Buda (1825). In *Philologica Jassyensia*, an IX, nr. 1 (17), pp. 9-16.
- Clim, M.R. (2015). La lexicografia rumana informatizada: tendencias obstáculos y logros. In *Lexicografía de las lenguas románicas. Aproximaciones a la lexicografía moderna y contrastiva. Volumen II*, Editores María José Domínguez Vázquez, Xavier Gómez Guinovart, Carlos Valcárcel Riveiro, Coordinadores María Dolores Sánchez Palomino, María José Domínguez Vázquez, Berlin/München/Boston, De Gruyter, pp. 95-110.
- Le Lexicon de la Buda (1825) = Lesicon romanescu-latinescu-ungurescu-nemtescu quare de mai mulți autori, in cursul a trideci, si mai multoru 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 realizată de Maria Aldea, Daniel Corneliu Leucuța, Lilla Marta Vremir, Vasilica Eugenia Cristea, Adrian Aurel Podaru. Coordonator Maria Aldea. Cluj-Napoca: 2013, <http://dspace.bcucluj.ro/jspui/handle/123456789/48723> [10/03/2016].
- Leucuța, D.C., Harhătă, B., Vremir, L.M., Aldea, M. (2012). The Romanian-Latin-Hungarian-German Lexicon - The Lexicon of Buda (1825). Informatics Challenges for an Emended and On-Line Ready Edition. In *Proceedings of the 15th EURALEX*

*International Congress 7-11 August 2012, Oslo*, Edited by Ruth Vatvedt and Julie Matilde Torjusen, Representralen, UIO, pp. 903-909, [http://www.euralex.org/elx\\_proceedings/Euralex2012/pp903-909%20Leucuta,%20Harhata,%20Vremir%20and%20Aldea.pdf](http://www.euralex.org/elx_proceedings/Euralex2012/pp903-909%20Leucuta,%20Harhata,%20Vremir%20and%20Aldea.pdf) [10/03/2016].

Tamba, E. (2014). Le lexicografia rumana. Historia y actualidad. In *Lexicografía de las lenguas románicas. Perspectiva histórica. Volumen I*, Edited by Félix Córdoba Rodríguez, Ernesto González Seoane, María Dolores Sánchez Palomino, Berlin/München/Boston, De Gruyter, pp. 265-282.

Tamba Dănilă, E., Clim, M.R., Catană-Spenchiu, A., Patrașcu, M. (2012). The Evolution of the Romanian Digitalized Lexicography. The Essential Romanian Lexicographic Corpus. In *Proceedings of the 15th EURALEX International Congress 7-11 August 2012, Oslo*, Edited by Ruth Vatvedt and Julie Matilde Torjusen, Representralen, UIO, pp. 1014-1017, [http://www.euralex.org/elx\\_proceedings/Euralex2012/pp1014-1017%20Danila,%20Clim,%20Patrascu%20and%20Catana-Spenchiu.pdf](http://www.euralex.org/elx_proceedings/Euralex2012/pp1014-1017%20Danila,%20Clim,%20Patrascu%20and%20Catana-Spenchiu.pdf) [10/03/2016].

*The Lexicon of Buda (1825). Emended and on-line ready edition*, page web du projet, [http://lexicon.dyndns.info/prezentare/index\\_files/Page558.htm](http://lexicon.dyndns.info/prezentare/index_files/Page558.htm) [10/03/2016].

Seche, M. (1966 / 1969). *Schiță de istorie a lexicografiei române*, vol. I-II, București : Editura Științifică.

## Remerciements

Nous tenons à remercier tous les membres de notre équipe pour leurs efforts intellectuels sans lesquels cette édition numérisée du *Lexicon de Buda* n'aurait pas pu voir le jour. De même, le soutien financier et institutionnel de diverses personnes et institutions nous a été à la fois salutaire et précieux.