

## Section 16. Projets en cours ; ressources et outils nouveaux

### L'édition informatisée de *Lexicon de Buda (1825)*

Maria Aldea, Bogdan Harhătă, Daniel-Corneliu Leucuța, Lilla-Marta Vremir

#### Résumé

Le but de notre article est la présentation de notre projet lexicographique développé au sein de l'Université « Babeș-Bolyai » de Cluj-Napoca, financé par l'Autorité Nationale Roumaine pour la Recherche Scientifique, CNCS-UEFISCDI, et intitulé *Lexiconul de la Buda (1825). Ediție emendată și preluată electronic pentru consultare on-line* (PN-II-RU-TE-2011-3-0170).

Dans notre article nous allons mettre en évidence les défis d'ordre linguistique et informatique de l'édition informatisée d'une œuvre lexicographique d'exception de l'Ecole Latiniste de Transylvanie, le *Lexicon de Buda (1825)*.

#### Introduction

Résultat de l'effort intellectuel de plusieurs érudits de Transylvanie, sur une période de plus de 30 ans, le *Lexicon de Buda (1825)* (désormais abrégé *LB*) représente le point de repère de la lexicographie roumaine moderne, étant apprécié par les traités de linguistique et d'histoire de la langue roumaine comme le premier dictionnaire explicatif et étymologique du roumain (voir Seche 1966 / 1969 ; Lupu 1999 ; Chivu 2007 ; Chivu 2008 ). Rédigé en quatre langues (roumain, latin, allemand et hongrois), le *LB*, au-delà de son enjeu idéologique, tel l'argumentation de la filiation latine du roumain, est un outil à valeur culturelle et linguistique remarquables, car il conserve la mémoire sémantique de l'espace pluriethnique et plurilingue de la Transylvanie de la fin du XVIII<sup>ème</sup> s. et le début du XIX<sup>ème</sup> s. (voir Aldea 2011 ; Harhătă et alii 2012 ). Malgré sa valeur inestimable, le *LB* n'a jamais fait l'objet d'une réédition. Pour quelles raisons ? Plusieurs ; il s'agit des difficultés d'ordre technique et linguistique, telle l'absence du caractère systémique de l'information, de l'orthographe étymologique et cyrillique du roumain, de lettres gothique spécifiques pour le texte allemand, de diacritiques hongrois pour le texte hongrois, etc.

#### Quelques propos sur le projet

Notre projet se déroule sur une période de 24 mois, plus précisément entre le 5 octobre 2011 et le 4 octobre 2013.

Par son objectif, notre projet s'aligne à d'autres projets réussis d'informatisation de dictionnaires roumains, notamment contemporains, tel le *DEX on-line* ([www.dexonline.ro](http://www.dexonline.ro)), l'*EDTLR – Le Dictionnaire de la Langue Roumaine en format électronique*, la réalisation du *Corpus Lexicographique Roumain* actuel, et le projet d'informatisation du *Dictionnaire Trésor de la Langue Roumaine*, sous la tutelle de l'Académie Roumaine. Mais, le mérite de notre projet est d'être le premier, dans l'espace culturel roumain, à réaliser une édition informatisée d'un dictionnaire ancien, bien que des éditions philologiques de glossaires bilingues roumains, en format papier, aient vu le jour sous la direction de Gheorghe Chivu, de Ladislau Gáldi, etc. Toutefois, nous devons dire que sur l'internet on trouve une version scanné et OCR-isé du *LB*. Mais une recherche dans le texte proposé par Google est restrictive et réductionniste, car la page scanné a des parties foncées, l'orthographe n'est pas reconnue, les graphèmes du texte original sont transposés dans la version électronique avec des symboles « parasites », les résultats ne sont pas affichés entièrement mais d'une façon limitée, etc. À cause de ces faits, l'information obtenue par le moteur Google aussi bien d'un point de vue linguistique, que scientifique n'est ni précise, ni correcte.

#### Quels défis ?

Dans notre effort de mener à bien ce projet, nous nous sommes affrontés à deux types de défis : d'une part, d'ordre linguistique et, d'autre part, d'ordre informatique.

Les défis informatiques sont : (1) la recherche et la création de nouvelles polices de caractères pour l'orthographe étymologique et cyrillique du roumain, pour les lettres gothiques spécifiques, et pour les diacritiques hongrois ; (2) rendre visibles ces polices de caractères dans l'interface web et dans le web browser ; (3) la création d'un site multilingue ; (4) la création de l'interface pour le travail simultané des

membres de l'équipe ; (5) la sécurité du système face à des attaques frauduleuses, etc. (voir Leucuța *et alii* 2012).

Les défis d'ordre linguistiques sont : (1) la réécriture manuelle du *LB* en format électronique, tout en respectant son contenu originel (c'est-à-dire la préface rédigée en latin et en roumain avec l'orthographe étymologique et cyrillique ; les principes orthographiques rédigés en latin, par P. Maior ; *Le Dialogue entre le neveu et son oncle*, par P. Maior, rédigé en roumain, avec l'orthographe étymologique et cyrillique ; les articles de dictionnaire) ; (2) établir les principes et les normes obligatoires pour l'approche linguistique du *LB*, étant donné le fait qu'à l'époque de sa rédaction le système linguistique et normatif n'était pas encore fixé ; (3) la traduction en roumain des principes orthographiques élaborés par P. Maior ; (4) la transcription interprétative de la préface et du *Dialogue ...* ; (5) la transcription interprétative de chaque langue présente dans l'article de dictionnaire ; (6) établir les champs informationnels et le type de recherche à faire aussi bien dans le contenu d'origine du *LB*, que dans la variante amendée et annotée, etc.

\*Cet proposition de communication est élaboré dans le cadre du projet no PN-II-RU-TE-2011-3-0170, financé par l'Autorité Nationale Roumaine pour la recherche scientifique, CNCS-UEFISCDI.

### Bibliographie

**Lexicon de Buda (1825)** = *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.

**Aldea 2011**, Maria Aldea, « Importanța Lexiconului de la Buda (1825) pentru lexicografia românească », dans *Astra blăjeană*. Publicație trimestrială editată de ASTRA – Despărțământul Timotei Cipariu Blaj, septembrie 2011, p. 42.

**Chivu 2007**, Gh. Chivu, « Este Lexiconul de la Buda un dicționar etimologic? », dans *Limba Română, limbă romanică. Omagiu acad. Marius Sala la împlinirea a 75 de ani*, Sanda Reinheimer Rîpeanu et Ioana Vintilă-Rădulescu (coordonare), București, Editura Academiei Române, 2007, pp. 61–65.

**Chivu 2008**, Gh. Chivu, « Lexiconul de la Buda – primul dicționar modern al limbii române », dans *Language and Literature. European Landmarks of Identity / Limbă și literatură. Repere identitare în context european. Lucrările celei de-a VI-a conferințe internaționale a Facultății de Litere, Pitești, 23-25 mai 2008, Volumul I*, Pitești, Editura Universității din Pitești, 2008, pp. 26–29.

**Harhătă et alii 2012**, Bogdan Harhătă, Maria Aldea, Lilla Marta Vremir & Daniel-Corneliu Leucuța, « The Lexicon of Buda. A Glimpse into the Beginnings of Mainstream Romanian Lexicography », dans *Proceedings of the 15<sup>th</sup> EURALEX International Congress 7-11 August 2012, Oslo*, Edited by Ruth Vatvedt and Julie Matilde Torjusen, Representralen, UIO, 2012, pp. 282-295.

**Leucuța et alii 2012**, Daniel-Corneliu Leucuța, Bogdan Harhătă, Lilla Marta Vremir & Maria Aldea, « The Romanian-Latin-Hungarian-German Lexicon - The Lexicon of Buda (1825). Informatics Challenges for an Emended and On-Line Ready Edition », dans *Proceedings of the 15<sup>th</sup> EURALEX International Congress 7-11 August 2012, Oslo*, Edited by Ruth Vatvedt and Julie Matilde Torjusen, Representralen, UIO, 2012, pp. 903-909.

**Lupu 1999**, Coman Lupu, *Lexicografia românească în procesul de occidentalizare latino-romanică a limbii române moderne (1780-1860)*, București, Logos, 1999.

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