

## **CLRE. Corpus lexicographique roumain essentiel.**

### **100 dictionnaires de la langue roumaine alignés au niveau de l'entrée et, partiellement, au niveau du sens**

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

Pendant les dernières années, plusieurs lexicographes et informaticiens roumains ont réalisé quelques projets interdisciplinaire de recherche linguistique roumaine, ayant pour but, d'une part, de valoriser les ressources existantes par leur acquisition en format électronique et, d'autre part, la création de nouvelles ressources et de nouveaux instruments pour l'étude de la langue (des dictionnaires informatisés, des corpus de textes écrits / parlés).

Ce travail porte sur l'accomplissement d'un projet qui s'est proposé de réaliser un corpus lexicographique qui inclut eDTLR (la version informatisée du *Dictionnaire «Trésor» de la Langue Roumaine. DLR* – réalisé sous l'égide de l'Académie Roumaine et qui comprend 14 tomes en 37 volumes, totalisant environ 20.000 pages et plus de 175.000 entrées, y compris les variantes) et les autres dictionnaires essentiels de la langue roumaine (dictionnaires anciens et nouveaux, dictionnaires généraux ou «spéciaux», etc.), alignés au niveau de l'entrée et, partiellement, au niveau du sens.

De cette manière, la langue roumaine va se comparer avec les autres langues qui possèdent déjà un corpus pareil: *Le rayon des dictionnaires* – <http://www.atilf.fr/spip.php?rubrique76> – collection de dictionnaires informatisés français, du XVI<sup>ème</sup> jusqu'au XX<sup>ème</sup> siècle; *Nuevo tesoro lexicográfico de la lengua española*, <http://buscon.rae.es/ntlle/SrvltGUILoginNtllle> – base de données qui inclut des versions facsimilées de tous les dictionnaires édités et publiés par la Real Academia Española; *Das Wörterbuchnetz*, <http://germazope.uni-trier.de/Projects/WBB/> – réseau de dictionnaires de langue allemande, créé à l'Université Trier d'Allemagne, etc.

Ce genre de corpus représente un instrument de travail indispensable pour les chercheurs linguistes, lexicographes ou non, car il rend accessible des informations comprises dans des divers dictionnaires, dès les premiers dictionnaires écrits en roumain jusqu'aux derniers.

On a eu des difficultés à réaliser une telle démarche, liées à la grande quantité de matériel qui doit être traité du point de vue lexicographique et informatique (100 dictionnaires – d'environ 120.000 pages de dictionnaires), à la situation des dictionnaires écrits en alphabet cyrillique ou en alphabet de transition, aux problèmes de copyright dans le cas des dictionnaires plus récents, etc.

La spécificité de notre projet est représentée par:

– récupération des dictionnaires anciens, en alphabet cyrillique par la transcription des entrées en alphabet latin;

– l'alignement de ces 100 dictionnaires pour illustrer l'évolution sémantique des mots et aussi morphologique et phonétique;

– le grand nombre de dictionnaires qu'on envisage (environ 100 dictionnaires de la Bibliographie de DLR – 120.000 pages de dictionnaire), qui doivent être inclus dans la base de données et alignées au niveau de l'entrée et, partiellement, du sens;

– la perspective historique sur la lexicographie roumaine et, implicitement, sur la langue roumaine dûe la sélection des dictionnaires qu'on a proposée;

– le fait que ce qu'on propose est un corpus suivant toutes les normes scientifiques, sa réalisation par une équipe mixte, formée par des linguistes et des informaticiens, et l'évaluation des résultats par des professionnels consacrés dans le domaine, assurant la qualité demandée par les exigences par une recherche compétente et adéquate.

Dans le projet on a utilisé tant des méthodes classiques / traditionnelles (par exemple, la translittération des entrées écrits en alphabet cyrillique ou en alphabet de transition) ou l'étude comparatif, au niveau sémantique des dictionnaires, que de méthodes nouvelles, de lexicographie informatisée.

Un autre but important pour le projet est la réalisation d'une liste de mots quasi-exhaustive, pour la langue roumaine, à partir du corpus aligné, ce qui constitue un autre instrument de travail pour les linguistes.

Ce projet suit les changements qui se sont produit dans la lexicographie roumaine actuelle. Les résultats de toutes ces démarches ouvrent une nouvelle étape dans la recherche lexicographique roumaine, ce qui permet l'alignement de la recherche roumaine aux standards internationaux de ce domaine.

L'importance de ce corpus, qui sera accessible à tous les chercheurs – et non seulement à ceux qui travaillent au *Dictionnaire* – est indiscutable. La version informatisée du dictionnaire (eDTLR) et le corpus lexicographique roumain essentiel faciliteront l'accès des romanistes à un instrument de travail indispensable pour les études comparatives.

### **Bibliographie**

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

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

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

TLIO = *Tesoro della lingua italiana delle origini* – <http://tlcio.ovi.cnr.it/TLIO/index2.html>

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

Tamba Dănilă, Elena, Clim, Marius-Radu, Pătrașcu, Mădălin, Catană-Spenchiu, Ana, *The Evolution of the Romanian Digitalized Lexicography. The Essential Romanian Lexicographic Corpus*, in *Proceedings of the 15 th EURALEX International Congress*, 7-11 august 2012, Oslo, eds. Ruth Vatvedt Fjeld, Julie Matilde Torjusen, , Press Representrales, UiO, p. 225.