

## ParCoGLiJe :

Corpus parallèle pour l'étude des grands classiques de la littérature de jeunesse

<b>Laboratoire pilote</b>	CLLE-ERSS, UMR 5263 (CNRS & Université de Toulouse 2 Jean Jaurès)
<b>Responsable du projet</b>	Dejan STOSIC, CLLE-ERSS (UMR 5263 CNRS & UT2J)
<b>Contact</b>	Dejan Stosic CLLE-ERSS (B512) Université de Toulouse 2 Jean Jaurès / MDR 5, allées Antonio Machado F-31058 Toulouse Cedex 9  Tél. 05 61 50 36 10 Mail: <a href="mailto:dejan.stosic@univ-tlse2.fr">dejan.stosic@univ-tlse2.fr</a>

La littérature de jeunesse est un des domaines fondamentaux de la littérature, d'où le fait qu'il soit à la fois fleurissant et extrêmement attractif autant pour les spécialistes de littérature que pour le grand public, y compris les plus jeunes lecteurs. Sa place dans l'enseignement primaire et secondaire n'est guère étonnante. Les spécialistes du domaine de la traductologie estiment cependant que la traduction en littérature pour la jeunesse reste un champ d'investigation très peu exploré. Lévêque (2015) souligne ainsi que « ce domaine est encore neuf en France, voire balbutiant » en dépit du fait qu'il se trouve à la croisée de nombreuses disciplines : littérature, traductologie, linguistique, didactique, histoire du livre et de l'édition, transferts culturels, etc.

L'objectif du projet ParCoGliJe est de fournir un premier corpus parallèle français-anglais de littérature de jeunesse à tous ceux qui en font leur objet d'étude ou d'intérêt : littéraires, linguistes, traductologues, didacticiens, enseignants, simples amateurs, etc. Ce projet s'inscrit dans une entreprise plus globale de constitution d'une base textuelle trilingue menée au sein de CLLE-ERSS. La ressource, intitulée ParCoLab, est consultable à l'adresse : <http://parcolab.univ-tlse2.fr/>. La constitution d'un corpus aligné français-anglais pour l'étude des grands classiques de la littérature de jeunesse est un préalable crucial pour aborder ces textes selon différentes perspectives, que celles-ci soient croisées ou pas.

Les textes originaux français et anglais sont alignés avec une de leurs traductions dans l'autre langue au niveau des paragraphes et des phrases. L'objectif étant de fournir une ressource textuelle librement téléchargeable, seuls les ouvrages libres de droits (fin du 19<sup>ème</sup> et début du 20<sup>ème</sup> siècles) ont été retenus dans cette première phase du travail. Le tableau ci-dessous présente les ouvrages intégrés à la plateforme ParCoLab et diffusés dans cette distribution.

Ouvrages en français		
Titre	Taille (en mots)	
	Original (FR)	Traduction (EN)
Daudet, Alphonse (1869), <i>Lettres de mon moulin</i>	46 592	47 706
Dumas, Alexandre (1844), <i>Les trois mousquetaires</i>	213 791	228 900
Comtesse de Ségur (1860), <i>Mémoires d'un âne</i>	54 662	42 040

Verne, Jules (1870), <i>Vingt mille lieues sous les mers</i>	142 959	141 936
Sous-total	458 004	460 582
<b>Ouvrages en anglais</b>		
Titre	Taille (en mots)	
	Original (EN)	Traduction (FR)
Dickens, Charles (1837), <i>Oliver Twist</i>	157 584	164 786
Hodgson Burnett, Frances (1911), <i>The Secret Garden</i>	80 558	76 940
Kipling, Rudyard (1894), <i>Jungle book</i>	51 334	55 913
Stevenson, Robert Louis (1883), <i>Treasure Island</i>	68 996	69 827
Sous-total	358 472	367 466
TOTAL	1 644 524	

Du point de vue quantitatif, cette première version du corpus comporte donc 1,6 millions de mots. Un accent particulier a été mis sur le traitement qualitatif des textes, conformément aux choix méthodologiques propres au projet ParCoLab. Un financement de la part du consortium CORLI nous a permis de finaliser et de diffuser cette ressource.

Les fichiers diffusés sont au format XML - normé TEI-P5, ce qui implique que chaque fichier individuel dispose de métadonnées détaillées. Quant à la structuration du contenu des ouvrages, de manière générale les chapitres correspondent aux éléments <div>, les paragraphes aux éléments <p>, et les phrases aux éléments <s>. Les fichiers dans les deux langues sont structurés de sorte à contenir le même nombre de ces éléments ; les segments parallèles d'un texte se trouvent donc au même endroit dans la structure XML.