

Étude du profil lexical d'un corpus de l'environnement

Patrick Drouin
Département de linguistique et de traduction
Université de Montréal
C.P. 6128, succursale Centre-ville
Montréal (QC) H3C 3J7
<http://termino.info>
patrick.drouin@termino.info

Il est généralement reconnu que les textes spécialisés comprennent trois couches lexicales principales : une terminologie liée au domaine (le lexique utilisé pour exprimer les connaissances spécifiques à ce domaine); des unités lexicales de langue générale (utilisée par l'ensemble des locuteurs d'une langue et que l'on peut trouver dans tous les types de textes); et une couche intermédiaire que l'on nomme généralement lexique transdisciplinaire (Tutin 2007, Drouin 2007, Tutin 2008, Hatier 2016).

Nous pensons qu'au sein de domaines très vastes tels que l'Environnement, qui englobe une grande variété de sujets (changement climatique, développement durable, énergies renouvelables, pollution de l'eau, etc.), la couche terminologique peut être divisée en deux couches distinctes. Une première partie du lexique est spécifique aux thématiques. Par exemple, les termes tels que chloration ou turbines marines sont respectivement spécifiques aux sous-domaines de la pollution de l'eau et des énergies renouvelables. Cependant, les travaux entrepris par l'Observatoire de linguistique Sens-texte (Drouin, L'Homme et Robichaud 2018) laissent penser qu'une autre partie du lexique relève de l'ensemble du domaine de l'environnement : écosystème, durable, énergie, développement, etc. On obtiendrait ainsi quatre couches lexicales différentes dans les textes spécialisés de vastes domaines plutôt que trois : le lexique général (LG), le lexique transdisciplinaire (LT), le lexique propre aux thématiques (LPT) et le lexique transdisciplinaire de domaine (LTD).

Dans cette présentation, nous nous intéresserons d'abord à la place et au rôle de chacune des couches lexicales. Nous présenterons par la suite une méthodologie semi-automatique ayant pour but d'isoler le LTD au sein d'un corpus de l'environnement. Le défi principal de son repérage réside dans la difficulté de le distinguer du LT et du LPT. Nous nous intéresserons aussi aux possibilités offertes par les outils de traitement automatique de la langue et aux limites de ces mêmes outils pour la prise en charge de la description du profil lexical des textes spécialisés.