

Quel sens prennent les mots anglais les plus fréquents du discours de la santé ? Une étude lexicographique du corpus ScienText

L'approche lexicographique dont nous souhaiterions rendre compte trouve son fondement dans la recherche d'équivalents de traduction dans quatre langues romanes (espagnol, français, italien, portugais) des mots anglais les plus fréquents du discours scientifique de la santé. La finalité est avant tout didactique : identifier des mots qui sont analogues dans ces langues, autrement dit dont la forme et le sens sont similaires, afin d'explorer ultérieurement la possibilité d'enseigner à des (futurs) personnels de la santé une langue romane à partir de leurs pré-acquis supposés en anglais dans ce domaine de spécialité. L'objectif a été également de mesurer quelle part du lexique anglais utilisé dans le discours scientifique de la santé est composée de termes de spécialité, et quelle part est plutôt à rattacher au lexique transdisciplinaire dans les catégories grammaticales et sémantiques étudiées. L'étude, qui met au jour certaines particularités des adjectifs, noms et verbes fréquents extraits, pointe des limites liées par exemple à la combinatoire de la langue anglaise, mais aussi aux outils statistiques.

Nous avons extrait des 35 millions de mots du corpus scientifique anglais de ScienText (Grossmann & Tutin, 2013) les 2000 plus fréquents. Nous expliquerons comment nous avons procédé pour écrémer, rectifier et homogénéiser les mots qui se rencontrent dans les écrits académiques de la santé. Cependant, une fois la liste à peu près stabilisée, nous avons rencontré bien des questions liées au sens des lemmes dans le discours. En effet, une polysémie procède de la spécialisation du sens d'un mot dans un domaine de spécialité, et une autre polysémie est liée à ce que le corpus regroupe des textes appartenant à plusieurs sous-spécialités (médecine, génétique, génomique, etc.). La délimitation des acceptions que peut prendre un même mot s'est opérée notamment au moyen du profil combinatoire (Blumenthal, 2002) dans le Lexicoscope (Kraif & Diwersy, 2012). 2208 acceptions fréquentes ont été isolées et définies.

Nous nous pencherons sur une limite principale que nous avons perçue, liée à la fréquence des cooccurrents d'un mot fréquent du lexique anglais. Nous indiquerons également les difficultés rencontrées lors du classement des acceptions parmi deux catégories sémantiques, le lexique de spécialité de la santé et le lexique scientifique transdisciplinaire (LST, Drouin, 2007).

Références

Références bibliographiques

BLUMENTHAL, PETER (2002). Profil combinatoire des noms. Synonymie distinctive et analyse contrastive, *Zeitschrift für Französische Sprache und Literatur*, 112 : 115-138.

GROSSMANN, FRANCIS & TUTIN, AGNÈS (2013). *L'écrit scientifique : du lexique au discours*. Rennes : Presses Universitaires de Rennes.

DROUIN, PATRICK (2007). Identification automatique du lexique scientifique transdisciplinaire. *Revue française de linguistique appliquée* 2007/2, vol. XII, p. 45-64.

KRAIF, OLIVIER & DIWERSY, Sacha (2012). Le Lexicoscope : un outil pour l'étude de profils combinatoires et l'extraction de constructions lexico-syntaxiques, *Actes de la conférence TALN 2012* : 399-406.

TUTIN, AGNÈS (2008). L'apport des corpus annotés pour l'élaboration semi-automatique d'une base de collocations de la langue scientifique générale, in François Maniez, Pascaline Dury, Nathalie Arlin & Claire Rougemont (Éd.) *Corpus et dictionnaires de spécialité*, 45-65. *Travaux du CRTT*. Grenoble : Presses Universitaires de Grenoble.

WILLIAMS, GEOFFREY C. (1999). *Les réseaux collocationnels dans la construction et l'exploitation d'un corpus dans le cadre d'une communauté de discours scientifique*. Thèse d'anglais. Nantes : Université de Nantes.

Références aux sites Internet

Lexicoscope, <http://phraseotext.u-grenoble3.fr/lexicoscope/> (nd). Consulté le 23 mars 2018.

ScienText, https://corpora.aiakide.net/scientext18/?do=SQ.setCorpus&corpus=sqCorpus_sctexts-en_sctexts-en&view=texts (nd). Consulté le 23 mars 2018.