

Caractérisation de la ville du futur dans un corpus de science-fiction

Sami Guembour^{1,2} Chuanming Dong^{1,3} Catherine Domingues^{1,4}

(1) LASTIG, Univ Gustave Eiffel, ENSG, IGN, France

(2) Université Paris Cité, Paris, France

(3) ADEME, Agence de l'Environnement et de la Maîtrise de l'Energie, France

(4) I-SITE FUTURE, Paris, France

{sami.guembour, catherine.domingues}@ign.fr
dongchuanming@yahoo.com

RÉSUMÉ

Ce résumé présente le travail de recherche détaillé dans l'article ([Guembour et al., 2023](#)). Ce travail s'intègre au projet PARoles de VilleS (PARVIS, <https://parvis.hypotheses.org/>); il se concentre sur la caractérisation de la ville du futur dans un corpus de romans de science-fiction et de fictions climatiques constitué par l'équipe PARVIS, en utilisant des techniques de traitement automatique des langues. L'objectif est d'analyser les images de la ville du futur à travers les éléments qu'elle contient (lieux, objets urbains, etc.) et les fonctions qui leur sont associées. Cependant, tous les romans du corpus ne parlent pas de la ville, d'où la nécessité d'identifier ceux dans lesquels elle constitue le cadre dans lequel s'intègrent l'action et les personnages. Pour cela, une ressource terminologique ([Topalov et al., 2010](#)) permettant d'identifier les éléments spécifiques de la ville a été utilisée. Elle regroupe 533 mots, en majorité des noms, désignés par la forme OOC (abréviation de "Object Of the City" pour "Objets de la ville").

Un algorithme de clustering fondé sur les fréquences des OOC dans les romans est utilisé pour identifier ceux de la ville, et ainsi construire un nouveau corpus qui est spécifique à la ville. Les OOC ayant permis de construire ce nouveau corpus n'ont pas tous les mêmes fréquences et donc pas le même poids dans la description de la ville du futur. Un algorithme de co-clustering a été appliqué sur les fréquences des OOC dans le nouveau corpus afin d'identifier ceux ayant un poids important dans la description de la ville; ces derniers sont appelés OOC discriminants. Un parseur de dépendance syntaxique est ensuite mis en oeuvre sur les phrases du nouveau corpus afin d'extraire les fonctions (verbes) associées aux OOC discriminants. Des algorithmes de clustering (et des techniques de réduction de dimension pour la visualisation des résultats) sont ensuite utilisés de manière itérative sur les cinq fonctions les plus fréquentes de chaque OOC discriminant pour regrouper ces OOC afin d'identifier d'éventuelles fonctions nouvelles ou anciennes associées à des OOC (anciens ou nouveaux objets de la ville).

Les résultats montrent que la ville du futur (dans le corpus PARVIS), comme celle d'aujourd'hui, vise à répondre principalement à deux problématiques principales : la circulation et l'habitation.

MOTS-CLÉS : TAL - corpus - clustering - réduction de dimensions - ville - urbain du futur - science-fiction.

KEYWORDS: NLP - corpus - clustering - dimension reduction - city - future urban - science fiction..

Références

GUEMBOUR S., DONG C. & DOMINGUÈS C. (2023). Characterization of the city of the future from a science fiction corpus. In E. MÉTAIS, F. MEZIANE, V. SUGUMARAN, W. MANNING & S. REIFF-MARGANIEC, Édts., *Natural Language Processing and Information Systems*, p. 313–325, Cham : Springer Nature Switzerland.

TOPALOV C., DE LILLE L. C., DEPAULE J.-C. & MARIN B. (2010). *L'aventure des mots de la ville*. Paris, France : Robert Laffont.