

*Aux représentantes
et représentants des médias*

COMMUNIQUÉ DE PRESSE

Twitter est bel et bien généré

Neuchâtel, le 16 mai 2023. Les hommes et les femmes s'expriment différemment sur Twitter, utilisant un vocabulaire et des émojis qui les distinguent. C'est l'une des conclusions de la thèse de doctorat de Catherine Ikae à l'Institut d'informatique de l'Université de Neuchâtel. La chercheuse en linguistique computationnelle a également exploré des techniques de *machine learning* permettant de savoir si un tweet donné a été rédigé par une femme ou par un homme. La présentation publique de sa thèse a eu lieu le 5 mai.

Qu'est-ce qui distingue les hommes et les femmes dans l'écriture de brefs messages ? Pour répondre à cette question, la doctorante Catherine Ikae, sous la direction du professeur Jacques Savoy, a passé au crible 338'000 tweets en langue anglaise, rédigés par une proportion égale d'autrices et d'auteurs.

On apprend ainsi que les femmes twittent en recourant plus souvent aux pronoms (*I/me, you, we/our*), et expriment plus souvent leurs émotions que les hommes (*happy, excited, etc.*). Elles portent également plus d'attention aux mots décrivant des liens sociaux (*girl, friend, family, xx*). Les hommes quant à eux préfèrent les pronoms de la troisième personne (*he, they*), citant plus de nombres et recourant à davantage de négations.

L'utilisation des émojis et des hashtags présente également des différences significatives entre les genres. Les femmes en font un usage plus récurrent que les hommes : une moyenne de 9 émojis sur mille mots pour elles, contre 6 sur mille mots pour eux. Les émojis les plus populaires ne sont pas les mêmes. Alors que les doigts féminins tapent plus volontiers des #love accompagnés d'émojis 😊, ❤️, 🥰, les pouces masculins ont une préférence pour 😏, 👍, 🙌, 🙏. S'agissant des thèmes abordés dans les messages écrits, les observations, là encore, confirment les clichés : les échanges sur le shopping et les marques sont plutôt prisés par les dames, tandis que les messieurs devisent volontiers sur le sport, les jeux ou la technologie.

Dans un autre volet de cette recherche, Catherine Ikae et Jacques Savoy ont testé dix modèles de *machine learning* pour apprendre à un logiciel à reconnaître si le texte analysé a été écrit par un homme ou une femme. Le plus performant d'entre eux est un réseau neuronal abrégé MLP qui a réussi à attribuer le bon genre caché derrière les tweets dans plus de 80% des cas. « Le principal résultat de cette étude a consisté à réduire drastiquement le nombre de mots et autres caractéristiques nécessaires à cette fin (densité des émojis, des pronoms ou des mots dénotant une émotion) passant de près de 140'000 à 300, sans diminuer la performance », note Jacques Savoy.

Encouragés par ces résultats en langue anglaise, Catherine Ikae et Jacques Savoy suggèrent de tester l'approche pour d'autres langues indo-européennes répandues (par exemple, l'espagnol) ou dans des idiomes très utilisés sur internet mais bien différents de l'anglais, comme le chinois ou le bengali.

En savoir plus :

Catherine Ikae & Jacques Savoy. "Gender identification on Twitter," Journal of the Association for Information Science & Technology, Association for Information Science & Technology, vol. 73(1), pages 58-69, 2022 <https://ideas.repec.org/a/bla/jinfst/v73y2022i1p58-69.html>

Contacts :

Prof. Jacques Savoy, Institut d'informatique

Tél. +41 32 718 13 75 / +41 79 593 55 71; jacques.savoy@unine.ch

Dre Catherine Ikae, Institut d'informatique

catherine.ikae@unine.ch