

# Le dire ou l'écrire sur les réseaux sociaux numériques : comparaison des messages écrits et des messages audios sur la chaîne Telegram des gilets jaunes de Haute-Garonne<sup>1</sup>

Pierre Ratinaud<sup>1</sup>, Brigitte Sebbah<sup>2</sup>

<sup>1</sup>LERASS, Université de Toulouse Jean Jaurès – pierre.ratinaud@univ-tlse2.fr

<sup>2</sup>LERASS, Université de Toulouse Paul Sabatier – brigitte.sebbah@iut-tlse3.fr

## Abstract

Those involved in the Yellow Vests movement have used numerous communication channels, most of which come under the heading of socio-numerical networks (Facebook, Twitter, WhatsApp...). We gained access to the Telegram feed of the Haute-Garonne yellow vests and set about analyzing the discussions held there. Initially, we looked at the 50,000 written messages, seeking to determine the different themes addressed based on an analysis with the Reinert method (Sebbah & Ratinaud, 2023) in the IRaMuTeQ software (Ratinaud, 2020). The availability of the Whisper model (Radfort & al., 2022), which enables the automatic transcription of oral discourse, also enabled us to analyze the 2201 audio messages present on the discussion threads. After highlighting some of the limitations of the automatic transcription produced by Whisper, we will focus on the differences and similarities between text and audio messages. Our first remarks will concern the statistics associated with these corpora (average message size, word frequencies, etc.), then we'll compare the preferred themes.

**Keywords:** Telegram, audio transcription, Whisper, oral, writing

## Résumé

Les acteurs du mouvement des gilets jaunes ont utilisé de nombreux canaux de communications relevant majoritairement des réseaux socionumériques (Facebook, Twitter, WhatsApp...). Nous avons eu accès au fil Telegram des gilets jaunes de Haute-Garonne et avons entrepris l'analyse des discussions qui s'y sont tenues. Nous nous sommes dans un premier temps intéressés aux 50 000 messages écrits en cherchant à déterminer les différentes thématiques abordées à partir d'une analyse avec la méthode Reinert (Sebbah & Ratinaud, 2023) dans le logiciel IRaMuTeQ (Ratinaud, 2020). La mise à disposition du modèle Whisper (Radfort & al., 2022) qui permet la transcription automatique du discours oral nous a permis d'envisager également l'analyse des 2201 messages audios présents sur les fils de discussion. Après avoir souligné quelques limites de la retranscription automatique produite par Whisper, nous nous attacherons à pointer les différences et les similarités entre les messages textuels et les messages audios. Nos premières remarques concerneront les statistiques associées à ces corpus (taille moyenne des messages, fréquences des mots, etc...), puis nous comparerons les thèmes privilégiés.

**Mots clés :** Telegram, transcription de l'audio, Whisper, oral, écrit

---

<sup>1</sup> Ce travail a été réalisé dans le cadre du Labex SMS, portant la référence ANR-11-LABX-0066, a bénéficié d'une aide de l'État gérée par l'Agence Nationale de la Recherche au titre du programme Investissements d'avenir portant la référence ANR-11-IDEX-0002-02.