

# La contribution de la linguistique dans l'analyse de la dynamique sociale : l'exemple des Gilets jaunes

Ana Zwitter Vitez  
UP FHŠ, Oddelek za uporabno jezikoslovje  
ana.zwitter@fhs.upr.si

UL FF, Oddelek za prevajalstvo  
ana.zwitter@ff.uni-lj.si

---

## Introduction

Dans la perception publique, la linguistique est souvent placée en dehors des sciences susceptibles de procurer des innovations proprement dites. Un groupe de linguistes qui développent un système d'annotation morpho-syntaxique automatique de corpus textuels sera traité différemment d'un groupe de chimistes qui développent un procédé permettant de lier des composites thermoplastiques avec des métaux sans adhésif. Cette perception différente est probablement due au fait qu'il est plus facile d'imaginer l'utilité des inventions des sciences de la nature que celle des inventions des sciences humaines.

L'objectif de cet article est de démontrer que l'analyse linguistique peut contribuer à une meilleure compréhension de la dynamique sociale. Nous allons analyser des messages par lesquels les utilisateurs du réseau social Twitter réagissent aux premières actions du mouvement des Gilets jaunes annonçant le début du combat contre le gouvernement d'Emmanuel Macron.

Comme ces messages révèlent des opinions et des émotions très polarisées, nous allons essayer de voir s'il existe des différences linguistiques par lesquelles on pourrait distinguer les messages exprimant le support de ceux qui expriment le désaccord avec les actions des Gilets jaunes. L'analyse comporte 30 tweets et porte sur les niveaux textuel, syntaxique et lexical.

## 1. La pertinence de l'analyse linguistique

Claude Hagège constate que la linguistique est « la seule science actuelle dont l'objet coïncide avec le discours qu'elle tient sur lui » (Hagège 1986, 62). En d'autres mots, un linguiste utilise la langue pour étudier la langue. Cette spécificité attribue à la linguistique un statut unique et imperceptible en même temps. La réflexion linguistique joue un rôle très important dans la formation des cultures, dans le domaine de l'apprentissage des langues et dans le domaine des technologies du langage qui rendent possible la communication entre l'homme et la machine. Parmi les domaines mentionnés, celui des technologies du langage reçoit le plus d'attention du public à cause de son potentiel économique. On bénéficie au niveau quotidien des résultats de recherches en linguistique. Il suffit de penser aux possibilités de faire corriger son texte par un correcteur automatique, de faire traduire automatiquement des textes par des outils de traduction et de se faire lire des textes écrits, oralisés par un logiciel de lecture automatique.

Dans les dernières années, le potentiel économique de la linguistique informatique et du traitement automatique du langage naturel ont influencé aussi les méthodes de l'analyse linguistique. Avec les nouveaux médias et les réseaux sociaux, la communication elle-même a changé. Autrefois, le discours public était réservé aux figures publiques. A l'aide des nouveaux médias, chaque individu peut publier des textes et des informations qui sont reçus par un public. L'analyse de ces textes peut apporter des informations sur leur auteur, sur son opinion et sur les thématiques qui l'intéressent. Wright (2009) constate que « pour de nombreuses entreprises, l'opinion en ligne est transformée en une sorte de monnaie virtuelle qui peut faire ou détruire un produit sur le marché ». Par conséquent, l'analyse de la communication humaine (parlée et écrite) est devenue un outil extrêmement important par lequel les entreprises ou les structures politiques cherchent à trouver leur chemin jusqu'à l'utilisateur final, ses orientations, ses désirs et ses angoisses (Zwitter Vitez et Pollak 2019).

La recherche de la production publique des utilisateurs des réseaux sociaux offre des possibilités presque illimitées et des moyens très puissants pour promouvoir leurs produits commerciaux ou battre leurs adversaires politiques. C'est pourquoi l'analyse de sentiment peut être utilisée pour examiner l'opinion publique et la dynamique sociale (Speriosu et al. 2011), prédire les résultats des élections (O'Connor et al. 2010; Smailović et al. 2015), mesurer la satisfaction des clients (Tedeschi et Benedetto 2015) et suivre la

dynamique des marchés boursiers (Rao et Srivastava 2014; Smailović et al. 2014).

## 2. Approches d'analyse de la dynamique sociale à travers la langue

Les débats autour de différents événements de la dynamique sociale produisent généralement des émotions polarisées : les usagers soutiennent ou détestent avec plus ou moins d'énergie les gestes des promoteurs des changements sociaux. Très souvent, les commentaires postés en tant que réactions aux mouvements sociaux dépassent la portée de ce geste isolé et se réfèrent à une problématique politique et sociétale plus large (Zwitter Vitez et Fišer, 2016). La recherche présentée dans cet article porte sur les domaines de la langue des réseaux sociaux, de l'analyse de sentiment et de l'analyse de l'opinion.

Le discours sur Twitter fait partie du discours électronique médié (Panckhurst 2014). Il est doté de certaines spécificités lexicales, graphiques et syntaxiques, comme émoticônes, abréviations et répétition de caractères (Farzindar et al. 2013) et connaît un certain nombre d'ajustements à cause de la limitation de la quantité de caractères dans un message (280 caractères maximum, espaces compris). Même si le discours sur Twitter est généralement perçu comme spontané et créatif (Vidak et Ackiewicz 2016), Longhi (2013) constate que « cette spontanéité n'est pas forcément maintenue dans le cadre du discours politique » à cause de l'impact énorme des tweets des politiciens auprès des usagers du réseau. Parmi les spécificités du discours sur Twitter, il semble pertinent de souligner également sa dimension technolangagière (Paveau 2013) représentée par le mot-dièse (hashtag) qui apporte une dimension affective aux messages (Cervulle et Pailler 2014) et, très souvent, modifie l'aspect pragmatique des messages (Longhi 2013).

Dans le cadre de l'analyse de sentiment qui fait partie du traitement automatique du langage (Pang et Lee 2008; Smailović et al. 2014), on utilise avant tout les méthodes d'apprentissage automatique. Celles-ci reposent sur un ensemble d'exemples annotés manuellement avec différentes catégories de sentiments (e. g. positif, négatif, neutre). Cet ensemble d'exemples sert de modèle pour les logiciels informatiques qui identifient les caractéristiques distinctives des différents sentiments. Dans l'étape suivante, les logiciels attribuent l'un des sentiments aux exemples non-catégorisés. Ces modèles sont très utiles sur de grands ensembles de données mais fournissent très souvent des résultats sous forme de n-grammes (Jalam et

Chauchat 2002) qui, d'ordinaire, ne permettent pas de faire des interprétations significatives.

Au niveau qualitatif de l'analyse linguistique, nous pouvons examiner des émotions beaucoup plus raffinées comme le bonheur (Stefanowitch 2004), la honte (Retzinger 1995) et même l'ironie (Haverkate 1990). L'analyse du discours permet d'observer la fonction performative du langage (Austin 1962), d'analyser les niveaux implicites de la communication (Ducrot 1972) et d'examiner les discours socialement déterminés (Butler 1997). Ces approches sont très intéressantes pour l'analyse qualitative mais ne peuvent pas être appliquées directement à l'identification des émotions sur des bases de données plus grandes.

Notre approche d'analyse des messages postés sur Twitter essaie de combiner l'approche quantitative et qualitative afin de pouvoir fonctionner sur des bases textuelles plus grandes et fournir des résultats sous forme de catégories linguistiques facilement interprétables.

### 3. Méthodologie

Nous présentons une analyse quantitative et qualitative de tweets qui reflètent les réactions face aux premières actions du mouvement des Gilets jaunes. L'objectif de l'analyse est d'examiner les caractéristiques linguistiques des messages qui expriment l'accord et le désaccord avec le mouvement des Gilets jaunes. Une fois que les messages ont été catégorisés manuellement selon le sentiment qu'ils véhiculent, ils ont été analysés aux niveaux textuel, syntaxique et lexical.

#### 3.1 *Compilation du corpus*

Les tweets analysés ont été publiés en tant que réactions à l'apparition du mouvement des Gilets jaunes. Ce mouvement de protestation non structuré s'est consolidé en octobre 2018 et avait au départ pour mission principale la diffusion d'appels à manifester contre l'augmentation du prix des carburants. Quelques semaines plus tard, le mouvement a commencé à se réaliser sous forme de blocages de routes et de ronds-points. Dans les mois qui suivaient, les actions du mouvement ont dépassé le domaine des carburants pour s'étendre aux domaines sociaux et politiques plus larges. Le canal principal de messages du mouvement étant les réseaux sociaux, plusieurs profils d'utilisateurs liés au mouvement des Gilets jaunes ont émergé. Quelques uns de ces profils, comme *@\_Gilets\_Jaunes\_*, *@GiletsJaunesGo*



Image 1. Le tweet du profil Les Gilets Jaunes

ou @GiletsToulouse, existent encore en 2020, mais certains parmi eux ne publient quasiment plus de tweets. Nous avons pris en analyse le profil *Les Gilets Jaunes @soutienj* qui a plus de 13 mille followers et qui semble communiquer avec le public le plus large (Image 1).

Nous avons pris en compte les 34 premiers messages publiés en tant que réactions au tweet du profil Les Gilets Jaunes. Ensuite, nous avons éliminé 4 messages qui étaient hors sujet ou qui faisaient partie d'un débat interne, comme démontré par l'exemple où un utilisateur pose la question sur les possibilités de joindre le mouvement à l'aide de plateformes différentes (1) :

- (1) *Comments adherer aux gilets jaunes ? Sur twitter je trouve pas, facebook je suis pas fan et google ne sort que des articles de presse...*

Le corpus, traité pour l'analyse, compte 30 tweets qui témoignent d'une grande polarisation émotionnelle des usagers.

### 3.2 Annotation de l'opinion

Nous avons manuellement attribué une opinion (positive ou négative) aux 30 tweets du corpus. Suivant la méthodologie présentée dans l'étude Zwitter Vitez (2020, sous presse), la catégorie positive a été nommée « Accord » et la catégorie négative « désaccord ». Nous avons consciemment opté pour

ces deux catégories contrastées parce que l'un des objectifs de notre étude était d'élaborer une méthodologie utilisable également pour des analyses de grandes quantités de textes. Nous nous rendons compte du fait qu'un seul annotateur ne peut pas assurer une objectivité totale de catégorisation, alors nous avons fait une attention particulière aux exemples qui n'étaient pas clairement polarisés. Dans 4 exemples (13%), nous avons eu des doutes avant de parvenir à la décision finale parce qu'ils impliquent des commentaires exprimant des sentiments différents (2) :

(2) *je soutiens les gilets jaunes ; mais PAS LES CASSEURS*

Dans ces 4 exemples, nous avons essayé de déterminer le sentiment prédominant du message. Nous savons que les décisions sur le sentiment prédominant restent subjectives et que les résultats auraient été plus fiables si plusieurs annotateurs avaient réalisé l'annotation de sentiments.

### 3.3 Analyse linguistique

Nous avons analysé chaque message aux niveaux textuel, syntaxique et lexical comme suite au travail présenté dans Zwitter Vitez (2020 sous presse). Au niveau textuel, nous avons examiné la longueur des tweets analysés (de 1 à 5 phrases). Au niveau syntaxique, nous avons d'abord examiné le type de phrase (déclarative, exclamative, interrogative et injonctive). Ensuite, nous avons déterminé la structure syntaxique de phrase (simple, complexe). Dans l'étape suivante, nous avons identifié l'éventuelle présence d'éléments lexicaux qui explicitent le sentiment de l'auteur (e. g. *heureux, imbécile* etc.). Le tableau (1) présente les annotations linguistiques aux niveaux de la structure textuelle, syntaxique et lexicale du commentaire choisi.

Tableau 1. Les annotations linguistiques du commentaire choisi.

Toute la France est derrière vous !	
Opinion	Accord
Nombre de phrases	1
Type de phrase	Exclamation
Structure	Simple
Vocabulaire	/

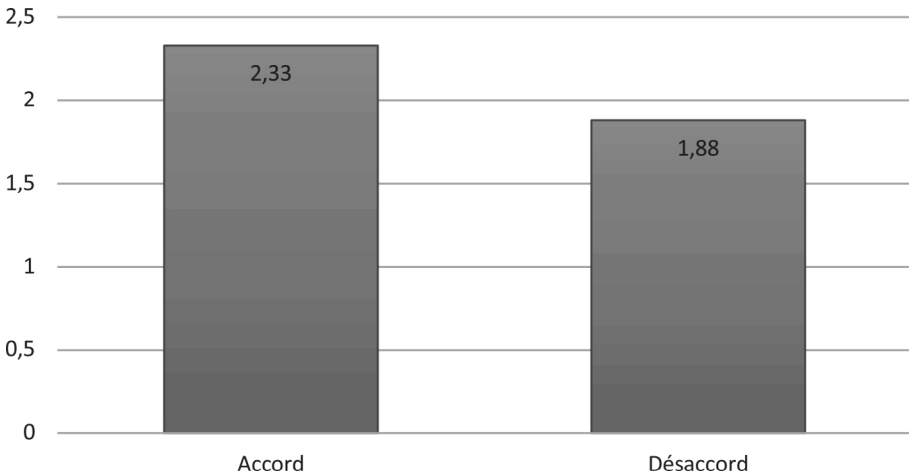
Une fois que tous les tweets ont été annotés, nous avons examiné les deux catégories de sentiments (positive et négative) à tous les niveaux d'an-

notation linguistique. L'objectif de cette analyse était de voir s'il existe, pour chaque catégorie de sentiment, des différences au niveau linguistique qui pourraient aider à distinguer les messages exprimant un sentiment positif de ceux qui véhiculent un sentiment négatif.

## 4. Analyse

### 4.1 Structure textuelle

Au niveau textuel, nous avons essayé de voir les spécificités concernant la structure des tweets qui expriment l'accord et le désaccord. Le graphique (1) montre le nombre moyen de phrases composant les tweets analysés.



Graphique 1. Le nombre moyen de phrases dans les tweets exprimant l'accord et le désaccord.

Les résultats montrent que les tweets exprimant l'accord comportent en moyenne 2,3 phrases tandis que les tweets exprimant le désaccord sont composés de 1,8 phrase en moyenne. Puisque dans les études précédentes (Zwitter Vitez, Fišer 2016, Zwitter Vitez 2020), nous n'avons pas fait l'analyse de la longueur des tweets, il est difficile de spéculer sur le caractère significatif de ce résultat. Cependant, il est intéressant de voir les exemples de tweets exprimant l'accord (3) et de tweets exprimant le désaccord (4) avec le message de l'auteur *Gilets Jaunes* :

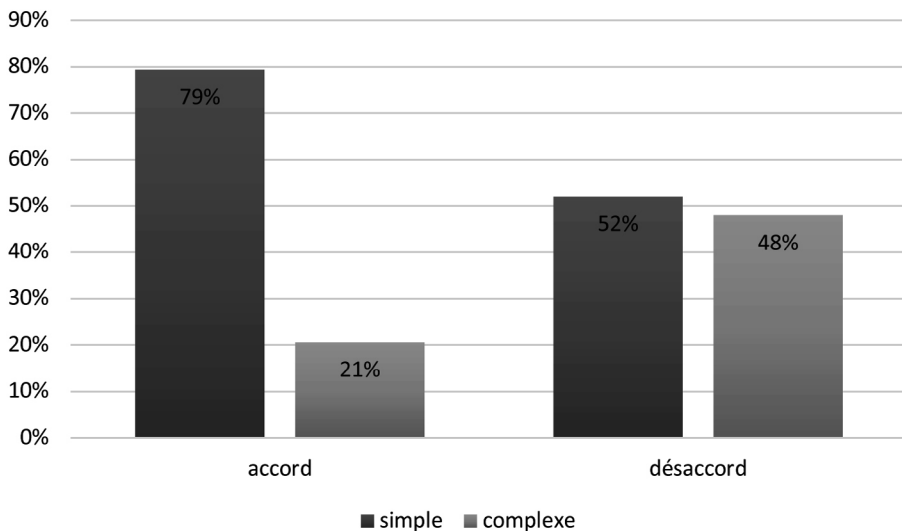
- (3) *Gilets jaunes !!! Pas de soumission, Le coup d'État! Débarrasser vous du dictateur et instauré un nouveau gouvernement .*

- (4) *86 millions d'€ pour réparer le radars que vous avez détruit, c'est ça de moins sur le budget de l'Etat et donc dans les mesures d'amélioration du pouvoir d'achat des personnes qui bossent, mais bon vous n'êtes peut-être pas concernés.*

Le premier résultat au niveau de la structure textuelle démontre que les auteurs des tweets de support au geste des Gilets jaunes semblent être composés de plus de phrases que ceux qui expriment une critique.

#### 4.2 Niveau syntaxique

Au niveau de la syntaxe, nous avons d'abord déterminé la structure de phrase utilisée (déclarative, exclamative, interrogative et injonctive) pour chaque tweet du corpus selon la classification de Riegel et al. (1994, 101) : phrase déclarative, exclamative, interrogative et injonctive. Le graphique (2) montre la distribution des structures syntaxiques dans le corpus analysé.



Graphique 2. La structure de phrase dans les tweets exprimant l'accord et le désaccord.

Les résultats montrent que dans les tweets qui expriment l'accord, la structure de la phrase simple (79%) prédomine sur la structure de la phrase complexe (21%). Ceci est exprimé par l'exemple (5) :

- (5) *je soutiens toujours les gilets jaunes*



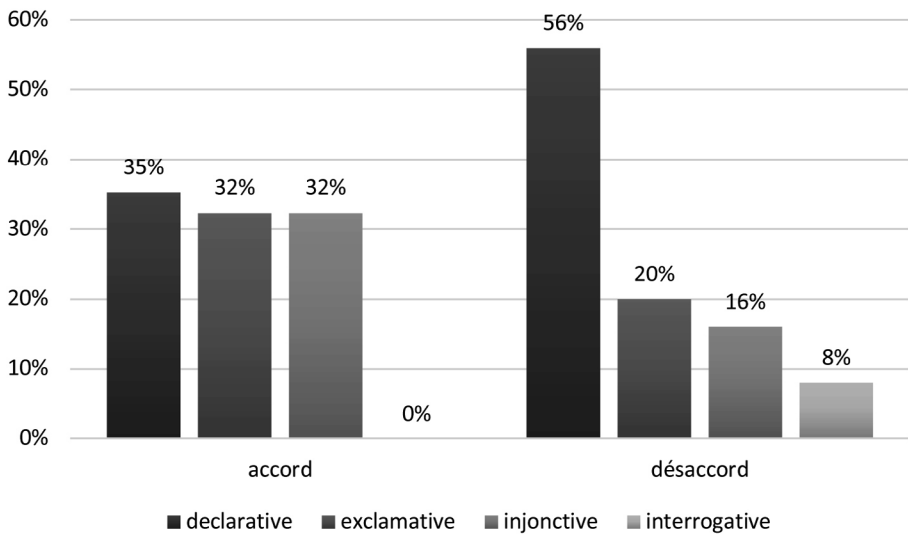
De l'autre côté, il semble que les tweets exprimant le désaccord ont une distribution égale de structure simple (52%, exemple 6) et de structure complexe (48%, exemple 7) :

(6) *Allez assumer la vie plutôt*

(7) *C sur que bloquer des gens honnêtes qui veulent aller travailler c très intelligent.*

Le premier résultat au niveau syntaxique démontre que les auteurs des tweets de support au mouvement des Gilets jaunes se contentent de présenter simplement et brièvement leur opinion tandis que ceux qui expriment une critique ressentent le besoin d'élaborer et d'argumenter davantage leur opinion.

La deuxième étape de l'analyse syntaxique est centrée sur la distribution des types de phrases auxquels appartiennent les tweets du corpus (Graphique 3).



Graphique 3. La distribution de types de phrases dans les tweets exprimant l'accord et le désaccord.

D'après le graphique (3), il semble que les messages de support au tweet des Gilets jaunes sont présentés dans des proportions égales par la phrase déclarative (35%), par la phrase exclamative (32%), et par la phrase injonctive (32%), ce qui est illustré par les exemples (8, 9 et 10) :

- (8) *Du fond de mon lit, je vous soutiens, de tout coeur avec vous.*
- (9) *Courage, ensemble vous vaccinerez le Capitalisme Sauvage !*
- (10) *Demandez la libération de citoyens arrêtés ce we!*

Les commentaires qui expriment le désaccord par rapport au tweet des Gilets Jaunes ont une structure déclarative prédominante (56%), comme le montre l'exemple (11):

- (11) *Tout ce que le peuple risque de récolter est l'arrivée de l'extrême droite au pouvoir et là ça sera vraiment le mépris des plus faibles.*

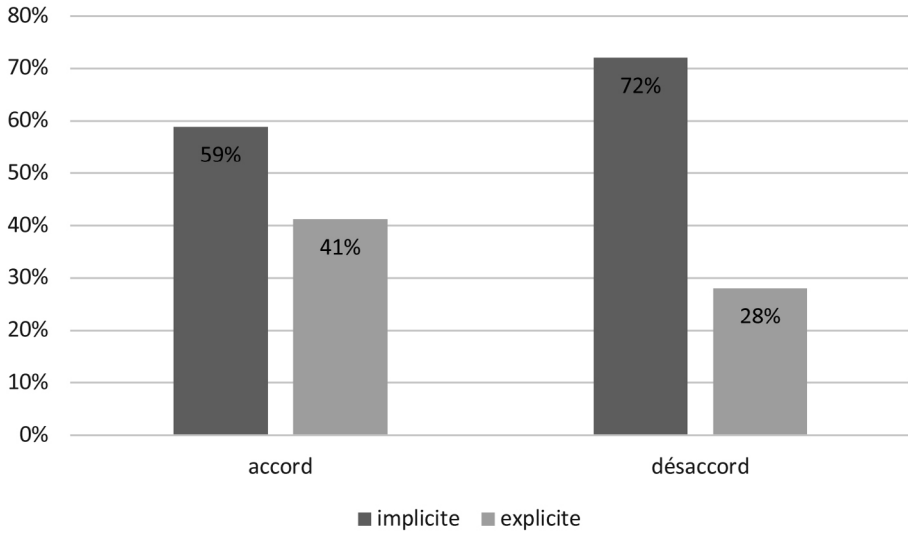
A part la prédominance de la phrase déclarative, les commentaires qui ne soutiennent pas le mouvement des Gilets jaunes révèlent une proportion importante de phrases interrogatives (22%), exclamatives (20%) et même injonctives (11%). Les exemples de structure interrogative présentent des questions cyniques qui ont la fonction de problématiser l'absurdité du mouvement que les auteurs sont en train de commenter (exemple 12), ou la fonction de démontrer le caractère agressif du mouvement (exemple 13) :

- (12) *Et mon crédit il se rembourse tout seul ?*
- (13) *Et si on le soutient pas, on peut leur dire sans se faire insulter ?*

D'après les résultats de l'analyse des types de phrases dans les tweets analysés, il semble que les commentaires de support présentent exclusivement une distribution déclarative et exclamative tandis que les messages qui expriment le désaccord, voire l'indignation pour le tweet des Gilets jaunes, ont une structure plus hétérogène comprenant les phrases déclaratives, exclamatives, interrogatives et injonctives. Comme démontré sur un corpus plus large (Zwitter Vitez 2020), l'interrogation semble être un indice sûr du désaccord.

#### 4.3 Niveau lexical

L'objectif primaire de l'analyse lexicale était d'examiner si les tweets étudiés contiennent ou non le vocabulaire explicite du sentiment ou de l'opinion de l'auteur (e. g. *merci* ou *irresponsables*). A partir de ce critère, nous avons classé les tweets en catégories de vocabulaire explicite (*soutenir*) et implicite (*Des gens taxés qui bloquent des gens taxés*). Le graphique (4) montre la distribution de tweets qui contiennent le vocabulaire explicite exprimant l'opinion de l'auteur.



Graphique 4. Le vocabulaire dans les tweets exprimant l'accord et le désaccord.

A partir du graphique 4, nous pouvons constater que dans un peu moins de la moitié de tweets (41%) qui expriment l'accord avec les protestataires, il est possible d'identifier la structure lexicale explicite qui porte l'opinion de l'auteur, ce que montre l'exemple (14). Les exemples typiques de structures exprimant l'accord sont *merci, courage, soutenir*.

(14) *je soutiens toujours les gilets jaunes*

Par contre, dans plus de la moitié de tweets exprimant l'accord avec les Gilets jaunes, le sentiment de l'auteur n'est pas évident au niveau lexical, ce que montre l'exemple (15)

(15) *De + en + de français sont avec vous !*

Dans les commentaires qui expriment le désaccord, il était encore plus difficile d'identifier la structure lexicale qui véhicule explicitement le sentiment parce que ces messages semblent assez neutres au niveau du vocabulaire (exemple 16). Parmi les structures lexicales qu'on a pu identifier (28%), on trouve par exemple les mots *agresser, irresponsables et polluer* (exemple 17) :

(16) *Pendant ce temps des enfants souffrent, meurent et des personnes rament tous les jours car eux ne sont pas assistés.*

- (17) *Si des gens se font agressés et tués par votre mouvement car ils utilisent leur voiture, c'est certain que ce sera mémorable*

Les résultats de l'analyse quantitative montrent que déjà dans les commentaires positifs et négatifs, il n'est pas toujours facile d'identifier le porteur lexical du sentiment de l'auteur, mais que dans les commentaires négatifs, c'est encore moins évident. Parmi les exemples de vocabulaire explicite, nous avons également fait une analyse qualitative qui se réfère au champ sémantique représenté par les structures lexicales examinées. Nous avons pu constater que les structures porteuses de l'opinion positive ne sont pas étroitement liées à la thématique abordée, et appartiennent à un champ sémantique assez large (e. g. *merci, courage, soutien*). Par contre, les structures porteuses de l'opinion négative appartiennent à un champ sémantique plus étroit qui indique tout de suite la thématique abordée (e. g. *diesel, polluer, irresponsables*). Nous avons également remarqué que parmi les commentaires négatifs, il y a un nombre important d'exemples ironiques où l'utilisateur se sert du vocabulaire positif afin de communiquer son sentiment négatif (exemple 18)

- (18) *Quelle belle et grande cause !!! Je pense que vous ne mesurez pas vos propos. Marquer l'histoire pour du diesel ! Mais quel ridicule combat. Pensez plutôt à notre belle planète et votre combat sera alors ambitieux.*

## 5. Potentiel pour une analyse linguistique systématique de la dynamique sociale

La partie analytique a démontré qu'il existe des différences d'ordre linguistique (structurel) entre les messages qui portent une opinion positive et ceux qui représentent une opinion négative de l'auteur. À l'époque où les méthodes d'analyse d'opinion sont utilisées principalement par de grandes entreprises et corporations, il semble pertinent de faire une réflexion sur leur utilisation dans le domaine des sciences humaines et sociales. Le potentiel des méthodes linguistiques est important surtout dans la perspective d'une meilleure compréhension de la société, de l'analyse des stéréotypes sociodémographiques et de la réponse actuelle au dynamisme social.

Afin de permettre une analyse systématique de la dynamique sociale à l'aide de méthodes linguistiques et computationnelles, nous proposons une recherche qui pourrait fournir des réponses à la question de l'opinion de différents profils d'auteurs concernant les diverses problématiques de la

société. Dans une première étape, il faudrait compiler des corpus de textes sur l'actualité sociale (e. g. la pandémie du coronavirus, la problématique des migrations ou celle de la législation familiale). Ensuite, on pourrait annoter ces corpus par des modèles computationnels et analyser comment les différents auteurs s'expriment par rapport à la problématique sociale. Cette technique de l'analyse de l'opinion pourrait être complétée par l'analyse automatique de différents profils d'auteurs, appelée le profilage d'auteur (Argamon et al. 2009, Prabhakara et al. 2014, Basile et al. 2017). Afin de réaliser le profilage d'auteur, il faudrait tout d'abord utiliser les corpus existants qui comportent les informations sur les différents profils d'auteurs de textes. Ce sont, pour le slovène, le corpus Janes (Fišer et al. 2016), le corpus OPA (Zwitter Vitez 2013) et, pour les textes parlés, le corpus Gos (Verdonik et al. 2013). Les corpus mentionnés comportent déjà des informations sur le genre, l'âge, la région et le niveau de formation des auteurs et sont utiles pour l'entraînement de modèles computationnels de classification. Ensuite, on pourrait appliquer les modèles computationnels de profilage d'auteurs sur les corpus de textes actuels.

Une fois que les techniques de l'analyse de l'opinion et du profilage d'auteur seraient combinées, on pourrait voir comment les différents profils d'auteurs s'expriment par rapport aux différents événements de la dynamique sociale. Cela pourrait fournir des réponses relatives aux réactions des hommes ou des femmes sur la problématique de la pandémie du coronavirus, leur rapport envers la problématique des migrations ou l'opinion des auteurs jeunes ou plus âgés à propos de la législation familiale etc.

## Discussion

L'objectif de la présente analyse était de dégager la structure linguistique des tweets positifs et négatifs par lesquels les utilisateurs réagissent au changement de la dynamique sociale. Nous avons analysé 30 tweets qui ont été publiés en tant que réactions au développement du mouvement des Gilets jaunes. L'analyse porte sur les niveaux textuel, syntaxique et lexical.

Les résultats montrent que les tweets positifs sont moins nombreux que les tweets négatifs. Les tweets de support des Gilets jaunes ont tendance à avoir une structure de phrases exclamatives simples, un vocabulaire de soutien explicite. De l'autre côté, les tweets qui révèlent les émotions négatives envers le mouvement des Gilets jaunes se caractérisent par une structure syntaxique complexe, des phrases interrogatives et un vocabulaire plutôt neutre.

S'il est vrai que l'échantillon analysé est petit et limité à un seul sujet, les résultats sont assez cohérents tout au long de l'analyse. Comme nous l'avons souligné dans Zwitter Vitez (2020 sous presse), une explication plausible pour ces résultats pourrait être liée à la fonction discursive des messages positifs et négatifs. Les auteurs de messages positifs n'hésitent pas à montrer leur affectation, qui révèle parfois leur appartenance à une certaine classe sociale ou régionale, alors que les auteurs de commentaires négatifs s'identifient plutôt avec une argumentation plus élaborée et non-affectée. Il nous semble également qu'un individu critique a tendance à éviter tout élément qui indiquerait son appartenance à une catégorie sociale ou régionale particulière.

Outre les résultats internes de cette analyse, notre objectif était de montrer que la linguistique peut contribuer une connaissance importante dans l'analyse de la dynamique sociale. Avec une analyse détaillée des messages publiés face à l'établissement du mouvement des Gilets jaunes, nous avons ouvert des pistes pour d'autres interprétations et démontré un nouveau potentiel de la coopération de la linguistique et du traitement automatique des langues. Dans les recherches à suivre, nous voudrions poursuivre le processus de l'analyse de l'opinion en combinaison avec les méthodes de profilage d'auteur et examiner de cette manière plusieurs thématiques de l'actualité européenne et mondiale.

## Références

### *Sources publiées*

Wright, A. «Mining the Web for Feelings, Not Facts.» *New York Times*, 23 août 2009. <https://www.nytimes.com/2009/08/24/technology/internet/24emotion.html> [consulté le 20 février 2020]

### *Bibliographie*

Austin, J. L. 1962. *How to do things with words: The William James Lectures delivered at Harvard University in 1955*. Oxford: Clarendon Press.

Basile, A., G. Dwyer, M. Medvedeva, J. Rawee, H. Haagsma, et M. Nissim. 2017. «N-GrAM: New Groningen Author-profiling Model.» Dans *Working Notes of CLEF 2017*, eds. L. Cappellato, N. Ferro, L. Goeriot, et T. Mandl, 1–11. Dublin: PAN Lab on Digital Text Forensics.

Butler, J. 1997. *The psychic life of power: Theories in subjection*. Stanford : Stanford University press.

- Cervulle, M., et F. Pailler. 2014. «#mariagepourtous: Twitter et la politique affective des hashtag.» *Revue française des sciences de l'information et de la communication* 4: 1–10.
- Ducrot, O. 1972. *Dire et ne pas dire. Principes de sémantique linguistique*, Paris: Herman et Cie.
- Farzindar, A., et M. Roche. 2013. «Les défis du traitement automatique du langage pour l'analyse des réseaux sociaux.» *Revue TAL – Traitement Automatique des langues* 54 (3): 7–16.
- Hagège, C. 1986. *L'Homme de paroles, contribution linguistique aux sciences humaines*. Paris: Gallimard.
- Haverkate, H. 1990. «A speech act analysis of irony.» *Journal of Pragmatics* 14 (1): 77–109.
- Jalam, R., et J.-H. Chauchat. 2002. «Pourquoi les n-grammes permettent de classer des textes? Recherche de mots-clefs pertinents à l'aide des n-grammes caractéristiques.» Dans *JADT Journées internationales d'Analyse statistique des Données Textuelles*, eds. A. Morin, P. Sébillot: 1–10. Rennes: INRIA.
- Longhi, J. 2013. «Essai de caractérisation du tweet politique.» *L'information grammaticale* 136 (1): 25–32.
- O'Connor, B., R. Balasubramanyan, B. R. Routledge, et N. A. Smith. 2010. «From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series.» *Proceedings of the International AAAI Conference on Weblogs and Social Media* 11 (1): 1–2.
- Pang, B. Lee, L. 2008. «Opinion Mining and Sentiment Analysis.» *Foundations and Trends® in Information Retrieval* 1–2: 1–135.
- Panckhurst, R. 2006. «Le discours électronique médié: bilan et perspectives.» *Lire, Écrire, Communiquer et Apprendre avec Internet* 1: 345–366.
- Paveau, M.-A. 2013. «Technodiscursivités natives sur Twitter. Une écologie du discours numérique.» *Épistémé (Revue internationale de sciences humaines et sociales appliquées)* 9: 139–176.
- Prabhakaran, V., E. E. Reid, et O. Rambow. 2014. «Gender and power: How gender and gender environment affect manifestations of power.» Dans *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* eds. A. Moschitti, B. Pang, et W. Daelemans, 1965–1976. Doha: Association for Computational Linguistics.

- Rao, T., et S. Srivastava. 2014. «Twitter sentiment analysis: How to hedge your bets in the stock markets.» Dans *State of the Art Applications of Social Network Analysis*, éd. F. Can, T. Ozyer et F. Polat, 227–247. Bâle: Springer.
- Retzinger, S. M. 1995. «Identifying Shame and anger in Discourse.» *American behavioral scientist* 38 (8): 1104–1113.
- Riegel, M., J.-C. Pellat, et R. Rioul. 1994. *Grammaire méthodique du français*. Paris: Quadrige.
- Smailović, J., M. Grčar, N. Lavrač, et M. Žnidaršič. 2014. «Stream-based active learning for sentiment analysis in the financial domain.» *Information Sciences* 285: 181–203.
- Smailović, J., J. Kranjc, M. Grčar, M. Žnidaršič, et I. Mozetič. 2015. «Monitoring the Twitter sentiment during the Bulgarian elections.» Dans *Proceedings of the IEEE International Conference on Data Science and Advanced Analytics*, éd. E. Gaussier, 1–10. Paris: Piscataway.
- Speriosu, M., N. Sudan, et S. Upadhyay. 2011. «Twitter polarity classification with label propagation over lexical links and the follower graph.» Dans *EMNLP ,11: Proceedings of the First Workshop on Unsupervised Learning in NLP*, éd. O. Abend, 53–63. Edinburgh: Association for Computational Linguistics.
- Stefanowitch, A. 2004. «Happiness in English and German: A metaphorical-pattern analysis.» Dans *Language, culture and mind*, éd. M. Achard, S. Kemmer, 137–149. Stanford: CSLI Publications.
- Tedeschi, A., et F. Benedetto. 2015. «A cloud-based big data sentiment analysis application for enterprises' brand monitoring in social media streams.» *Research and Technologies for Society and Industry Leveraging a better tomorrow*: 186–191. New Jersey: Institute of Electrical and Electronics Engineers.
- Verdonik, D., I. Kosem, A. Zwitter Vitez, S. Krek, et M. Stabej. 2013. «Compilation, transcription and usage of a reference speech corpus : the case of the Slovene corpus GOS.» *Language resources and evaluation* 47 (4): 1031–1048.
- Vidak, M., et A. Ackiewicz. 2016. «Les outils multimodaux de Twitter comme moyens d'expression des émotions et des prises de position.» *Cahiers de praxématique* 66: 1–18.
- Zwitter Vitez, A., et D. Fišer. 2016. «Linguistic Analysis of Emotions in Online News Comments - an Example of the Eurovision Song Contest.» Dans *Proceedings of the 4th Conference on CMC and Social Media Corpora*



for the Humanities, eds. M. Beiswenger, et D. Fišer, 74–76. Ljubljana: Filozofska fakulteta.

Zwitter Vitez, A. 2013. «Le décryptage de l'auteur anonyme: l'affaire des électeurs en survêtements.» *Linguistica* 53 (1): 91–101.

Zwitter Vitez, A., et S. Pollak. 2019. «Profiliranje avtorjev besedil: viri, metode in uporaba na področju humanistike in družboslovja.» Dans *Znanost in družbe prihodnosti*, éd. A. Kanjuo Mrčela, 168–172. Ljubljana: Fakulteta za družbene vede.

Zwitter Vitez, A. 2020. «Le discours politique et l'expression de l'opinion sur Twitter: analyse syntaxique, lexicale et orthographique.» *Ars et humanitas* (sous presse). Ljubljana: Filozofska fakulteta.

### *Sources électroniques*

Argamon, S., J. B. Goulain, R. Horton, et M. Olsen. 2009. «Vive la différence! Text mining gender difference in French literature.» *Digital Humanities Quarterly* 3 (2). <http://www.digitalhumanities.org/dhq/vol/3/2/000042/000042.html>

Fišer, D., T. Erjavec, et N. Ljubešič. 2016. «JANES vo.4. 2016. Korpus slovenskih spletnih uporabniških vsebin.» *Slovenščina 2.0* 4 (2): 67–100. <http://nl.ijs.si/janes/wp-content/uploads/2014/08/fiserothers16.pdf>

Les Gilets Jaunes @soutienjg. Twitter. Réactions au message publié le 22 novembre 2018.

<https://twitter.com/soutienjg/status/1065565949053595648> [consulté le 20 janvier 2020].

### **Povzetek**

*Vloga jezikoslovja pri razumevanju družbene dinamike: primer francoskega gibanja rumenih jopičev*

V prispevku analiziramo jezikovno zgradbo sporočil, s katerimi so se uporabniki družbenega omrežja Twitter odzvali na gibanje rumenih jopičev, ki simbolizira upor proti vladi Emmanuela Macrona. Analiza zajema tri jezikovne ravni: na besedilni ravni smo preučevali dolžino objav, na skladenski ravni sta nas zanimali zgradba in naklon povedi, na leksikalni ravni pa smo poskušali ugotoviti, ali je avtorjevo mnenje izraženo z eksplicitnim besediščem ali ne. Rezultati kažejo, da so za tvite, ki podpirajo gibanje rumenih jopičev, značilni enostavna zgradba povedi, vzklični naklon in besedišče, ki eksplicitno izraža podporo. Pri tvitih, ki ne podpirajo giba-

nja rumenih jopičev, pa je opaziti večstavčno skladenjsko zgradbo, pogosto rabo vprašalnega naklona in nevtralno besedišče. Pričujoča razprava izpostavlja relevantnost jezikoslovnih metod, ki lahko prispevajo k preseganju družbenih stereotipov in omogočijo globlje razumevanje aktualnega družbenega dogajanja.

## Abstract

*The role of linguistics in understanding social dynamics:  
the example of the Yellow vests movement in France*

This article aims to identify the linguistic structure of messages written by French Twitter users reacting to the establishment of the Yellow vests movement protesting the government of Emmanuel Macron. The analysis includes three levels: textual level analysing the length of the posts, syntactic level inspecting the sentence structure and complexity, and lexical level observing whether the users include explicit vocabulary revealing their opinion or not. The results show that the tweets supporting the Yellow vests movement are more likely to perform simple sentence structure, to use more exclamative structures and a more explicit vocabulary. On the other hand, the tweets that reveal negative sentiments towards the Yellow vests movement are characterized by a more complex syntactic structure, frequent use of interrogative sentences and a rather neutral vocabulary. This contribution underlines the potential of linguistic methods at overcoming social stereotypes and understanding social dynamism.