

Ana Zwitter Vitez

## **Le discours politique et l'expression de l'opinion sur Twitter : analyse syntaxique, lexicale et orthographique**

### 1 Introduction

À l'époque des nouveaux médias, les utilisateurs de forums, de réseaux sociaux et de portails d'actualité ont la possibilité d'exprimer publiquement leur opinion sur le dynamisme politique, social ou personnel. Le discours politique, vivement présent sur les réseaux sociaux, suscite des réactions presque immédiates des récepteurs. Les chercheurs de diverses disciplines, dont la linguistique, l'informatique, la sociologie, essaient d'analyser ces réactions afin de saisir des informations sur l'état émotionnel des auteurs des textes examinés (Pang et Lee, 2008, 7 ; O'Dea et al., 2017, 319).

L'analyse de l'opinion et l'identification des émotions dans le langage représentent des domaines de recherche émergents parce qu'il existe un lien étroit entre l'excitation physiologique d'une émotion et son affichage social (Mygovych, 2013, 206). Si nous comprenons ce que ressentent les gens, nous pouvons analyser ou même prédire comment ils réagiront dans certaines situations. C'est pourquoi l'analyse de sentiment peut être utilisée pour examiner l'opinion publique et la dynamique sociale (Speriosu et al., 2011, 53), prédire les résultats des élections (O'Connor et al., 2010, 1-2 ; Smailović et al., 2015), mesurer la satisfaction des clients (Tedeschi et Benedetto, 2015, 186) et suivre la dynamique des marchés boursiers (Rao et Srivastava, 2014, 227 ; Smailović et al., 2014, 2).

Dans l'exemple des réseaux sociaux numériques (RSN), l'analyse de l'opinion est pertinente parce qu'il s'agit d'un environnement qui fournit à leurs usagers de formuler leurs propres opinions et de trouver leur propre identité indépendamment du contenu médiatique institutionnel. Ainsi, les RSN sont devenus une plateforme à travers laquelle les usagers expriment leur opinion sur les différents produits, sur leur vie quotidienne et sur la dynamique sociale. Wright (2009) constate dans *New York Times* 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 peurs (Zwitter Vitez et Pollak, 2019, 168).

Les débats autour des gestes politiques représentent un matériel d'analyse particulièrement attachant parce qu'ils évoquent généralement des opinions assez polarisées : les usagers soutiennent ou détestent fortement les messages présentés par les hommes



politiques. Très souvent, les commentaires postés en tant que réactions à un geste politique dépassent la portée de ce geste isolé et se réfèrent à une problématique politique et sociétale plus large (Zwitter Vitez et Fišer, 2016, 76).

L'objectif de la présente analyse est de dégager la structure linguistique des tweets, exprimant l'opinion positive et négative, par lesquels les utilisateurs réagissent au discours politique et qui révèlent des émotions très polarisées. Ont été analysés 270 tweets publiés en tant que réactions au tweet du président français Emmanuel Macron déclarant avoir régularisé la situation d'un migrant qui avait sauvé la vie d'un enfant. L'analyse porte sur les niveaux de la syntaxe, du vocabulaire et de l'orthographe.

## 2 Analyse des émotions sur les réseaux sociaux numériques

Sur le plan théorique, l'analyse de l'opinion sur les réseaux sociaux numériques couvre trois domaines de recherche : l'analyse du langage des RSN, l'analyse des sentiments et l'analyse linguistique de l'expression de l'opinion. Chacun de ces trois domaines a développé ces propres approches d'analyse.

Le discours sur Twitter fait partie du discours électronique médié (Panckhurst, 2006, 345). Il est doté de certaines spécificités lexicales, graphiques et syntaxiques, comme les émoticônes, abréviations et répétitions de caractères (Farzindar et al., 2013, 7) et connaît un certain nombre d'ajustements à cause de la limitation de la quantité de caractères dans un message (140 caractères maximum, espaces comprises). Même si le discours sur Twitter est généralement perçu comme spontané et créatif (Vidak et Ackiewicz, 2016, 2), Longhi (2013, 25) constate que « cette spontanéité n'est pas forcément maintenue dans le cadre du discours politique » à cause de l'impact énorme des tweets des politiques auprès des usagers du réseau. Parmi les spécificités du discours sur Twitter, il semble pertinent de souligner également sa dimension technolangagière (Paveau, 2013, 139) représentée par le mot-dièse (hashtag) qui apporte une dimension affective aux messages (Cervulle et Paillet, 2014) et, très souvent, modifie l'aspect pragmatique des messages (Longhi, 2013, 26).

Dans le cadre de l'analyse de sentiment qui fait partie du traitement automatique du langage (Pang et Lee, 2008 ; Smailović et al., 2013, 181), on utilise surtout 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 des résultats très souvent sous forme de n-grammes (Jalam et Chauchat, 2002, 1) qui, le plus souvent, 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, 1975), 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 les approches 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 la dynamique politique en France en mai 2018. L'objectif de l'analyse était d'examiner les caractéristiques linguistiques des messages qui expriment l'opinion positive et négative. Une fois que les messages ont été catégorisés manuellement selon le sentiment qu'ils véhiculent, ils ont été analysés aux niveaux syntaxique, lexical et orthographique.

#### 3.1 Compilation du corpus

Les tweets analysés ont été provoqués par un message publié par le profil du Président français Emmanuel Macron. Il a déclaré le 28 mai 2018 avoir régularisé la situation d'un migrant qui a sauvé la vie d'un enfant (Image 1).



Image 1 : Le tweet d'Emmanuel Macron sur la naturalisation de M. Gassama.  
Source : Twitter (<https://twitter.com/emmanuelmacron/status/1001028433437184000>)

Nous avons pris en compte les 300 premiers messages publiés en tant que réactions au tweet d'Emmanuel Macron. Ensuite, nous avons éliminé 30 messages qui étaient hors sujet ou qui faisaient partie d'un débat interne, comme démontré par l'exemple (1) :

- (1) *Les racistes Français que des chiens et et des embecile et surtout des batard même pas capable de reconnaître*

Le corpus, traité pour l'analyse, compte 270 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 270 tweets du corpus. 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 tout à fait polarisés. Dans quatre exemples (1,5%), nous avons hésité quant à la décision finale parce que ces commentaires impliquent des commentaires exprimant des sentiments différents (2) :

- (2) *Très encourageant présid@, car nos chef ne prennent jamais nos valeurs en compte!*

Dans ces quatre exemples, nous avons essayé de déterminer le sentiment prédominant du message.

### 3.3 Analyse linguistique

Nous avons analysé chaque message aux niveaux de la syntaxe, du vocabulaire et de l'orthographe comme suite au travail présenté dans Zwitter Vitez et Fišer (2016, 74). 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 la présence éventuelle d'éléments lexicaux exprimant explicitement le sentiment de l'auteur (*heureux, imbécile, etc.*). Au niveau de l'orthographe, nous avons déterminé si le commentaire analysé a été rédigé dans une langue qui obéit à la norme standardisée de l'orthographe française

---

1 Les messages analysés sont cités sous leur forme authentique.

ou non (Kannas, 2012). Le tableau (1) présente les annotations linguistiques aux niveaux syntaxique, lexical et orthographique du commentaire choisi.

Tableau 1 : Les annotations linguistiques du commentaire choisi.

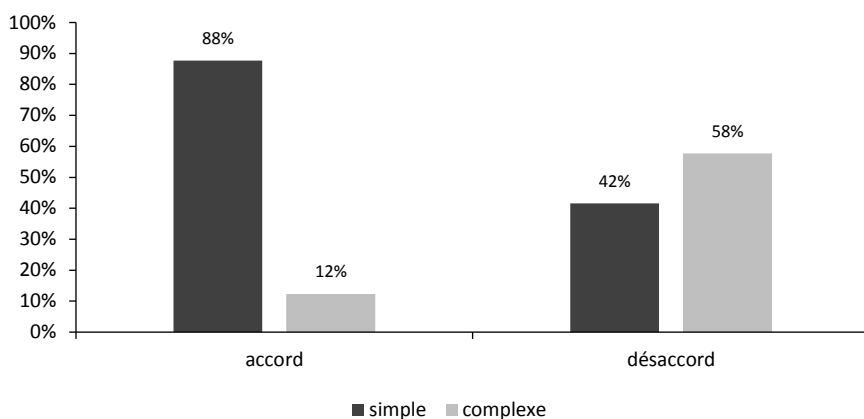
<i>Je suis tres heureuse de cette decision !</i>	
Opinion	Accord
Type de phrase	Exclamation
Structure	Simple
Vocabulaire	Heureuse
Orthographe	Standard

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'annotation 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 qui expriment un sentiment positif de ceux qui véhiculent un sentiment négatif.

## 4 Analyse et résultats

### 4.1 Syntaxe

Au niveau de la syntaxe, nous avons d'abord déterminé la structure de phrase utilisée pour chaque tweet du corpus selon la classification de Riegel et al. (1994, 101) : phrase déclarative, exclamative, interrogative et injonctive. Le graphique (1) montre la distribution des structures syntaxiques dans le corpus analysé.



Graphique 1 : 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 (88%) prédomine sur la structure de la phrase complexe (12%). Ceci est exprimé par l'exemple (3) :

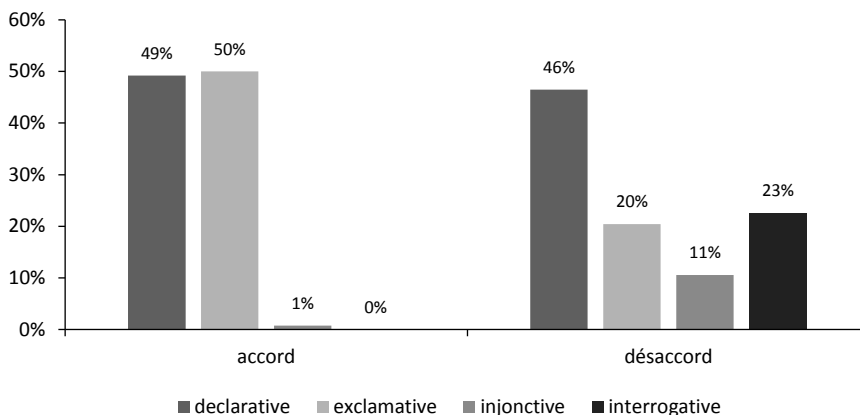
(3) *Belle décision M. Le président.*

La structure prédominante des tweets de désaccord est celle d'une phrase complexe (59%), même si les exemples de phrase simple ne sont pas rares (41%), ce qui est illustré par l'exemple (4) :

(4) *Et oui il vaut mieux honorer un jeune migrant qu'honorer ces vieux retraités qui ont donné leur vie de labeur à la France ...*

Le premier résultat au niveau syntaxique démontre que les auteurs des tweets de support au geste de M. Macron 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 observe la distribution des types de phrases auxquels appartiennent les tweets du corpus (Graphique 2).



Graphique 2 : La distribution des types de phrases dans les tweets exprimant l'accord et le désaccord.

D'après le graphique (2), il semble que les messages de support au tweet d'E. Macron sont présentés exclusivement dans une phrase déclarative (49%) et exclamative (50%) ce qui est illustré par l'exemple (5)

(5) *Bravo et merci a cet homme !!*

Les commentaires qui expriment le désaccord par rapport au tweet du Président ont une structure déclarative prédominante (47%), comme le montre l'exemple (6) :

(6) *La naturalisation ne releve pas des compétences du président de la république.*

La distribution des commentaires qui ne supportent pas la décision du Président est plus uniformément répartie parce qu'elle connaît également une proportion importante de phrases interrogatives (22%), exclamatives (20%) et même injonctives (11%). Les exemples de structure interrogative présentent souvent des questions cyniques qui ont la fonction de dépeindre l'absurdité du geste politique que les auteurs sont en train de commenter (exemple 7), ou bien ils proposent une autre solution qui semble plus raisonnable (exemple 8) :

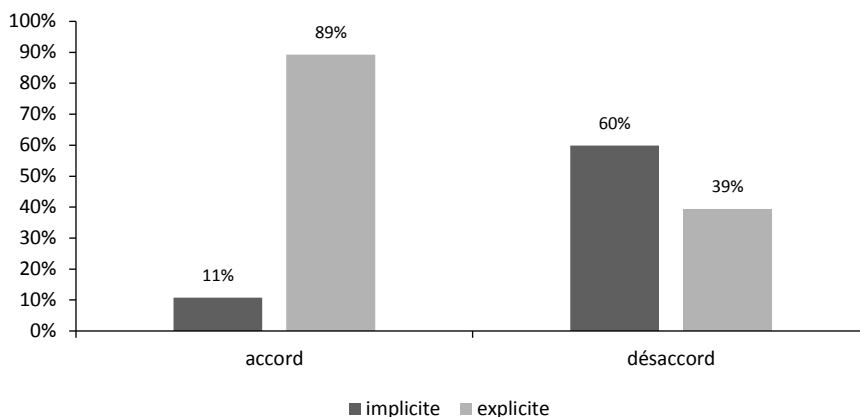
(7) *Du coup on rajoute une case « nombre de bébés sauvés » dans les formulaires de demande de naturalisation?*

(8) *Mais pourquoi n'avez-vous jamais reçu le jeune lyonnais #Marin qui s'est fait fracassé et qui en gardera des séquelles à vie pour avoir sauvé un couple pris à partie par une racaille parce qu'ils s'embrassaient ?*

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 de M. Macron, ont une structure plus hétérogène comprenant des phrases déclaratives, exclamatives, interrogatives et injonctives. L'interrogation semble être un indice sûr de désaccord.

## 4.2 Vocabulaire

L'objectif primaire de l'analyse lexicale était d'examiner si les tweets étudiés contiennent ou non un vocabulaire relevant explicitement du sentiment ou de l'opinion de l'auteur (*félicitations* ou *terrorisme*, par exemple). A partir de ce critère, nous avons classé les tweets en catégories de vocabulaire explicite (*bravo*) et implicite (*d'autres leaders devraient en prendre exemple*).



Graphique 3 : Le vocabulaire dans les tweets exprimant l'accord et le désaccord.

À partir du graphique 3, nous pouvons constater que dans la plupart de tweets (89%) qui expriment l'accord avec le geste présidentiel, il est possible d'identifier la structure lexicale explicite qui porte l'opinion de l'auteur, ce que montre l'exemple (9). Les exemples typiques de structures exprimant l'accord sont *merci, bon/bien, félicitations, heureux, récompense*.

(9) *Bravo Monsieur Président. Je suis très heureuse de cette décision !*

Dans les commentaires qui expriment le désaccord, il était beaucoup plus difficile d'identifier la structure lexicale qui véhicule explicitement le sentiment parce que ces messages semblent être assez neutres au niveau du vocabulaire (exemple 10). Parmi les structures lexicales qu'on a pu identifier (39%), on trouve, par exemple, les mots *mépris, écoeurant, clandestins* et *mascarade*.

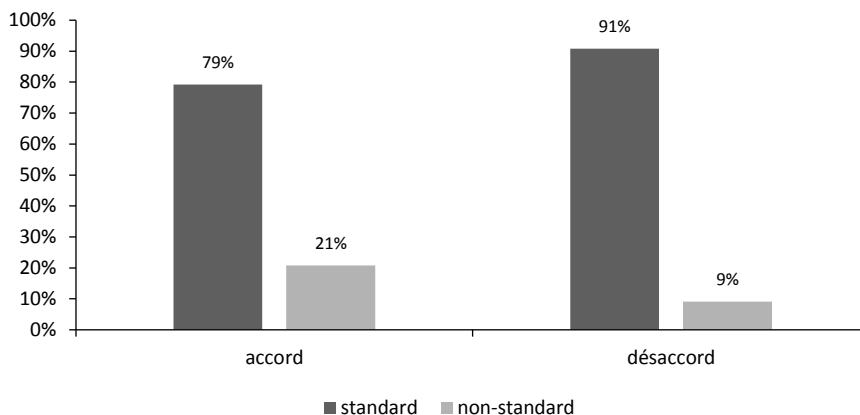
(10) *Consultez un peu le peuple Mr Macron consultez le!!*

Les résultats de l'analyse quantitative montrent que, dans les commentaires positifs, il est plus facile d'identifier le porteur lexical du sentiment de l'auteur. 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 (par exemple, *bravo, magnifique, bien fait*). Par contre, les structures porteuses de l'opinion négative appartiennent à un champ sémantique plus étroit qui indique tout de suite la thématique abordée (*raciste, clandestin, terrorisme, etc.*).



### 4.3 Orthographe

Le dernier niveau d'analyse porte sur le niveau de l'orthographe. Nous avons examiné si les tweets exprimant l'accord ou le désaccord ont été rédigés en utilisant une orthographe standard ou non (Graphique 4).



Graphique 4 : L'orthographe des tweets exprimant l'accord et le désaccord.

Nous pouvons voir sur le graphique (4) que la majorité des tweets exprimant l'accord et le désaccord sont rédigés en français standard. Pourtant, parmi les tweets exprimant l'accord, 20% ne se conforment pas à la norme standard (exemple 11) tandis que les tweets qui véhiculent le désaccord obéissent dans la plupart des cas (91%) à la norme standard (exemple 12) :

(11) *emmanuel macron cest tres tres gentille brvo*♡♡♡

(12) *Il faut être en situation régulière et en France depuis 5 ans pour pouvoir demander la naturalisation. Art. 21-17 du code civil.*

Au niveau de l'orthographe, nous pouvons donc voir que même si la majorité des tweets sont rédigés en langue standard, les tweets positifs ont une proportion plus importante d'orthographe non-standard (20%), tandis que cette proportion est moins importante dans les tweets négatifs (9% seulement).<sup>2</sup> Ce résultat est inattendu parce qu'on aurait cru que les auteurs qui manifestent leur désaccord seraient plus enclins à utiliser la langue non-standard.

2 Pour déterminer si la différence entre les deux catégories est suffisamment grande pour être statistiquement significative, nous avons fait le test du chi carré. Le chi carré étant de 8,1 et la valeur du  $p$  étant de .004, ce résultat est significatif.

## 5 Discussion

L'objectif de la présente analyse était de dégager la structure linguistique des tweets positifs et négatifs par lesquels les utilisateurs réagissent au discours politique. Nous avons analysé 270 tweets qui ont été publiés en tant que réactions au tweet du président français Emmanuel Macron déclarant avoir régularisé la situation d'un migrant qui a sauvé la vie d'un enfant. L'analyse porte sur les niveaux de la syntaxe, du vocabulaire et de l'orthographe.

Les résultats montrent que les tweets positifs sont moins nombreux que les tweets négatifs. Les tweets de support au message d'E. Macron ont davantage tendance à avoir une structure de phrases exclamatives simples, un vocabulaire de support explicite et une orthographe non-conventionnelle. De l'autre côté, les tweets qui révèlent les émotions négatives 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. 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 émotion, qui révèle parfois leur appartenance à une certaine classe sociale ou régionale, alors que les auteurs des commentaires négatifs s'identifient plutôt avec une argumentation plus élaborée et sans marque d'affect. 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.

Enfin, il nous semble pertinent de discuter les points forts, les limitations et les potentiels ouverts de cette recherche. Parmi les qualités, on peut sans doute évoquer que les catégories que nous avons analysées (structure et type de phrase, lexique explicite ou neutre, orthographe standard ou non-standard) dérivent du fondement du système linguistique français. Suite à leur caractère facilement déterminable, elles pourraient être reprises par des analyses automatiques, ce qui permettrait d'analyser des bases de données considérablement plus vastes. Contrairement aux analyses computationnelles, dont les résultats ne permettent pas toujours de conclure sur des interprétations discursives, les résultats de cette analyse s'adressent davantage aux linguistes et autres chercheurs en sciences humaines et sociales.

Nous nous rendons compte du fait que la méthodologie proposée ne pourra pas englober les niveaux pragmatique et cognitif de l'analyse textuelle. Pourtant, dans les recherches à suivre, nous voudrions entreprendre une analyse plus nuancée des éléments émotifs, détectables par les possibilités de l'analyse qualitative, ce qui pourrait donner un aperçu encore plus approfondi de l'expression des sentiments. Au niveau technodiscursif englobant les structures d'enrichissement les messages de Twitter, il serait également intéressant d'étudier la fonction du lien interne, représenté par le

mot-dièse (hashtag), et les émoticônes illustrant le propos du message. Nous voudrions examiner plusieurs thématiques de l'actualité européenne et mondiale afin de voir si les tendances linguistiques que nous avons détectées par rapport aux différentes expressions du sentiment sont liées aux thématiques spécifiques ou bien éventuellement applicables plus généralement à l'analyse des pratiques discursives.

## Références

- Austin, J. L., *How to do things with words*, Oxford 1975.
- Butler, J., *The psychic life of power: Theories in subjection*, Stanford 1997.
- Cervulle, M., Pailler, F., #mariagepourtous : Twitter et la politique affective des hashtags, *Revue française des sciences de l'information et de la communication* 4, 2014, pp. 1-10.
- Ducrot, O., *Dire et ne pas dire. Principes de sémantique linguistique*, Paris 1972.
- Farzindar, A., Roche, M., Les défis du traitement automatique du langage pour l'analyse des réseaux sociaux, *Revue TAL – Traitement Automatique des langues* 54/3, 2013, pp. 7-16.
- Fišer, D., Erjavec, T., Ljubešić, N., JANES v0.4: Korpus slovenskih spletnih uporabniških vsebin, *Slovenščina 2.0* 4/2, 2016, pp. 67-100.
- Haverkate, H., A speech act analysis of irony, *Journal of Pragmatics* 14/1, 1990, pp. 77-109.
- Jalam, R., Chauchat, J.-H., Pourquoi les n-grammes permettent de classer des textes ? Recherche de mots-clefs pertinents à l'aide des n-grammes caractéristiques, *JADT Journées internationales d'Analyse statistique des Données Textuelles*, 2002, pp. 1-10.
- Kannas, C., *Bescherelle : L'orthographe pour tous*, Hatier 2012.
- Liu, B., *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*, Cambridge 2015.
- Longhi, J., Essai de caractérisation du tweet politique, *L'information grammaticale* 136/1, 2013, pp. 25-32.
- Mygovitch, I. Secondary nomination in the modern English language: affective lexical units, *Visnik LNU imeni Tarasa Ševčenka* 1/1, 2013, pp. 206-214.
- O'Connor, B., Balasubramanyan, R., Routledge, B. R., Smith, N. A., From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series, *Proceedings of the International AAAI Conference on Weblogs and Social Media* 11(122-129), 2010, pp. 1-2.
- O'Dea, B., Larsen, M. E., Batterham, P. J., Calear, A., Christensen, H., A Linguistic Analysis of Suicide-Related Twitter Posts, *Crisis* 38, 2017, pp. 319-329.
- Pang, B. Lee, L., Opinion Mining and Sentiment Analysis, *Foundations and Trends in Information Retrieval* 1-2, pp. 1-135.
- Panckhurst, R., Le discours électronique médié : bilan et perspectives, *Lire, Écrire, Communiquer et Apprendre avec Internet* 1, 2006, pp. 345-366.
- Paveau, M.-A., Technodiscursivités natives sur Twitter. Une écologie du discours numérique, *Épistémé (Revue internationale de sciences humaines et sociales appliquées)* 9, 2013, pp. 139-176.
- Rao, T., Srivastava S., Twitter sentiment analysis: How to hedge your bets in the stock markets. Dans: *State of the Art Applications of Social Network Analysis* (éds. Can, F., Ozyer, T., Polat, F.), Zurich 2014, pp. 227-247.
- Retzinger, S. M., Identifying Shame and anger in Discourse, *American behavioral scientist* 38/8, 1995, pp. 1104-1113.

- Riegel, M., Pellat, J.-C., Rioul, R., *Grammaire méthodique du français*, Quadriga 1994.
- Ritchie, G., The Linguistic Analysis of Jokes, *Journal of Literary Semantics* 33/2, 2004, pp. 196-197.
- Smailović, J., Grčar, M., Lavrač, N., Žnidaršič, M., Stream-based active learning for sentiment analysis in the financial domain, *Information Sciences* 285, 2014, pp. 181-203.
- Smailović, J., Kranjc, J., Grčar, M., Žnidaršič, M., Mozetič, I., Monitoring the Twitter sentiment during the Bulgarian elections, Dans : *Proceedings of the IEEE International Conference on Data Science and Advanced Analytics* (éd. Eric Gaussier), 2015. pp. 1-10.
- Speriosu, M., Sudan, N., Upadhyay, S., 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* (dir. Abend, O. et al.), 2011, pp. 53-63.
- Stefanowitch, A., Happiness in English and German: A metaphorical-pattern analysis. Dans : *Language, culture and mind* (dir. Achard, M., Kemmer, S.), 2004, pp. 137-149.
- Tedeschi, A., Benedetto, F., 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*, 2015, pp. 186-191.
- Retzinger, M. S., *Violent Emotions: Shame and Rage in Marital Quarrels*, Newbury Park 1998.
- Vidak, M., Ackiewicz, A., Les outils multimodaux de Twitter comme moyens d'expression des émotions et des prises de position, *Cahiers de praxématique* 66, 2016, pp. 1-18.
- Wright, A., *Mining the Web for Feelings, Not Facts*, New York Times, 23. août 2009.
- Zwitter Vitez, A., Fišer, D., 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* (éds. Beiswenger, M., Fišer, D.), 2016, pp. 74-76.
- Zwitter Vitez, A., Pollak, S. Profiliranje avtorjev besedil: viri, metode in uporaba na področju humanistike in družboslovja. Dans : *Znanost in družbe prihodnosti* (éd. Kanjuo Mrčela, A.), Ljubljana 2019, pp. 168-172.

Ana Zwitter Vitez

## **Le discours politique et l'expression de l'opinion sur Twitter : analyse syntaxique, lexicale et orthographique**

**Mots-clés** : discours politique, nouveaux médias, analyse linguistique, analyse des émotions, Twitter

À l'époque des nouveaux médias, les utilisateurs de forums, de réseaux sociaux et de portails d'actualité ont la possibilité d'exprimer publiquement leur opinion sur le dynamisme politique, social ou personnel. Le discours politique, vivement présent sur les réseaux sociaux, suscite des réactions des récepteurs presque immédiates. L'objectif de la présente analyse est de dégager la structure linguistique des tweets par lesquels les utilisateurs réagissent au discours politique et qui révèlent des opinions très polarisées. Nous avons analysé 270 tweets qui ont été publiés en tant que réactions au tweet du président français Emmanuel Macron déclarant avoir régularisé la situation d'un migrant qui a sauvé la vie d'un enfant. L'analyse porte sur les niveaux de la

syntaxe, du vocabulaire et de l'orthographe. Les résultats montrent que les tweets de support au message de Macron ont davantage tendance à avoir une structure de phrase exclamative simple, un vocabulaire de support explicite et une orthographe non-conventionnelle. De l'autre côté, les tweets qui révèlent les émotions négatives se caractérisent par une structure syntaxique complexe, des phrases interrogatives et un vocabulaire plutôt neutre. Cette recherche pourrait approfondir les connaissances existantes sur la structure linguistique de l'expression de l'opinion et compléter les méthodes computationnelles de l'analyse de sentiment.

Ana Zwitter Vitez

### **Political discourse and expressing opinion on Twitter: syntactic, lexical and orthographic analysis**

**Keywords:** political discourse, new media, linguistic analysis, sentiment analysis, Twitter

In the era of new media, users of forums, social networks and news portals can publicly express their opinions about various political, social or personal topics. Political discourse, strongly present on social networks, arouses almost immediate reactions from users. The objective of this analysis is to identify the linguistic structure of tweets by which users react to political discourse and consequently express highly polarized opinions. We analysed 270 tweets published as reactions to a tweet from the French President Emmanuel Macron claiming to have regularized the situation of a migrant who saved a child's life. The linguistic analysis includes the levels of syntax, vocabulary, and spelling. The results show that the tweets that support Macron's political message are more likely to use simple exclamatory syntactic structures, an explicitly supportive vocabulary, and non-standard spelling. On the other hand, tweets expressing a negative opinion are more likely to have a complex syntactic structure, include interrogative clauses and contain a more neutral vocabulary. This research can add to the existing knowledge on the linguistic structure of opinion expression and complement automatic methods of sentiment analysis.

Ana Zwitter Vitez

### **Politični diskurz in izražanje mnenja na Twitterju: skladenjska, leksikalna in pravopisna analiza**

**Ključne besede:** politični diskurz, novi mediji, jezikovna analiza, analiza sentimenta, Twitter

Uporabniki forumov, družbenih omrežij in novinarskih portalov vsakodnevno javno izražajo svoje mnenje o različnih političnih, družbenih ali osebnih tematikah. Pogosto se zelo burno odzovejo na politični diskurz. Cilj raziskave je prepoznati jezikovno strukturo tvitov, s katerimi uporabniki reagirajo na politični diskurz in pri tem izražajo močno polarizirana mnenja.

Analizirali smo 270 tvitov, ki se odzivajo na sporočilo francoskega predsednika Emmanuela Macrona, da je legaliziral status migranta, ki je rešil življenje otroka. Jezikoslovna analiza zajema ravni skladnje, besedišča in pravopisa. Rezultati kažejo, da tviti, ki podpirajo Macronovo politično sporočilo, pogosteje uporabljajo enostavne vzključne povedi, z besediščem eksplicitno izražajo podporo in pogosteje odstopajo od pravopisne norme. Po drugi strani imajo tviti, ki izražajo negativno mnenje, bolj kompleksno skladenjsko strukturo in pogosto vprašalni naklon ter vsebujejo bolj nevtralnno besedišče. Raziskava lahko dopolni obstoječe znanje o jezikovnih specifikah izražanja mnenja in prispeva k izboljšanju avtomatskih metod analize javnega mnenja.

## O avtorici

**Ana Zwitter Vitez** je docentka za francoski jezik na Oddelku za Uporabno jezikoslovje Fakultete za humanistične študije Univerze na Primorskem. Raziskovalno se ukvarja s področji korpusnega jezikoslovja, računalniško posredovane komunikacije in ugotavljanja avtorstva besedil, pri čemer jo posebej zanima povezava med kvantitativnimi in kvalitativnimi metodami analize. V okviru projekta *Sporazumevanje v slovenskem jeziku* je koordinirala gradnjo slovenskega govornega korpusa Gos, vodila podoktorski projekt *Jezikovnotehnoške analize za ugotavljanje osebnega profila avtorja besedila*, v okviru projekta Jezikoslovna analiza nestandardne spletne slovenščine Janes pa je vodila sklop *Govor in spletna slovenščina*.

## About the author

**Ana Zwitter Vitez** is Assistant Professor of French at the Department of Applied Linguistics, Faculty of Humanities, University of Primorska. Her research interests include corpus linguistics, computer-mediated communication and authorship attribution. She is particularly interested in the connection between quantitative and qualitative methods of analysis. Within the project "Communication in Slovene" she coordinated the compilation of the national corpus of spoken Slovene. She completed the postdoctoral project "Language technology analyses for author profiling", and conducted the work package "Speech and Online Slovene" within the Janes project.