



**HoGent**

*Geassocieerde Faculteit Toegepaste Taalkunde*

**L'analyse des sentiments appliquée sur des tweets  
politiques: une étude de corpus**

Cynthia Van Hee

Scriptie voorgedragen tot het bekomen van de graad van

Master in de meertalige communicatie

Masterproefbegeleider:  
Prof. Dr. Veronique Hoste

Academiejaar 2012-2013

## AVANT-PROPOS

Avant tout, je tiens à exprimer ma plus vive reconnaissance à ma directrice de mémoire, Madame le Professeur Véronique Hoste, pour m'avoir donné l'opportunité d'entreprendre cette étude intéressante. Je lui adresse mes remerciements particuliers pour son soutien tout au long de ce projet de recherche, pour ses conseils judicieux et pour sa grande disponibilité.

Je souhaite exprimer ma gratitude envers Monsieur Wouter Cochie pour m'avoir fourni les conseils langagiers nécessaires durant le procès d'écriture. J'exprime également ma reconnaissance à Madame Marjan Van de Kauter pour les nombreux conseils qu'elle m'a apportés tout au long de l'étape d'annotation du corpus. Je remercie Monsieur Bart Desmet pour son aide lors de la collection des données de cette étude.

Enfin, je souhaite exprimer ma plus vive gratitude à mes parents, à ma sœur et à Bert pour leur soutien sans faille tout au long de ces quatre années d'étude et en particulier lors de la réalisation de ce mémoire.

## TABLE DES MATIERES

1. INTRODUCTION .....	7
1.1. Les médias sociaux .....	7
1.2. Objectif de ce mémoire .....	9
1.3. Description des chapitres .....	10
2. LITTERATURE .....	11
2.1. L'échange d'opinions .....	11
2.2. L'analyse des sentiments .....	12
2.3. Deux approches différentes du système d'analyse automatique des sentiments .....	13
2.3.1. L'analyse automatique des sentiments basée sur un lexique .....	14
2.3.2. L'analyse automatique des sentiments basée sur des corpus .....	18
2.4. Twitter dans le domaine politique .....	24
3. METHODE .....	28
3.1. Description des corpus .....	28
3.2. Le logiciel d'annotation Brat .....	30
3.3. Le schéma d'annotation de Van de Kauter et Desmet (2012) .....	31
3.4. Procédure d'annotation .....	34
3.4.1. L'expression subjective .....	34
3.4.2. Les modificateurs .....	36
3.4.3. La source et l'expression linguistique qui la lie au sentiment exprimé .....	37
3.4.4. L'objet, la polarité et l'intensité du sentiment exprimé .....	38
3.4.5. La cause .....	39
3.4.6. Les expressions subjectives incorporées dans une autre .....	39
3.4.7. Les relations coréférentielles et les composants de l'objet .....	40
3.4.8. Les relations causales neutres .....	40
4. DESCRIPTION DE LA PREMIERE ANALYSE .....	41
5. DESCRIPTION DE LA DEUXIEME ANALYSE .....	41
6. DESCRIPTION DE LA TROISIEME ANALYSE .....	42
7. RESULTATS .....	43
7.1. Résultats de l'analyse des sentiments .....	43
7.1.1. Distribution des sentiments et le sentiment prépondérant dans tout tweet .....	44
7.1.2. Rapport entre le sexe de l'auteur et la polarité du sentiment exprimé .....	47

7.1.3.	Intensité des sentiments exprimés .....	48
7.1.4.	Explicitation des sentiments exprimés .....	50
7.1.5.	Modificateurs .....	51
7.1.6.	Sujets abordés.....	52
7.1.7.	Evaluation du schéma d'annotation .....	54
7.2.	Résultats de la mesure de fréquence des partis et personnages politiques .....	57
7.2.1.	Analyse des tweets publiés par les hommes.....	58
7.2.2.	Analyse des tweets publiés par les femmes .....	59
7.2.3.	Analyse des autres tweets dans le corpus .....	61
7.2.4.	Analyse globale du corpus .....	62
7.2.5.	Comparaison de l'analyse avec les résultats des élections de 2012 .....	64
7.3.	Résultats de l'analyse à l'aide du logiciel TExSIS .....	67
8.	CONCLUSIONS.....	71
8.1.	L'analyse des sentiments .....	71
8.2.	L'analyse des orientations politiques .....	74
8.3.	L'extraction terminologique à l'aide du logiciel TExSIS .....	75
9.	BIBLIOGRAPHIE.....	77
10.	ANNEXES .....	i
10.1.	Tableau général de l'analyse des sentiments .....	i
10.2.	Corpus de 200 tweets.....	xli
10.3.	Comparaison du résultat des élections de 2012 et de 2006 .....	lxiv
10.4.	Liste de la terminologie extraite à l'aide du logiciel TExSIS.....	lxv
10.5.	Corpus de 4 629 tweets.....	lxv
10.6.	Corpus de 15 952 tweets.....	lxv
10.7.	Fiche mémoire .....	lxv

## LISTE DES TABLEAUX ET DES FIGURES

Tableau 1 : attributs des expressions private state et des actes de parole explicites (Wiebe, 2002).....	20
Tableau 2 : attributs des expressions subjectives sans explicitation (Wiebe, 2002). ....	21
Tableau 3: attributs des actes de parole objectifs (Wiebe, 2002). ....	21
Tableau 4 : attributs des sources des expressions private state et des actes de parole (Wiebe, 2002).....	21
Tableau 5 :attributs des attitudes exprimées dans une expression private state (Wilson, 2008). .....	23
Tableau 6 : attributs de l'annotation de l'objet d'une attitude (Wilson, 2008). ....	24
Tableau 7 : comparaison entre la prédiction de Twitter et celle des sondages PB et MdH (Tjong Kim Sang et Bos, 2012). ....	27
Tableau 8 : nouvelle comparaison entre la prédiction de Twitter et celle des sondages traditionnels (Tjong Kim Sang et Bos, 2012). ....	27
Figure 1: exemple d'annotation à l'aide du logiciel Brat. ....	30
Figure 2 : distribution des 335 sentiments exprimés dans le corpus de 200 tweets. ....	45
Figure 3 : proportion des tweets dans le corpus selon le sentiment préponderant dans tout tweet. ....	45
Tableau 9 : les tweets comportant une combinaison de sentiments. ....	46
Tableau 10 : rapport entre le sexe de l'auteur du tweet et la polarité du sentiment exprimé...	47
Tableau 11 : rapport entre le sexe de l'auteur du tweet et l'intensité / la polarité du sentiment exprimé.....	48
Tableau 12 : rapport entre le sexe de l'auteur et l'intensité / la polarité du sentiment exprimé en pourcentages. ....	49
Tableau 13 : sentiments explicites et implicites dans le corpus. ....	50
Tableau 14 : modificateurs utilisés par les hommes et les femmes dans le corpus.....	51
Tableau 15 : sujets les plus abordés par les femmes et le sentiment exprimé à leur égard.....	53
Tableau 16: sujets les plus abordés par les hommes et le sentiment exprimé à leur égard.....	53
Tableau 17: total de mentions des principaux partis politiques flamands dans les tweets des hommes. ....	58
Tableau 18 : total de mentions de onze personnages politiques flamands dans les tweets des hommes. ....	58

Tableau 19 : mentions des principaux partis politiques flamands dans les tweets des femmes. .....	59
Tableau 20 : mentions de onze personnages politiques flamands dans les tweets des femmes. .....	60
Tableau 21 : mentions des principaux partis politiques flamands dans les tweets dont le sexe de l'auteur est inconnu. ....	61
Tableau 22 : mentions de onze politiques flamands dans les tweets dont le sexe de l'auteur est inconnu. ....	61
Tableau 23 : total mentions des principaux partis politiques dans le corpus. ....	62
Tableau 24 : total de mentions de onze personnages politiques dans le corpus.....	62
Tableau 25 : mentions positives, négatives et neutres dans le corpus.....	64
Tableau 26 : résultat des huit principaux partis flamands lors des élections de 2012.....	64
Tableau 28 : top cinquante des termes, classés par leur valeur termhood. ....	67
Tableau 27 : top cinquante des termes, classés par leur fréquence. ....	67

## 1. INTRODUCTION

### 1.1. Les médias sociaux

De nos jours, l'Internet se révèle plus que jamais un outil indispensable d'échange d'informations. Il nous offre une quantité considérable d'informations à une vitesse inédite et ses services s'adaptent de plus en plus aux besoins des internautes. Ceux-ci peuvent consulter l'Internet pour trouver de l'information, envoyer des e-mails, acheter des produits, lire des journaux en ligne, etc.

Pendant les dernières années, l'Internet a connu encore une plus vaste portée grâce au développement des médias sociaux. Basés sur des techniques de communication faciles et accessibles pour tous, ces médias favorisent les interactions sociales à travers l'Internet. Les médias sociaux se distinguent des médias traditionnels tels que les journaux, la télévision et la radio parce que leur utilisation est peu coûteuse et libre, de façon à permettre à tout le monde d'accéder à ou de publier de l'information. Ils subviennent aux besoins des individus d'échanger des opinions, de demander des conseils et de communiquer de façon rapide et facile. Les possibilités sont nombreuses : envoyer des messages et des photos, réaliser des vidéoconférences, télécharger des documents, etc. Offrant un accès libre et gratuit, les médias sociaux ont considérablement favorisé la communication de masse et ils ont déclenché le débat public sur Internet.

Les médias sociaux qui ont récemment pu bénéficier d'un considérable essor sont les réseaux sociaux tels que Facebook, LinkedIn et Twitter. Ce sont des sites web qui rassemblent des identités sociales telles que des individus, des entreprises et des organisations qui peuvent échanger de l'information à travers des interactions sociales. Grâce à leur caractère maniable et leur accès libre, les réseaux sociaux bénéficient d'un succès croissant auprès du grand public.

La popularité des nouveaux médias est d'autant plus grande que la demande d'informations est devenue plus importante dans notre société. En général, les gens aiment consulter les avis d'autres personnes avant de passer à l'action ou de se faire une opinion. Autrefois, ces avis

provenaient surtout de leur environnement social direct. Aujourd'hui, les gens aiment également consulter les opinions objectives de personnes qui leur sont inconnues.

Ce besoin pressant d'informations dans la société n'a pas échappé aux nombreux commerçants qui se réfèrent au Web afin de savoir ce qui est dit à l'égard de leurs produits. De la même façon, la vie politique est de plus en plus dominée par le flux informationnel sur Internet et plus concrètement sur les réseaux sociaux. Depuis quelques années, les hommes politiques étendent leurs campagnes électorales jusqu'à Facebook et Twitter. Tel est le cas de Barack Obama, l'actuel président des États-Unis. Celui-ci a, en vue des élections présidentielles de 2012, utilisé Twitter comme l'un de ses principaux outils pour faire campagne<sup>1</sup>.

Tous les récents développements dans le domaine d'échange d'informations et d'opinions ont donné un coup de volant aux applications informatiques conçues pour l'analyse et la détection de sentiments exprimés sur Internet. Présentée dans la littérature sous le nom de *opinion mining* ou *sentiment analysis*, l'analyse des sentiments s'utilise entre autres pour la détection d'opinions sur des sites web et des réseaux sociaux, l'éclaircissement sur le comportement des consommateurs, la recommandation de produits et l'explication du résultat des élections. Elle consiste à rechercher des textes évaluatifs sur Internet tels que des critiques et des recommandations et à analyser de façon automatique ou manuelle les sentiments qui y sont exprimés afin de mieux comprendre l'opinion publique.

Il a déjà été démontré par des études antérieures que l'analyse des sentiments s'avère particulièrement intéressante pour ceux qui ont intérêt à connaître l'opinion publique, que ce soit pour des raisons personnelles, commerciales ou politiques. Ainsi, de nombreux systèmes autonomes ont déjà été développés pour l'analyse automatique des sentiments. Généralement, ces systèmes étaient entraînés aux textes évaluatifs traditionnels tels que les comptes rendus cinématographiques ou les critiques d'un livre. Toutefois, depuis quelques années se sont graduellement ajoutés à ces textes traditionnels les textes non traditionnels tels que les messages envoyés via les réseaux sociaux. Ceux-ci constituent une source précieuse d'opinions échangées parmi les multiples internautes.

---

<sup>1</sup> Goret, C. (29.10.2012). De eerste sociale campagne. *Metro*, p. 6.

Par conséquent, il importe de concevoir des systèmes automatiques aptes à rechercher et à analyser les sentiments qui sont exprimés sur les réseaux sociaux. A cet effet, une grande partie de cette étude sera consacrée à l'analyse manuelle des sentiments exprimés dans des tweets et à l'examen des possibilités qu'offre une telle analyse.

## 1.2. Objectif de ce mémoire

Trois corpus comportant des tweets seront analysés lors de la présente étude. Les tweets sont des messages de 140 caractères au maximum qui sont envoyés via le réseau social Twitter<sup>2</sup>. A partir d'une base de données de tweets politiques, la présente étude poursuit un triple but. L'objectif principal est de faire, à partir d'un schéma d'annotation détaillée, une analyse manuelle des sentiments exprimés dans les tweets du premier corpus. Ensuite, les résultats de l'analyse seront utilisés pour évaluer les possibilités et les défauts du schéma d'annotation. Le deuxième objectif de cette étude est d'examiner quels partis et personnages politiques sont principalement évoqués sur Twitter en vue des élections. Les résultats de l'analyse seront comparés avec les résultats des élections communales et provinciales de 2012 pour examiner si un rapport pourrait être établi entre les paroles échangées sur Twitter et le résultat des élections. Le troisième objectif consiste à détecter les mots les plus évoqués dans les tweets pour révéler la terminologie spécifiquement utilisée dans un contexte politique.

La principale partie de cette étude comprend l'analyse du premier corpus. L'étendue du corpus est limitée à deux cents tweets afin de permettre une analyse manuelle détaillée des sentiments qui y sont exprimés. A partir de l'analyse, nous essaierons de formuler une réponse aux questions suivantes : combien de sentiments sont généralement exprimés par tweet ? Domine-t-il, en général, un sentiment négatif ou positif dans les tweets ? Existe-t-il un rapport entre le sexe de l'auteur du tweet et la polarité / l'intensité du sentiment exprimé ? Des femmes expriment-elles leurs opinions d'une autre façon que les hommes ? Ensuite, il sera faite une évaluation des possibilités qu'offre le schéma d'annotation utilisé et de ses éventuels défauts.

---

<sup>2</sup> Twitter est un outil de microblogage libre et gratuit qui permet aux utilisateurs d'envoyer des tweets (des messages de 140 caractères au maximum), des photos et des vidéos sur internet.

La deuxième partie de cette recherche est consacrée à l'analyse du deuxième corpus comportant 4 629 tweets politiques. L'objectif est d'identifier les partis et les personnages politiques les plus mentionnés dans le corpus et les évaluations qui sont faites d'eux. Il sera également examiné quels partis politiques sont généralement évoqués par les hommes et par les femmes dans le corpus. En comparant les résultats de l'analyse avec le résultat des élections, il sera examiné si un rapport pourrait être établi entre les mentions d'un parti politique dans notre corpus et le nombre de voix que ce parti a enfin obtenu. Des études précédentes de Tumasjan et al. (2010) et Tjong Kim Sang et Bos (2012) ont démontré que Twitter s'avère un instrument de prédiction politique assez fiable. Toutefois, jusqu'à présent, aucune étude ne s'est concentrée sur les élections en Belgique. Il est vraisemblable que l'analyse du corpus ne dévoilera que quelques orientations politiques de l'électorat, étant donné que l'étendue du corpus est relativement limitée. Néanmoins, si les résultats de l'expérience sont satisfaisants, cette étude pourrait ouvrir quelques perspectives de recherche sur la possibilité d'apporter les informations des réseaux sociaux aux sondages politiques traditionnels en Belgique.

La troisième partie de la présente étude comprend l'analyse du corpus le plus étendu. Comportant 15 952 tweets, le troisième corpus permettra d'effectuer une extraction terminologique et l'étiquetage morpho-syntaxique sur une quantité considérable de tweets. De cette façon, l'analyse révélera les mots les plus mentionnés dans le corpus et les catégories grammaticales attribuées à tous les mots. Les résultats révéleront de précieuses informations sur la terminologie spécifiquement utilisée dans un contexte politique. Cette analyse pourrait encourager de futures recherches consacrées à la comparaison entre les mots souvent utilisés sur les médias sociaux et les médias traditionnels en vue des élections.

### 1.3. Description des chapitres

Ce mémoire comprend dix chapitres au total. Le prochain chapitre (le chapitre 2) présente une étude théorique des diverses applications de l'analyse des sentiments dans le domaine des médias sociaux. Le chapitre 3 comprend la description des corpus (3.1) et du logiciel d'annotation utilisé (3.2). Ensuite, il présente le schéma d'annotation de Van de Kauter et Desmet (2012) (3.3) et la procédure d'annotation suivie lors de cette étude (3.4). Le chapitre 4 comprend une présentation détaillée de l'analyse des sentiments du premier corpus, le

chapitre 5 décrit l'analyse du deuxième corpus et l'analyse du troisième corpus est détaillée dans le chapitre 6. Les résultats des trois analyses sont présentés dans le chapitre 7. Le chapitre 8 renferme la conclusion et les perspectives de recherche qu'ouvre la présente étude. Le chapitre 9 comprend la bibliographie et les annexes sont incluses dans le chapitre 10.

## 2. LITTERATURE

### 2.1. L'échange d'opinions

Avant de se concentrer sur l'analyse des sentiments, il importe de définir le mot « sentiment ». Le terme couvre plusieurs acceptions en fonction du domaine dans lequel il est appliqué. Dans la littérature, un sentiment est « un jugement, une opinion qui se fonde sur une appréciation subjective (et non sur un raisonnement logique) → avis, idée, point de vue »<sup>3</sup>.

Suivant Pang et Lee (2008), l'opinion des autres a toujours été une pièce d'information très précieuse au moment de se faire une opinion ou de prendre une décision. En effet, avant l'apparition du Web et l'Internet, les gens avaient intérêt à connaître les opinions de leurs amis ou de leur famille. Il leur était demandé de recommander un mécanicien automobile ou de faire savoir quel parti politique recevrait leur voix lors des prochaines élections. Grâce à l'essor considérable qu'ont connu le Web et l'Internet à partir des années quatre-vingt-dix, il est devenu possible pour tous de consulter l'opinion d'un vaste groupe de personnes à travers le Web. L'opinion de ce groupe est d'autant plus précieuse que les personnes dedans ne sont ni des membres de la famille, ni des critiques professionnels intéressés à promouvoir un produit. Leur opinion est considérée comme plus objective et par conséquent, plus valable (Pang et Lee, 2008).

Il existe une demande croissante d'informations évaluatives qui a fait naître le phénomène des nouveaux médias ou médias sociaux. Les messages SMS, les forums en ligne, les blogs et les réseaux sociaux ne sont que quelques exemples de ces nouveaux médias. Ils connaissent une immense popularité grâce à la vitesse à laquelle les messages sont envoyés et à leur grande accessibilité (Pang et Lee, 2008).

---

<sup>3</sup> Définition selon Le Nouveau Petit Robert de la langue Française 2009.

A part l'intérêt que les individus ont à consulter les opinions d'autrui, il existe d'autres personnes et organisations qui veulent connaître les opinions échangées à leur égard sur le Web. En premier lieu, il existe l'intérêt commercial auquel pourrait servir un système autonome d'analyse de sentiments. Pang et Lee (2008) ont donné l'exemple d'un commerçant qui aimerait savoir pourquoi son nouveau modèle de portatif ne se vend pas bien. Il lui convient donc de connaître les jugements personnels des consommateurs à l'égard du portatif : la conception n'est pas gentille ? Le service client était mauvais ? Ce ne sont que quelques questions dont la réponse est très précieuse pour le commerçant afin qu'il puisse adapter ses produits aux impératifs du marché. En deuxième lieu, il est récemment né un intérêt du monde politique pour les paroles échangées sur le Web. A part les commerçants, les hommes politiques ont également intérêt à savoir ce que l'électorat dit à leur égard. Comme les gens sont de plus en plus présents sur les réseaux sociaux afin d'échanger leurs idées avec autrui, il importe que les hommes politiques observent attentivement et continûment l'évolution des conversations en ligne (Pang et Lee, 2008).

## 2.2. L'analyse des sentiments

L'année 2001 marque le début de la demande croissante de systèmes automatiques d'analyse des sentiments. Cette demande a émané d'une combinaison de développements : l'essor de méthodes d'apprentissage automatique, la disponibilité d'ensembles de données grâce à l'expansion du Web et la naissance de sites Web qui recueillent des critiques. Il existe dans la littérature plusieurs termes qui renvoient à l'analyse des sentiments : *opinion mining*, *sentiment analysis*, *subjectivity analysis*, *review mining*, *appraisal extraction*, etc. La différence entre les dénominations est insignifiante, étant donné que toutes font référence à l'analyse des opinions et des sentiments exprimés dans des textes subjectifs. Toutefois, les termes *opinion mining* et *sentiment analysis* s'avèrent les plus univoques et s'utilisent par conséquent très fréquemment (Pang et Lee, 2008).

Suivant Pang et Lee (2008), l'analyse des sentiments d'un texte, qu'il soit écrit ou électronique, consiste en plusieurs étapes. Au moment de concevoir un système autonome d'analyse automatique des sentiments, il importe que celui-ci soit capable de suivre ces différentes étapes de l'analyse de façon indépendante. Tout d'abord, les textes subjectifs doivent être distingués des textes neutres ou objectifs. Ensuite, le système doit détecter les

paragraphe ou les phrases comportant des jugements évaluatifs. Finalement, toutes les informations doivent être présentées dans une analyse globale du sentiment exprimé dans le texte. Considérant les différentes étapes de l'analyse des sentiments, la conception d'un système autonome devrait être telle que l'analyse des sentiments du système approche le plus possible celle qui est faite par le cerveau humain (Pang et Lee, 2008).

Jusqu'à présent, les systèmes d'analyse automatique des sentiments ont souvent été entraînés aux textes traditionnels tels que les critiques en ligne et les articles de journaux. Toutefois, il se manifeste dans notre société un intérêt croissant pour les réseaux sociaux et les opinions qui y sont exprimées. Il est par conséquent primordial de concevoir un système automatique dont le fonctionnement soit adapté à ce nouveau type de textes provenant des réseaux sociaux.

### 2.3. Deux approches différentes du système d'analyse automatique des sentiments

La catégorisation de textes remonte aux années soixante et a connu un véritable essor à partir des années quatre-vingt-dix grâce à la présence d'investigateurs spécialisés dans le domaine et aux nombreux développements sur le plan informatique. L'analyse des sentiments est un produit de toutes les innovations dans le domaine de la catégorisation de textes. En effet, depuis les dix dernières années, les linguistes s'intéressent de plus en plus à l'analyse des sentiments dans des textes écrits ou électroniques. Cet intérêt accru est principalement dû aux évolutions dans le domaine des méthodes d'apprentissage automatique, une meilleure accessibilité aux bases de données et l'expansion du Web (Pang et Lee, 2008).

Généralement, il existe deux méthodes différentes afin de procéder à l'analyse automatique des sentiments. La première méthode est basée sur un lexique construit à partir de dictionnaires existants (*lexicon-based approach*). La deuxième méthode est basée sur des corpus comportant des textes évaluatifs dont le langage est généralement subjectif (*corpus-based approach*).

### 2.3.1. L'analyse automatique des sentiments basée sur un lexique

Jusqu'à présent, la plupart des études de l'analyse des sentiments se sont basées sur des lexiques. Ceux-ci permettent d'évaluer la polarité d'un texte subjectif à l'aide de deux groupes de mots : ceux qui expriment un sentiment positif et ceux qui expriment un sentiment négatif. Le système de traitement extrait du texte tous les mots positifs et négatifs et les met dans le groupe correspondant. La somme des mots représente une évaluation globale du sentiment dans le texte ; si la quantité de mots positifs l'emporte sur celle de mots négatifs, le logiciel tend à dire que le texte exprime un sentiment négatif. Inversement, le texte est considéré comme positif.

Pang et Lee (2008) ont insisté sur les difficultés liées à la méthode basée sur un lexique. Ils ont remarqué cinq problèmes qui peuvent surgir lors de l'analyse et qui entravent la fiabilité des résultats. La première difficulté est que le sentiment d'un texte peut être exprimé d'une façon très subtile sans que des mots explicitement subjectifs soient utilisés, comme dans l'exemple suivant (extrait de Pang et Lee, 2008) :

- « Jane Austen's books madden me so that I can't conceal my frenzy from the reader. Every time I read 'Pride and Prejudice' I want to dig her up and beat her (..) »

Parfois, il s'avère également difficile de distinguer les faits des opinions. Ainsi, les exemples suivants (extraits de Pang et Lee, 2008) présentent des informations similaires. Néanmoins, la première phrase est considérée comme factuelle tandis que la deuxième est considérée comme subjective, présentant l'opinion de l'auteur.

- « Miss Austen is not a poetess »  
 - « (...) none of the ravishing enthusiasm of poetry »

Une troisième difficulté est liée au contexte d'une phrase. Celui-ci est susceptible d'altérer le sentiment exprimé au sein de la phrase. En plus, l'acception d'un mot varie parfois selon le contexte dans lequel il figure. La phrase suivante (extraite de Pang et Lee, 2008) illustre qu'un mot ou une phrase peut exprimer des opinions différentes selon le contexte.

- « Go read the book »

Si la phrase figure dans la critique d'un livre, elle exprime un sentiment positif, étant donné qu'elle incite la lecture du livre. En revanche, si la phrase figure dans la critique d'un film, elle exprime un sentiment négatif à l'égard du film, car elle incite la lecture du livre. De cette façon, l'auteur de la phrase suggère que le livre est meilleur que son adaptation (Pang et Lee, 2008).

Pang et Lee (2008) ont signalé une quatrième difficulté, qui est liée à la structure du texte. Ainsi, il se peut que, bien qu'un texte comprenne plusieurs mots positifs, le sentiment global exprimé soit négatif à cause d'un mot ou une phrase en particulier. Dans l'exemple suivant (extrait de Pang et Lee, 2008), les mots « brilliant », « great plot », « first grade » et « good » suggèrent que le sentiment exprimé à l'égard du film est positif. Cependant, la toute dernière phrase (« however, it can't hold up ») révèle que l'auteur n'a pas du tout apprécié le film.

- « This film should be brilliant. It sounds like a great plot, the actors are first grade, and the supporting cast is good as well, and Stallone is attempting to deliver a good performance. However, it can't hold up. »

Un dernier problème qu'ont évoqué Pang et Lee (2008) sont les éléments dans un texte qui modifient le sentiment exprimé tels que l'ironie, le sarcasme, la négation, les tournures de phrase et la position des mots. La phrase suivante (extraite de Pang et Lee, 2008) illustre bien de ce problème :

- « I hate the Spice Girls. (...) Why I saw this movie is a really, really, really long story, but I did, and one would think I'd despise every minute of it. But... Okay, I'm really ashamed of it, but I enjoyed it (...). The plot is such a mess that it's terrible. But I loved it. »

En énumérant quelques problèmes cruciaux que présente une méthode d'analyse des sentiments basée sur un lexique, Pang et Lee (2008) ont déjà suggéré que seule la somme des mots positifs et négatifs ne suffit pas afin de déterminer la polarité d'un texte et d'analyser son sentiment de façon satisfaisante, car ni le contexte ni la position des mots dans une phrase sont pris en considération. Pourtant, ce sont des éléments qui peuvent altérer l'acception d'un mot.

Ces problèmes paraissent encore davantage compliquer l'analyse des sentiments à mesure que le texte à analyser est raccourci jusqu'à ce qu'il ne contienne que quelques mots. Tel est le cas des tweets, des messages de 140 caractères au maximum qui sont envoyés via Twitter. Les études synthétisées dans les paragraphes suivants illustrent les difficultés qui surgissent lorsque l'analyse des sentiments basée sur un lexique est appliquée sur des textes exceptionnellement courts tels que des tweets.

Moreno-Ortiz et Pérez Hernández (2012) ont appliqué la méthode basée sur un lexique sur des textes extrêmement courts provenant des réseaux sociaux tels que Twitter. Leur étude cadre dans l'atelier scientifique TASS<sup>4</sup> de 2012 qui avait pour objectif de promouvoir l'utilisation des systèmes automatiques pour l'analyse de sentiments de textes courts. Les investigateurs ont été demandés d'analyser automatiquement un corpus de 60 000 tweets et de déterminer pour tout tweet la polarité et l'intensité du sentiment exprimé. Ils ont utilisé le logiciel *Sentitext*<sup>5</sup> basé sur un lexique espagnol semi-automatiquement collecté et manuellement révisé. À partir des mots négatifs et positifs dans un texte, le logiciel sait calculer la valeur sentimentale globale du texte sur une échelle de zéro à dix. Les résultats les plus importants de leur expérimentation sont détaillés dans les paragraphes suivants.

En premier lieu, l'évaluation a affirmé l'hypothèse que la performance des systèmes d'analyse basés sur un lexique dépend fortement de la quantité de mots que comporte un texte. Bien que les systèmes soient conçus d'une telle manière qu'un texte un peu plus court qu'un texte standard ne pourrait pas altérer le résultat, les tweets se sont avérés trop courts afin de garantir l'efficacité du logiciel *Sentitext*. Ainsi, l'analyse a révélé que les tweets comportaient moyennement 14,1 unités lexicales, dont 5,5 exprimaient un sentiment et 8,6 étaient neutres. Ce résultat implique que 39,3 % des mots que comportaient les tweets étaient évaluatifs, ce qui a produit une valeur sentimentale globale tellement élevée qu'elle n'était plus tout à fait fiable. Tenant compte du fait qu'un pourcentage de 25 % est déjà exceptionnellement élevé pour des textes traditionnels, les investigateurs ont constaté que Twitter s'utilise surtout pour exprimer ses opinions. Par conséquent, il importe qu'un système autonome d'analyse des

---

<sup>4</sup> Le TASS (*Taller de Análisis de Sentimientos en la SEPLN*) est un atelier scientifique espagnol qui a promu la recherche dans le domaine de l'analyse automatique des sentiments. Il a eu lieu le 7 septembre 2012 et il a été organisé par la société SEPLN (*Sociedad Española para el Procesamiento del Lenguaje Natural*). Tous les résultats des études expérimentales sont publiés sur le site web <http://www.daedalus.es/TASS>.

<sup>5</sup> Le logiciel *Sentitext* est un système autonome d'analyse des sentiments qui est basé sur un lexique. Le site web officiel du logiciel est <http://www.sentitext.com>.

sentiments soit adapté à de telles textes fortement subjectifs (Moreno-Ortiz et Pérez Hernández, 2012).

En deuxième lieu, l'expérimentation avait pour objectif d'évaluer la performance du logiciel *Sentitext* en ce qui concerne l'analyse des sentiments. L'étude a révélé une performance assez faible du logiciel par comparaison aux annotateurs humains, en particulier pour l'indication de l'intensité d'un sentiment et pour l'analyse du sentiment global d'un texte. Le calcul de cette valeur sentimentale globale étant établi pour les textes traditionnels de taille moyenne, il ne s'avère pas approprié pour être appliqué aux tweets (Moreno-Ortiz et Pérez-Hernández, 2012).

Long Jiang et al. (2011) ont également affirmé que le tweet lui-même, avec un maximum de 140 caractères, s'avère trop court afin d'y appliquer la méthode basée sur un lexique sans ajouter des modifications pour la rendre plus efficace. Les investigateurs ont constaté que l'objet d'un sentiment exprimé est trop souvent négligé par un classificateur automatique. Ainsi, il s'avère difficile pour un système automatique de distinguer l'objet de l'expression « better » dans la phrase « Windows 7 is much better than Vista » (exemple tiré de Long Jiang et al, 2011). Constatant que seule une analyse syntaxique de la phrase permet de déterminer si l'objet de « better » est « Windows 7 » ou « Vista », les investigateurs ont proposé une analyse des sentiments qui prenne en considération l'objet du sentiment exprimé et les autres tweets (= le contexte) qui affectent le sentiment afin d'éviter une fausse interprétation d'une telle phrase ambiguë.

A cet effet, les investigateurs ont proposé d'incorporer dans l'analyse des sentiments l'analyse de la structure syntaxique des tweets et de leur contexte. Ce dernier consiste en trois types de tweets : les *retweets* ; les tweets qui sont republiés par d'autres personnes, les *replying tweets* ; les tweets répondant au premier tweet (ils sont précédés du caractère « @ ») et les autres tweets qui sont postés par la même personne. Afin que le classificateur automatique rende compte des relations syntaxiques dans les tweets, l'étiquetage morpho-syntaxique (*POS-tagging*) des mots et la décomposition analytique (*syntactic parsing*) ont été appliqués. Le langage dans les tweets a été normalisé de façon que le classificateur automatique rende compte des différentes orthographes d'un mot, par exemple « luvé » pour « love » et « goooooood » pour « good ». Afin que le classificateur prenne en considération le contexte des tweets, un graphique a été créé représentant tous les tweets traitant d'un objet particulier. A

partir de ce graphique a été calculé le sentiment global exprimé à l'égard de cet objet (Long Jiang et al. 2011).

Avant de procéder aux expérimentations, Long Jiang et al. (2010) avaient analysé la performance globale du classificateur automatique par comparaison aux annotateurs humains. Cette comparaison avait révélé deux principaux types d'erreurs du classificateur automatique : 1) le système d'analyse automatique marquait les tweets neutres comme des tweets subjectifs et 2) le système marquait le bon sentiment, mais il ne le savait pas lier à l'objet correspondant. Trois expérimentations ont été faites afin d'évaluer la performance du classificateur automatique pour quelques étapes spécifiques de l'analyse des sentiments. Elles ont révélé que le classificateur automatique avait une performance moyenne de 66 % par comparaison aux annotateurs humains. Toutefois, après que le contexte des tweets et les objets des sentiments exprimés avaient été pris en considération, la performance du classificateur s'est élevée à 68,3 % (Long Jiang et al. 2011).

En conclusion, il s'avère très important de tenir compte de l'objet et le contexte d'un sentiment exprimé lors de l'analyse d'un texte subjectif. Néanmoins, les études de Moreno-Ortiz & Pérez-Hernández (2012) et Long Jiang et al. (2011) ont révélé que les deux éléments sont négligés par un système d'analyse basée sur un lexique. Ces défauts entravent considérablement la performance du système, surtout dans des textes exceptionnellement courts tels que des tweets.

Considérant les défaillances d'une méthode d'analyse des sentiments basée sur un lexique, nous avons préféré utiliser lors de cette étude une méthode basée sur des corpus. Etant basée sur des annotations humaines, cette méthode permettra une analyse profonde des tweets dans laquelle sont incluses tant l'analyse syntaxique que l'analyse du contexte des tweets.

### 2.3.2. L'analyse automatique des sentiments basée sur des corpus

Contrairement à la méthode basée sur un lexique, l'analyse automatique des sentiments basée sur des corpus ne s'appuie pas sur un vocabulaire de mots positifs et négatifs. En revanche, elle requiert l'élaboration de deux corpus annotés manuellement. Le premier corpus constitue le corpus d'apprentissage qui s'utilise afin d'entraîner un système automatique. Il comporte

des notes ajoutées par des annotateurs humains. A partir de ces notes, le système automatique devrait être capable de procéder à une analyse pareille de façon autonome. Le deuxième corpus constitue le corpus de test. Celui-ci est élaboré afin de vérifier la performance du système automatique. Dans un scénario idéal, les résultats de l'analyse faite par le système automatique correspondraient cent pour cent avec ceux du corpus d'apprentissage. Afin que la performance du système automatique soit maximale, il importe que le corpus d'apprentissage soit représentatif pour le corpus de test.

Avant de détailler la méthode basée sur des corpus, il importe de préciser la différence entre deux méthodes d'apprentissage différentes : l'apprentissage supervisé et l'apprentissage non supervisé. L'apprentissage supervisé implique l'élaboration de deux corpus : un corpus d'apprentissage et un corpus de test (*supra*). La méthode s'appelle « supervisée » parce que le système automatique est entraîné à traiter une base de données en se basant sur un corpus d'apprentissage qui comporte des modèles déjà traités (dans la présente étude, ces modèles sont des tweets). En revanche, la méthode non supervisée ne requiert qu'un corpus. La méthode implique que le système autonome doit lui-même structurer les informations au sein du corpus en les divisant en groupes. Ainsi, il doit organiser la base de données d'une telle manière que les données les plus similaires soient associées dans un groupe et les données différentes dans un autre.

Les annotations faites lors de cette étude pourront servir, lors de futures recherches, de corpus d'apprentissage pour l'entraînement d'un système autonome d'analyse automatique des sentiments. Ainsi, elles constitueront la base d'une méthode d'apprentissage supervisée. Wiebe (2002) a été la première investigatrice dans le domaine de catégorisation de textes à proposer d'utiliser des annotations manuelles de textes évaluatifs comme corpus d'apprentissage pour l'entraînement d'un système autonome d'analyse des sentiments. A cet effet, elle a élaboré un manuel d'annotation comportant des instructions détaillées pour l'analyse des sentiments.

Dans son mémoire, Wiebe (2002) a fait la distinction entre deux groupes d'expressions subjectives qui forment l'objet des annotations : les expressions subjectives explicitées, qui comportent les expressions *private state* et les actes de parole, et les expressions subjectives non explicitées. Les phrases suivantes (extraites de Wiebe, 2002) illustrent les trois types d'expressions subjectives :

- « John hates Bill » (= expression private state)
- « Mary said she would be home late » (= acte de parole)
- « What an idiot » (= expression subjective sans explicitation)

Suivant Wiebe (2002 : 4), une expression *private state* est « un état émotionnel qui s'exprime par une opinion, une conviction, une pensée, un sentiment, une émotion, une évaluation, un objectif ou un jugement ». Elle se compose d'une source, d'un objet et d'un sentiment. Un acte de parole est utilisé par un locuteur afin d'agir sur un autre interlocuteur. Les verbes *affirmer*, *demander* et *convaincre* sont des exemples d'actes de parole. Les expressions subjectives sans explicitation expriment une opinion sans qu'elles soient introduites par une expression linguistique, telle que la phrase « What an idiot! » (exemple extrait de Wiebe, 2002). Bien que la phrase exprime un sentiment à l'égard de quelqu'un, aucune expression linguistique n'est détectée qui lie l'énonciation à sa source. Il se peut également que la source d'une expression subjective soit l'auteur de la phrase. Dans l'exemple « John thinks that that moron is brilliant » (exemple de Wiebe, 2002), « John » est la source de l'expression subjective « brilliant », mais l'expression « moron » provient de l'auteur de la phrase.

La description de la procédure d'annotation de Wiebe (2002) étant très large et détaillée, nous nous référons à un travail plus récent de la même investigatrice (Wiebe et al. 2005) dans lequel elle a synthétisé en quatre schémas tous les éléments essentiels (= *les attributs*) de la procédure d'annotation pour les trois types d'expressions subjectives et la source d'un sentiment exprimé.

Tableau 1 : attributs des expressions *private state* et des actes de parole explicites (Wiebe, 2002).

- |   |
|---|
| <ul style="list-style-type: none"> <li>• <b>text anchor:</b> a pointer to the span of text that represents the speech event or explicit mention of a private state.</li> <li>• <b>source:</b> the person or entity that is expressing the private state, possibly the writer.</li> <li>• <b>target:</b> the target or topic of the private state, i.e., what the speech event or private state is about.</li> <li>• <b>properties:</b> <ul style="list-style-type: none"> <li>– <b>intensity:</b> the intensity of the private state; values <i>low</i>, <i>medium</i>, <i>high</i>, or <i>extreme</i>.</li> <li>– <b>expression intensity:</b> the contribution of the speech event or private state expression itself to the overall intensity of the private state; values <i>neutral</i>, <i>low</i>, <i>medium</i>, <i>high</i>, or <i>extreme</i>.</li> <li>– <b>insubstantial:</b> true, if the private state is not substantial in the discourse. For example, a private state in the context of a conditional often has the value <i>true</i> for attribute <i>insubstantial</i>.</li> </ul> </li> </ul> |
|---|

- **attitude type:** this attribute currently represents the polarity of the private state. The possible values are *positive*, *negative*, *other*, or *none*.

Tableau 2 : attributs des expressions subjectives sans explicitation (Wiebe, 2002).

- **text anchor:** a pointer to the span of text that represents the speech event or explicit mention of a private state.
- **source:** the person or entity that is expressing the private state, possibly the writer.
- **properties:**
  - **intensity:** the intensity of the private state; values *low*, *medium*, *high*, or *extreme*.
  - **attitude type:** this attribute currently represents the polarity of the private state. The possible values are *positive*, *negative*, *other*, or *none*.

Tableau 3: attributs des actes de parole objectifs (Wiebe, 2002).

- **text anchor:** a pointer to the span of text that denotes the speech event.
- **source:** the speaker or writer.
- **target:** the target or topic of the speech event, i.e., the content of what is said. To date, targets of objective speech event frames are not yet annotated in our corpus.

Tableau 4 : attributs des sources des expressions private state et des actes de parole (Wiebe, 2002).

- **text anchor:** a pointer to the span of text that denotes the speech event.
- **source:** the speaker or writer.

Il importe de remarquer que les schémas n'ont été développés que pour l'annotation de sentiments explicitement exprimés. Dans les actes de parole objectives ne sont indiqués que la source et l'objet de l'énonciation. Le schéma de Wiebe (2002) ne prévoit pas que de telles phrases objectives comportent une évaluation de quelqu'un ou quelque chose. En revanche, suivant le schéma de Van de Kauter et Desmet (2012), ces phrases si peuvent comporter des sentiments (*infra*).

Wiebe et al. (2005) ont mené une étude afin de vérifier en quelle mesure le schéma permet une annotation consistante sans que les résultats varient beaucoup selon les différents annotateurs qui l'utilisent. Un corpus comportant 10 657 phrases extraites de la presse mondiale a été établi et toutes les phrases ont manuellement été annotées par trois annotateurs qui se sont entraînés à l'aide des directives d'annotation conçues par Wiebe (2002). Le test du Kappa ( $\kappa$ ) de Cohen (Cohen, 1960) a été appliqué afin de mesurer l'accord entre les

annotateurs. Celui-ci a révélé un score  $\kappa$  de 0,77 pour la distinction entre les phrases subjectives et objectives et un score  $\kappa$  de 0,81 pour la distinction entre les expressions *private state* et les actes de parole objectifs. Une comparaison avec d'études précédentes a révélé que le schéma d'annotation de Wiebe (2002) a provoqué un meilleur accord entre les annotateurs. Ces résultats ont confirmé l'hypothèse que, plus le schéma d'annotation est détaillé, plus les annotateurs sont d'accord (Wiebe et al. 2005).

Cette conclusion a également été déduite par Wilson (2008). Dans sa dissertation, elle a proposé une extension au schéma d'annotation des expressions *private state* conçu par Wiebe (2002) et repris par Wiebe et al. (2005). L'extension apportée par Wilson (2008) consistait en une analyse plus détaillée du type de sentiment exprimé et de ses objets. Rien n'a été changé au schéma de base conçu par Wiebe (2002), sauf qu'au premier et au troisième schéma (présentant respectivement les attributs des expressions *private state* et des actes de parole explicites et les attributs des actes de parole objectifs) a été ajoutée la caractéristique *implicit*. La phrase suivante (tirée de Wilson, 2008) est un exemple d'un acte de parole marqué de l'attribut *implicit* :

- « This [the nomination of Miers] is a missed opportunity of historic proportions »

Le sentiment exprimé dans la phrase est considéré comme implicite, car aucune expression linguistique n'est détectée qui lie le sentiment à sa source. Nous verrons dans la section 3.3 du chapitre 3 que le schéma d'annotation conçu par Van de Kauter et Desmet (2012) part d'une autre conception de sentiments implicites, à savoir des sentiments implicitement exprimés dans des phrases qui comportent de l'information factuelle et objective.

Wilson (2008) a étendu le schéma d'annotation pour les expressions *private state* afin de mieux procéder à l'analyse du type de sentiment exprimé et de son ou ses objets. Ainsi, elle a conçu des schémas d'annotation séparés pour l'attitude<sup>6</sup> et l'objet d'une phrase subjective alors que Wiebe (2002) avait incorporé ces attributs dans le schéma d'annotation des expressions *private state* et des actes de paroles explicités. Selon Wilson (2008), des schémas d'annotation séparés permettent d'identifier plusieurs attitudes et objets au sein d'une expression subjective, ce qui rend l'annotation plus détaillée et par conséquent, plus efficace.

---

<sup>6</sup> Wilson (2008) a utilisé le mot *attitude* pour désigner ce que Wiebe (2002) avait appelé *sentiment*. Elle a préféré le mot *attitude* étant donné que le terme est plus général. Il recouvre entre autres les sentiments.

Tableau 5 :attributs des attitudes exprimées dans une expression *private state* (Wilson, 2008).

<ul style="list-style-type: none"> <li>• <b>id:</b> a unique, alphanumeric ID for identifying the attitude annotation. The ID is used to link the attitude annotation to the private state that it is a part of.</li> <li>• <b>text anchor:</b> a pointer to the span of text that captures the attitude being expressed.</li> <li>• <b>attitude type:</b> type of attitude being expressed.</li> <li>• <b>target link:</b> list of one or more target frame IDs, or the string none.</li> <li>• <b>intensity:</b> <i>low, low-medium, medium, medium-high, high, high-extreme.</i></li> <li>• <b>properties:</b> <ul style="list-style-type: none"> <li>– <b>inferred:</b> true, if the attitude is inferred.</li> <li>– <b>sarcastic:</b> true, if the attitude is realized through sarcasm.</li> <li>– <b>repetition:</b> true, if the attitude is realized through the repetition of words, phrases, or syntax.</li> <li>– <b>contrast:</b> true, if the attitude is realized only through contrast with another attitude.</li> </ul> </li> </ul>
---

Le tableau 5 présente les attributs des attitudes exprimées dans une expression *private state*. L'attribut *attitude type* comprend l'extension de la catégorie de types d'attitudes proposée par Wilson (2008). En effet, Wiebe (2002) et Wiebe et al. (2005) n'avaient distingué que trois types d'attitudes : les attitudes positives, les attitudes négatives et les attitudes du type *autre*. En revanche, Wilson (2008) a distingué six types d'attitudes : les sentiments (positifs/négatifs), les accords (positifs/négatifs), les argumentations (positives/négatives), les intentions (positives/négatives), les spéculations et les attitudes du type *autre*. L'attribut *target link* permet de lier une ou plusieurs attitudes à un ou plusieurs objets correspondants, alors que le schéma de Wiebe (2002) ne tenait compte que d'un seul objet par expression subjective. L'attribut *intensity* a été étendu de quatre intensités distinguées par Wiebe (2002) à six distinguées par Wilson (2008).

Il vaut également la peine d'éclaircir les quatre caractéristiques (*properties*) de l'attitude que Wilson (2008) a ajoutées au schéma. *Inferred* veut dire qu'une attitude est suggérée au lieu d'être explicitement mentionnée. La caractéristique *sarcastic* comprend les attitudes qui sont exprimées en maniant du sarcasme. La caractéristique *repetition* comprend les attitudes qui sont renforcées en les répétant et la caractéristique *contrast* désigne les attitudes qui constituent une opposition à une autre attitude exprimée dans la phrase.

Etant donné que l'annotation de l'objet (*target*) est déjà en grande partie incluse dans l'annotation de l'attitude, le nombre d'attributs de l'objet est plutôt limité. Les deux attributs sont présentés dans le tableau 6.

Tableau 6 : attributs de l'annotation de l'objet d'une attitude (Wilson, 2008).

- **id:** a unique alphanumeric ID for identifying the target annotation. The ID is used to link the target to the attitude frame.
- **text anchor:** a pointer to the span of text that denotes the target.

Il importe de signaler que les principales modifications apportées par Wilson (2008) consistent en l'extension du type d'attitude exprimée (*attitude type*), le lien entre les différents objets d'une attitude (*target link*) et les caractéristiques qui peuvent être attribuées à une attitude (*properties*) telles que la déduction, le sarcasme, la répétition et l'opposition.

L'hypothèse de Wilson (2008) que les extensions au schéma de base conçu par Wiebe (2002) permettraient une annotation fiable a été confirmée par les résultats d'une mesure d'accord entre les annotateurs. Le score  $F^7$  pour l'identification des attitudes était de 0,885 et celui pour l'identification des objets se situait autour de 0,885. Le test du Kappa ( $\kappa$ ) de Cohen (Cohen, 1960) a révélé un score  $\kappa$  de 0,80 pour la détermination du type d'attitude. Ces résultats élevés ont démontré que les modifications apportées par Wilson (2008) au schéma de base de Wiebe (2002) ont rendu le schéma d'annotation de sentiments plus détaillé, ce qui a généré un meilleur accord entre les différents annotateurs et par conséquent, une annotation plus fiable.

#### 2.4. Twitter dans le domaine politique

Dans notre société, les réseaux sociaux tels que Facebook, Twitter et LinkedIn constituent de plus en plus des moyens de débat politique par lesquels les gens peuvent exprimer leurs opinions sur la vie politique<sup>8</sup>. Depuis quelques années, les campagnes politiques se transforment de plus en plus en des campagnes sociales<sup>9</sup>. En effet, l'échange d'opinions et de sentiments sur les réseaux sociaux est tel qu'il éveille l'intérêt des autorités politiques et qu'il inspire des recherches à la possibilité d'intégrer les réseaux sociaux dans les sondages politiques traditionnels.

Tumasjan et al. (2010) ont mené une étude comparable à la présente en se basant sur un corpus de 104 003 tweets néerlandophones afin de pronostiquer les résultats des élections

<sup>7</sup> En statistique, le score F1 (aussi score F) s'utilise afin de mesurer la précision d'un test.

<sup>8</sup> Herregods, J. (10.10.2012). Verkiezingen : 'Vind ik leuk'. *Metro*, p. 4.

<sup>9</sup> Goret , C. (29.10.2012). De eerste sociale campagne. *Metro*, p. 6.

fédérales qui ont eu lieu en Allemagne en 2009. Les investigateurs ont voulu répondre à trois questions :

1. Twitter est-il devenu un forum de discussion politique en ligne ?
2. A quel point Twitter reflète-t-il le sentiment politique de l'électorat ?
3. Twitter constituerait-il un indicateur fiable pour prédire le résultat des élections dans l'avenir?

La méthodologie de la recherche consistait en une analyse des tweets mentionnant les noms des six partis politiques qui sont représentés au parlement allemand. Afin d'identifier les sentiments exprimés dans les tweets, les investigateurs ont utilisé le logiciel LIWC2007<sup>10</sup> qui permet d'évaluer les éléments émotionnels, cognitifs et structurels des tweets.

En se basant sur les résultats de leur recherche, les investigateurs ont conclu que Twitter s'utilise effectivement comme un forum de discussion politique en ligne : plus d'un tweet sur trois faisait partie d'une conversation et près de 20 % de tous les messages étaient des *retweets*, à savoir des tweets qui répondent à un autre tweet. Ces résultats ont démontré que Twitter ne s'utilise pas que pour exprimer ses opinions, mais également pour discuter avec d'autres internautes. L'étude a également révélé que 0,6 % des utilisateurs dans le corpus avaient publié 21,2 % de tous les tweets tandis que 50,3 % des utilisateurs avaient posté 10,2 % de tous les tweets. Ces chiffres ont indiqué que, bien que Twitter constitue un forum de débat politique en ligne, ce forum est dominé par un nombre d'utilisateurs relativement limité (Tumasjan et al. 2010).

Afin de vérifier si Twitter reflète le sentiment politique de l'électorat, Tumasjan et al. (2010) ont comparé les sentiments exprimés sur Twitter à l'égard de neuf hommes et femmes politiques (dont cinq présidents et quatre autres) avec leurs images généralement construites dans les médias traditionnels. Une mesure de fréquence de mots à l'aide du logiciel LIWC2007 a produit un profil virtuel de tout homme ou femme politique. La valeur  $d$  a été calculée à l'aide d'une formule mathématique. Elle représentait les ressemblances entre le profil virtuel des hommes politiques et leur image réelle. Les valeurs étant très proche de zéro ( $d = 0$  signifiait que les deux images correspondaient 100 %) :  $d = 0,10$  pour les présidents des partis et  $d = 0,24$  pour les autres hommes et femmes politiques, les investigateurs ont

---

<sup>10</sup> LIWC (*Linguistic Inquiry and Word Count*) est un logiciel de catégorisation de textes développé par James W. Pennebaker, Roger J. Booth et Martha E. Francis. Le site web officiel du logiciel est <http://www.liwc.net/>.

constaté que Twitter donne une réflexion fiable du sentiment politique de l'électorat (Tumasjan et al. 2010).

Finalement, la fréquence avec laquelle les partis politiques avaient été mentionnés dans les tweets a été comparée avec le nombre de sièges que ces partis ont obtenu après les élections. Cette comparaison a révélé un taux d'erreur de prédiction moyen de 1,7 % pour Twitter. Une synthèse des cinq sondages politiques traditionnels en Allemagne avait révélé que le taux d'erreur du meilleur sondage traditionnel était de 0,8 % et celui du moins fiable était de 1,5%. Ainsi, Tumasjan et al. (2010) ont conclu que Twitter, avec un taux d'erreur moyen de 1,7 %, peut être considéré comme un outil de prédiction assez fiable.

En conclusion, les résultats de la recherche de Tumasjan et al. (2010) ont révélé qu'en se basant sur le nombre de tweets mentionnant un certain parti politique, il est possible de prédire approximativement le nombre de voix qu'obtiendra ce parti politique lors des élections.

Une étude pareille de Tjong Kim Sang et Bos (2012) s'est basée sur les résultats de Tumasjan et al. (2010) afin de vérifier si leurs conclusions peuvent être généralisées à d'autres pays tels que les Pays-Bas. En premier lieu, les investigateurs ont examiné quelles adaptations s'avéraient nécessaires afin qu'une prédiction politique faite par Twitter soit aussi fiable qu'une prédiction faite par un moyen de sondage traditionnel. En deuxième lieu, ces adaptations ont été prises en considération afin de prédire les résultats des élections du 2 mars 2011.

Tjong Kim Sang et Bos (2012) ont établi un corpus de 7 000 tweets politiques néerlandophones. Considérant toute mention d'un parti politique dans le corpus comme une voix pour ce parti, l'étude a révélé que la prédiction de Twitter avoisinait celles des sondages néerlandais traditionnels *Politieke Barometer (PB)* et *Maurice de Hond (MdH)*. Les résultats de la comparaison sont présentés dans le tableau 7.

Tableau 7 : comparaison entre la prédiction de Twitter et celle des sondages PB et MdH (Tjong Kim Sang et Bos, 2012).

Parti politique	Nombre de sièges selon Twitter	Nombre de sièges selon PB	Nombre de sièges selon MdH	Moyen des deux sondages
PVV	18	12	12	12
VVD	13	14	16	15
CDA	12	9	10	9,5
PvdA	9	13	13	13

Constatant que la différence entre Twitter et les sondages traditionnels était parfois assez grande (comme pour la prédiction des résultats des partis PVV et PvdA), les investigateurs ont introduit deux étapes de normalisation pour améliorer la prédiction de Twitter. Ainsi, un seul tweet par utilisateur a été pris en considération -car tout électeur ne peut voter qu'une fois- et seuls les tweets exprimant un sentiment positif à l'égard d'un parti ont été gardés -car seule une opinion positive peut générer une voix.

Ensuite, 3 356 tweets ont été annotés manuellement afin d'identifier les sentiments y exprimés. Puis, un taux de sentiment a été calculé pour tout parti politique afin de permettre une prédiction du nombre de sièges de tout parti. Ces résultats montrant encore une différence considérable avec ceux des sondages traditionnels, les investigateurs ont chiffré un taux de population pour tout parti de manière à réduire l'écart démographique entre la population active sur Twitter et l'électorat des Pays-Bas. Les résultats sont présentés dans le tableau 8.

Tableau 8 : nouvelle comparaison entre la prédiction de Twitter et celle des sondages traditionnels (Tjong Kim Sang et Bos, 2012).

Parti politique	Taux de population	Nombre de sièges selon Twitter	Nombre de sièges selon les sondages traditionnels
PVV	0,93	12	12
VVD	1,23	15	15
CDA	0,80	10	9,5
PvdA	1,76	13	13

En fin, Tjong Kim Sang et Bos (2012) ont constaté que les adaptations et les mesures de normalisation dans le traitement des données ont considérablement réduit l'écart entre la prédiction de Twitter et celle des outils de sondage traditionnels. Finalement, ils ont tenté de prédire les résultats des élections à partir de 28 704 tweets collectés une semaine avant les élections. Les résultats ont révélé qu'à l'exception des trois principaux partis (le VVD, le

CDA et le PvdA), tous les pourcentages prédits par Twitter ne différaient que 1,7 % au maximum du pourcentage réel des élections.

Conformément aux résultats de Tumasjan et al. (2010), Tjong Kim Sang et Bos (2012) ont démontré que Twitter s'avère un indicateur fiable pour prédire le résultat des élections, à un tel point qu'il pourrait faire partie des moyens de prédiction traditionnels. Une partie de cette étude sera consacrée à vérifier si la constatation vaut également pour la Belgique. La section 7.2 donnera les résultats de cette vérification.

### 3. METHODE

Ce chapitre présente la description des trois corpus établis lors de cette étude et la procédure d'annotation suivie lors de l'analyse des sentiments du premier corpus. Il donne également une description du logiciel d'annotation Brat<sup>11</sup> et du schéma d'annotation établi par Van de Kauter et Desmet (2012) afin d'illustrer la procédure d'annotation suivie lors de l'analyse des sentiments.

#### 3.1. Description des corpus

La présente étude est basée sur trois corpus comportant des tweets. Afin qu'il soit possible d'examiner le sentiment politique exprimé sur Twitter en vue des élections communales et provinciales du 14 octobre 2012, tous les tweets traitent de la politique. A cet effet, un script Python<sup>12</sup> a été utilisé afin de collecter automatiquement des tweets politiques qui ont été publiés sur Twitter entre le jeudi 13 septembre 2012 et le lundi 15 octobre 2012. Pour la présente étude, des tweets ont été collectés à l'aide de différents mots-dièse politiques comportant un parti politique (par exemple *#Open Vld* et *#CD&V*), le président d'un parti politique (par exemple *#Tobback* et *#DeCroo*) ou un terme politique plus général (par exemple *#vk2012*, *#villapolitica* et *#gemeenteraadsverkiezingen*). Seuls les partis politiques

---

<sup>11</sup> Le logiciel Brat est un logiciel d'annotation utilisée pour l'analyse syntaxique profonde d'un texte. Le site web officiel du logiciel est <http://brat.nlplab.org/>.

<sup>12</sup> Un script Python est un logiciel qui est particulièrement utilisé comme langage de programmation pour automatiser des actions répétitives. Il se compose entre autres d'un code fourni par Twitter qui permet de rechercher sur le réseau social tous les tweets comportant un mot-dièse particulier. Le site web officiel du logiciel Python est <http://www.python.org/>.

flamands ont été pris en considération pour assurer que la plupart des tweets soient néerlandophones.

Le premier corpus se compose de deux cents tweets politiques automatiquement collectés entre le 13 septembre et le 15 octobre 2012 à l'aide de différents mots-dièse politiques. Du total de tweets, 17 sont des *retweets* et 17 autres tweets sont des *mentions*<sup>13</sup>, dont quatre sont également des *replying tweets*. Du total de 200 tweets, 97 ont été publiés par une femme et 103 par un homme.

Le deuxième corpus se compose de tweets politiques qui ont été collectés entre le 21 septembre et le 15 octobre 2012. Du total de 4 629 tweets, 3 200 ont été publiés par un homme et 718 par une femme. De 668 tweets n'a pas pu être identifié le sexe de l'auteur en raison d'un manque d'informations. Dans la plupart des cas, ces tweets avaient été publiés par une organisation ou une entreprise et les informations sur le profil correspondant ne permettaient pas d'identifier le sexe de l'auteur du tweet. Des 43 tweets restants n'a non plus pu être identifié le sexe de l'auteur parce que le profil dont provenait le tweet ne pouvait plus être retrouvé sur Twitter. Tous les tweets du corpus comportent le mot-dièse politique *#vk2012* qui a été proposé par la VRT<sup>14</sup> afin d'inviter l'électorat à aborder sur Twitter divers sujets politiques en vue des élections communales et provinciales de 2012. Ayant remarqué ce mot-dièse sur le site web de du journal de la VRT<sup>15</sup> le 21 de septembre, nous avons décidé de collecter à partir de ce moment-là tous les tweets comportant ce mot-dièse. Grâce à son caractère général, il a permis de collecter les tweets comportant tout type d'information sur les élections en général sans que l'attention soit fixée sur un parti ou un personnage politique en particulier.

Finalement, le troisième corpus est composé de 15 952 tweets collectés entre le 13 septembre et le 15 octobre 2012 à l'aide d'une multiplicité de mots-dièse politiques comportant tant des noms de partis politiques que des noms des présidents des partis et des termes politiques généraux : *#NVA*, *#VanBesien*, *#Gemeenteraadsverkiezingen2012*, *#DeCroo*, etc. Tenant compte de l'étendue considérable du corpus, le sexe des auteurs des tweets n'a pas pu être

---

<sup>13</sup> Les mentions sont des tweets contenant une arobase et un nom propre ou un mot quelconque pour signaler qu'ils traitent d'une certaine personne ou chose.

<sup>14</sup> La VRT (*Vlaamse Radio- en Televisieomroeporganisatie*) est l'entreprise publique belge qui organise la diffusion de la radio et la télévision pour la Communauté flamande.

<sup>15</sup> [Http://www.deredactie.be](http://www.deredactie.be).

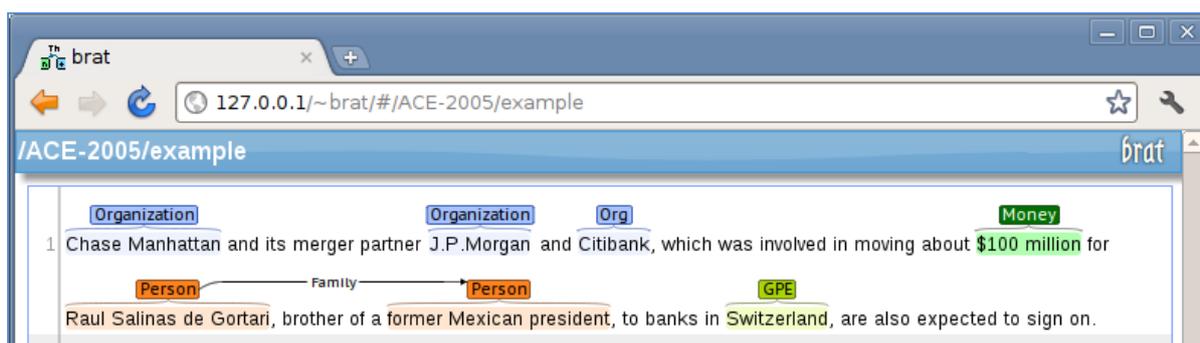
identifié, étant donné que ces informations doivent être recherchées de façon manuelle sur le profil des auteurs des tweets.

### 3.2. Le logiciel d'annotation Brat

Le logiciel Brat<sup>16</sup> est un outil d'annotation en ligne qui a été conçu pour l'annotation manuelle de textes écrits. Le logiciel permet aux annotateurs d'ajouter dans les textes à analyser des indications de structure. Le nombre de ces indications est limité et leur représentation est fixée par le logiciel pour qu'elles puissent automatiquement être analysées et traitées par un ordinateur lors d'une prochaine étape. Le logiciel a été développé afin de procéder à une analyse profonde de la structure d'une phrase ou d'un texte. Il permet de marquer, en plus des indications de structure elles-mêmes, les relations établies entre les différentes indications ajoutées au sein d'une phrase.

La figure 1 illustre une phrase annotée à l'aide du logiciel Brat<sup>17</sup>. Certains syntagmes de la phrase sont marqués d'une catégorie grammaticale spécifique telle que *organization* et *person* et des liens ont été apportés entre les différentes annotations afin d'identifier la relation qui les unit.

Figure 1: exemple d'annotation à l'aide du logiciel Brat.



Deux catégories d'annotation sont visualisées dans l'exemple : les annotations de syntagmes individuels (tels que les syntagmes qui sont marqués par les catégories *organization* et *person*) et les annotations qui représentent les relations entre deux annotations telles que le

<sup>16</sup> [Http://brat.nlplab.org/](http://brat.nlplab.org/).

<sup>17</sup> Exemple tiré du site web officiel du logiciel Brat : <http://brat.nlplab.org>.

lien *family*, indiquant que le premier syntagme (« Raul Salinas de Gortari ») entretient des relations familiales avec le deuxième (« former Mexican president »).

L'annotation d'un texte ne se limite pas à celle d'une seule phrase. Ainsi, le logiciel permet de lier les indications ajoutées au sein d'une phrase à celles qui sont apportées dans une phrase précédente ou suivante. De cette façon, le logiciel permet d'intégrer le contexte d'une phrase dans son analyse grammaticale.

### 3.3. Le schéma d'annotation de Van de Kauter et Desmet (2012)

Pour l'annotation manuelle du premier corpus, la présente étude s'est basée sur les directives établies par Van de Kauter et Desmet (2012). Les directives sont principalement basées sur le manuel d'annotation conçu par Wiebe (2002) et étendu par Wilson (2008). Toutefois, le nouveau schéma comprend quelques modifications pour rendre l'analyse des sentiments plus détaillée.

La modification la plus importante est que le schéma d'annotation de Van de Kauter et Desmet (2012) distingue trois types d'expressions subjectives : les expressions *private state*, les expressions *polar fact* et les expressions *polar resultative causative*. En revanche, le schéma de Wiebe (2002) n'avait été conçu que pour la détection d'expressions *private state*. L'addition des deux catégories d'expressions subjectives implique que le schéma de Van de Kauter et Desmet (2012) permet de détecter également des sentiments dans des textes qui ne contiennent que de l'information objective et factuelle, alors que ces textes étaient considérés comme neutres suivant le schéma d'annotation de Wiebe (2002). Cette extension au schéma de base conçu par Wiebe (2002) s'avère indispensable étant donné qu'elle permet l'identification de sentiments implicitement exprimés. De cette façon, il est possible de faire une analyse encore davantage profonde des sentiments exprimés dans un texte.

L'objectif de l'annotation d'un texte évaluatif est d'en extraire toutes les expressions subjectives qu'il comporte. Les expressions subjectives sont des expressions linguistiques qui expriment un sentiment positif ou négatif à l'égard d'un ou plusieurs objets. Dans le manuel d'annotation de Van de Kauter et Desmet (2012) sont distingués trois types d'expressions

subjectives. Chacun des trois types est illustré à l'aide des exemples suivants (tirés de Van de Kauter et Desmet, 2012) :

1. **Les expressions *private state***: les expressions linguistiques explicites qui expriment un sentiment positif ou négatif à l'égard d'un ou plusieurs objets.

– « De liquiditeit en solvabiliteit van KBC zijn gewoon goed »

Dans la phrase, l'auteur exprime un sentiment positif (« zijn gewoon goed ») à l'égard de l'objet « de liquiditeit en solvabiliteit van KBC ». Le syntagme « zijn gewoon goed » constitue l'expression *private state* de la phrase.

2. **Les expressions *polar fact***: les expressions implicites qui expriment un sentiment positif ou négatif à l'égard d'un ou plusieurs objets. Bien que les expressions *polar fact* représentent des informations factuelles, elles comportent une évaluation subjective d'un ou plusieurs objets.

– « Kim Clijsters heeft de kwartfinale van het olympisch tennistornooi verloren »

Dans l'exemple, les mots « heeft de kwartfinale van het olympisch tennistornooi verloren » expriment un sentiment négatif à l'égard de l'objet « Kim Clijsters ». La phrase comporte un sentiment implicitement exprimé : une description factuelle qui mène à une évaluation (dans ce cas-ci négative) de son objet. Les expressions *polar fact* se distinguent des expressions *private state* parce que l'information qu'elles contiennent peut objectivement être vérifiée en comparant des résultats ou en mesurant une certaine valeur (Van de Kauter et Desmet, 2012).

3. **Les expressions *polar resultative causative*** : les expressions subjectives qui établissent une relation causale subjective entre deux syntagmes de la phrase. Elles indiquent que la cause a un effet positif ou négatif sur l'objet.

– « Twee zelfmoordterroristen pleegden een aanslag op een militair konvooi »

Les expressions *polar resultative causative* se distinguent des expressions *private state* parce que leur expression subjective consiste en l'effet que la cause a sur son objet. Dans la phrase

suivante, « twee zelfmoordterroristen » est la cause d'un effet négatif sur l'objet « een militair konvooi », car le résultat, un attentat (« een aanslag »), marque une évaluation négative de cet objet. Il se peut également qu'une relation causale soit établie sans qu'un sentiment soit exprimé. Dans ce cas-là, la phrase contient une relation causale neutre (Van de Kauter et Desmet, 2012).

Notre corpus comportant très peu d'expressions *polar resultative causative*, la catégorie ne sera pas décrite en détail. Il suffit de signaler que les expressions *polar resultative causative*, telles que les expressions *private state* et *polar fact* constituent la base du schéma d'annotation conçu par Van de Kauter et Desmet (2012).

La structure des trois types d'expressions subjectives étant pareille, leur annotation suit la même procédure. Celle-ci implique une annotation profonde sur tous les niveaux de la phrase. Afin qu'aucune expression subjective ne soit négligée, l'annotation commence toujours au niveau syntaxique le plus bas de la phrase et finit au niveau le plus complexe. Le point de départ du procès d'annotation est la détection de l'expression subjective. Après que le type de l'expression est déterminé, tous les éléments de base qui sont liés à cette expression sont désignés (Van de Kauter et Desmet (2012) :

1. **L'objet (*target*)** : le syntagme qui est affecté par le sentiment exprimé. Une phrase peut comporter plusieurs objets.
2. **La source (*source*)** : la personne ou le sujet qui exprime le sentiment. Elle n'est pas toujours explicitée dans la phrase.
3. **L'expression linguistique qui lie l'expression subjective à la source (*source expression*)** : cette expression n'est présente que si la source est explicitée dans la phrase.
4. **Le modificateur (*modifier*)** : l'élément lexical qui apporte des modifications au sentiment exprimé, telles qu'un renforcement ou un affaiblissement du sentiment. Une phrase ne contient pas toujours de modificateur.
5. **La cause (*cause*)** : la cause est l'élément lexical qui provoque un sentiment positif ou négatif à l'égard de l'objet. La cause n'est présente que dans les expressions *polar resultative causative* ou dans les expressions causales neutres.

Ces éléments de base sont les mêmes pour les trois types d'expressions subjectives. L'ordre dans lequel ils sont détectés est fixé par la procédure d'annotation. Celle-ci consiste en cinq

étapes différentes dont la cinquième n'est pas prise en considération lors de cette étude (Van de Kauter et Desmet, 2012).

1. Détecter l'expression subjective et déterminer son type ;
2. indiquer (s'ils sont présents) tous les modificateurs qui altèrent l'expression subjective ;
3. désigner la source de l'expression subjective et, si elle est explicitée, l'expression linguistique qui unit les deux ;
4. marquer le ou les objets de l'expression subjective et déterminer la polarité ainsi que l'intensité du sentiment exprimé ;
5. si l'expression subjective est du type *polar resultative causative*, signaler la cause du sentiment exprimé.

### 3.4. Procédure d'annotation

Pour l'annotation manuelle du premier corpus, la présente étude s'est basée sur les directives établies par Van de Kauter et Desmet (2012) qui prescrivent une procédure d'annotation appropriée pour les trois types d'expressions subjectives. Cette section présente la procédure d'annotation suivie lors de cette étude. Après les cinq étapes de base sont détaillées les consignes d'annotation pour les expressions subjectives incorporées dans d'autres, les relations coréférentielles et les composants de l'objet et finalement les expressions causales neutres.

#### 3.4.1. L'expression subjective

L'expression subjective comprend tous les mots qui forment le sentiment exprimé dans une phrase. Elle se compose d'un mot au minimum et d'une phrase complète au maximum et elle peut comporter un ou plusieurs modificateurs qui altèrent son sentiment. L'expression subjective ne comporte jamais d'objets et de sources.

La forme grammaticale d'une expression subjective peut varier ; elle peut être un adjectif, un syntagme nominal, un groupe verbal ou une phrase complète. Les captures d'écran (1) et (2)



### 3.4.2. Les modificateurs

Les modificateurs altèrent le sentiment exprimé dans une phrase, par exemple en renforçant ou en affaiblissant son intensité. Ils peuvent être incorporés au sein de l'expression subjective ou ils peuvent constituer des éléments isolés dans la phrase.

Onze catégories de modificateurs sont distinguées dans le manuel de Van de Kauter et Desmet (2012). Nous avons ajouté quelques exemples afin d'illustrer toute catégorie :

1. Modificateurs de négation : « niet », « nooit », « niets », etc.
2. Modificateurs de renforcement : « veel », « erg », « ! », etc.
3. Modificateurs d'affaiblissement : « relatief », « weinig », « zelden », etc.
4. Modificateurs de modalité : « waarschijnlijk », « lijken », « schijnen », etc.
5. Modificateurs de question : « ? », « niet waar ? », etc.
6. Modificateurs de condition : « als », « op voorwaarde dat », etc.
7. Modificateurs de référence à l'avenir : « zullen », « volgende week », etc.
8. Modificateurs de référence au passé : « gisteren », « maakte », etc.
9. Modificateurs de perspective : « voor de gezondheid », « voor kinderen », etc.
10. Modificateurs de spécification : « 14 miljoen euro », « 250 dollar », etc.
11. Autres modificateurs : « zoals gewoonlijk », « tiens », etc.

La capture d'écran (3) montre que plusieurs modificateurs peuvent être détectés au sein d'une phrase :

(3)



1. « blijft » = modificateur de renforcement.
2. « meermaals » = modificateur de renforcement.

3. « zijn gaan » = modificateur de référence au passé.
4. « begint .. te » = modificateur d'affaiblissement.

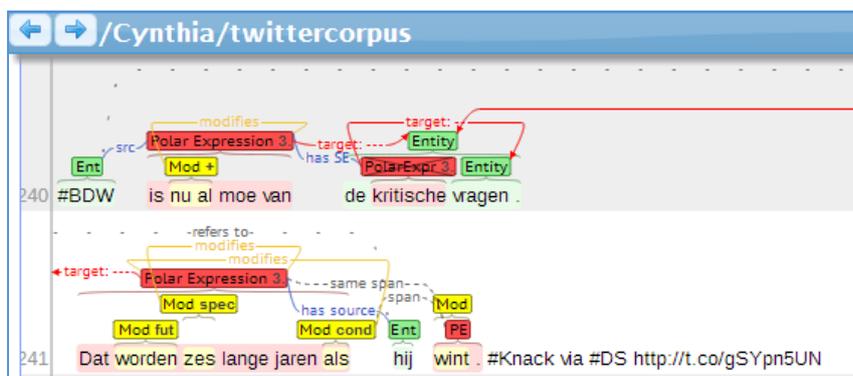
Dans l'exemple (3), le point d'interrogation n'est pas considéré comme un modificateur de la première expression subjective, étant donné que l'expression subjective consiste précisément en cette question. La condition pour marquer un syntagme comme un modificateur, c'est qu'après avoir enlevé le modificateur, il reste toujours une expression subjective dans la phrase. Cela n'est pas le cas ici. Si le point d'interrogation était supprimé, la phrase serait neutre.

### 3.4.3. La source et l'expression linguistique qui la lie au sentiment exprimé

Bien que les expressions subjectives aient toujours une source, celle-ci n'est pas toujours explicitée dans la phrase. Elle peut être tacitement présente si la source d'une expression subjective est l'auteur du texte ou une personne mentionnée dans une phrase précédente ou suivante. Si plusieurs références à la source sont détectées, celle qui se situe syntaxiquement le plus proche de l'expression subjective doit être marquée (Van de Kauter et Desmet, 2012).

Parfois, la source de l'expression subjective est explicitée au sein du tweet, mais il se note quand-même la présence de l'auteur. Dans l'exemple (4), le syntagme « is nu al moe van » est à la fois l'expression subjective et l'expression linguistique qui unit le sentiment et sa source « #BDW ». Bien que l'auteur du tweet ne soit pas la source du sentiment exprimé, les mots « nu al » suggèrent sa présence.

(4)



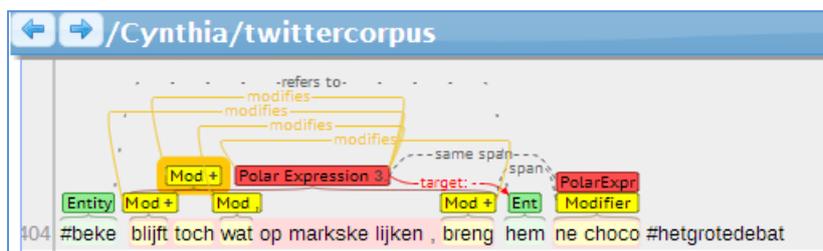
Dans certains cas, la source de l'expression subjective est implicite : elle ne se situe pas au sein de la phrase analysée, mais dans une phrase précédente ou suivante. Or, notre corpus ne comporte pas d'exemple d'une source implicite, étant donné que dans la majorité des cas, la source des sentiments exprimés est l'auteur du tweet lui-même. En outre, étant donné qu'un tweet est généralement entouré de peu de contexte, il est rare que la source du sentiment exprimé se situe hors du tweet lui-même.

Il se peut également que la source d'une expression subjective soit impersonnelle. Dans ce cas-là, la source est le peuple ; tout le monde, indiquée par le pronom personnel indéfini « on » (« men » en néerlandais). Dans la plupart des cas, la source impersonnelle figure dans des phrases passives (Van de Kauter et Desmet, 2012).

#### 3.4.4. L'objet, la polarité et l'intensité du sentiment exprimé

Toute expression subjective a nécessairement un objet. Celui-ci peut figurer au sein de la phrase analysée ou dans une phrase précédente ou suivante. Si la phrase comporte plus d'un objet, l'objet qui se situe syntaxiquement le plus proche de l'expression subjective doit être marqué. Dans l'exemple (5), le pronom personnel « hem » est marqué comme l'objet de l'expression « blijft toch wat op markske lijken, breng ... ne choco ». En effet, le mot « hem » se situe syntaxiquement plus proche de l'expression subjective que le mot « #beke » auquel il renvoie.

(5)



Après que tous les objets sont déterminés, la polarité et l'intensité de l'opinion exprimée sont marquées. Trois polarités différentes sont distinguées : *positive*, *négative* et *autre*. La polarité *autre* est marquée dans le cas où le sentiment exprimé n'est ni positif, ni négatif, comme la surprise, la crainte et le doute. L'intensité du sentiment exprimé peut être spécifiée par trois

degrés : *faible*, *moyenne* ou *grande*. Ainsi, toute expression subjective est marquée d'une des catégories suivantes : *faiblement négative* (-), *modérément négative* (--), *fortement négative* (---), *faiblement positive* (+), *modérément positive* (++) , *fortement positive* (+++), *faiblement autre* (~), *modérément autre* (~~) et *fortement autre* (~~~). Cependant, la classification d'intensité de la polarité est loin d'être ferme ; elle dépend en grande partie de l'interprétation personnelle de l'annotateur (Van de Kauter et Desmet, 2012).

Comme il a été affirmé par Van de Kauter et Desmet (2012)<sup>18</sup>, il se peut également que l'expression subjective d'une phrase soit non substantielle. Dans ce cas-là, le sentiment exprimé n'est pas réel parce qu'il constitue une incertitude, un avenir, un doute, une crainte, etc.

#### 3.4.5. La cause

Une phrase évaluative contient une cause si l'expression subjective est du type *polar resultative causative*. Faute d'exemples de ce type d'expression dans notre corpus, nous n'entrerons pas en détail sur cette catégorie.

#### 3.4.6. Les expressions subjectives incorporées dans une autre

Comme il a déjà été signalé dans ce chapitre, il importe de commencer l'annotation au niveau syntaxique le plus bas de la phrase, étant donné qu'une expression subjective peut être intégrée dans une autre. Si deux expressions subjectives sont détectées dans une phrase, l'expression subjective située à un niveau syntaxique plus complexe que la première ne peut être annotée que si après que la première expression subjective en est enlevée, elle exprime toujours un sentiment.

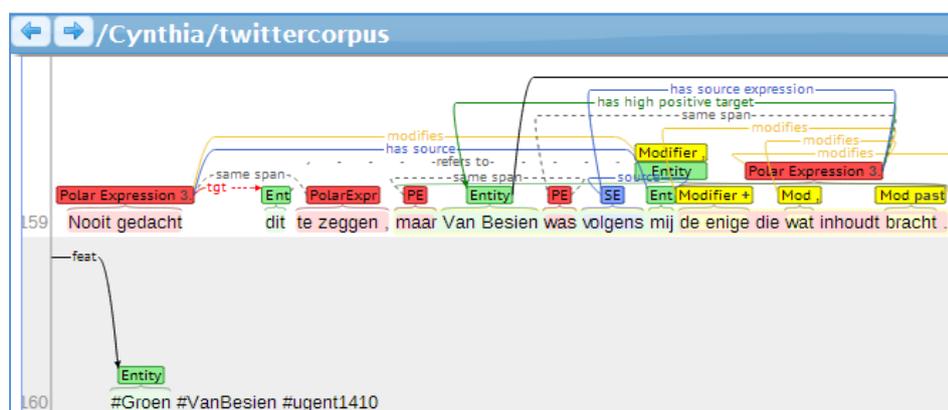
---

<sup>18</sup> Suivant Wiebe et al. (2005).

### 3.4.7. Les relations coréférentielles et les composants de l'objet

Si une phrase évaluative contient plusieurs références à l'expression subjective, celle qui se situe syntaxiquement le plus proche de l'expression subjective doit être marquée. Dans la plupart des cas, ces références sont des pronoms personnels tels que *hij* et *zij*, des adjectifs possessifs tels que *mijn* et *hun* ou des adjectifs démonstratifs tels que *die* et *dat* (Van de Kauter et Desmet, 2012).

(6)



Dans le tweet de l'exemple (6), le pronom démonstratif « dit » est marqué comme l'objet de l'expression subjective « nooit gedacht ... te zeggen » parce qu'il fait référence à l'objet réel de l'expression subjective : « maar Van Besien was volgens mij de enige die wat inhoudt (*sic*) bracht ».

Il se peut que le sentiment exprimé dans une phrase évaluative n'affecte non seulement son objet, mais également un ou plusieurs autres éléments de la phrase parce que ce sont des composants ou des caractéristiques de l'objet. Dans ce cas-là, ces éléments doivent également être marqués comme objets de l'expression subjective.

### 3.4.8. Les relations causales neutres

Faute d'exemples de relations causales neutres dans notre corpus, nous n'entrerons pas en détail sur cette catégorie.

#### 4. DESCRIPTION DE LA PREMIERE ANALYSE

Ce chapitre présente la méthode appliquée lors de l'analyse du premier corpus. Celle-ci consistait en une analyse des sentiments de deux cents tweets politiques. L'objectif des annotations était de procéder à une analyse profonde des sentiments exprimés dans le corpus et d'établir une comparaison entre les tweets publiés par des hommes et ceux publiés par des femmes.

Avant de procéder à l'analyse, tous les tweets ont été normalisés. Cette normalisation consistait à séparer les mots qui avaient été adhérents à d'autres mots sans que le terme créé soit un mot existant, tels que « nietleuk » et « doezovoort ». L'analyse des sentiments du premier corpus a été faite à l'aide du logiciel Brat<sup>19</sup> (voir la section 3.2) et en se basant sur les directives d'annotation établies par Van de Kauter et Desmet (2012) (voir la section 3.3).

La première étape de l'analyse consistait à présenter en tableau le résumé de l'analyse des sentiments dans tout tweet (voir la section 10.1 des annexes). Ce résumé comporte le nombre de sentiments trouvés dans tout tweet, l'intensité et la polarité de tout sentiment, ainsi que son niveau d'explicitation et les modificateurs qui ont altéré le sentiment exprimé. Le tableau comporte également les informations de base de tout tweet telles que le sexe de l'auteur du tweet, le sujet du message et le mot-dièse au moyen duquel le tweet a été collecté. La deuxième étape consistait à déduire, à partir du tableau général, des conclusions en ce qui concerne les sujets souvent abordés, le sentiment prépondérant dans tout tweet, le rapport entre le sexe de l'auteur du tweet et la polarité / l'intensité / l'explicitation des sentiments, ainsi que l'emploi de modificateurs. Les résultats de ces recherches permettront de tirer des conclusions sur les possibilités qu'offre le schéma d'annotation de Van de Kauter et Desmet (2012) et de révéler d'éventuels défauts. Ils sont présentés dans la section 7.1.

#### 5. DESCRIPTION DE LA DEUXIEME ANALYSE

Ce chapitre présente la description du traitement du deuxième corpus. Avant de procéder à l'analyse du corpus, deux mesures de normalisation ont été prises. En premier lieu, les tweets datant du 15 octobre ont été supprimés, car à ce moment-là, les résultats des élections étaient

---

<sup>19</sup> [Http://brat.nlplab.org/](http://brat.nlplab.org/).

déjà connus. Par conséquent, les tweets de ce jour ne comportaient plus de prédictions ou de tentatives pour récolter des voix. En deuxième lieu, seuls les tweets néerlandophones ont été gardés pour faciliter l'analyse. Après ces deux mesures de normalisation, le corpus comportait 4 246 tweets dont 2 908 avaient été postés par des hommes et 661 par des femmes. Des 677 tweets restants n'a pas pu être identifié le sexe de l'auteur en raison d'un manque d'informations.

Après la normalisation, le corpus entier a été soumis à une mesure de fréquence manuelle afin de révéler les partis et les hommes politiques les plus évoqués sur Twitter en vue des élections et de détecter les sentiments exprimés à leur égard. Il a également été examiné si le sexe de l'auteur du tweet a joué un rôle dans la fréquence des mentions de certains partis ou personnages politiques et leur évaluation. Ensuite, il a été vérifié si les sentiments exprimés à l'égard des partis et des personnages politiques pourraient contribuer à l'explication du résultat des élections communales et provinciales de 2012. Plus concrètement, la fréquence avec laquelle un parti politique flamand avait été mentionné dans le corpus a été comparée avec le pourcentage de voix que ce parti a obtenu.

L'analyse s'est concentrée sur les huit principaux partis politiques flamands : la N-VA, le CD&V, le sp.a, l'Open Vld, le Vlaams Belang, Groen, la LDD et le PVDA et leurs présidents actuels, respectivement : Bart De Wever, Wouter Beke, Bruno Tobback, Gwendolyn Rutten, Gerolf Annemans, Wouter Van Besien, Jean-Marie Dedecker et Peter Mertens. Trois autres personnes ont également été prises en considération lors de l'analyse : Alexander De Croo (président de l'Open Vld jusqu'aux élections), Bruno Valkeniers (président du Vlaams Belang jusqu'aux élections) et Patrick Janssens (bourgmestre d'Anvers jusqu'aux élections, succédé par Bart De Wever). La popularité de ces trois hommes politiques de notre corpus a été comparée avec celle de leurs successeurs.

## 6. DESCRIPTION DE LA TROISIEME ANALYSE

Dans ce chapitre est incluse la méthodologie de l'analyse du troisième corpus de tweets politiques. L'analyse consistait à extraire une liste de termes à partir des tweets dans le corpus et de déterminer la catégorie grammaticale de tous les mots des termes. L'objectif de l'extraction terminologique était de révéler la terminologie spécifiquement utilisée dans un

contexte politique. L'objectif de l'étiquetage grammatical était de dévoiler les catégories grammaticales souvent utilisées dans un langage politique.

Le corpus se compose de 15 952 tweets politiques comportant une multitude de mots-dièse différents. Par conséquent, les sujets politiques abordés dans le corpus sont très divers. Aucune information n'a été ajoutée en ce qui concerne le sexe de l'auteur des tweets, étant donné le nombre considérable de données. En effet, l'étendue du corpus s'est avérée trop large afin de procéder à une analyse manuelle des tweets. Aussi le corpus entier a-t-il été soumis à une mesure de fréquence à l'aide de l'extracteur de termes *TExSIS*<sup>20</sup>, un logiciel qui permet l'extraction terminologique à partir d'une base de données spécialisée. Lors de l'analyse, le logiciel a collecté tous les termes les plus évoqués dans le corpus selon leur fréquence et leur valeur *termhood*<sup>21</sup> et il a attribué à tous les mots de ces termes la catégorie grammaticale à laquelle ils appartiennent.

## 7. RESULTATS

Ce chapitre donne un aperçu des principaux résultats de l'analyse des trois corpus de tweets politiques. Il répondra à quelques questions cruciales qui ont été posées dans la section 1.2. Au total, trois analyses différentes ont été faites, chacune à partir d'un corpus différent. Les sections 7.1, 7.2 et 7.3 présentent les résultats des trois analyses. Étant donné que le chapitre 3 présente plusieurs exemples de l'annotation des tweets, la procédure d'annotation suivie lors de l'analyse du premier corpus n'est plus détaillée dans ce chapitre.

### 7.1. Résultats de l'analyse des sentiments

Dans cette section sont présentés les résultats de l'analyse du premier corpus comportant deux cents tweets. Tous les tweets ont été annotés de façon manuelle à partir des directives conçues par Van de Kauter et Desmet (2012). L'objectif des annotations était de faire une analyse profonde des sentiments exprimés dans le corpus et d'établir une comparaison entre les tweets

---

<sup>20</sup> Le logiciel *TExSIS* permet d'extraire d'un texte un thésaurus de mots clés ou de termes pertinents (noms de personnes, d'organisations, de lieux et d'autres mots-clés) d'une façon automatique ou semi-automatique. Le site web officiel du logiciel est <http://lt3.hogent.be/en/projects/texsis/>.

<sup>21</sup> La valeur *termhood* d'un terme indique le degré de spécialisation de son sens dans le domaine à l'étude (Kageura et Umino, 1996). (Voir la section 7.3).

publiés par les hommes et ceux publiés par les femmes du corpus. Cette comparaison concernait surtout l'explicitation des sentiments exprimés, la polarité, l'intensité et l'emploi de modificateurs.

#### 7.1.1. Distribution des sentiments et le sentiment prépondérant dans tout tweet

Comme il a été démontré par Long Jiang et al. (2011), il importe de prendre en considération l'objet et le contexte d'un sentiment exprimé afin de procéder à une analyse profonde des sentiments dans un texte. La détection de l'objet de tout sentiment exprimé était d'autant plus importante lors de cette étude que beaucoup de tweets comportaient plus d'un sentiment, ce qui rend l'analyse davantage équivoque.

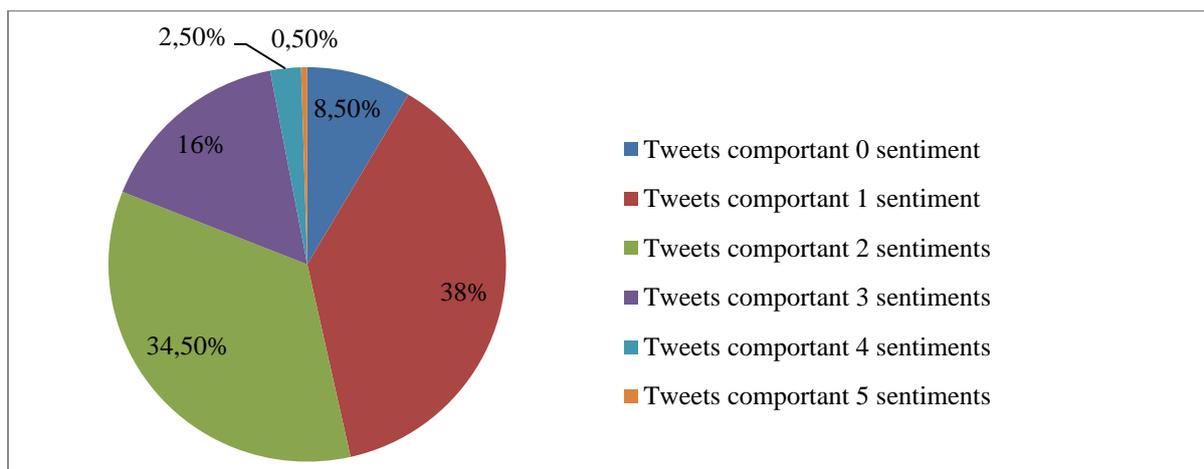
Au total, 335 sentiments sont exprimés dans le corpus de 200 tweets. Du point de vue statistique, cela implique qu'en moyenne, près de deux sentiments (1,68) sont exprimés par tweet. Néanmoins, 17 tweets du corpus ont été qualifiés de neutre. Dans la plupart des cas, ces tweets ne comportaient que de l'information factuelle et avaient pour objectif principal informer ses lecteurs. L'exemple (7) de notre corpus est un tweet qui est considéré comme neutre :

(7)

- « Vandaag krijgen de partijen die geen nationale verankering hebben hun lijstnummer. Nog 20 dagen voor de verkiezingen. #vk2012 »

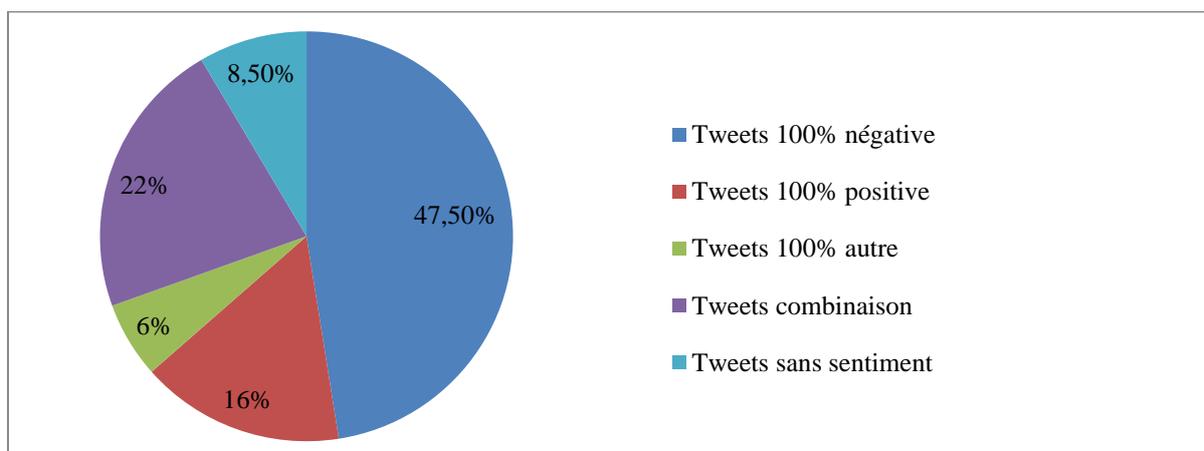
Le nombre de tweets neutres est relativement limité dans ce corpus. Toutefois, la majorité des tweets ne comportaient pas plus d'un ou deux sentiments et le nombre maximum de sentiments identifiés par tweet est cinq. La figure 2 illustre la distribution des sentiments dans le corpus de deux cents tweets.

Figure 2 : distribution des 335 sentiments exprimés dans le corpus de 200 tweets.



Considérant la possibilité qu'un tweet comporte plusieurs sentiments, une analyse manuelle a été faite afin de trouver le sentiment général exprimé dans tout tweet. La figure 3 présente la proportion des tweets dans le corpus, basée sur le sentiment global exprimé dans tout tweet.

Figure 3 : proportion des tweets dans le corpus selon le sentiment prépondérant dans tout tweet.



L'analyse a révélé que près de la moitié des tweets comportaient un sentiment 100 % négatif, 16 % des tweets dans le corpus étaient entièrement positifs et 22 % des tweets comportaient une combinaison de sentiments positifs et négatifs. Au total, 6 % des tweets ont été marqués de la polarité *autre*. Cela veut dire que le sentiment exprimé n'était ni positif, ni négatif, tel que la surprise, le doute ou le suspense. Parfois, l'étiquette *autre* a été attribuée parce que l'expression subjective consistait en un adjectif qualificatif dont la polarité était ambiguë tels que les mots « raar » (*étrange*), « grappig » (*drôle*), « gewaagd » (*osé*) et « pittig » (*vif*). Dans la plupart des cas, le contexte du tweet était trop limité afin de déterminer si l'adjectif comportait un sens négatif ou positif. Finalement, dans 8,5 % des tweets du corpus n'a été

exprimé aucun sentiment. Ces tweets ne comportaient que de l'information factuelle et vérifiable telle que des citations ou des faire-part.

Afin d'éclaircir la catégorie de tweets comportant une combinaison de sentiments, une analyse supplémentaire a été faite afin de trouver le sentiment prépondérant dans ces tweets. Une fois de plus, une analyse profonde de l'objet de tout sentiment exprimé s'est avérée nécessaire.

Tableau 9 : les tweets comportant une combinaison de sentiments.

Combinaison de sentiments	Nombre de tweets	Représentation dans le corpus (%)
50 % positif ; 50 % négatif (*)	15	7,5 %
33,3 % positif ; 66,7 % négatif	8	4 %
66,7 % positif ; 33,3 % négatif	6	3 %
25 % positif ; 75 % négatif	4	2 %
50 % négatif ; 50 % autre	4	2 %
33,3 % positif; 33,3 % négatif ; 33,3 % autre	2	1 %
50 % positif; 50 % autre	2	1 %
40 % positif; 60 % négatif	1	0,5 %
66,7 % positif; 33,3 % autre	1	0,5 %
66,7 % négatif ; 33,3 % autre	1	0,5 %

(\*) Dans cinq cas des quinze, le tweet en question ne comportait qu'un seul sentiment dont la polarité variait selon l'objet qu'il affectait.

La catégorie la plus importante comprend les tweets dans lesquels le sentiment exprimé était 50 % négatif et 50 % positif. Ces tweets constituent 7,5 % du total de tweets dans le corpus. L'astérisque dans la première catégorie indique que cinq tweets de cette catégorie comportaient un sentiment qui affectait plusieurs objets et qui avait une polarité différente selon l'objet affecté. Dans l'exemple (8), l'expression subjective « was interessanter dan » exprime un sentiment positif à l'égard du premier objet « twitterfeed over #hetgrotedebat » et un sentiment négatif à l'égard du deuxième objet « het debat zelf ».

(8)

- « Conclusie: twitterfeed over #hetgrotedebat was interessanter dan het debat zelf. »

L'analyse du sentiment prépondérant dans tout tweet a révélé que dans la plupart des tweets de notre corpus était exprimé un sentiment négatif à l'égard de l'objet. Ce résultat pourrait

mener à la conclusion que dans un contexte politique, Twitter est souvent utilisé afin d'exprimer des sentiments défavorables à l'égard de quelqu'un ou quelque chose et moins souvent pour faire partager des opinions positives. Toutefois, l'étendue du corpus ne permet pas de généraliser cette constatation pour tous les tweets qui ont été publiés en vue des élections.

### 7.1.2. Rapport entre le sexe de l'auteur et la polarité du sentiment exprimé

Du total de 335 sentiments exprimés, 191 ont été exprimés par des hommes et 144 par des femmes. Comme le révèle le tableau croisé dynamique ci-dessous, du total de 335 sentiments, 214 étaient négatifs, 96 étaient positifs et 25 ont été marqués de la polarité *autre* (il a été expliqué dans la section 7.1.1 ce que l'on entend par la polarité *autre*).

Tableau 10 : rapport entre le sexe de l'auteur du tweet et la polarité du sentiment exprimé.

Polarité du sentiment exprimé	Sexe		Total
	f	m	
Autre/négative/positive			
Autre	13	12	25
Négative	83	131	214
Positive	48	48	96
<b>Total</b>	<b>144</b>	<b>191</b>	<b>335</b>

Les résultats des hommes ne peuvent être comparés avec ceux des femmes que si le total de sentiments exprimés par les deux groupes est pris en considération. Considérant que les hommes ont exprimé 191 sentiments au total, l'on peut conclure qu'ils se sont plus souvent exprimés de façon négative (dans 68,6 % des cas) que les femmes (dans 57,6 % des cas). La différence est moins remarquable lorsque l'on compare les sentiments positifs exprimés par les hommes avec ceux des femmes. Dans 25,1 % des cas, les hommes dans le corpus ont exprimé un sentiment positif dans leur tweet alors que les femmes l'ont fait dans 33,3 % des cas. Les femmes ont exprimé plus de sentiments de polarité *autre* (dans 9 % des cas) que les hommes (dans 6,3 % des cas). De la même façon que l'analyse discutée dans la section 7.1.1, cette analyse a révélé que les sentiments négatifs prédominent dans le corpus.

### 7.1.3. Intensité des sentiments exprimés

Le logiciel d'annotation Brat<sup>22</sup> permet d'indiquer pour toute expression subjective l'intensité des sentiments exprimés. Ainsi, un sentiment peut être qualifié d'une intensité faible, moyenne ou grande. Cette section présente l'intensité indiquée pour tous les sentiments exprimés dans le corpus. Celle-ci a été attribuée en se basant sur les formes verbales, les adjectifs qualificatifs et les modificateurs dans les tweets.

Tableau 11 : rapport entre le sexe de l'auteur du tweet et l'intensité / la polarité du sentiment exprimé.

Sexe	Autre			Négative			Positive			total			
	faible	grande	moyenne	faible	grande	moyenne	faible	grande	moyenne				
f	1	6	6	13	4	33	46	83	4	16	28	48	144
m	2	1	9	12	12	56	63	131	1	18	29	48	191
<b>Total</b>	<b>3</b>	<b>7</b>	<b>15</b>	<b>25</b>	<b>16</b>	<b>89</b>	<b>109</b>	<b>214</b>	<b>5</b>	<b>34</b>	<b>57</b>	<b>96</b>	<b>335</b>

En se référant aux données du tableau croisé dynamique, l'on constate que la plupart des sentiments exprimés étaient négatifs (214) et que l'intensité la plus attribuée dans toute catégorie est *moyenne*. Une intensité *moyenne* veut dire que le sentiment exprimé n'est pas renforcé ni affaibli par un autre syntagme dans la phrase. Les expressions suivantes sont extraites de notre corpus et comportent toutes un sentiment de moyenne intensité : « een aangename debatnamiddag », « ik ben blij eens geen Antwerps debat te moeten aanhoren ». Dans ces expressions, les syntagmes soulignés ne sont ni renforcés, ni affaiblis par des éléments lexicaux tels que « zeer » et « heel ».

L'intensité *faible* est l'intensité la moins retrouvée dans le corpus. Elle implique souvent la présence des mots tels que « een beetje », « relatief » et « wat ». Les expressions suivantes du corpus contiennent des sentiments de faible intensité : « ik word altijd een beetje week van dat straatgebeuren op verkiezingszondag », « dat CD&V-verhaal moest nochtans pas klaar zijn ná de verkiezingen toch? ».

Les sentiments de forte intensité comportent souvent des mots tels que « zeer », « erg » et « echt ». Les expressions suivantes de notre corpus illustrent bien de cette catégorie de

<sup>22</sup> [Http://brat.nlplab.org/](http://brat.nlplab.org/).

sentiments : « ik kan echt zijn kop niet meer zien », « BDW is nu al moe van de kritische vragen ».

Considérant le nombre total de sentiments exprimés par les femmes et les hommes (respectivement 144 et 191), les femmes ont dans 38,2 % des cas opté pour un sentiment de grande intensité, dans 6,3 % des cas pour un sentiment de faible intensité et dans 55,6 % des cas pour un sentiment de moyenne intensité. Les hommes ont opté pour un sentiment fort dans 39,3 % des cas, pour un sentiment de faible intensité dans 7,9 % des cas et pour un sentiment de moyenne intensité dans 52,9 % des cas. En termes de pourcentage, les femmes ont exprimé plus de sentiments de moyenne intensité que les hommes alors que les hommes ont plus souvent opté pour des sentiments extrêmes de faible ou de grande intensité.

En général, la comparaison entre les sentiments exprimés par les femmes et les hommes ne révèle pas de différences remarquables, bien que cela ne ressorte pas très clairement du tableau 11. Présentant les résultats en pourcentages, le tableau 12 donne une meilleure vue de la comparaison hommes-femmes en ce qui concerne l'intensité des sentiments exprimés.

Tableau 12 : rapport entre le sexe de l'auteur et l'intensité / la polarité du sentiment exprimé en pourcentages.

sexe	Polarité autre			Polarité négative			Polarité positive			Total de sentiments
	Faible intensité	Grande intensité	Moyenne Intensité	Faible intensité	Grande intensité	Moyenne intensité	Faible intensité	Grande intensité	Moyenne intensité	
f	0,7%	4,2%	4,2%	2,8%	22,9%	31,9%	2,8%	11,1%	19,4%	144
m	1%	0,5%	4,7%	6,3%	29,3%	33%	0,5%	9,4%	15,2%	191

Il ressort du tableau 12 que la différence la plus remarquable est détectée dans la catégorie des sentiments négatifs de grande intensité. En effet, 29,3 % de tous les sentiments exprimés par les hommes sont classés dans cette catégorie, alors que seulement 22,9 % des sentiments exprimés par les femmes y appartiennent. Une autre différence assez remarquable entre les hommes et les femmes est à situer dans la catégorie de sentiments positifs de moyenne intensité.

#### 7.1.4. Explicitation des sentiments exprimés

Outre l'intensité du sentiment exprimé dans un tweet, le logiciel d'annotation permet d'indiquer pour tout sentiment son niveau d'explicitation. Un sentiment est considéré comme implicite si la phrase dans laquelle il est exprimé ne contient que de l'information objective ou factuelle.

Tableau 13 : *sentiments explicites et implicites dans le corpus.*

<b>Explicitation</b>	<b>Sexe</b>		<b>Total</b>
	<b>f</b>	<b>m</b>	
Explicite	141	186	327
Implicite	3	5	8
<b>Total</b>	<b>144</b>	<b>191</b>	<b>335</b>

Le tableau croisé dynamique révèle que seulement huit sentiments dans le corpus ont été exprimés de façon implicite, dont cinq ont été exprimés par des hommes et trois par des femmes. Traduits en pourcentages, les résultats indiquent que la différence entre les hommes et les femmes est infime en ce qui concerne l'explicitation de leurs sentiments : les femmes ont choisi dans 2,1 % des cas de ne pas expliciter le sentiment qu'ils voulaient exprimer alors que les hommes ont préféré les sentiments tacites aux sentiments explicites dans 2,6 % des cas.

Il ressort du tableau 13 que peu de sentiments implicitement exprimés ont été retrouvés dans le corpus de deux cents tweets. Dans le but de trouver une possible explication à cette constatation, il importe de remarquer que Twitter est un réseau social qui fait partie des nouveaux médias. Ce qui caractérise les nouveaux médias, c'est qu'ils sont accessibles pour tous et qu'ils constituent des moyens pour favoriser la communication de masse dont les émetteurs sont aussi nombreux que les récepteurs. En outre, ils ne font pas encore partie des moyens de communication officiels. Une conséquence logique est que le langage utilisé sur les réseaux sociaux est beaucoup plus informel et plus direct que le langage des médias traditionnels tels que la radio et la télévision. De plus, la politique est un sujet qui a soulevé pendant les dernières années une vive controverse dans notre pays. Cela pourrait mener à ce que les gens aient des opinions tranchées sur le sujet qui peuvent difficilement être exprimées de façon neutre et objective. En conclusion, il est possible que tant le caractère informel des tweets que le sujet controversé favorisent l'explicitation des sentiments.

### 7.1.5. Modificateurs

Les différents types de modificateurs ont été présentés dans la section 3.4.2. Les modificateurs constituent les éléments lexicaux qui altèrent le sentiment exprimé dans une phrase. Cela implique que, s'ils étaient enlevés de la phrase, le sentiment y exprimé serait (parfois légèrement) différent. Lors de l'analyse, tous les modificateurs ont été identifiés dans tout tweet afin de révéler le nombre total de modificateurs dans le corpus. Le sexe des auteurs des tweets a également été pris en considération afin de pouvoir détecter un possible rapport entre le sexe de l'auteur du tweet et le nombre et le type de modificateur(s) que le tweet comporte. Le résultat de l'analyse est présenté dans le tableau croisé dynamique ci-dessous.

Tableau 14 : modificateurs utilisés par les hommes et les femmes dans le corpus.

Sexe	Type											Total
	Affaiblissement	Autre	Avenir	Condition	Renforcement	Modalité	Négation	Passé	Perspective	Question	Spécification	
f	17	11	4	2	73	1	7	6	1	4		126
m	14	12	10	3	72	3	14	8		5	2	143
<b>Total</b>	<b>31</b>	<b>23</b>	<b>14</b>	<b>5</b>	<b>145</b>	<b>4</b>	<b>21</b>	<b>14</b>	<b>1</b>	<b>9</b>	<b>2</b>	<b>269</b>

Le 14 illustre qu'au total, 269 modificateurs ont été détectés dans le corpus. Tous les modificateurs ont été classifiés en onze catégories différentes dont la catégorie la plus étendue est celle du renforcement. En effet, du total de 269 modificateurs identifiés, plus de la moitié avaient une fonction de renforcement. Dans ce cas-là, les modificateurs étaient très souvent des mots tels que « heel », « erg », « nog » et les points d'exclamation. Considérant le nombre total de modificateurs employés par les hommes et les femmes, l'on constate que les femmes avaient plus tendance à ajouter à leur sentiment un modificateur de renforcement (dans 57,9 % des cas) que les hommes (dans 50,4 % des cas).

La deuxième catégorie la plus importante est celle de l'affaiblissement. La catégorie comprend les modificateurs qui atténuent le sentiment exprimé dans la phrase tels que les mots « een beetje », « tamelijk », « relatief » et « wat ». Les modificateurs d'affaiblissement sont, de la même façon que les modificateurs de renforcement, plus souvent utilisés par les femmes du corpus : ils constituent 13,5 % du total de modificateurs utilisés par les femmes alors qu'ils ne constituent que 9,8 % du total de modificateurs utilisés par les hommes.

La troisième catégorie la plus significative, c'est la catégorie qui comprend les modificateurs du type *autre*. Ces modificateurs n'appartiennent à aucune des autres dix catégories et leur fonction est plutôt phatique. Bien que leur signification soit parfois ambiguë, ils sont pris en considération parce que leur présence altère effectivement le sentiment exprimé dans le tweet. Quelques exemples de ces modificateurs sont des interjections telles que « *eigenlijk* », « *zoals verwacht* » et « *ach ja* ». Ils ont plus souvent été utilisés par les femmes de notre corpus que par les hommes.

Nous ne discuterons plus en détail les autres catégories de modificateurs. Prenant en considération le nombre total de sentiments exprimés par les femmes et les hommes, respectivement 144 et 191 (voir le tableau 14), les femmes avaient généralement plus tendance à ajouter des modificateurs que les hommes. Dans 87,5 % des sentiments exprimés, elles ont ajouté au moins un modificateur. En revanche, les hommes ont ajouté des modificateurs à 74,9 % des sentiments exprimés.

En général, l'analyse a révélé que 269 modificateurs ont été retrouvés dans le total de 200 tweets. Ce résultat implique une moyenne de 1,3 modificateurs par tweet. Au total, 66 tweets du corpus ne comportaient aucun modificateur. Prenant en considération les 17 tweets neutres (*supra*), seulement 49 des tweets comportant au moins un sentiment ne contenaient aucun modificateur. Traduite en pourcentages, cette constatation implique que 73,2 % de tous les tweets non neutres dans le corpus contenaient au moins un modificateur.

#### 7.1.6. Sujets abordés

L'étendue du premier corpus s'avère trop limitée afin de tirer des conclusions représentatives sur les principaux sujets abordés sur Twitter dans un contexte politique. Toutefois, il était intéressant d'examiner de quels thèmes parlaient principalement les femmes dans le corpus et de quels thèmes parlaient les hommes. Les tableaux 15 et 16 présentent les dix sujets les plus abordés respectivement par les femmes et les hommes du corpus et la polarité du sentiment exprimé à l'égard des sujets.

Tableau 15 : sujets les plus abordés par les femmes et le sentiment exprimé à leur égard.

Classement	Sujet	Polarité du sentiment
1.	Les élections communales et provinciales	71,4 % négative 28,3 % positive
2.	Bart De Wever (le président du parti N-VA)	85,7 % négative 14,3 % positive
3.	Het Grote Debat (l'émission-débat de la VRT qui a été diffusée le 7 octobre 2012)	100 % négative
4.	Les affiches de campagne dans la rue	71,4 % négative 28,6 % positive
5.	Open Vld	20 % négative 80 % positive
6.	Wouter Beke (le président du parti CD&V)	80 % négative 20 % positive
7.	Vlaams Belang	100 % négative
8.	CD&V	40 % négative 60 % positive
9.	Vincent Van Quickenborne (membre du parti Open Vld)	75 % négative 25 % positive
10.	Wouter Van Besien (le président du parti Groen)	25 % négative 75 % positive

Tableau 16: sujets les plus abordés par les hommes et le sentiment exprimé à leur égard.

Classement	Sujet	Polarité du sentiment
1.	Bart De Wever	76,9 % négative 23,1 % positive
2.	Groen	60 % négative 40 % positive
3.	Wouter Beke	100 % négative
4.	Open Vld	66,7 % négative 33,3 % positive
5.	N-VA	73,3 % négative 26,7 % positive
6.	Sp.a	66,7 % négative 33,3 % positive
7.	Bruno Tobback (le président du parti sp.a)	100 % négative
8.	Les élections communales et provinciales	100 % négative
9.	Vincent Van Quickenborne	75 % négative 25 % positive
10.	Alexander De Croo (ancien président du parti Open Vld)	100 % négative

Prenant en considération les deux tableaux, une comparaison entre les hommes et les femmes révèle quelques différences en ce qui concerne la polarité du sentiment exprimé à l'égard de ces sujets. Ainsi, lorsque les femmes ont parlé des élections, elles étaient plus souvent positives (dans 28,3 % des cas) que les hommes qui n'ont rien dit de positif là-dessus. Le parti Open Vld a obtenu des femmes plus d'évaluations positives que négatives alors que les hommes ont généralement exprimé un sentiment négatif à l'égard du parti (dans 66,7 % des cas). Dans les deux groupes, les évaluations de Vincent Van Quickenborne étaient négatives dans 75 % des cas. Un autre homme politique qui a souvent été évoqué dans le corpus est Bart De Wever. Il ressort des tableaux 15 et 16 que les hommes avaient plus tendance à dire des choses positives du président de la N-VA (dans 23,1 % des cas) que les femmes (dans 14,3 % des cas). En revanche, les femmes ont exprimé plus de sentiments positifs à l'égard de Wouter Beke (dans 20 % des cas) que les hommes (dans aucun cas).

Toutefois, il importe de souligner une fois de plus que l'étendue du corpus s'avère trop limitée afin de tirer des conclusions représentatives sur les sujets souvent abordés dans des contextes politiques. Le deuxième corpus comportant 4 629 tweets est plus approprié pour y baser une telle analyse. En outre, le deuxième corpus ne comporte que des tweets contenant le mot-dièse #vk2012. Etant très général, ce mot-dièse couvre toute une série de sujets politiques sans se limiter à quelques noms de partis ou de personnages politiques, comme il était le cas dans ce premier corpus.

#### 7.1.7. Evaluation du schéma d'annotation

L'analyse des sentiments du premier corpus a révélé les possibilités qu'offre le schéma d'annotation de Van de Kauter et Desmet (2012). Il est également ressorti de l'analyse que le schéma comporte quelques extensions essentielles par rapport au schéma d'annotation de base conçu par Wiebe (2002).

En premier lieu, le schéma de Van de Kauter et Desmet (2012) permet d'identifier les sentiments implicitement exprimés, c'est-à-dire, les sentiments exprimés dans des textes qui ne comportent que de l'information objective ou factuelle. Afin d'éclaircir encore davantage la principale différence entre le schéma d'annotation de Wiebe (2002) et celui de Van de Kauter et Desmet (2012), il importe d'expliquer la notion « implicite ». Suivant Wiebe

(2002), un sentiment est considéré comme implicite s'il n'est pas introduit par une expression linguistique qui le lie à sa source. Ainsi, une phrase telle que « Quel homme insupportable ! » est considérée comme contenant un sentiment implicite, car elle manque d'une expression linguistique telle que « .. dit-il ». Toutefois, suivant le schéma d'annotation de Van de Kauter et Desmet (2012), ce type de phrase est considéré comme l'expression explicite d'un sentiment, étant donné la présence du mot subjectif « insupportable ». Dans ce schéma, les sentiments qui sont considérés comme implicites sont les sentiments qui ne sont pas explicitement mentionnés, sinon suggérés dans des phrases factuelles et objectives. Le tweet suivant de notre corpus est considéré comme comportant un sentiment implicitement exprimé :

- « @GoedeleWachters Wat u ook niet zach (*sic*) op het scherm...Kleinere partijen zoals Pvd+ #hetgrote debat ».

En effet, la phrase comporte de l'information factuelle ; il peut être vérifié si le PVDA était ou n'était pas présent dans l'émission-débat *Het Grote Debat* et si le parti est plus petit que les autres partis flamands. Toutefois et d'une manière subtile, la phrase exprime un sentiment négatif à l'égard de l'émission-débat *Het Grote Debat* auquel les représentants du PVDA n'avaient apparemment pas été invités. En conclusion, contrairement au schéma de Wiebe (2002), le schéma d'annotation utilisé lors de cette étude ne se borne pas à l'analyse de phrases subjectives ; il prend également en considération les phrases factuelles et objectives qui peuvent comporter des sentiments implicitement exprimés.

En deuxième lieu, le schéma de Van de Kauter et Desmet (2012) permet une nette distinction entre trois types d'expressions subjectives : les expressions *private state*, les expressions *polar fact* et les expressions *polar resultative causative*. En revanche, le schéma d'annotation de Wiebe (2002) n'avait été conçu que pour l'annotation d'expressions *private state*. Présentant les différentes formes que peuvent adopter les expressions subjectives, le nouveau schéma rend l'identification des expressions subjectives plus facile.

En troisième lieu, le nouveau schéma d'annotation présente une liste considérable de modificateurs qui permet d'identifier tous les syntagmes qui altèrent le sentiment exprimé dans une phrase. Au total, onze types de modificateurs sont distingués (voir la section 3.4.2). Finalement, le manuel d'annotation de Van de Kauter et Desmet (2012) comprend une

délimitation des désignations dans les catégories *intensité du sentiment* et *polarité du sentiment*. Ainsi, les sous-groupes *intensité extrême* et *aucune polarité* qui avaient été ajoutés par Wiebe (2002) et Wilson (2008) ont été supprimés à cause de leur caractère ambigu. Ainsi, toutes les modifications au schéma d'annotation de base de Wiebe (2002) ont contribué à une analyse des sentiments plus détaillée et davantage univoque.

En général, le schéma d'annotation de Van de Kauter et Desmet (2012) a permis de procéder à une analyse profonde de notre corpus de tweets politiques. Les résultats l'analyse des sentiments permettent de tirer des conclusions sur le sentiment global exprimé dans tout tweet, le nombre de sentiments exprimés par tweet, la polarité et l'intensité du sentiment ainsi que son niveau d'explicitation et la fréquence avec laquelle les différents types de modificateurs ont été utilisés dans le corpus.

En conclusion, le schéma d'annotation de Van de Kauter et Desmet (2012) a permis une annotation manuelle détaillée des sentiments exprimés dans le premier corpus. Toutefois, le schéma d'annotation de Van de Kauter et Desmet (2012) pourrait permettre une analyse encore davantage précise des sentiments si la catégorisation dans certaines phases de la procédure d'annotation était encore plus nette. Il est par exemple conseillé de ne prévoir que deux catégories pour indiquer la polarité d'un sentiment : *positive* et *négative*. En revanche, le présent schéma d'annotation prévoit également une troisième catégorie : *autre polarité*. Toutefois, l'ambiguïté de cette catégorie rend l'analyse globale des résultats plus équivoque.

La même remarque vaut pour la catégorie des modificateurs. Bien que la subdivision *autres modificateurs* ait un caractère équivoque, elle représente une partie considérable (8,6 %) du total de modificateurs retrouvés dans le corpus. Etant donné l'ambiguïté du sous-groupe, il vaut mieux ajouter encore davantage de subdivisions bien déterminées à la catégorie générique *modificateurs* de façon à réduire l'étendue du sous-groupe *autres modificateurs*. La majorité des modificateurs classés dans ce sous-groupe constituent des interjections et des éléments de concession ou d'opposition. Par conséquent, ces trois sous-groupes pourraient être ajoutés à la catégorie générique *modificateurs*.

De la même façon, la classification du type d'expression subjective pourrait être désambiguïsée davantage si sur l'échelle de subjectivité n'étaient indiquées que trois valeurs différentes : la valeur *objective* (désignée par le chiffre « 0 ») pour les expressions *polar fact*,

la valeur *subjective* (désignée par le chiffre « 3 ») pour les expressions *private state* et la valeur *causale* pour les expressions *polar resultative causative*. D'après nos expériences, les catégories « 1 » et « 2 » désignant les expressions subjectives qui ne sont ni véritablement des expressions *polar fact* ni des expressions *private state* rendent l'analyse plus équivoque.

Un dernier élément du schéma d'annotation de Van de Kauter et Desmet (2012) que nous avons considéré comme une faiblesse, c'est l'annotation des expressions subjectives qui sont incorporées dans d'autres expressions subjectives. Suivant le schéma, une deuxième expression subjective qui se situe à un niveau plus complexe de la phrase que la première ne peut être annotée que si elle constitue toujours une évaluation subjective après que la première expression en est enlevée (voir la section 3.3.7). Par exemple ; dans la phrase « Mooi betoog van #Groen », l'expression subjective « mooi » exprime un sentiment positif à l'égard de l'objet « betoog ». A un niveau plus complexe de la phrase, l'on tendrait à dire que le syntagme « mooi betoog van » exprime un sentiment positif à l'égard de « Groen ». Toutefois, suivant le schéma d'annotation de Van de Kauter et Desmet (2012), le syntagme ne peut pas être annoté de cette manière, étant donné que la suppression de l'expression « mooi », rendrait la phrase neutre. Néanmoins, il nous paraît essentiel d'indiquer que la phrase comporte tant une évaluation positive de l'exposé (« het betoog ») du parti Groen que du parti politique lui même.

## 7.2. Résultats de la mesure de fréquence des partis et personnages politiques

Cette section présente les résultats de l'analyse du deuxième corpus comportant 4 629 tweets collectés entre le 21 septembre et le 15 octobre 2012. Tous les tweets dans le corpus sont néerlandophones et comportent le mot-dièse #vk2012.

En premier lieu, il a été examiné quels partis et hommes politiques sont les plus évoqués dans le corpus et quel sentiment est généralement exprimé à leur égard. Etant donné la vive controverse autour du maïorat d'Anvers, une comparaison a été établie entre la popularité du candidat Bart De Wever et celle du candidat Patrick Janssens, afin de vérifier si le gagnant était également la personne la plus populaire dans des tweets du corpus. Il a également été établie une comparaison entre les tweets postés par des femmes et ceux publiés par des hommes afin de comparer les sentiments exprimés par les deux groupes.

En deuxième lieu, les sentiments exprimés à l'égard des personnages et partis politiques dans le corpus ont été comparés avec le résultat des élections communales et provinciales de 2012. L'objectif était de vérifier si Twitter peut refléter le sentiment politique de l'électorat.

### 7.2.1. Analyse des tweets publiés par les hommes

Près de la moitié des tweets publiés par les hommes (1 285 du total de 2 908) mentionnaient un des partis ou des personnages politiques évoqués dans le paragraphe précédent. Les tableaux 17 et 18 présentent pour tout parti et tout personnage politique le total de mentions trouvées dans les tweets des hommes et le nombre de tweets positifs et négatifs sur l'objet en question. Parfois, une mention était neutre, comme l'exemple suivant : « Vlaams Belang #Buggenhout stelt zijn kandidaten voor ». Dans ce cas-là, aucune conclusion n'a pu être déduite de la préférence politique de l'auteur du tweet.

Tableau 17: total de mentions des principaux partis politiques flamands dans les tweets des hommes.

<b>Parti politique</b>	<b>Mentions positives</b>	<b>Mentions négatives</b>	<b>Mentions neutres</b>	<b>Total de mentions</b>
N-VA	82	115	64	261
Groen	96	23	46	165
CD&V	43	55	35	133
Sp.a	44	54	26	124
Open Vld	40	46	33	119
Vlaams Belang	12	31	16	59
PVDA	31	15	11	57
LDD	3	10	0	13

Tableau 18 : total de mentions de onze personnages politiques flamands dans les tweets des hommes.

<b>Personnage politique</b>	<b>Mentions positives</b>	<b>Mentions négatives</b>	<b>Mentions neutres</b>	<b>Total de mentions</b>
Bart De Wever	36	110	25	171
Patrick Janssens	11	36	21	68
Peter Mertens	27	14	1	42
Wouter Beke	2	22	5	29
Jean-Marie Dedecker	2	8	2	12
Bruno Tobback	2	7	1	10
Alexander De Croo	3	5	0	8
Wouter Van Besien	5	0	2	7
Bruno Valkeniers	0	2	1	3
Gerolf Annemans	0	1	2	3

Gwendolyn Rutten	0	1	0	1
------------------	---	---	---	---

Il ressort du tableau 17 que les partis politiques les plus mentionnés par les hommes sont la N-VA, Groen et le CD&V. Tenant compte des mentions positives et négatives de tout parti, l'analyse a révélé que les partis politiques les plus populaires<sup>23</sup> auprès les hommes du corpus sont Groen, le PVDA et l'Open Vld. En effet, ce sont les partis dont le total d'évaluations positives était, en termes de pourcentage, le plus élevé. Les partis Vlaams Belang et LDD sont les moins populaires, étant donné que leurs évaluations étaient généralement négatives (dans respectivement 72 % et 77 % des cas).

Le tableau 18 révèle que les personnages politiques les plus évoqués par les hommes sont Bart De Wever, Patrick Janssens et Peter Mertens. Tenant compte des sentiments attribués à tout personnage, l'on constate que les personnages politiques les plus populaires auprès les hommes sont Wouter Van Besien, Peter Mertens et Alexander De Croo. Bruno Valkeniers, Gerolf Annemans et Gwendolyn Rutten sont les moins populaires, n'ayant reçu aucune évaluation positive.

Une prédiction sur le maïorat d'Anvers révélerait que Bart De Wever est plus apprécié par les hommes que Patrick Janssens, étant donné qu'il a obtenu plus d'évaluations positives que ce dernier. Toutefois, la différence est infime (25 % par opposition à 23 %).

### 7.2.2. Analyse des tweets publiés par les femmes

Près de 40 % des tweets publiés par les femmes (249 du total de 661) comportaient un parti ou personnage politique mentionné dans la liste. Les tableaux 19 et 20 présentent le total de références trouvées dans les tweets des femmes.

Tableau 19 : mentions des principaux partis politiques flamands dans les tweets des femmes.

Parti politique	Mentions positives	Mentions négatives	Mentions neutres	Total de mentions
Groen	48	7	26	81

<sup>23</sup> La popularité d'un parti ou d'un personnage politique a été calculée en comparant les mentions positives avec les mentions négatives : les mentions positives et négatives ont été divisées par la somme des deux catégories afin d'obtenir un pourcentage. Les partis et les personnages politiques dont le pourcentage de mentions positives était le plus élevé, sont considérés comme les plus populaires. La catégorie *neutre* n'a pas été prise en considération pour ce calcul.

N-VA	14	11	11	36
CD&V	4	9	15	28
Open Vld	11	5	7	23
Sp.a	5	1	5	11
Vlaams Belang	0	5	2	7
PVDA	1	1	1	3
LDD	0	3	0	3

Tableau 20 : mentions de onze personnages politiques flamands dans les tweets des femmes.

Personnage politique	Mentions positives	Mentions négatives	Mentions neutres	Total de mentions
Bart De Wever	2	12	4	18
Patrick Janssens	2	5	4	11
Wouter Van Besien	2	0	6	8
Wouter Beke	1	3	2	6
Peter Mertens	3	1	1	5
Bruno Tobback	0	1	2	3
Alexander De Croo	1	1	1	3
Bruno Valkeniers	0	1	1	2
Gwendolyn Rutten	0	0	1	1
Jean-Marie Dedecker	0	0	0	0
Gerolf Annemans	0	0	0	0

Il ressort du tableau 19 que Groen, la N-VA et le CD&V sont les partis politiques les plus mentionnés par les femmes. L'évaluation des partis Groen, sp.a et Open Vld étant positive dans respectivement 87 %, 83 % et 69 % des cas, l'on pourrait constater que ce sont les partis les plus populaires auprès les femmes du corpus. En revanche, le Vlaams Belang et la LDD sont les partis les moins populaires, n'ayant reçu aucune évaluation positive.

Le tableau 20 révèle que Bart De Wever, Patrick Janssens et Wouter Van Besien sont les hommes politiques les plus évoqués par les femmes. Les évaluations de Wouter Van Besien, Peter Mertens et Alexander De Croo étaient généralement positives (dans respectivement 100 %, 75 % et 50 % des cas), ce qui démontre leur popularité auprès les femmes du corpus. Aucune mention positive n'a été retrouvée de Bruno Valkeniers, Gwendolyn Rutten, Jean-Marie Dedecker et Gerolf Annemans.

Considérant le tableau 20, une prédiction sur le maïorat d'Anvers révélerait que Patrick Janssens est le candidat de prédilection des femmes. En effet, 29 % de ses évaluations étaient positives, alors que seulement 14 % des évaluations de Bart De Wever étaient positives.

### 7.2.3. Analyse des autres tweets dans le corpus

Près de la moitié (323 du total de 677) des autres tweets du corpus comportaient des références aux partis et personnages politiques mentionnés au début du chapitre. Parmi ces 677 tweets se situent les 640 tweets dont le sexe de l'auteur n'a pas pu être identifié et les 37 tweets qui provenaient d'un profil inexistant ou non retrouvé. Les résultats de l'analyse sont présentés dans les tableaux 21 et 22.

*Tableau 21 : mentions des principaux partis politiques flamands dans les tweets dont le sexe de l'auteur est inconnu.*

<b>Parti politique</b>	<b>Mentions positives</b>	<b>Mentions négatives</b>	<b>Mentions neutres</b>	<b>Total de mentions</b>
N-VA	33	10	37	80
Groen	28	2	46	76
Sp.a	4	3	26	33
CD&V	8	13	11	32
Open Vld	12	0	11	23
Vlaams Belang	7	7	6	20
PVDA	0	0	5	5
LDD	1	0	1	2

*Tableau 22 : mentions de onze politiques flamands dans les tweets dont le sexe de l'auteur est inconnu.*

<b>Personnage politique</b>	<b>Mentions positives</b>	<b>Mentions négatives</b>	<b>Mentions neutres</b>	<b>Total de mentions</b>
Bart De Wever	7	7	6	20
Peter Mertens	5	0	3	8
Patrick Janssens	2	3	2	7
Wouter Beke	2	2	0	4
Jean-Marie Dedecker	2	0	2	4
Wouter Van Besien	0	1	3	4
Bruno Tobback	0	2	0	2
Alexander De Croo	0	1	1	2
Gerolf Annemans	0	0	1	1
Bruno Valkeniers	0	0	0	0
Gwendolyn Rutten	0	0	0	0

Il ressort du tableau 21 que la N-VA, Groen et le sp.a sont les partis politiques les plus évoqués dans cette partie du corpus. Les partis les plus populaires sont le PVDA, la LDD et Groen, étant donné que la majorité de leurs appréciations étaient positives (respectivement 100 %, 100 % et 93 %). Toutefois, il importe de nuancer l'idée que le sentiment exprimé à l'égard de la LDD était 100 % positive, tenant compte du fait qu'il n'a été trouvée qu'une seule mention subjective du parti. Les partis les moins évoqués sont la LDD et le PVDA et les partis les moins populaires sont le Vlaams Belang et le CD&V.

Les personnages politiques les plus évoqués sont Bart De Wever, Peter Mertens et Patrick Janssens. De la même façon que l'analyse des tweets des hommes, l'analyse de cette partie du corpus révèle que Bart De Wever est plus populaire que Patrick Janssens. En revanche, la comparaison entre la popularité des candidats au maïorats d'Anvers dans les tweets des femmes avait révélé que celles-ci ont préféré Patrick Janssens à Bart De Wever.

#### 7.2.4. Analyse globale du corpus

Pour l'analyse globale du corpus, tous les résultats des mesures de fréquence dans les trois catégories ont été pris en considération. Ils sont résumés dans les tableaux 23 et 24.

Tableau 23 : total mentions des principaux partis politiques dans le corpus.

<b>Parti politique</b>	<b>Mentions positives</b>	<b>Mentions négatives</b>	<b>Mentions neutres</b>	<b>Total de mentions</b>
N-VA	129	136	112	377
Groen	172	32	118	322
CD&V	55	77	61	193
Sp.a	53	58	57	168
Open Vld	63	51	51	165
Vlaams Belang	19	43	24	86
PVDA	32	16	17	65
LDD	4	13	1	18

Tableau 24 : total de mentions de onze personnages politiques dans le corpus.

<b>Personnage politique</b>	<b>Mentions positives</b>	<b>Mentions négatives</b>	<b>Mentions neutres</b>	<b>Total de mentions</b>
Bart De Wever	45	129	35	209
Patrick Janssens	15	44	27	86
Peter Mertens	35	15	5	55

Wouter Beke	5	27	7	39
Wouter Van Besien	7	1	11	19
Jean-Marie Dedecker	4	8	4	16
Bruno Tobback	2	10	3	15
Alexander De Croo	4	7	2	13
Bruno Valkeniers	0	3	2	5
Gerolf Annemans	0	1	3	4
Gwendolyn Rutten	0	1	1	2

L'analyse globale a révélé que la N-VA, Groen et le CD&V sont les partis politiques les plus mentionnés dans le corpus. A l'exception des partis Groen, Open Vld et PVDA, tous les partis politiques ont obtenu plus d'évaluations négatives que positives dans le corpus. Tenant compte du fait que 84 % des évaluations de Groen étaient positives, ce parti peut être considéré comme le parti le plus populaire dans le corpus. En revanche, le Vlaams Belang et la LDD sont les partis les moins populaires, étant donné que leurs appréciations étaient généralement négatives.

Il ressort du tableau 24 que les personnages politiques les plus évoqués dans le corpus sont Bart De Wever, Patrick Janssens et Peter Mertens. En termes de pourcentages, Wouter Van Besien a reçu le plus grand nombre d'évaluations positives (88 %). Par conséquent, il peut être considéré comme l'homme politique le plus populaire dans le corpus. En revanche, tous les sentiments exprimés à l'égard de Bruno Valkeniers, Gerolf Annemans et Gwendolyn Rutten étaient négatifs, ce qui entraîne la conclusion que ce sont les personnes les moins populaires du corpus. Pourtant, il importe de nuancer cette constatation en signalant que le nombre de mentions de ces personnes était infime. Wouter Beke et Bruno Tobback peuvent également être considérés comme deux personnages peu populaires : respectivement 84 % et 83 % des références à eux comportaient un sentiment négatif.

Une comparaison entre les deux candidats au maïorat d'Anvers a révélé une légère prédilection pour Bart De Wever (26 % d'évaluations positives) au détriment de Patrick Janssens (25 % d'évaluations positives).

Le tableau 25 illustre pour tout groupe de tweets du corpus la proportion de mentions positives, négatives et neutres.

Tableau 25 : mentions positives, négatives et neutres dans le corpus.

Auteur des tweets	% de mentions positives	% de mentions négatives	% de mentions neutres
Femmes	31,3 %	32,1 %	36,8 %
Hommes	31,3 %	47,9 %	20,9 %
Sexe inconnu	34,5 %	21,9 %	43,7 %

Il ressort du tableau 25 que la plupart des références aux partis et personnages politiques étaient neutres dans les tweets des femmes et des tweets dont le sexe de l'auteur est inconnu. Ces mentions neutres consistaient souvent en des annonces objectives et factuelles telles que « Groen Waarschoot op campagne met tolk Vlaamse Gebarentaal ». En revanche, les tweets des hommes comportaient surtout des évaluations négatives.

La proportion de mentions neutres est la plus significative dans le groupe de tweets dont le sexe de l'auteur est inconnu. Cela pourrait s'expliquer par le fait que beaucoup de tweets dans ce groupe ont été publiés par des organisations neutres telles que *Actua TV* et *VRT Journaal* ou par des partis régionaux tels que *sp.a Wichelen* et *N-VA Maarkedal*. Une partie considérable de ces tweets comportaient de l'information neutre telle que des faire-part et des programmes politiques.

#### 7.2.5. Comparaison de l'analyse avec les résultats des élections de 2012

Les résultats de l'analyse du deuxième corpus étant connus, ils peuvent être comparés avec les résultats des élections communales et provinciales d'octobre 2012. Le tableau 26 présente les résultats que les huit principaux partis flamands ont obtenu dans les cinq provinces de Flandre. Bruxelles n'étant pas de province, elle n'est pas prise en considération.

Tableau 26 : résultat des huit principaux partis flamands lors des élections de 2012.<sup>24</sup>

Province	N-VA	CD&V	sp.a	Groen	Open Vld	PVDA+ *	Vlaams Belang	LDD **
Anvers	35,9 %	16,8 %	12,8 %	9,3 %	10,1 %	3,4 %	10,9 %	-
Limbourg	26,1 %	27,5 %	20,1 %	20,1 %	14,1 %	2,2 %	9,1 %	-
Flandre-Orientale	26,1 %	19,8 %	12,7 %	9 %	19,3 %	1,7 %	9,3 %	-
Brabant flamand	25,8 %	19,5 %	12,1 %	9,6 %	16,8 %	1,2 %	6,7 %	-

<sup>24</sup> Source: <http://www.vlaanderenkiest.be/verkiezingen2012/index.html>.

Flandre-Occidentale	25,3 %	27,6 %	15,8 %	7,5 %	13,4 %	1,3 %	7,7 %	-
---------------------	--------	--------	--------	-------	--------	-------	-------	---

\* La liste du PVDA s'appelle *PVDA+* parce sur la liste peuvent figurer des personnes qui ne sont pas membres du parti. \*\* Les résultats de la LDD ne sont pas mentionnés sur le site web.

Il ressort du tableau 26 que la N-VA et le CD&V ont obtenu le plus grand nombre de voix lors des élections en Flandre. La N-VA était le parti le plus grand à Anvers, en Flandre-Orientale et en Brabant flamand, suivi par le CD&V. La situation est inverse en Flandre-Occidentale et à Limbourg, où le CD&V est ressorti comme le parti le plus grand, suivi par la N-VA. Les résultats de notre analyse avaient révélé que la N-VA et le CD&V étaient deux des partis les plus mentionnés sur Twitter en vue des élections. Toutefois, ce n'étaient pas les partis les plus populaires, étant donné que leur évaluation était généralement négative.

Il est remarquable que les résultats de Groen et du PVDA se situent sensiblement plus bas que ceux des autres partis à l'exception de la LDD et du Vlaams Belang. Pourtant, l'analyse du deuxième corpus avait révélé que ces partis se sont avérés, avec l'Open Vld, les plus populaires, ayant reçu une évaluation généralement plus positive que les autres partis. Afin de mieux comprendre cet écart, les résultats des élections de 2012 ont été comparés avec ceux des élections de 2006 (voir la section 10.2 des annexes). Cette comparaison a révélé que les partis Groen et PVDA sont les seuls partis flamands (à l'exception de la N-VA à Anvers) à avoir enregistré une augmentation de voix dans presque toutes les provinces (Groen a seulement perdu quelques voix dans la province de Flandre-Orientale) par comparaison aux élections de 2006. Cette comparaison explique mieux les résultats obtenus lors de l'analyse du corpus. En effet, les sentiments positifs exprimés à l'égard des partis Groen et PVDA dans notre corpus correspondent à leur popularité augmentée par comparaison aux élections de 2006.

Le tableau 26 révèle que les partis qui ont obtenu le moins de voix lors des élections sont la LDD, qui n'a obtenu un résultat considérable qu'à Middelkerke<sup>25</sup>, le PVDA et le Vlaams Belang. En effet, il est ressorti de notre analyse que la LDD et le Vlaams Belang étaient les partis les moins populaires. En revanche, le parti PVDA était l'un des partis les plus populaires dans le corpus. Cet écart entre le résultat de notre analyse et le résultat des élections peut être atténué en considérant la sensible augmentation de voix que le parti PVDA

<sup>25</sup> Dedecker: "LDD niet van de kaart geveegd". (14.10.2012). *De Morgen*. [En ligne] <http://www.demorgen.be/dm/nl/3625/Verkiezingen-2012/article/detail/1516811/2012/10/14/Dedecker-LDD-niet-van-de-kaart-geveegd.dhtml> [24.04.2013].

a enregistré en octobre 2012 par comparaison aux élections de 2006. De ce point de vue, le PVDA a obtenu de bons résultats lors des dernières élections, ce qui correspond aux nombreuses évaluations positives du parti dans notre corpus.

Il s'avère difficile de comparer la popularité des personnages politiques telle qu'elle ressort du corpus avec leur popularité dans la vie réelle. Toutefois, des trois événements suivants peut être analysé en quelle mesure ils pourraient avoir été prédits à partir de notre corpus :

1. La succession de Bart De Wever à Patrick Janssens comme bourgmestre d'Anvers.
2. La succession de Gwendolyn Rutten à Alexander De Croo comme présidente de l'Open Vld.
3. La succession de Gerolf Annemans à Bruno Valkeniers comme président du Vlaams Belang.

Une comparaison entre les deux candidats au maïorats d'Anvers a révélé qu'en général, Bart De Wever était plus populaire que Patrick Janssens dans le corpus. Toutefois, la différence entre les deux est infime : 26 % des évaluations de Bart De Wever étaient positives alors que le pourcentage d'évaluations positives de Patrick Janssens s'élève à 25 % (voir le tableau 24). La différence infime entre les deux reflète la réalité des élections ; la course très disputée entre les deux candidats a été affirmée dans plusieurs journaux, entre autres dans le quotidien néerlandophone *De Standaard* <sup>26</sup>.

Il est également ressorti du tableau 24 que les évaluations de Alexander De Croo étaient généralement négatives : du total de 11 sentiments exprimés à son égard (les mentions neutres ne sont pas prises en considération), 7 étaient négatives (= 63 %). La même constatation vaut pour Gwendolyn Rutten. Cependant, elle n'a été mentionnée que deux fois, dont l'une mention était neutre et l'autre exprimait un sentiment négatif (= 100 %). En se basant sur ce résultat, l'on pourrait constater que Gwendolyn Rutten est moins populaire qu'Alexander De Croo. Toutefois, il s'avère très difficile de comparer les sentiments exprimés à l'égard des deux personnes, étant donné les rares mentions de Gwendolyn Rutten.

---

<sup>26</sup> Nek-aan-nekrace naar Antwerpse stadhuis. (01.09.2012). *De Standaard*. [En ligne] [http://www.standaard.be/artikel/detail.aspx?artikelid=DMF20120901\\_017](http://www.standaard.be/artikel/detail.aspx?artikelid=DMF20120901_017) [12.04.2013].

La même constatation vaut pour la comparaison entre l'ancien président du parti Vlaams Belang Bruno Valkeniers et son successeur Gerolf Annemans. Tous les deux n'ont reçu aucune évaluation positive dans le corpus. En se basant sur les mentions subjectives de chacun, il n'est pas possible de constater quelle personne des deux est la plus populaire.

### 7.3. Résultats de l'analyse à l'aide du logiciel TExSIS

Dans cette section sont présentés les résultats de l'analyse du troisième corpus contenant 15 952 tweets politiques. L'étendue assez large du corpus a permis d'y effectuer une mesure de fréquence des termes automatique à l'aide du logiciel TExSIS<sup>27</sup>. Le principal objectif de cette extraction terminologique était de dévoiler les principaux mots évoqués sur Twitter à l'occasion des élections et d'identifier les catégories grammaticales souvent utilisées dans un langage politique.

L'extraction terminologique du logiciel a produit une liste contenant tous les termes pertinents du corpus, leur fréquence dans le corpus, leur attribut *termhood*<sup>28</sup> et la catégorie grammaticale à laquelle appartiennent les mots du terme. La liste étant très large, nous avons décidé de ne présenter que les cinquante mots les plus évoqués dans le corpus. Les tableaux 27 et 28 présentent la liste de ces mots, classés respectivement par leur fréquence dans le corpus et par leur valeur *termhood*, et la catégorie grammaticale attribuée à tous les mots.

Tableau 27 : top cinquante des termes, classés par leur fréquence.

Class-ement	Terme	Fréquence	Cat. gram.
1	RT	5767	N
2	vk	5301	N
3	groen	1555	ADJ
4	BDW	1450	N
5	NVA	1415	N
6	nva	942	N
7	hetgrotedebat	821	N
8	bdw	745	N
9	Groen	699	ADJ
10	antwerpen	677	N

Tableau 28 : top cinquante des termes, classés par leur valeur *termhood*.

Class-ement	Terme	Valeur <i>termhood</i>	Cat. gram.
1	RT	849.4986	N
2	hetgrotedebat	638.95451	N
3	openvld	535.50859	V
4	villapolitica	526.61132	N
5	vk	437.20551	N
6	NVA	344.16675	N
7	ÿV	303.16439	N
8	OpenVLD	287.61425	V
9	nva	280.81231	N
10	BDW	274.39495	N

<sup>27</sup> [Http://lt3.hogent.be/en/projects/texsis](http://lt3.hogent.be/en/projects/texsis).

<sup>28</sup> La valeur *termhood* d'un mot / terme indique le degré de spécialisation de son sens dans le domaine à l'étude (Kageura et Umino, 1996).

11	verkiezingen	607	N
12	villapolitica	573	N
13	openvld	572	V
14	VK	508	N
15	debat	467	N
16	stemmen	378	N
17	cdenv	375	N
18	Wever	333	N
19	Bart	316	N
20	eens	316	N
21	stemmen	276	N
22	burgemeester	264	N
23	partij	253	N
24	gent	246	N
25	verandering	235	N
26	vragen	234	N
27	vlaamsbelang	227	N
28	VlaamsBelang	221	N
29	gemeenteraads- verkiezingen	218	N
30	ÿV	206	N
31	campagne	206	N
32	kruitfabriek	201	N
33	tijd	200	N
34	DeWever	194	N
35	zoersel	191	N
36	be	181	N
37	terzaketv	173	N
38	LDD	169	N
39	OpenVLD	165	V
40	deweever	160	N
41	stad	158	N
42	politiek	153	N
43	partijen	149	N
44	coalitie	146	N
45	vr	145	N
46	programma	141	N
47	één	136	N
48	kracht	134	N
49	verhaal	134	N
50	pas	133	N

11	terzaketv	273.51562	N
12	OpenVld	245.27826	V
13	cdenv	242.6199	N
14	één	238.54878	N
15	bdw	196.68455	N
16	brusselkiest	195.04586	N
17	deochtend	186.3683	N
18	begov	183.71263	N
19	atvwoz	182.18208	N
20	vlaamsbelang	178.32266	N
21	VlaamsBelang	175.95019	N
22	openVLD	170.52279	V
23	ÿv	162.24472	N
24	PVDA+	160.78471	N
25	groen	159.06587	ADJ
26	LDD	155.1006	N
27	PatrickJanssens	153.28801	N
28	VK	135.34387	N
29	CDenV	130.2035	N
30	hé	126.73511	N
31	kruitfabriek	126.25594	N
32	patrickjanssens	125.69745	N
33	vrtjournaal	124.30984	N
34	HetGroteDebat	124.15929	N
35	grün	122.37633	N
36	canvastv	121.6314	N
37	A'pen	120.55303	N
38	vlareg	119.93457	N
39	villapolitica canvastv	118.42737	N, N
40	laatsteronde	118.39556	N
41	én	118.23543	N
42	dekruitfabriek	117.83148	N
43	debattle	116.48361	N
44	vlapar	115.08192	N
45	belastingregerin g	114.92191	N
46	turno	109.37211	N
47	ovld	108.92306	N
48	vlabra	108.19442	N
49	DeWever	107.8905	N
50	dtv	107.78408	N

Il ressort des tableaux 27 et 28 que la plupart des mots présentés dans les deux classifications sont les mêmes. Toutefois, l'ordre dans lequel ils sont présentés est différent. Cette différence est due à la façon dont les mots sont classés. Dans le tableau 27, ils sont classés selon leur fréquence dans le corpus, c'est-à-dire selon le nombre total de mentions des mots. Le tableau révèle que les mots « RT », « vk », « groen », « BDW » et « NVA » sont les cinq mots les

plus récurrents dans le corpus. Les données du tableau 28 sont classées selon leur valeur *termhood*. La valeur *termhood* d'un mot est calculée à l'aide d'une formule mathématique dont le résultat est une comparaison de la fréquence absolue du mot dans le corpus spécifique à un domaine avec sa fréquence dans un corpus plus général. En conclusion, la valeur *termhood* d'un mot indique son degré de spécialisation dans un corpus spécifique à un domaine par comparaison à un corpus de référence plus général (Vintar, 2010). La classification selon la valeur *termhood* révèle que les cinq mots les plus utilisés sont « RT », « hetgrotedebat », « openvld », « villapolitica » et « vk ».

L'importance d'une classification basée sur la valeur *termhood* peut être illustrée à l'aide d'une comparaison entre les deux tableaux. Bien que tous les deux comportent des mots politiques, le tableau 27 contient plusieurs mots qui pourraient figurer également dans des contextes plus généraux : « stad », « tijd », « pas », « verhaal », « eens », etc. En revanche, le tableau 28 étant classifié selon la valeur *termhood*, il ne comporte que les mots spécifiques du domaine de la politique. Tous les mots qui sont susceptibles de figurer également dans des contextes plus larges sont classés en bas de la liste. En conclusion, le tableau 28 s'avère plus approprié que le tableau 27 à nous donner une idée de la terminologie couramment utilisée dans un contexte politique. Néanmoins, le tableau 27 est digne d'être pris en considération, car dans son top dix des mots les plus évoqués figure entre autres le mot « antwerpen ». En effet, la ville d'Anvers était l'une des villes les plus discutées dans les médias traditionnels et sur Twitter<sup>29</sup> lors des élections. Le terme étant susceptible de figurer également dans des contextes non politiques, il est classé en bas de la liste du tableau 28.

Il ressort du tableau 28 que les *retweets* comprenant le terme « RT » sont très abondants dans le corpus. Une autre partie considérable du corpus se compose de termes désignant les émissions-débat diffusées en vue des élections telles que *Het Grote Debat*, *Brussel Kiest* et *De Laatste Ronde*. Le tableau présente également les principaux partis politiques flamands : l'Open Vld, la N-VA, Groen, le CD&V, le sp.a, le Vlaams Belang, la LDD et le PVDA. Toutefois, la liste ne comporte que quelques hommes politiques : Bart De Wever, le président de la N-VA ; Patrick Janssens, membre du parti sp.a et ancien bourgmestre d'Anvers et Veli Yüksel (jV), membre du CD&V. Le mot « vk » est l'abréviation du mot « verkiezingen » en néerlandais. Il a également obtenu une valeur *termhood* assez élevée, ce qui indique que

---

<sup>29</sup> De Lobel, P. (13.10.2012). Over de parking geen tweets. *De Standaard*, p. 8.

beaucoup de tweets comportaient des références explicites aux élections en général. Le tableau 28 comprend également un nombre considérable de mots qui ont été mentionnés plusieurs fois d'une orthographe différente.

Le tableau 27 présente une fréquence élevée des mots « RT », « vk », des différentes émissions-débat et des principaux partis politiques flamands. Contrairement au tableau 28 comportant trois hommes politiques, le tableau 27 n'en contient que deux ; Bart De Wever et Veli Yüksel. Le tableau 27 comporte également moins de références doubles (c'est-à-dire, deux ou plusieurs références à une entité, toutes d'une orthographe différente) que le tableau 28, mais il contient plus de termes généraux qui n'appartiennent pas spécifiquement à un vocabulaire politique.

En ce qui concerne la catégorie grammaticale des termes, il ressort des tableaux que la plupart des mots sont classifiés de nom (*N*). Le terme « Groen » a été considéré comme un adjectif (*ADJ*), bien qu'il soit probable qu'il ne référerait pas au couleur correspondant mais au parti politique portant ce nom. Au terme « Open Vld » a toujours été attribuée la catégorie *verbe* (*V*), bien que ce soit un nom propre. Il n'est pas tout à fait clair pourquoi TExSIS<sup>30</sup> ait classé ce parti dans la catégorie *verbe*, mais cela pourrait s'expliquer par la présence du mot « open » dans certains verbes en néerlandais, tels que « openen », « openmaken » et « opendoen ».

Prenant en considération les dix premiers mots des deux classifications, l'on constate que les mots les plus évoqués dans le corpus de 15 952 tweets politiques sont « RT », « vk », « N-VA », « Groen », « Open Vld », « Bart De Wever », « Het Grote Debat », « Antwerpen » et « Veli Yüksel ». Cette liste révèle que beaucoup de *retweets* ont été envoyés, ce qui indique un certain niveau de conversation et de discussion entre les utilisateurs de Twitter au sujet des élections. Les mentions nombreuses du mot « vk » indique que le sujet des élections en général (le système de vote, les affiches électorales, l'attention qui est accordée aux élections dans les médias, etc.) a souvent été évoqué dans le corpus. Il ressort également de la liste que l'émission-débat *Het Grote Debat* a souvent été un sujet de conversation, de la même façon que les partis politiques N-VA, Groen et Open Vld. Dans ce top dix des mots pertinents des deux tableaux se situent également les mêmes hommes politiques, à savoir Bart De Wever et

---

<sup>30</sup> [Http://lt3.hogent.be/en/projects/texsis](http://lt3.hogent.be/en/projects/texsis).

Veli Yüksel. En effet, ce sont deux personnages qui ont reçu, au cours des semaines avant les élections, beaucoup d'attention dans les médias, provoquant une vive controverse à certains moments<sup>31</sup>.

## 8. CONCLUSIONS

Ce chapitre présente les conclusions qui peuvent être déduites des analyses des trois corpus. L'objectif de cette étude était triple : le principal objectif était d'effectuer une analyse des sentiments manuelle d'une série de tweets politiques, d'analyser les résultats et d'évaluer le schéma d'annotation conçu par Van de Kauter et Desmet (2012) en indiquant ses points forts et ses faiblesses. Le deuxième objectif était d'examiner quels partis et personnages politiques sont les plus évoqués dans un corpus de quatre mille tweets environ et quelles orientations politiques ressortent de ce corpus. Il a également été vérifié s'il existe des différences entre l'orientation politique des hommes et celle des femmes. Ensuite, nous avons voulu comparer les résultats de l'analyse avec le résultat des élections d'octobre 2012. Le troisième objectif était d'examiner à partir d'un corpus assez large de tweets politiques quelle terminologie est spécifiquement utilisée dans un contexte politique.

Trois corpus ont été établis lors de cette étude, contenant respectivement 200, 4 629 et 15 952 tweets politiques. Au total, trois expérimentations ont été effectuées, toutes à partir d'un corpus différent. Les sections suivantes fournissent les conclusions qui peuvent être déduites de toute analyse.

### 8.1. L'analyse des sentiments

L'analyse des sentiments du premier corpus a révélé qu'en moyenne, un sentiment a été exprimé par tweet. Cela mène à la constatation qu'en général, les tweets sont des messages subjectifs. Considérant le pourcentage élevé de tweets négatifs dans le corpus, l'on peut constater que dans un contexte politique, Twitter s'utilise surtout pour exprimer des opinions défavorables à l'égard de quelqu'un ou quelque chose ou de se plaindre de quelque chose.

---

<sup>31</sup> Le président du parti N-VA a parfois été controversé à cause de son programme électoral et ses finalités politiques au niveau fédéral. Veli Yüksel du parti CD&V a soulevé une vive controverse à cause de sa lettre au Parlement flamand dans lequel il a fait une requête personnelle au juge.

En général, les hommes ont exprimé plus de sentiments négatifs dans le corpus que les femmes, surtout au sujet des élections en général. En outre, l'intensité avec laquelle ils ont exprimé leurs sentiments était souvent plus forte que celle des femmes. Celles-ci étaient généralement plus prudentes et plus modérées au moment d'exprimer leurs opinions. De plus, les femmes avaient plus tendance à atténuer le sentiment exprimé par des modificateurs d'affaiblissement que les hommes.

Généralement, au moins un modificateur a été exprimé par tweet et le type de modificateur le plus utilisé est celui du renforcement. Ce résultat implique que les sentiments exprimés dans les tweets ont presque toujours été altérés. Dans la plupart des cas, cette altération consistait en un renforcement du sentiment exprimé.

Finalement, le corpus comportait très peu de sentiments implicitement exprimés, ce qui entraîne comme conséquence que Twitter est généralement utilisé pour exprimer ses opinions de façon explicite, à travers un langage subjectif et direct. La quantité de sentiments implicitement exprimés n'a pas permis d'établir une comparaison entre les hommes et les femmes en ce qui concerne l'explicitation des sentiments qu'ils ont exprimés.

L'évaluation du schéma d'annotation de Van de Kauter et Desmet (2012) a révélé quelques différences importantes par rapport au schéma d'annotation de base conçu par Wiebe (2002). En premier lieu, le nouveau schéma permet de distinguer trois types d'expressions subjectives, par opposition à Wiebe (2002) qui n'en avait distingué qu'un. En plus, cette extension de types d'expressions subjectives le rend possible de détecter des sentiments dans des phrases qui ne comportent que de l'information objective ou factuelle. De telles phrases étaient considérées comme neutres par Wiebe (2002), par conséquent, elles n'ont pas été prises en considération pour l'analyse des sentiments. En deuxième lieu, le schéma de Van de Kauter et Desmet (2012) permet de détecter onze types de modificateurs qui altèrent le sentiment exprimé dans une phrase (voir la section 3.4.2). En troisième lieu, Van de Kauter et Desmet (2012) ont supprimé les sous-groupes ambigus *intensité extrême* et *aucune polarité* des catégories génériques *intensité du sentiment* et *polarité du sentiment*.

En général, le schéma d'annotation de Van de Kauter et Desmet (2012) a permis de procéder à une analyse profonde de notre corpus de tweets politiques. Les résultats l'analyse des sentiments ont permis de tirer des conclusions sur le sentiment global exprimé dans tout

tweet, le nombre de sentiments exprimés par tweet, la polarité et l'intensité du sentiment ainsi que son niveau d'explicitation et la fréquence avec laquelle les différents types de modificateurs ont été utilisés dans le corpus.

Bien que le schéma d'annotation de Van de Kauter et Desmet (2012) se soit avéré très approprié afin de procéder à l'analyse des sentiments dans un texte, quelques suggestions ont été faites lors de cette étude afin de rendre l'analyse encore davantage précise. Ainsi, il a été proposé de supprimer les sous-groupes ambigus *autre polarité* et *autres modificateurs* des catégories génériques correspondants *polarité du sentiment exprimé* et *modificateurs*. Il a également été suggéré de ne prévoir que trois valeurs pour qualifier les expressions subjectives : la valeur objective, la valeur subjective et la valeur causale. Une dernière proposition qui a été faite implique une annotation encore davantage profonde des expressions subjectives incorporées dans d'autres.

Dans le cadre de la conception d'un système autonome d'analyse automatique des sentiments et en vue de futures recherches, les annotations manuelles faites lors de cette étude pourraient être utilisées pour l'entraînement d'un système autonome d'analyse automatique des sentiments (voir la section 2.3.2). Il serait d'autant plus intéressant d'utiliser ces annotations pour l'entraînement d'un système automatique qu'elles comportent des informations sur les sentiments exprimés dans des tweets, qui constituent des textes non traditionnels.

Jusqu'à présent, la plupart des systèmes autonomes d'analyse automatique des sentiments étaient basés sur un lexique, tels que le logiciel *Sentitext*<sup>32</sup> utilisé lors de l'étude de Moreno-Ortiz et Pérez Hernández (2012) et le logiciel *LIWC2007*<sup>33</sup> utilisé lors de la recherche de Tumasjan et al. (2010). Néanmoins, la présente étude a révélé qu'une analyse des sentiments basée sur des corpus annotés de façon manuelle permet une détection davantage complète des sentiments exprimés dans un texte. En effet, l'analyse des sentiments à l'aide du schéma d'annotation de Van de Kauter et Desmet (2012) était tellement détaillée qu'elle a permis de détecter de multiples modificateurs qui avaient altéré le sentiment exprimé dans certains tweets. L'identification de ces modificateurs n'aurait pas été possible en ne se basant que sur un lexique de mots positifs et négatifs.

---

<sup>32</sup> [Http://www.sentitext.com/](http://www.sentitext.com/).

<sup>33</sup> [Http://www.liwc.net/](http://www.liwc.net/).

## 8.2. L'analyse des orientations politiques

L'analyse générale du deuxième corpus a révélé que les partis politiques N-VA, Groen et CD&V ont le plus souvent été évoqués dans le corpus, ainsi que les hommes politiques Bart De Wever, Patrick Janssens et Peter Mertens. Tenant compte des évaluations positives et négatives de tout parti et personnage politique, l'analyse a dévoilé que les partis les plus populaires dans le corpus étaient Groen, PVDA et Open Vld et les hommes politiques les plus populaires étaient Wouter Van Besien, Peter Mertens et Alexander De Croo. En revanche, les partis Vlaams Belang et LDD peuvent être considérés comme les partis les moins populaires dans le corpus, étant donné qu'ils ont reçu le moins d'évaluations positives. Avec le PVDA, le Vlaams Belang et la LDD sont les partis les moins évoqués dans le corpus.

Si sur ces résultats était basée une prédiction politique, elle révélerait que les partis Groen, PVDA et Open Vld obtiendraient le plus grand nombre de voix lors des élections. Néanmoins, l'analyse des résultats des élections du 14 octobre 2012 (voir la section 7.2.5) a révélé que la N-VA et le CD&V ont obtenu le plus de voix. En conclusion, en ne se basant que sur l'analyse du corpus, il n'aurait pas été possible de prédire quels partis obtiendraient le plus de voix lors des élections communales et provinciales de 2012.

Toutefois, il importe de nuancer cette constatation par deux observations. En premier lieu, bien que les partis N-VA et CD&V ne soient pas ressortis comme les partis les plus populaires du corpus, ils ont très fréquemment été évoqués dans les tweets du corpus. Cela soutient l'idée que les partis ont en tout cas suscité beaucoup d'attention. En deuxième lieu, les partis Groen et PVDA sont ressortis comme les partis les plus populaires du corpus. En effet, la comparaison entre les résultats des élections de 2012 et 2006 a révélé que ce sont ces partis qui ont enregistré une considérable augmentation de voix par rapport aux autres partis. De ce point de vue, le PVDA et Groen peuvent effectivement être considérés comme deux *gagnants* des élections de 2012.

En ce qui concerne la comparaison entre les tweets des hommes et ceux des femmes, l'analyse a révélé que l'orientation politique des hommes dans le corpus est semblable à celle des femmes. Dans les deux groupes, les hommes politiques les plus populaires étaient Wouter Van Besien, Peter Mertens et Alexander De Croo. Les partis politiques qui ont obtenu le plus d'évaluations positives des hommes sont Groen, PVDA et Open Vld et ceux qui ont obtenu le

plus d'évaluations positives des femmes sont Groen, sp.a et Open Vld. Dans les deux groupes, les partis PVDA, Vlaams Belang et LDD ont le moins souvent été évoqués. En plus, ces deux derniers sont ressortis comme les partis les moins populaires.

En conclusion, l'analyse a révélé que l'orientation politique des hommes du corpus n'est guère différente de celle des femmes. Toutefois, une exception doit être faite en comparant la popularité de Bart De Wever avec celle de Patrick Janssens. Il est ressorti des tweets des hommes que Bart De Wever a reçu plus d'évaluations positives que Patrick Janssens. Par contre, les tweets des femmes ont révélé que l'évaluation globale de Patrick Janssens était plus positive que celle de Bart De Wever.

Prenant en considération que Twitter peut être utilisé comme un outil de prédiction fiable, ce qui a déjà été démontré par des études précédentes de Tumasjan et al. (2010) et de Tjong Kim Sang et Bos (2012), le résultat de notre analyse entraîne la conclusion que l'étendue du corpus est insuffisante afin d'y baser une prédiction fiable du résultat des élections. Par conséquent, si l'on veut généraliser la constatation de Tumasjan et al. (2010) et Tjong Kim Sang et Bos (2012) en Belgique, de nouvelles recherches seront nécessaires à partir d'une base de données plus large.

### 8.3. L'extraction terminologique à l'aide du logiciel TExSIS

L'extraction terminologique du troisième corpus a révélé que les mots les plus évoqués dans le corpus étaient « RT », « Het Grote Debat », « vk », « Bart De Wever », « Patrick Janssens », « Veli Yüksel » et les principaux partis politiques flamands.

A partir de ces résultats peut être déduite la conclusion que beaucoup de *retweets* sont envoyés dans un contexte politique, ce qui implique que l'électorat aime discuter et entamer des conversations sur Twitter en vue des élections. Il est également ressorti de l'analyse que l'émission-débat *Het Grote Debat* a très souvent été évoqué dans le corpus, ce qui entraîne comme conséquence qu'il a probablement suscité une vive controverse. Tous les principaux partis politiques flamands ont été mentionnés dans le corpus, bien que les partis N-VA, Groen et Open Vld aient une fréquence davantage élevée que les autres partis. Ce résultat correspond

approximativement à la mesure de fréquence effectuée sur le deuxième corpus qui avait révélé un total de mentions très élevé des partis N-VA, Groen et CD&V.

Il se note également une analogie entre les mentions des personnages politiques dans le deuxième et le troisième corpus. La mesure de fréquence du troisième corpus avait révélé que les politiques les plus évoqués étaient Bart De Wever, Patrick Janssens et Veli Yüksel. Dans le corpus de 4 629 tweets, c'étaient Bart De Wever, Patrick Janssens et Peter Mertens. La différence s'explique par le fait que l'analyse du deuxième corpus n'avait pris en considération qu'une liste limitée de personnages politiques, dont la majorité étaient les présidents des principaux partis flamands. Par conséquent, les mentions d'autres personnages politiques tels que Veli Yüksel n'avaient pas été prises en considération.

L'analyse morpho-syntaxique du troisième corpus a révélé que tous les termes souvent évoqués dans le corpus sont des noms. Cependant, les partis politiques Open Vld et Groen ont respectivement été classifiés de verbe et d'adjectif. Cela est probablement dû au fait que le mot « open » dans le nom « Open Vld » est le préfixe de certains verbes en néerlandais tels que « openmaken ». De la même façon, le mot « groen » est un adjectif dans son sens original, désignant la couleur verte. Ce résultat mène à la conclusion que, dans le cas où un mot prend plusieurs formes grammaticales selon le sens qu'il a dans un certain contexte, le logiciel TExSIS ne sait pas toujours attribuer la forme grammaticale appropriée.

Dans le cadre de futures recherches, il serait intéressant de comparer les mots politiques souvent évoqués sur les réseaux sociaux avec ceux qui sont généralement utilisés dans les médias traditionnels tels que la presse écrite, la télévision et la radio afin de vérifier s'il existe des différences remarquables.

## 9. BIBLIOGRAPHIE

- Brooke, J., Tofiloski, M., & Taboada, M. (2009). Cross-Linguistic Sentiment Analysis: From English to Spanish. *Proceedings of RANLP 2009, Recent Advances in Natural Language Processing*. Présenté à la conférence RANLP 2009, Borovets, Bulgarie.
- Cohen, J. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37–46.
- Dedecker: "LDD niet van de kaart geveegd". (14.10.2012). *De Morgen*. [En ligne] <http://www.demorgen.be/dm/nl/3625/Verkiezingen2012/article/detail/1516811/2012/10/14/Dedecker-LDD-niet-van-de-kaart-geveegd.dhtml> [24.04.2013].
- De Lobel, P. (13.10.2012). Over de parking geen tweets. *De Standaard*, p. 8.
- El Jardín de Junio. (2011). Sentitext, el tesoro escondido en las palabras. [En ligne] <http://www.sentitext.com/bienvenida.html> [08.04.2013].
- Goret, C. (29.10.2012). De eerste sociale campagne. *Metro*, p. 6.
- Herregods, J. (10.10.2012). Verkiezingen : ‘Vind ik leuk’. *Metro*, p. 4.
- Kageura, Kyo et Bin Umino. (1996). Methods for automatic term recognition : A review. *Terminology*, 3(2), 259–289.
- Language and translation technology team (s.d.). Terminology Extraction for Semantic Interoperability and Standardization, TExSIS. [En ligne] <http://lt3.hogent.be/en/projects/texsis> [24.04.2013].
- Long J., Mo Y., Ming Z., Xiaohua L. & Tiejun Zhao. (2011). Target-dependent Twitter Sentiment Classification. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, 151–160.
- Moreno-Ortiz, A. & Pérez Hernández, C. (2012). Lexicon-Based Sentiment Analysis of Twitter Messages in Spanish. Etude extraite de <http://www.daedalus.es/TASS>. [12.04.2013].
- Nek-aan-nekrace naar Antwerpse stadhuis. (01.09.2012). *De Standaard*. [En ligne] [http://www.standaard.be/artikel/detail.aspx?artikelid=DMF20120901\\_017](http://www.standaard.be/artikel/detail.aspx?artikelid=DMF20120901_017) [12.04.2013].
- Pang, B. & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2), 1-135.
- Pennebaker, J.W., Francis, M.E. & Booth, R.J. (2007). Linguistic Inquiry and Word Count (LIWC): LIWC2007. Mahwah, NJ: Erlbaum Publishers.
- Python software foundation. (1990-2013). Python Programming Language – Official Website. [En ligne] <http://python.org/> [15.03.2013].
- Sentiment. (2009). Dans *Le Nouveau Petit Robert de la langue Française en ligne* [11.02.2013].
- TASS 2013.(s.d.). Workshop on Sentiment Analysis at SEPLN. [En ligne] <http://www.daedalus.es/TASS2012/about.php> [18.02.2013].
- The Brat contributors. (2010-2012). Brat rapid annotation tool. [En ligne] <http://brat.nlplab.org/> [14.04.2013].
- Tjong Kim Sang, E. & Bos, J. (2012) Predicting the 2011 Dutch Senate Election Results with Twitter, *Proceedings of SASN 2012, the EACL 2012 Workshop on Semantic Analysis in Social Networks*, Avignon, France, 2012.
- Tumasjan, A. et al. (2010). Predicting Elections with Twitter: What 140 Characters Reveal

- about Political Sentiment, *Proceedings of the Fourth AAAI conference on Weblogs and Social Media*, 2010.
- Vintar, S. (2010). "Bilingual term recognition revisited. The bag-of-equivalents term alignment approach". *Terminology*, 16(2), 141-158.
- Vlaanderen kiest. [Base de données du résultat des élections de 2012].  
[Http://www.vlaanderenkiest.be/verkiezingen2012/index.html](http://www.vlaanderenkiest.be/verkiezingen2012/index.html) [12.04.2013].
- Wiebe, J. (2002). Instructions for annotating opinions in newspaper articles. Département de Computer Science Technical Report TR-02-101, Université de Pittsburgh, Pittsburgh.
- Wiebe, J., Wilson, T. & Cardie C. (2005). Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation (formerly Computers and the Humanities)*, 39(2/3), 164-210.
- Wilson, A. (2008). Fine-grained Subjectivity and Sentiment Analysis: Recognizing the Intensity, Polarity, and Attitudes of Private States. Dissertation non publiée. Université de Pittsburgh, Pittsburgh.

## 10. ANNEXES

10.1. Tableau général de l'analyse des sentiments

<b>Tweet</b>	<b>Sexe</b>	<b>Mot-dièse</b>	<b>Sujet</b>	<b>Sentiment</b>	<b>#sentiments</b>	<b>Explicitation</b>	<b>Polarité</b>	<b>Intensité</b>	<b>Modificateurs</b>
1	f	#verkiezingen2012	stemrecht gedetineerden	v	3	explicite	négative	forte	renforcement (groot)
						explicite	négative	forte	
						explicite	négative	forte	
2	f	#verkiezingen2012	verkiezingen	v	2	explicite	négative	moyenne	passé (was), négation (geen)
						explicite	négative	moyenne	passé (was)
3	m	#gemeenteraadsverkiezingen	discriminatie van kandidaten die niet digitaal zijn	v	2	explicite	négative	forte	
						explicite	négative	moyenne	négation (niet eens)
4	f	#gemeenteraadsverkiezingen	gemeenteraadsverkiezingen	x	0				

5	f	#gemeenteraadsverkiezingen	campagneborden		3	explicite	négative	moyenn e	
						explicite	négative	forte	autre (ondertussen wel)
						explicite	négative	forte	renforcement (op elke hoek van de straat)
6	m	#VanQuickenborne	beleidsniveaus/Vincent Van Quickenborne/Paul Magnette/Annemie Turtelboom	v	2	explicite	positive	moyenn e	modalité (blijkbaar)
						explicite	négative	moyenn e	modalité (blijkbaar)
7	f	#verkiezingen2012	verkiezingsslogans	v	1	explicite	négative	moyenn e	négation (niet)
8	m	#gemeenteraadsverkiezingen	Steven Vanackere/de begroting	v	1	explicite	négative	moyenn e	autre (nu echt), question (?)
9	m	#NVA	verkiezingsfolder Bart De Wever	v	1	explicite	négative	forte	
10	m	#NVA	N-VA/allochtonen	v	2	explicite	autre	moyenn e	

						explicite	négative	moyenn e	affaiblissement (normaal), autre (tegenwoordig)
11	f	#gemeenteraadsverkiezingen	stemmen/verkiezingszondag	v	1	explicite	négative	basse	autre (met andere woorden), avenir (die zondag)
12	m	#verkiezingen2012	Milka Malfait/OpenVld/Bart De Wever/Wouter Van Besien	v	1	explicite	positive	moyenn e	
13	m	#NVA	Bart De Wever/de Kruitfabriek/N-VA	v	3	explicite	négative	moyenn e	
						explicite	positive	forte	renforcement (weeral, 100%, !), passé (gescoord)
						explicite	positive	forte	négation (gewoon niet), renforcement (ondanks verwoede pogingen)
14	m	#vk2012	Antwerpen/burgemeesterschap/Bart De Wever/Patrick Janssens	v	3	explicite	négative	forte	

						explicite	négative	moyenn e	
						explicite	négative	moyenn e	
15	f	#gemeenteraadsverkiezingen	verkiezingen	v	2	explicite	négative	forte	renforcement (veel)
						explicite	négative	moyenn e	
16	f	#vk2012	socialisme Brussel	x	0				
17	m	#NVA	Bart De Wever/journalistiek	v	2	explicite	autre	forte	affaiblissement (maar wel onafhankelijke kritische journalistiek!), renforcement (het is duidelijk dat), négation (niet)
						explicite	négative	forte	renforcement (het is duidelijk dat)
18	m	#CDenV	campagnemateriaal/Asse/CD&V	x	0				

19	f	#vk2012	burgemeesterschap Antwerpen	×	0				
20	f	#vk2012	Milka Malfait	×	0				
21	f	#VlaamsBelang	Filip Dewinter/islam/Vlaams Belang	v	1	explicite	négative	forte	condition (als...verklaring van i.s.l.a.m. mag geloven)
22	f	#vk2012	14 oktober/CD&V	v	3	explicite	autre	forte	renforcement (!)
						implicite	positive	moyenn e	
						explicite	positive	forte	renforcement (!), avenir (op 14 oktober)
23	m	#OpenVld	OpenVld/verkiezingen	v	2	explicite	négative	forte	
						explicite	négative	forte	autre (tja)

24	m	#bdw	Patrick Janssens/Bart De Wever/De Kruitfabriek	V	3	explicite	négative	moyenn e	
						explicite	négative	forte	
						explicite	négative	moyenn e	
25	m	#bdw	Bart De Wever/Calimero/ATV-debatten	V	1	explicite	négative	forte	renforcement (blijft, o.a.)
26	m	#gemeenteraadsverkiezingen	Annemie Turtelboom/derde weg/Antwerpen	V	1	explicite	négative	moyenn e	autre (tiens)
27	f	#vk2012	Open Vld/Tervuren	V	1	explicite	positive	forte	renforcement (alvast, !)
28	m	#NVA	N-VA/Guy Verhofstadt	V	3	explicite	négative	moyenn e	
						explicite	négative	moyenn e	
						explicite	négative	moyenn e	
29	m	#Openvld	OpenVld	V	3	explicite	positive	forte	

						explicite	négative	forte	renforcement (veilig)
						explicite	positive	forte	renforcement (!)
30	m	#Openvld	OpenVld/N-VA	v	2	explicite	négative	forte	
						explicite	positive	moyenn e	
31	f	#vk2012	Dendermonde/N-VA/	x	0				
32	f	#vk2012	verkiezingslijsten	v	2	explicite	négative	forte	renforcement (de enige)
						explicite	négative	moyenn e	affaiblissement (vooral), renforcement (blijft)
33	m	#NVA	Guy Verhofstadt/N-VA/OpenVld	v	2	explicite	négative	forte	affaiblissement (Beste Guy daarmee zijn we nog niet vergeten dat jij Openvld naar de verdoemenis hebt geholpen!)

						explicite	négative	forte	négation (nog niet), passé (hebt geholpen), renforcement (!)
34	m	#NVA	N-VA/containerwoord extreem-rechts	V	2	explicite	négative	forte	
						explicite	négative	forte	négation (?)
35	m	#gemeenteraadsverkiezingen	Sint-Joost/gemeenteraadsverkiezingen/campagne	V	2	explicite	négative	forte	autre (wel), négation (nog geen), renforcement (enkel mogelijk)
						explicite	négative	basse	
36	m	#Groen	Groen/N-VA	V	2	explicite	négative	moyenne	
						explicite	négative	forte	renforcement (:-))
37	m	#NVA	De zevende dag/N-VA/burgemeesterschap/verkiezingen	V	3	explicite	autre	moyenne	question (?)

						explicite	négative	basse	question (?)
						explicite	négative	moyenn e	avenir (na VK2012)
38	m	#NVA	N-VA/Vlaams Belang/Antwerpen	v	2	explicite	négative	moyenn e	négation (niet, geen)
						explicite	négative	forte	renforcement (!)
39	m	#Groen	Groen/Antwerpse ring	v	3	explicite	négative	moyenn e	
						explicite	négative	forte	autre (dan ook), renforcement (graag, houden)
						explicite	positive	moyenn e	
40	m	#NVA	Bart De Wever/Villa Vanthilt/N-VA	v	2	explicite	positive	forte	renforcement (losjes)
						explicite	négative	forte	renforcement (losjes)
41	m	#CDenV	CD&V	v	1	explicite	négative	moyenn e	affaiblissement (meestal)

42	m	#Beke	Wouter Beke/CD&V	v	3	explicite	positive	moyenn e	
						explicite	négative	moyenn e	
						explicite	négative	forte	renforcement (al een paar jaar)
43	f	#CDenV	Affiche Hilde Crevits/CD&V	v	1	explicite	autre	forte	renforcement (wel heel)
44	f	#vk2012	regionale partijen	x	0				
45	f	#vk2012	campagneborden/Astene/verkiezingen	x	0				
46	m	#NVA	N-VA	v	1	explicite	négative	moyenn e	avenir (op 14/10)
47	f	#vk2012	Jan Peumans/Federale regering/Villa Politica	v	2	explicite	négative	forte	renforcement (na amper twee minuten, al), passé (gaf)
						explicite	négative	moyenn e	affaiblissement (nog)

48	f	#vk2012	verkiezingen	V	2	explicite	positive	moyenn e	affaiblissement (gewoon)
						explicite	négative	moyenn e	négation (niet echt)
49	f	#Groen	verkiezingsmagazine Patrick Janssens/Groen	V	1	explicite	négative	forte	renforcement (ouch), modalité (moet...zijn)
50	f	#NVA	N-VA/verkiezingen	V	1	explicite	négative	forte	renforcement (!)
51	m	#Openvld	OpenVld	V	1	explicite	positive	forte	renforcement (!!)
52	f	#VlaamsBelang	Vlaams Belang	V	2	explicite	négative	forte	renforcement (allemaal in 1 zin)
						explicite	négative	moyenn e	
53	m	#Groen	betog van Groen	V	1	explicite	positive	moyenn e	

54	m	#Groen	Wouter Van Besien/Groen	v	2	explicite	négative	forte	affaiblissement (maar Van Besien was volgens mij de enige die wat inhoudt (sic) bracht)
						explicite	positive	forte	renforcement (de enige), affaiblissement (wat), passé (bracht)
55	f	#VlaamsBelang	boerkaverbod/campagne Vlaams Belang	v	2	explicite	négative	moyenn e	
						explicite	négative	moyenn e	
56	m	#SP_A	verkiezingsflyers sp.a Gent	x	0				
57	f	#CDenV	CD&V/Wouter Beke	v	2	explicite	positive	basse	affaiblissement (nochtans, pas)
						explicite	positive	forte	renforcement (zeeën van, ;-)

58	m	#VanBesien	Wouter Van Besien	v	1	explicite	positive	moyenn e	
59	f	#VanBesien	Wouter Van Besien	v	3	explicite	positive	moyenn e	
						explicite	positive	moyenn e	négation (geen)
						explicite	positive	moyenn e	
60	f	#VanBesien	Wouter Van Besien/Antwerpen	v	2	explicite	positive	basse	affaiblissement (maar toch moet het weer over Antwerpen gaan)
						explicite	négative	moyenn e	renforcement (toch, weer)
61	f	#Tobback	Tobback	v	1	explicite	négative	moyenn e	condition (als je in VL twee stenen neerlegt), renforcement (het enige)
62	m	#Tobback	Tobback/cijnskiesrecht/sp.a	v	3	implicite	négative	basse	
						explicite	négative	moyenn e	renforcement (gezien SP tgn)

						implicite	négative	forte	vrouwenstemrecht was) passé (was tegen..)
63	m	#Valkeniers	Bruno Valkeniers	v	2	explicite	négative	moyenn e	
						explicite	négative	moyenn e	
64	m	#vk2012	woningprijzen	v	1	explicite	autre	moyenn e	renforcement (nog altijd)
65	f	#gemeenteraadsverkiezingen	verkiezingscampagnes	v	2	explicite	positive	forte	autre (trouwens), renforcement (erg), affaiblissement (maar ik ben het drama errond zo beu!) renforcement (zo, !)
						explicite	négative	forte	
66	m	#Groen	Groen/Sonsbeek	x	0				

67	m	#Groen	Groen	v	1	explicite	négative	forte	renforcement (!)
68	m	#NVA	Bart De Wever/media	v	4	explicite	négative	forte	renforcement (echt)
						explicite	négative	moyenn e	affaiblissement (wa)
						explicite	négative	forte	
						explicite	négative	forte	
69	m	#Openvld	OpenVld/sp.a/CD&V	v	2	explicite	négative	forte	avenir (gaan, op 14/10)
						explicite	autre	moyenn e	renforcement (;-))
70	m	#NVA	CD&V/N-VA	v	2	explicite	négative	basse	autre (toch)
						explicite	négative	moyenn e	
71	m	#Groen	Groen	v	5	explicite	positive	forte	
						explicite	négative	forte	renforcement (letterlijk)

						explicite	positive	forte	
						explicite	négative	forte	
						explicite	négative	forte	
72	m	#VillaPolitica	Groen/Hermes Sanctorum	V	1	explicite	positive	forte	renforcement (nochtans, onvoorwaardeli jk)
73	f	#verkiezingen2012	aanhangborden/campagne-affiches	V	1	explicite	négative	moyenn e	négation (niet)
74	m	#VillaPolitica	Calimerogehalte	V	3	explicite	négative	moyenn e	
						explicite	négative	moyenn e	renforcement (man, man, man)
						explicite	positive	moyenn e	
75	f	#vk2012	affiches Vlaams Belang en N-VA	V	3	explicite	négative	moyenn e	
						explicite	positive	moyenn e	

						explicite	négative	forte	
76	f	#VillaPolitica	Ingrid Lieten/NIP en PIP/OpenVld/Villa Politica	v	2	explicite	positive	moyenn e	
						explicite	positive	moyenn e	
77	m	#Openvld	PIP en NIP/OpenVld	v	1	explicite	négative	basse	question (?)
78	m	#bdw	Bart De Wever	v	3	explicite	négative	moyenn e	
						explicite	négative	forte	renforcement (nu al)
						explicite	négative	forte	avenir (worden), condition (als hij wint), spécification (zes)
79	f	#vk2012	verkiezingen	x	0				
80	f	#LDD	Vlaams parlement/Lode Vereeck/LDD	v	2	explicite	autre	forte	renforcement (echt, té, !)

						explicite	positive	forte	renforcement (altijd, !)
81	m	#Openvld	analyse OpenVld in Het Laatste Nieuws	V	4	explicite	négative	forte	
						explicite	négative	moyenn e	affaiblissement (maar lijkt correct in z'n diepgang)
						explicite	positive	moyenn e	modalité (lijkt)
						explicite	négative	basse	avenir (eerst 14/10, daarna)
82	f	#verkiezingen2012	OpenVld	V	1	explicite	autre	moyenn e	
83	f	#vk2012	Patrick Janssens	V	2	explicite	négative	forte	
						explicite	négative	moyenn e	
84	f	#Openvld	Open Vld	V	1	explicite	positive	forte	
85	m	#SP_A	sp.a	V	3	explicite	autre	basse	

						explicite	positive	moyenn e	renforcement (ook)
						explicite	négative	moyenn e	avenir (gaat..worden), autre (nog, dan)
86	m	#NVA	N-VA	v	1	explicite	négative	moyenn e	
87	f	#VanQuickenborne	Terzaketv/Vincent Van Quickenborne/ Stefaan De Clerck	v	2	explicite	négative	moyenn e	
						explicite	négative	forte	autre (eigenlijk), renforcement (ook, nog)
88	m	#gemeenteraadsverkiezingen	Banenverlies	v	1	explicite	négative	moyenn e	
89	m	#Openvld	jeugdhuis Lint/Open Vld	v	3	explicite	positive	moyenn e	
						explicite	négative	moyenn e	
						explicite	positive	moyenn e	

90	f	#CDenV	CD&V/Steven Dietvorst	v	2	explicite	positive	forte	renforcement (!!)
						explicite	positive	forte	renforcement (!)
91	m	#NVA	Wouter Beke/N-VA	v	1	explicite	négative	moyenn e	
92	m	#Beke	Wouter Beke	v	1	explicite	négative	forte	
93	m	#Beke	Wouter Beke	v	2	explicite	négative	moyenn e	renforcement (blijft, meermaals), passé (zijn gaan)
						explicite	négative	moyenn e	affaiblissement (begint te..)
94	m	#Openvld	Herman De Croo/OpenVld/Manneke Paul	v	3	explicite	négative	forte	
						explicite	négative	moyenn e	
						explicite	négative	forte	

95	f	#vk2012	media/Antwerpen	v	1	explicite	négative	moyenn e	renforcement (niet om iets maar), avenir (op 14 oktober)
96	m	#vk2012	Veiligheid	v	2	explicite	positive	forte	
						explicite	positive	forte	
97	m	#OpenVld	OpenVld	v	3	explicite	négative	moyenn e	
						explicite	négative	forte	renforcement (enkel), passé (warren ( <i>sic</i> ))
						explicite	négative	moyenn e	passé (hebben..gezien )
98	m	#Tobback	Tobback	v	1	implicite	négative	basse	
99	f	#vk2012	Bea Cantillon/welvaartsstaat	v	3	explicite	positive	moyenn e	
						explicite	positive	moyenn e	
						explicite	négative	moyenn e	renforcement (oh zo)

100	f	#CDenV	Elio Di Rupo/pensioensparen/OpenVld/CD&V/sp.a	v	1	explicite	négative	moyenn e	
101	m	#VillaPolitica	Mercedes Van Volcem/N-VA/Brugge/Pol Van den Driessche	v	1	explicite	positive	forte	renforcement (ook zonder Pol Van den Driessche)
102	m	#VillaPolitica	Jo Vandeurzen/Pascal Smet	v	2	explicite	positive	forte	renforcement (alle)
						explicite	négative	forte	renforcement (man man man)
103	m	#VanQuickenborne	besparingsmaatregelen Vincent Van Quickenborne	v	1	explicite	négative	moyenn e	spécification (10 km)
104	f	#Beke	debat/Daniel Termont/Bart De Wever/Wouter Beke	v	1	explicite	autre	forte	renforcement (zeer)
105	f	#VillaPolitica	CD&V/Veli Yüksel/Kris Peeters	x	0				

106	m	#Groen	Groen	V	2	explicite	positive	moyenn e	
						explicite	positive	moyenn e	
107	m	#SP_A	sp.a/verkiezingsreclame	V	2	explicite	négative	moyenn e	affaiblissement (wat), renforcement (toch, te)
						explicite	négative	moyenn e	
108	m	#Beke	Wouter Beke	V	1	explicite	négative	moyenn e	
109	m	#SP_A	Ieders Stem Telt/Ronse/sp.a	V	3	explicite	positive	moyenn e	
						explicite	négative	moyenn e	
						explicite	positive	moyenn e	renforcement (!)
110	f	#Groen	Groen	V	2	explicite	positive	forte	
						explicite	positive	moyenn e	affaiblissement (voor de rest)

111	m	#hetgrotedebat	sp.a/flitspalen	v	2	explicite	négative	moyenn e	
						explicite	négative	forte	condition (als het nu eens drie randcriminelen waren?)
112	f	#Valkeniers	intercommunales/Wouter Beke/Bruno Valkeniers/verkiezingsdebat	v	1	explicite	positive	forte	
113	m	#Groen	Groen	v	1	explicite	positive	moyenn e	
114	f	#hetgrotedebat	Wouter Beke/N-VA/de kracht van verandering	v	1	explicite	négative	basse	
115	f	#hetgrotedebat	humor Bart De Wever	v	2	explicite	autre	moyenn e	affaiblissement (maar..heeft humor)
						explicite	positive	moyenn e	
116	f	#hetgrotedebat	onbesliste kiezers	v	1	explicite	négative	forte	autre (uiteindelijk), négation (nog..niet), renforcement (steeds)

117	f	#hetgrotedebat	Bart De Wever	v	3	explicite	positive	moyenn e	
						explicite	négative	forte	renforcement (intussen)
						explicite	négative	moyenn e	
118	f	#hetgrotedebat	Het Grote Debat	v	2	explicite	positive	moyenn e	passé (was)
						explicite	négative	moyenn e	passé (was)
119	f	#vk2012	Bart De Wever/ATV	v	2	explicite	autre	moyenn e	
						explicite	négative	moyenn e	renforcement (wéér)
120	f	#hetgrotedebat	Goedele Liekens/Het Grote Debat	v	2	explicite	négative	moyenn e	autre (tja), renforcement (kennelijk, !)
						explicite	négative	moyenn e	
121	m	#hetgrotedebat	N-VA/Bart De Wever/Het Grote Debat	v	2	explicite	négative	forte	renforcement (het is overduidelijk)

						explicite	négative	forte	renforcement (!)
122	f	#hetgrotedebat	voorakkoorden/Deadline 14/10	v	2	explicite	négative	moyenn e	renforcement (écht)
						explicite	positive	moyenn e	renforcement (écht)
123	f	#hetgrotedebat	GAS-boetes	v	2	explicite	négative	forte	
						explicite	négative	moyenn e	affaiblissement (aub)
124	f	#hetgrotedebat	de dorpsstraat	x	0				
125	m	#hetgrotedebat	Bart De Wever	x	0				
126	m	#hetgrotedebat	Verkiezingscampagne	v	2	explicite	négative	moyenn e	
						explicite	négative	moyenn e	

127	f	#hetgrotedebat	Wouter Beke	v	3	explicite	négative	moyenn e	
						explicite	négative	moyenn e	renforcement (maar)
						explicite	négative	moyenn e	autre (blijkbaar), renforcement (héél), affaiblissement (of de arme man kent maar één spreekwoord)
128	m	#hetgrotedebat	Het Grote Debat/veiligheid/Wouter Beke	v	3	implicite	positive	moyenn e	
						explicite	positive	moyenn e	
						explicite	négative	forte	passé (heeft..ingetrapt ) , renforcement (nog maar eens, alle)
129	f	#hetgrotedebat	GAS-boetes	v	1	explicite	négative	moyenn e	avenir (komt), renforcement (nog meer), perspective (voor de wijkinspecteur)

130	f	#hetgrotedebat	Het Grote Debat/PVDA+	v	1	implicite	négative	moyenn e	renforcement (ook)
131	f	#hetgrotedebat	sp.a/Bruno Tobback	x	0				
132	m	#LDD	Het Grote Debat/LDD	v	1	explicite	négative	moyenn e	
133	m	#gemeenteraadsverkiezingen	Drukwerkafval	v	1	explicite	négative	basse	
134	m	#hetgrotedebat	GAS-boetes	v	2	explicite	négative	forte	renforcement (dringend), avenir (zal)
						explicite	négative	forte	renforcement (echt), négation (géén)
135	m	#hetgrotedebat	Wouter Beke	v	2	explicite	autre	moyenn e	
						explicite	autre	moyenn e	
136	m	#Tobback	Bruno Tobback/files/openbaar vervoer	v	2	explicite	négative	moyenn e	renforcement (elke dag)

						explicite	négative	basse	autre (dan), négation (niet)
137	m	#Beke	Wouter Beke/Markske van De Kampioenen	V	1	explicite	négative	moyenn e	renforcement (blijft, toch, brenge..ne choco), affaiblissement (wat)
138	f	#VanBesien	Wouter Van Besien/probleemanalyse/Alexander De Croo	V	3	explicite	positive	forte	renforcement (zeer)
						explicite	négative	moyenn e	renforcement (alleen maar)
						explicite	négative	forte	renforcement (zeer)
139	m	#bdw	Jean-Marie Dedecker /LDD/Het Grote Debat	V	2	explicite	négative	forte	renforcement (weer eens, totale)
						explicite	négative	moyenn e	
140	m	#Valkeniers	Bruno Valkeniers	V	1	explicite	négative	forte	

141	f	#hetgrotedebat	Het Grote Debat/verkiezingsangst	V	1	explicite	négative	forte	renforcement (alsook de eigenwaarde en hoffelijkheid)
142	m	#NVA	Bart De Wever/Geert Bourgeois/inburgering/afbraakbeleid/ N-VA	V	1	explicite	positive	moyenn e	renforcement (iets te)
143	f	#vk2012	Drukwerkafval	V	3	explicite	positive	moyenn e	
						explicite	positive	forte	
						explicite	positive	forte	
144	f	#gemeenteraadsverkiezingen	verkiezingsfolders/verkiezingscampagne	V	3	explicite	positive	moyenn e	
						explicite	négative	forte	renforcement (!)
						explicite	négative	forte	renforcement (!)
145	f	#vk2012	Bart De Wever/Patrick Janssens	V	2	explicite	négative	moyenn e	renforcement (zowel..als)
						explicite	négative	moyenn e	renforcement (zowel..als)

146	f	#gemeenteraadsverkiezingen	de Kruitfabriek/West-Vlaanderen	v	2	explicite	positive	moyenn e	
						explicite	positive	moyenn e	
147	f	#vk2012	stadshal Gent	v	1	explicite	autre	forte	passé (kon), renforcement (natuurlijk, !)
148	f	#vk2012	Antwerpen	v	2	explicite	négative	moyenn e	
						explicite	négative	moyenn e	
149	f	#vk2012	campagneborden/affiches sp.a/affiches Groen	v	2	explicite	positive	moyenn e	question (?)
						implicite	négative	moyenn e	
150	f	#vk2012	verkiezingen Schoten	x	0				
151	m	#VanBesien	Wouter Van Besien/Groen	v	2	explicite	positive	moyenn e	

						explicite	positive	forte	renforcement (!)
152	m	#bdw	Bart DeWever , Vincent Van Quickenborne , Elio Di Rupo , Laurette Onkelinckx	V	1	explicite	négative	moyenn e	question (?)
153	f	#VanQuickenborne	vergrijzing/brugpensioen/industrie en bedrijven	V	2	explicite	positive	basse	affaiblissement (maar hou misschien eerst de industrie hier en de bedrijven open hé ?)
						explicite	positive	moyenn e	affaiblissement (misschien), renforcement (eerst), question (?)
154	f	#vk2012	Ter Zake/Brussel	V	1	explicite	négative	moyenn e	
155	m	#Groen	fietspaden Gent/fietsers/Groen/verkeersovertredingen	V	2	explicite	négative	moyenn e	renforcement (natuurlijk)
						explicite	négative	forte	

156	m	#Groen	open ruimte/Groen	v	2	explicite	positive	moyenn e	
						explicite	positive	moyenn e	négation (niet)
157	f	#vk2012	Verkiezingen	v	3	explicite	négative	moyenn e	
						explicite	négative	forte	renforcement (egt (sic), geen zak), autre (e)
						explicite	négative	forte	
158	f	#VillaPolitica	Vincent Van Quickenborne	v	1	explicite	négative	moyenn e	autre (toch), renforcement (zo)
159	f	#VlaamsBelang	Vlaams Belang	v	2	explicite	négative	forte	
						explicite	négative	forte	
160	f	#VlaamsBelang	de Kruitfabriek/Vlaams Belang	v	1	explicite	négative	forte	négation (geen), renforcement (letterlijk)

161	f	#VlaamsBelang	Vlaams Belang	v	1	explicite	négative	forte	
162	m	#VlaamsBelang	de Kruitfabriek/Vlaams Belang	v	1	explicite	négative	forte	renforcement (nog maar eens bewezen)
163	m	#Openvld	OpenVld	v	2	explicite	positive	moyenn e	
						explicite	négative	moyenn e	affaiblissement (een beetje)
164	m	#VlaamsBelang	Vlaams Belang/de Kruitfabriek	v	4	explicite	négative	forte	
						explicite	négative	forte	
						explicite	positive	basse	affaiblissement (maar waarom zo'n duidelijk gefrustreerde, agressieve achterban als publiek?)
						explicite	négative	forte	
165	m	#Beke	Leopoldsburg/Wouter Beke	v	1	explicite	négative	forte	renforcement (ola!)

166	f	#VillaPolitica	Villa Politica/debat over wonen	v	1	explicite	négative	forte	
167	f	#VillaPolitica	Eric Van Rompuy/NIP	v	1	explicite	négative	moyenn e	
168	f	#Groen	Ronse/Groen	v	2	explicite	positive	moyenn e	
						explicite	positive	forte	renforcement (!)
169	m	#SP_A	De Laatste Ronde/Bruno Tobback/sp.a	v	4	explicite	autre	moyenn e	
						explicite	négative	forte	
						explicite	négative	forte	renforcement (verre van, en egoïstisch)
						explicite	négative	forte	renforcement (!)
170	f	#VillaPolitica	N-VA/Eric Van Rompuy	x	0				

171	m	#bdw	Bart De Wever/Twitter	v	1	explicite	négative	basse	avenir (vandaag)
172	m	#bdw	pensioenlijn 1765/Vincent Van Quickenborne	v	1	explicite	négative	moyenn e	renforcement (alvast maar)
173	f	#Tobback	Tobback/fietspaden/fietsers	v	4	explicite	négative	moyenn e	
						explicite	négative	forte	
						explicite	négative	forte	
						explicite	positive	forte	renforcement (so)
174	m	#VillaPolitica	Federaal parlement	v	1	explicite	positive	moyenn e	
175	m	#vk2012	geen sanctie voor wie niet gaat stemmen/N-VA	v	2	explicite	positive	moyenn e	
						explicite	positive	moyenn e	négation (niet)
176	m	#VillaPolitica	Ben Weyts/CD&V	v	1	explicite	autre	moyenn e	négation (niet), renforcement (maar blijft met voetjes op de

									grond voor zondag)
177	f	#vk2012	werkloosheid/armoede/Antwerpen/Wouter Van Besien/Groen	V	1	explicite	positive	moyenn e	
178	m	#VillaPolitica	bezuinigingsplan Elio Di Rupo	V	3	explicite	positive	moyenn e	renforcement (tenminste, maar op de staat), négation (niet)
						explicite	autre	moyenn e	renforcement (wel)
						explicite	négative	moyenn e	
179	m	#VlaamsBelang	Bart De Wever/Vlaams Belang	V	2	explicite	négative	moyenn e	condition (wat als BDW wint?), renforcement (!)
						explicite	positive	forte	renforcement (!)
180	f	#vk2012	onbesliste kiezers Leuven	V	2	explicite	négative	basse	
						explicite	négative	moyenn e	affaiblissement (enkele), renforcement

(meer)

181	m	#Tobback	Tobback	v	1	explicite	négative	forte	
182	m	#vk2012	N-VA	v	2	explicite	négative	moyenn e	renforcement (toch)
						explicite	positive	forte	
183	f	#vk2012	Heule/OpenVld/burgemeesterschap/Vi ncent Van Quickenborne	v	1	explicite	positive	moyenn e	
184	f	#Openvld	Wouter Beke	v	1	explicite	négative	basse	affaiblissement (een Beke)
185	m	#gemeenteraadsverkiezin gen	Burgerplicht	v	1	explicite	autre	basse	avenir (zal..gaan), autre (dan eens)
186	f	#vk2012	Verkiezingszondag	v	2	explicite	positive	basse	renforcement (altijd), affaiblissement (een beetje)

						explicite	positive	moyenn e	
187	f	#vk2012	VN-waarnemers/verkiezingen	V	1	explicite	négative	moyenn e	autre (eigenlijk), renforcement (ook)
188	f	#vk2012	Charleroi/verkiezingen over de taalgrens	V	1	explicite	positive	moyenn e	renforcement (ook)
189	f	#gemeenteraadsverkiezin gen	Gemeenteraadsverkiezingen	V	1	explicite	autre	moyenn e	renforcement (!), autre (wel)
190	f	#vk2012	Brugge	V	1	explicite	autre	forte	renforcement (veruit, !!)
191	f	#vk2012	sp.a	V	1	explicite	positive	forte	
192	f	#vk2012	Bart De Wever/burgemeesterschap/Antwerpen	V	1	explicite	négative	forte	renforcement (!)
193	f	#vk2012	stemplicht/OpenVld	V	2	explicite	positive	moyenn e	

						explicite	positive	moyenn e	
194	m	#DeCroo	Alexander De Croo	V	3	explicite	négative	moyenn e	
						explicite	négative	moyenn e	affaiblissement (normal)
						explicite	négative	moyenn e	
195	f	#gemeenteraadsverkiezingen	Gemeenteraadsverkiezingen	V	1	explicite	autre	moyenn e	renforcement (amai amai)
196	m	#DeCroo	Alexander De Croo/Bruno Tobbac/Wouter Beke	V	3	explicite	positive	moyenn e	
						explicite	négative	moyenn e	renforcement (ondertussen, al)
						explicite	négative	basse	
197	m	#vk2012	Kuurne/coalitie N-VA/Bart De Wever	V	2	implicite	positive	forte	
						explicite	négative	basse	

198	f	#gemeenteraadsverkiezingen	Bart De Wever/Patrick Janssens	V	1	explicite	négative	moyenn e	question (?), affaiblissement (maar goed, geen stemrecht hier)
199	m	#vk2012	Franstalige belastingregering/Vlaamse regering	V	1	explicite	négative	moyenn e	renforcement (steeds weer), autre (dan)
200	f	#DeCroo	Alexander DeCroo/Bruno Tobback/ Elio Di Rupo/Wouter Beke	V	2	explicite	autre	basse	question (?)
						explicite	autre	moyenn e	renforcement (én)

## 10.2. Corpus de 200 tweets

N°	Date	Identifiant	Tweet	Sexe de l'utilisateur	Hashtag
1	donderdag 13 september 2012	LoesvanWeert	Hoe het zit met stemrecht van gedetineerden? Via machtiging of tegoed verlofuren. Groot dilemma... #verkiezingen2012	f	#verkiezingen2012
2	donderdag 13 september 2012	BabettevanFJ FM	Het was geen kiezen...het was delen... #verkiezingen2012	f	#verkiezingen2012

3	zaterdag 15 september 2012	DirkDeCock1	#Gemeenteraadsverkiezingen: discriminatie van kandidaten die niet digitaal zijn. Krijgen niet eens een gedrukte versie van de #kiezerslijst.	m	#gemeenteraadsverkiezingen
4	zondag 16 september 2012	Nicolle1408	Elk nadeel heeft zo zijn voordeel #gemeenteraadsverkiezingen in België	f	#gemeenteraadsverkiezingen
5	maandag 17 september 2012	StevenLiefde	Ben die lelijke politieborden op elke hoek van de straat ondertussen wel beugezien. #gemeenteraadsverkiezingen	f	#gemeenteraadsverkiezingen
6	maandag 17 september 2012	geert_d8	Iedereen wordt blijkbaar liever burgemeester dan minister blijven. Veranderde impact beleidsniveaus? #VanQuickenborne #Magnetite#Turtelboom	m	#VanQuickenborne
7	dinsdag 18 september 2012	evelynelanneau	Samen werkt, samen sterk... Hoeveel partijnamen met #samen in oktober? #verkiezingen2012 #Nietorigineel	f	#verkiezingen2012
8	woensdag 19 september 2012	steff_wouters	Denkt Vanackere nu echt dat die 5 dagen na de #gemeenteraadsverkiezingen klaar is met de begroting ?	m	#gemeenteraadsverkiezingen
9	woensdag 19 september 2012	SamvanRooy1	Verkiezingsfolder van #DeWever: "Brussel is van iedereen". Hoe ongeloofwaardig kun je als partij zijn? #NVA <a href="http://t.co/KVv9bS94">http://t.co/KVv9bS94</a>	m	#NVA
10	donderdag 20 september 2012	BartStandaert	Het is raar dat de #NVA nog niet genoemd werd in de "allochtonen polemiek. Normaal is tegenwoordig alles de schuld van #NVA.	m	#NVA
11	donderdag 20 september 2012	Herselv	Ik moet dus met andere woorden niet uit mijn bed komen die zondag om te gaan stemmen... <a href="http://t.co/gNpIFQkY">http://t.co/gNpIFQkY</a> #gemeenteraadsverkiezingen2012	f	#gemeenteraadsverkiezingen

12	donderdag 20 september 2012	ClaesJeroen	Ik hoop dat dit Bart De Wever of Wouter Beke niet op ideeën brengt <a href="http://t.co/Hh5PAhc2">http://t.co/Hh5PAhc2</a> #verkiezingen2012	m	#verkiezingen2012
13	donderdag 20 september 2012	Drklwrs	#bdw #kruitfabriek weeral 100% gescoord! Gewoon niet van zijn stuk te krijgen ondanks verwoede pogingen #NVA	m	#NVA
14	donderdag 20 september 2012	D_Kjartan	Zal het nu zo veel uitmaken of #BDW of PJ burgemeester is van de onvermijdelijke N-VA/SP.a/CD&V coalitie in #Antwerpen2012? Nee. #vk2012	m	#vk2012
15	donderdag 20 september 2012	nathaliewillems	zie hier veel gekap van links naar rechts maar niet echt andersom. Hoe zou dat komen? #gemeenteraadsverkiezingen #spa #openvld #nva #groen	f	#gemeenteraadsverkiezingen
16	vrijdag 21 september 2012	MaiteMorren	Brusselse socialisten hebben 1.000 voorstellen klaar voor hoofdstad <a href="http://t.co/gM5cghUg">http://t.co/gM5cghUg</a> via @Knack #brussel #vk2012	f	#vk2012
17	vrijdag 21 september 2012	aStatement	Het is duidelijk dat #BDW niet de politiekers vreest maar wel onafhankelijke kritische journalistiek! #NVA #atv #karlapers	m	#NVA
18	vrijdag 21 september 2012	krisvanlaethem	een tros campagnemateriaal is in #asse toegekomen... en dat is nog maar het 1ste deel #cdenv #vlabra2010 <a href="http://t.co/9HwZNnEy">http://t.co/9HwZNnEy</a>	m	#CDenV
19	vrijdag 21 september 2012	els_ver	RT @DaveSinardet: De Antwerpse presidentsverkiezingen zijn nu helemaal losgebarsten. <a href="http://t.co/SoQ6dPan">http://t.co/SoQ6dPan</a> #antwerpen2012 #vk2012	f	#vk2012
20	vrijdag 21 september 2012	ElineJoukes	Een blote poep? #vk2012 #antwerpen	f	#vk2012

21	vrijdag 21 september 2012	kim_de_rouck	@vrtderedactie Als ik #filipdewinter zijn verklaring van i.s.l.a.m. mag geloven, is islam het synoniem voor #vlaamsbelang. #één #journaal	f	#VlaamsBelang
22	vrijdag 21 september 2012	GerdaZenner	RT @BollaertA: Ik beken! Ik ben kandidaat voor de provincie. Kleur op 14 oktober het vierde bolletje van cdenv! #vk2012	f	#vk2012
23	vrijdag 21 september 2012	hoegin	Tja, voor een partij op retour is élfke verkiezing er één te veel... #openvld #dso <a href="http://t.co/3QiAJaxo">http://t.co/3QiAJaxo</a>	m	#Openvld
24	vrijdag 21 september 2012	guillaumevds	Mooi hoe #PJ en #BDW elkaar verwijten dat ze negatieve campagne voeren en beiden gelijk hebben. #kruitfabriek	m	#bdw
25	vrijdag 21 september 2012	samvoeten	Ondertussen blijft #BDW Calimero spelen, oa door weg te blijven op de ATV-debatten. Bocht ingezet.	m	#bdw
26	zaterdag 22 september 2012	StephenDawn	#DS:"#Turtelboom is derde weg in #Antwerpen". Tiens, moet dat niet de 14de zijn? #gemeenteraadsverkiezingen #vk2012	m	#gemeenteraadsverkiezingen
27	zaterdag 22 september 2012	ShariDepre	Het zonlicht is ons alvast gunstig! Plaktocht door Tervuren! #Tervuren2012 #vk2012 #openvld <a href="http://t.co/j0ecrpwi">http://t.co/j0ecrpwi</a>	f	#vk2012
28	zaterdag 22 september 2012	Kenny_DB1983	De #NVA praat over zichzelf. Alle andere partijen praten over de NVA. Interessant zo'n verkiezingscampagne <a href="http://t.co/qyNm4RG">http://t.co/qyNm4RG</a> #Verhofstadt	m	#NVA
29	zaterdag 22 september 2012	AnciauxKoen	#OpenVld coryfeeën die veilig in Europese paleizen vertoeven moeten eindelijk eens beseffen: "Zwijgen is goud waard!" #2012 #vtm #rtv #vrt	m	#Openvld

30	zaterdag 22 september 2012	fdevriendt	#Openvld staart naar #nva zoals een konijn naar een lichtbak	m	#Openvld
31	zaterdag 22 september 2012	annemiedemuyt	<a href="http://t.co/upNPnTh2">http://t.co/upNPnTh2</a> #nva #dender stelt kandidaten voor de provincie #dendermonde #denderm2012 #vk2012 #oostvl2012 Ik als lijsttrekker	f	#vk2012
32	zaterdag 22 september 2012	AnnDeHondt	Wedden dat ik de enige eenvrouwspartij heb? RT Lijsten trekken blijft vooral een mannenzaak <a href="http://t.co/YpCQKVkk">http://t.co/YpCQKVkk</a> #vk2012 #katelijne2012	f	#vk2012
33	zaterdag 22 september 2012	janvergote	#Verhofstadt richt pijlen op #NVA. Beste Guy daarmee zijn we nog niet vergeten dat jij #Openvld naar de verdoemenis hebt geholpen!	m	#NVA
34	zondag 23 september 2012	Karl_Drabbe	Wat als we containerwoord 'extreemrechts' 'afschaffen'? #nva	m	#NVA
35	zondag 23 september 2012	LudwigDumont	13 oktober #gemeenteraadsverkiezingen in St. Joost, wel nog geen enkel mogelijk gemeenteraadslid gezien....#campagne?	m	#gemeenteraadsverkiezingen
36	zondag 23 september 2012	Ivandingenen	#Groen zegt aan #NVA: "De tijd van sensibiliseren is voorbij." Om in te kaderen. :) #7dag	m	#Groen
37	zondag 23 september 2012	StephRumm	Een kleine stoute voorspelling ? na #VK2012 heeft #NVA minder burgemeesters dan nu .... #7dag (want door alle anderen uitgesloten ?)	m	#NVA
38	zondag 23 september 2012	Johan_Br	Welke garantie is er dat #nva niet in coalitie gaat met #vb en we geen extreem rechtse stad krijgen? Denk hierover na Antwerpenaar !	m	#NVA

39	zondag 23 september 2012	davidneyskens	#Antwerpen2012, met #groen wordt de Antwerpse Ring een fietsboulevard, laten we hen dan ook graag in de oppositie houden, Vlaanderen dankt u	m	#Groen
40	zondag 23 september 2012	BDRaedemaeker	Bart De Wever praat iedereen losjes onder tafel #villavanhilt #nva	m	#NVA
41	zondag 23 september 2012	nozzlechief	Een vernieuwingsactie voor de #CDenV staat meestal synoniem voor een ander merk choco op hetzelfde brood. #JustSaying	m	#CDenV
42	zondag 23 september 2012	HedwigN	zo'n #WouterBeke die komt vertellen dat @CD&V een duidelijk verhaal mist, zou die beseffen dat hij dat vehikel al een paar jaar leidt?	m	#Beke
43	maandag 24 september 2012	MerckxKristel	RT @JDeMeulemeester: Deze affiche van Hilde Crevits #cdenv is op een wel heel originele manier gebruikt #liefde <a href="http://t.co/layNrbmq">http://t.co/layNrbmq</a>	f	#CDenV
44	maandag 24 september 2012	maaikedeporre	Vandaag krijgen de partijen die geen nationale verankering hebben hun lijstnummer. Nog 20 dagen voor de verkiezingen. #vk2012	f	#vk2012
45	maandag 24 september 2012	Celinevdc	Vanaf dit weekend: nog meer borden in Astene {astene inbegrepen} #astene #vk2012 #deinze2012 #celineop19 <a href="http://t.co/L18JFhnn">http://t.co/L18JFhnn</a>	f	#vk2012
46	maandag 24 september 2012	layoutdokter	Voorspelling: terugval #nva op 14/10.	m	#NVA
47	maandag 24 september 2012	LeslieLaureys	Na amper twee minuten, gaf Peumans al een sneer naar de federale regering. Dat duurde nog lang. #VillaPolitica	f	#vk2012

48	maandag 24 september 2012	LiesjeNi	Dus de kiezers moeten gewoon hopen dat er na 14/10 wel een verhaal komt #nietechtovertuigend #vk2012 <a href="http://t.co/RsSDVrA4">http://t.co/RsSDVrA4</a>	f	#vk2012
49	maandag 24 september 2012	MollyVerbeek	Ouch, dat moet pijnlijk zijn. Ook magazine klaar hebben voor de verkiezingen en dan net voorgestoken worden door Patrick...#antwerpen #groen	f	#Groen
50	maandag 24 september 2012	inevanwezemael	N-Va de kracht van veel blabla en weinig doen! <a href="http://t.co/EJ3bzpsV">http://t.co/EJ3bzpsV</a> @destandaard #gemeenteraadsverkiezingen #nva	f	#NVA
51	maandag 24 september 2012	johandevroe	Mooi initiatief!! >> 1.200 bezoekers en elf modezaken op D-Mode - Het Nieuwsblad <a href="http://t.co/Hprh11Td">http://t.co/Hprh11Td</a> via @nieuwsblad_be #Deerlijk#OpenVLD	m	#Openvld
52	dinsdag 25 september 2012	leensercu	Onveiligheidsgevoel,criminaliteit, allochtonen allemaal in 1 zin. Van integratie wordt nt gesproken. #vlaamsbelang #ugent1410	f	#VlaamsBelang
53	dinsdag 25 september 2012	brionnick	Mooi betoog van #groen #ugent1410	m	#Groen
54	dinsdag 25 september 2012	PieterjanLauwer	Nooit gedacht dit te zeggen, maar Van Besien was volgens mij de enige die wat inhoudt bracht. #Groen #VanBesien #ugent1410	m	#Groen
55	dinsdag 25 september 2012	klaraknaeps	Als de boerka verboden is, waarom mogen die van het #VlaamsBelang dat dan wel? Actie.. Haha <a href="http://t.co/hRjKKNVL">http://t.co/hRjKKNVL</a>	f	#VlaamsBelang
56	dinsdag 25 september 2012	joschelstraete	Nu in de Gentse brievenbussen: " #Gent is geen provinciestad" De Gentse kandidaten #sp_a stellen zich aan u voor.	m	#SP_A

57	dinsdag 25 september 2012	lutdebisschop	dat #cdenv -verhaal moest nochtans pas klaar zijn ná de verkiezingen toch ? Zeeën van tijd ;) @wbeke #UGent1410	f	#CDenV
58	dinsdag 25 september 2012	remybonnafe	Inhoud. Toegepast op gemeentepolitiek, en toch globaal. I like. #vanbesien #ugent1410	m	#VanBesien
59	dinsdag 25 september 2012	IsoldeBuysse	Van Besien aangenaam om naar te kijken, geen stijve hark. Ligt aangenaam in het oor. #woutervanbesien #ugent1410	f	#VanBesien
60	dinsdag 25 september 2012	KimberlyMoen	Inhoudelijk, maar toch moet het weer over Antwerpen gaan... #ugent1410 #vanbesien	f	#VanBesien
61	dinsdag 25 september 2012	itgirlkatia	Het enige wat mij is bijgebleven, als je in VL 2 stenen neerlegt, ze ofwel in de weg staan of in een ander z'n tuin. #ugent1410 #tobback	f	#Tobback
62	dinsdag 25 september 2012	davdberg	#Tobback praat over #cijnskiesrecht. Pijnlijk gezien #SP tgn vrouwenstemrecht was. #Ugent1410	m	#Tobback
63	dinsdag 25 september 2012	jaspervervaert	Inhoud vs partijvz 0-2 #helaas #jmdd #valkeniers #ugent1410	m	#Valkeniers
64	woensdag 26 september 2012	SWanten	De prijs van een woning wordt nog altijd bepaalt door locatie, locatie en locatie. #vk2012	m	#vk2012
65	woensdag 26 september 2012	JolienLaenen	Ik kijk trouwens erg uit naar de verkiezingen, maar ik ben het drama errond zo beu! #verkiezingen2012 #gemeenteraadsverkiezingen	f	#gemeenteraadsverkiezingen

66	woensdag 26 september 2012	francisherbers	De toekomst van #groen in de stad? cc @BCSonsbeek <a href="http://t.co/6ZAcejN">http://t.co/6ZAcejN</a>	m	#Groen
67	woensdag 26 september 2012	paulverbist	Te voet naar een klant! Plezaaaaant! #groen	m	#Groen
68	woensdag 26 september 2012	simonderycke	Bart De Wever... ik kan echt zijn kop niet meer zien. Hij zou beter wa minder in de media komen #misselijk #bedrog #nva	m	#NVA
69	woensdag 26 september 2012	eprenen	Traditionele partijen #OpenVLD , #CD&V en #SPa gaan afgaan als ne gieter op 14/10. Watch my words ;)	m	#Openvld
70	woensdag 26 september 2012	nestvanginderen	Wat zijn CD&V en N-VA toch lief voor elkaar. De operatie wederzijdse beschadiging is een succes te noemen. #cdenv #nva	m	#NVA
71	woensdag 26 september 2012	BobSalesBob	Doe de sociale stemtest: bij letterlijk iedereen die ik ken staat #groen op de laatste plaats. #win #GeitenwollenSokken #dom	m	#Groen
72	woensdag 26 september 2012	delesi35	Goh, weet groen dan niet waarover die staatshervorming gaat? Ze steunt ze nochtans onvoorwaardelijk #sanctorum #villapolitica	m	#VillaPolitica
73	woensdag 26 september 2012	spottermarie	Niet alle aanhangborden zijn geschikt voor affiches... #knokkeheist #verkiezingen2012 <a href="http://t.co/MVT0LKBE">http://t.co/MVT0LKBE</a>	f	#verkiezingen2012
74	woensdag 26 september 2012	godspeedspike	Daar gaan we weer. Dat Calimerogehalte zeg, man man man. Die heeft de goede keuze gemaakt. #villapolitica	m	#VillaPolitica

75	woensdag 26 september 2012	jcoboa	Al die lelijke koppen op die affiches van VB en NVA. Een normale mens wordt er bang van... #VK2012	f	#vk2012
76	woensdag 26 september 2012	mboschnl	Goed begin van mevrouw Lieten met NIP en PIP. Zijn de woorden al trending topic? #nip #pip #villapolitica	f	#VillaPolitica
77	woensdag 26 september 2012	nilsvber	Pip? Nip? Peppi en Kokki? Ideetje voor campagneposter Open VLD #openvld #vk2012 <a href="http://t.co/PdKanobO">http://t.co/PdKanobO</a>	m	#Openvld
78	woensdag 26 september 2012	pierrerambox	#BDW is nu al moe van de kritische vragen. Dat worden zes lange jaren als hij wint. #Knack via #DS <a href="http://t.co/gSYpn5UN">http://t.co/gSYpn5UN</a>	m	#bdw
79	donderdag 27 september 2012	Vrouwliefde	Zopas een persbericht geschreven voor 14 oktober. Gemaild ter nazicht en binnenkort publiceren. #vk2012	f	#vk2012
80	donderdag 27 september 2012	cdelaruelle	Lode Vereeck #LDD is echt té grappig! Altijd iets om naar uit te kijken! #vlapar	f	#LDD
81	donderdag 27 september 2012	jasperpillen	"#OpenVld als Titanic" - reportage in #HLN is hard in z'n analyse maar lijkt correct in z'n diepgang. Eerst 14/10, daarna werk aan winkel.	m	#Openvld
82	donderdag 27 september 2012	sofiedr	Gewaagd... #openvld #verkiezingen2012 <a href="http://t.co/i3Tpap1z">http://t.co/i3Tpap1z</a>	f	#verkiezingen2012
83	vrijdag 28 september 2012	vuurdistel	Ik kan mij niet van de indruk ontdoen dat Patrick Janssens de tactiek van de verschroeide aarde aan het toepassen is. Jammer. #VK2012	f	#vk2012

84	vrijdag 28 september 2012	KarinaVerhoeven	@abeloosf : absoluut ook gevonden toen ik nog in #Oostende woonde! succes met jullie #Jong #VLD ACTIE	f	#Openvld
85	vrijdag 28 september 2012	edevrie1	@alexanderdecroo #SP_a komt ook op voor de rechtste socialist. Dat gaat nog moeilijk worden voor jullie dan. <a href="http://t.co/dQRVNUGE#PSgov">http://t.co/dQRVNUGE#PSgov</a>	m	#SP_A
86	vrijdag 28 september 2012	sarcist	De kracht van verandering is...hetzelfde blijven. #NVA #vk2012 <a href="http://t.co/gQf37nmT">http://t.co/gQf37nmT</a>	m	#NVA
87	vrijdag 28 september 2012	manno_c	Gaat politiek eigenlijk ook over nog iets anders dan over de grote ego's van de politici zelf? #VanQuickenborne #DeClerck #terzakeTV	f	#VanQuickenborne
88	zaterdag 29 september 2012	paulpauwels4	Als #gemeenteraadsverkiezingen toch een #nationale test zijn; Heren, Dames wat doet u hier mee? <a href="http://t.co/pKX82nZV">http://t.co/pKX82nZV</a> +1350 werklozen bij!	m	#gemeenteraadsverkiezingen
89	zondag 30 september 2012	GBteam	Is een nieuw jeugdhuis terug mogelijk in #Lint ? Politiekers draaien rond de pot, #openvld geeft een correct antwoord.	m	#Openvld
90	zondag 30 september 2012	sallyvangorp	Topprestatie van cd&v op debattle!! Proficiat! @orry_vdw @StevenDietvorst #Schilde2012 #cdenv <a href="http://t.co/VGr4uUmd">http://t.co/VGr4uUmd</a>	f	#CDenV
91	zondag 30 september 2012	wimpeeters78	Ik stel voor dat #NVA stopt met campagne voeren, dat doen de anderen wel voor hen @wbeke #7dag	m	#NVA
92	zondag 30 september 2012	evertvanwijk	Bij iedere zin die #wouterbeke uitspreekt wordt zijn neus langer #7dag	m	#Beke

93	zondag 30 september 2012	truckdiver	Waarom blijft #WouterBeke meermaals terug komen op diegene die aan de kant zijn gaan staan? Begint afgezaagd te worden ... #dezevendedag	m	#Beke
94	maandag 1 oktober 2012	krijnjonckheere	Homo's zijn dus geen 'normale' mensen. Bedankt meneer Decroo. #openvld #geenstem #mannekepaul #vtm	m	#Openvld
95	maandag 1 oktober 2012	elkedecruynaere	Hallo media. Niet om iets maar op 14 oktober zijn het overall gemeenteraadsverkiezingen. Niet enkel in Antwerpen. #vk2012	f	#vk2012
96	maandag 1 oktober 2012	jonaswillebe	Ik hoop dat veiligheid het belangrijkste verkiezingsthema is in de gemeente waar #thuis zich afspeelt. @een #vk2012	m	#vk2012
97	dinsdag 2 oktober 2012	SonnyBlieck	#openvld kwam net langs op het werk, ze warren enkel geïnteresseerd in de grote bazen, op de werkvloer hebben we ze niet gezien.	m	#Openvld
98	dinsdag 2 oktober 2012	jorisphilips	Krijgt een idee van hoe de doordeweekse gemeenteraad eruit ziet #tobback aan het woord in monoloog	m	#Tobback
99	dinsdag 2 oktober 2012	beatever	Oh zo toevallige timing. #vk2012 "Een goede samenleving veronderstelt een hoog niveau van sociale herverdeling" <a href="http://t.co/bPGI1FOH">http://t.co/bPGI1FOH</a>	f	#vk2012
100	woensdag 3 oktober 2012	Barbara_Pas	Di Rupo is weer bezig: 210 miljoen € voorafname op belasting pensioensparen. Dank u #CDENV #openvld en #spa ... <a href="http://t.co/aOWJCI1f">http://t.co/aOWJCI1f</a>	f	#CDenV
101	woensdag 3 oktober 2012	ksavjee	Mercedes Van Volcem voorspelt 20 procent voor N-VA in Brugge, ook zonder Pol Van den Driessche #villapolitica	m	#VillaPolitica

102	woensdag 3 oktober 2012	koensteenkiste	Alle respect voor Vandeurzen ...heb maar eens Pascal Smet als collega ...man man man wat een ego-tripper #VillaPolitica	m	#VillaPolitica
103	vrijdag 5 oktober 2012	postbus24	@groen: we worden nu ten tijde van besparen verplicht om met de auto 10km verderop te gaan winkelen dankzij #vanquickenborne #edenred	m	#VanQuickenborne
104	zaterdag 6 oktober 2012	CoolsKat	Een zeer pittig debatje gisteren #termont #deweever #beke Terzake 2012 <a href="http://t.co/QMRvIXy">http://t.co/QMRvIXy</a>	f	#Beke
105	zaterdag 6 oktober 2012	lindewin	Zo meteen Kris Peeters over Veligate en de blunders en wie is de echte leider van CD&V. In #villapolitica, na middagjournaal op Een.	f	#VillaPolitica
106	zaterdag 6 oktober 2012	Freekvlooveren	Wie problemen analyseert, ontdekt meer problemen. Wie kansen analyseert ziet meer kansen #groen is het positieve alternatief	m	#Groen
107	zaterdag 6 oktober 2012	ceetjeden	Een niet-nadergenoemde socialistische partij neemt verkiezingsreclame toch wat te persoonlijk op... #onbeleefd #SP_A#blankenberge2012	m	#SP_A
108	zaterdag 6 oktober 2012	hgillijns	@PaulPauwels4 Nee, als je in huiselijke kring vloekt, is de straf 1 onzevader en 3 weesgegroetjes. Toch volgens #WouterBeke. @PaBriBrus	m	#Beke
109	zondag 7 oktober 2012	jeltenbaguet	'Ieders Stem Telt': een aangename debatnamiddag over de sociale problemen waar de Ronsenaar van wakker ligt. #sp_a zet in op deze thema's!	m	#SP_A
110	zondag 7 oktober 2012	LeenHonings	Voor de rest geen kwaad woord over mijn favoriete partij #groen	f	#Groen

111	zondag 7 oktober 2012	J_o_C_M_e_Y	goekoop van sp-a. 3 kinderen minder dood? En als het nu eens drie randkriminelen waren? dan geen flitspalen? #hetgrotedebat	m	#hetgrotedebat
112	zondag 7 oktober 2012	Pergolakaatje	Van buikdansen naar de intercommunales switchen - Jieha, zie je Wouter Beke denken... #Valkeniers #verkiezingsdebat	f	#Valkeniers
113	zondag 7 oktober 2012	GregLievens	één week voor de verkiezingen, beken ik kleur: #groen	m	#Groen
114	zondag 7 oktober 2012	IndraDewitte	'De kracht vd verandering is copy-paste van ons programma' Wouter Beke over N-va #hetgrotedebat #Eén	f	#hetgrotedebat
115	zondag 7 oktober 2012	Eva121	Neen. RT @Bogiellke: Zegt wat je wil over De Wever maar de mens heeft humor. #hetgrotedebat	f	#hetgrotedebat
116	zondag 7 oktober 2012	ElisaSonnevill e	Uiteindelijk weet ik nog steeds niet op wie ik moet stemmen #hetgrotedebat #vk2012	f	#hetgrotedebat
117	zondag 7 oktober 2012	KimRaes	intussen is #BDW een sterke merknaam, ik zou eens met SOS Piet bellen en een lijn kookpotten op de markt brengen #hetgrotedebat #ijzersmeden	f	#hetgrotedebat
118	zondag 7 oktober 2012	Inge_vdb	Conclusie: twitterfeed over #hetgrotedebat was interessanter dan het debat zelf	f	#hetgrotedebat
119	zondag 7 oktober 2012	saschaluyckx	Sorry Vlaanderen, Bart heeft ruzie met ATV en daarom moeten jullie het nu via de VRT wéér over ons hebben #hetgrotedebat #vk2012	f	#vk2012

120	zondag 7 oktober 2012	EmilieKortrijk	Tijd om te slapen #hetgrotedebat tja sneeuw witje (Goedele :- ) en de 7 dwergen - vrouwen waren kennelijk niet welkom! Spijtig #2012	f	#hetgrotedebat
121	zondag 7 oktober 2012	Ivaniox	Het is overduidelijk, alle partijen hebben schrik van de nva. Precies bende vorming tegen BDW. Zelig! #hetgrotedebat	m	#hetgrotedebat
122	zondag 7 oktober 2012	Leen_DeSchutter	RT @dominiek_be: De realiteit van voorakkoorden en waar het echt om draait, blijkt duidelijker uit #deadline1410 dan uit #hetgrotedebat.	f	#hetgrotedebat
123	zondag 7 oktober 2012	noorke	Heb ik het goed begrepen: GAS-boetes = naar politiestaat? als het van sommige partijen afhangt. Laat kinderen kind zijn aub #hetgrotedebat	f	#hetgrotedebat
124	zondag 7 oktober 2012	JolienDeClercq	Vanavond weinig gehoord over de #dorpsstraat op #hetgrotedebat #vk2012	f	#hetgrotedebat
125	zondag 7 oktober 2012	DavidRMKS	De Wever vs 6 voorzitters.. #hetgrotedebat	m	#hetgrotedebat
126	zondag 7 oktober 2012	fonsvandyck	Voor sommigen begint de campagne zwaar te wegen. En wordt het nog een lange laatste week #hetgrotedebat	m	#hetgrotedebat
127	zondag 7 oktober 2012	_EvelienC	Wouter Beke zet blijkbaar héél graag de tering naar de nering. Of de arme man kent maar één spreekwoord. #hetgrotedebat	f	#hetgrotedebat
128	zondag 7 oktober 2012	jefreyvdstraete	Conclusie #hetgrotedebat: iedereen is voor veiligheid. En Wouter Beke heeft nog maar eens alle open deuren ingetrapt.	m	#hetgrotedebat

129	zondag 7 oktober 2012	marleendehen au	RT @PaulPauwels4: #Hetgrotedebat #politie minder administratie met die #GAS_boetes komt er nog meer administratie voor de #Wijkinspecteur	f	#hetgrotedebat
130	zondag 7 oktober 2012	Kathyvandamme	@GoedeleWachters Wat u ook niet zach op het scherm...Kleinere partijen zoals PvdA+ #hetgrotedebat	f	#hetgrotedebat
131	zondag 7 oktober 2012	BrittLaenen	RT @sp_a: Bruno Tobback: "De essentie van sociaaldemocratie is dat individuen alleen vooruitgeraken door samen te werken." #hetgrotedebat	f	#hetgrotedebat
132	zondag 7 oktober 2012	moerbzzz	Wat doet #ldd in #hetgrotedebat	m	#LDD
133	zondag 7 oktober 2012	svflight	Hoe kleiner de garnaal hoe meer drukwerk-afval hij maakt #gemeenteraadsverkiezingen #dtz	m	#gemeenteraadsverkiezingen
134	zondag 7 oktober 2012	devliegherej	Wie denkt dat GAS-boetes jongeren zal helpen of heropvoeden mag zich dringend bezinnen. Fin. boete is echt géén oplossing#hetgrotedebat	m	#hetgrotedebat
135	zondag 7 oktober 2012	stain_d	Elke keer Wouter Beke "De tering naar de nering zetten" zegt, heffen we het glas. Afgesproken? #hetgrotedebat	m	#hetgrotedebat
136	zondag 7 oktober 2012	michaelvetters	#Tobback zit er elke dag zelf in, zegt hij. Waarom rijdt hij dan niet met het openbaar vervoer? #file #hetgrotedebat	m	#Tobback
137	zondag 7 oktober 2012	pieterdries	#beke blijft toch wat op markske lijken, breng hem ne choco #hetgrotedebat	m	#Beke

138	zondag 7 oktober 2012	loesvancleemput	zeer terecht van #vanbesien. U maakt alleen maar de probleemanalyse. #decroo zeer zwak #hetgrotedebat	f	#VanBesien
139	zondag 7 oktober 2012	grote_boodschap	Jean-Marie Dedecker bewijst weer eens zijn totale overbodigheid #LDD #infantiel	m	#bdw
140	zondag 7 oktober 2012	VlookupMast	Plaatsvervangende schaamte #valkeniers #hetgrotedebat	m	#Valkeniers
141	maandag 8 oktober 2012	elslea	RT @hbvi: #hetgrotedebat verkiezingsangst verlamt de redenaarskunst van menigeen. Alsook de eigenwaarde en de hoffelijkheid.	f	#hetgrotedebat
142	maandag 8 oktober 2012	Couverts	BDW heeft het iets te graag over inburgering om het afbraakbeleid van Bourgeois op dat vlak te verantwoorden... #NVA	m	#NVA
143	maandag 8 oktober 2012	FreyaPiryns	RT @ingejoris: Cool: oproep aan verkiezingskandidaten om drukwerk te compenseren met nieuwe bomen <a href="http://t.co/neuzSmXo">http://t.co/neuzSmXo</a> #vk2012 #ikdoemee	f	#vk2012
144	maandag 8 oktober 2012	Liesbethje1984	RT @evigeysels: #Gemeenteraadsverkiezingen ok, maar hou die folders bij! Denk aan het milieu!	f	#gemeenteraadsverkiezingen
145	maandag 8 oktober 2012	SofieSas	Het feest van de sociale verkiezingen en zowel Janssens als De Wever sturen hun kat.. #iedersstemtelt #vk2012	f	#vk2012
146	maandag 8 oktober 2012	Isoldevde	#kruitfabriek ik ben blij eens geen Antwerps debat te moeten aanhoren. West-Vlaanderen is een welkome afwisseling #gemeenteraadsverkiezingen	f	#gemeenteraadsverkiezingen

147	maandag 8 oktober 2012	LiesjeMarie	En de #stadshal kon natuurlijk niet onbesproken blijven tijdens deze verkiezingen! #Gent #vk2012 #delaatsteronde	f	#vk2012
148	maandag 8 oktober 2012	ilsehackethal	RT @Youssef_Kobo: Het Stad is van iedereen...(met een bovengemiddeld inkomen)#Antwerpen2012 #vk2012	f	#vk2012
149	maandag 8 oktober 2012	Daphne_Van Hecke	sp.a-groen affiches regenen van campagneborden, positief voorteken voor zondag? #gent2012 #vk2012	f	#vk2012
150	maandag 8 oktober 2012	EvelynNuyens	Nog even en we zijn er... #Schoten2012 #vk2012	f	#vk2012
151	maandag 8 oktober 2012	angelo_meuleman	Ik kreeg gistere SMS van niet-Groen-stemmer: "Jullie hebben geluk met #vanbesien" Vind ik ook!	m	#VanBesien
152	maandag 8 oktober 2012	CeesBoer	#DeWever , #VanQuickenborne , #DiRupo , #Onkelinckx . enz enz Moeten die mensen dit land niet besturen ?	m	#bdw
153	maandag 8 oktober 2012	MartineVDA1	#vergrijzing langer werken, ja, OK, maar hou misschien eerst de industrie hier en de bedrijven open hé ? #brugpensioen #vanquickenborne	f	#VanQuickenborne
154	maandag 8 oktober 2012	ilesabeth	Die ene keer dat het over Brussel gaat, is het over de impact ervan op Vlaanderen en de rand ... zucht #terzaketv #vk2012	f	#vk2012
155	dinsdag 9 oktober 2012	KasperDemol	Natuurlijk is er geen plaats voor de fietser als die midden op straat rijdt ipv op het fietspad #Groen #Gent #fail <a href="http://t.co/mUdrPI6z">http://t.co/mUdrPI6z</a>	m	#Groen

156	dinsdag 9 oktober 2012	samvdberghe	Tijd voor open ruimte, de open ruimte buiten het centrum bouwen we niet vol. #groen	m	#Groen
157	dinsdag 9 oktober 2012	LisaDecoster	RT @LaurentCooool: Die verkiezingen interesseren me egt geen zak eDas 1 grote speeltuin #vk2012	f	#vk2012
158	dinsdag 9 oktober 2012	anmoerenhout	Waarom klinken zoveel politici toch zo hees, een week voor de verkiezingen? #quickie #villapolitica	f	#VillaPolitica
159	dinsdag 9 oktober 2012	Loes_Colpaert	Zouden ze snappen dat niemand ze serieus neemt? #VlaamsBelang #Ergernissen @DeKruitfabriek	f	#VlaamsBelang
160	dinsdag 9 oktober 2012	EvaTSeyen	Geen water bij de wijn doen: samenwerken is letterlijk onmogelijk bij hun... #kruitfabriek #vlaamsbelang	f	#VlaamsBelang
161	dinsdag 9 oktober 2012	xSofieJulie	RT @MustiiQue: #VlaamsBelang kustt mijn kluuuteee	f	#VlaamsBelang
162	dinsdag 9 oktober 2012	npeeters	Walgelijke bende, nog maar eens bewezen. #VlaamsBelang #Kruitfabriek #VIER @opVIER	m	#VlaamsBelang
163	dinsdag 9 oktober 2012	wolfr_	liefste #openvld, ik zou graag jullie standpunten lezen, hier een beetje CSS om jullie site bruikbaar te maken	m	#Openvld
164	dinsdag 9 oktober 2012	kvanbeversluy s	snap dat je #vlaamsbelang uitnodigt, maar waarom zo'n duidelijk gefrustreerde, agressieve achterban als publiek? #dekruitfabriek#ergerlijk	m	#VlaamsBelang

165	dinsdag 9 oktober 2012	meur	Ola! Intriges in de leopoldsburgse politiek #leopoldsburg #wouterbeke <a href="http://t.co/g0WilC7u">http://t.co/g0WilC7u</a>	m	#Beke
166	woensdag 10 oktober 2012	sdelrue	RT @lindewin: 'T zijn verkiezingen inderdaad , iedereen schiet op iedereen #debatwonen #villapolitica	f	#VillaPolitica
167	woensdag 10 oktober 2012	Marliesvdwall e	RT @ksavje: Eric Van Rompuy is een NIP. #villapolitica	f	#VillaPolitica
168	woensdag 10 oktober 2012	CelineFoulon	Het positief alternatief voor Ronse! #groen <a href="http://t.co/ik3ecUkT">http://t.co/ik3ecUkT</a>	f	#Groen
169	woensdag 10 oktober 2012	driesvandecruys	Walgelijke prestatie in #laatsteronde van Tobback. Bewijs dat #sp_a verre van mensenlievend is en egoïstisch. Waar zijn de echte sossen?!	m	#SP_A
170	woensdag 10 oktober 2012	pdossche	"ja ik ben scherp want N-VA staat ons naar het leven" Eric Van Rompuy #villapolitica	f	#VillaPolitica
171	woensdag 10 oktober 2012	pvdkiebo	Vandaag houd ik een BDW-loze twitterdag #bdw	m	#bdw
172	woensdag 10 oktober 2012	stevenius	#1765 #vanquickenborne reserveer alvast maar 1777 voor onze generatie	m	#VanQuickenborne
173	woensdag 10 oktober 2012	MarleenCoppens	So true wat #tobback zegt op #een : met auto vervloek ik die fietsers, op mijn fiets de fietsers (overbevolkte fietspaden) en voetgangers	f	#Tobback

174	donderdag 11 oktober 2012	JDeMeulemester	Het valt misschien wat onder de radar, maar straks is het plenaire in het federaal parlement #villapolitica #hoera	m	#VillaPolitica
175	donderdag 11 oktober 2012	TomImberechts	Is dit hopen dat een deel van de #NVA stemmers gaat thuisblijven? #vk2012 "Wie niet gaat stemmen, wordt niet vervolgd" <a href="http://t.co/RkrCwBuT">http://t.co/RkrCwBuT</a>	m	#vk2012
176	donderdag 11 oktober 2012	Kldtje	#BenWeyts niet onder de indruk van kritiek van cd&v maar blijft met voetjes op de grond voor zondag #villapolitica	m	#VillaPolitica
177	donderdag 11 oktober 2012	ingejooris	Meer dan 33.000 werklozen en 1 op 4 kinderen leven in armoede in antwerpen. #vanbesien vn @groen wil dat aanpakken #vk2012 #terzaketv	f	#vk2012
178	donderdag 11 oktober 2012	gertr	#villapolitica Di Rupo bezuinigt tenminste niet over de hoofden van de bevolking, maar op de staat. Dat is in Nederland wel anders. Helaas.	m	#VillaPolitica
179	vrijdag 12 oktober 2012	belmad13	Wat als #BDW wint? Een #VlaamsBelang'er op #OCMW! Alles geven, Bart! #Antwerpen2012	m	#VlaamsBelang
180	vrijdag 12 oktober 2012	BrouwersKarin	Nog nooit zo laat zoveel onbesliste kiezers tegen gekomen #VK2012 #leuven2012 #Leuven En het gevoel nog enkele dagen meer nodig te hebben.	f	#vk2012
181	vrijdag 12 oktober 2012	dwwdevriese	De heer #Tobback zit lang genoeg in de politiek om te weten dat hoop de hoer der wanhopigen is. Meer hierover morgen in mijn HLN-column.	m	#Tobback
182	zaterdag 13 oktober 2012	petertossyn	Grappig toch dat een consevatieve partij "De kracht van verandering" propageert #vk2012	m	#vk2012

183	zaterdag 13 oktober 2012	kimdebue	De #openvld bus in #Heule klaar voor een nieuwe Burgemeester @VincentVQ	f	#vk2012
184	zaterdag 13 oktober 2012	JokeSluydts	Een Beke vuil tussen de tanden? #DeStandaard #Verkiezingen #WouterBeke <a href="http://t.co/UX0yvSMS">http://t.co/UX0yvSMS</a>	f	#Openvld
185	zondag 14 oktober 2012	MounierMalfa it	Zal dan eens de burgerplicht gaan vervullen #gemeenteraadsverkiezingen	m	#gemeenteraadsverkiezingen
186	zondag 14 oktober 2012	katrienboon	Ik word altijd een beetje week van dat straatgebeuren op verkiezingszondag. De democratie op pad. #vk2012 <a href="http://t.co/VUSGIyg4">http://t.co/VUSGIyg4</a>	f	#vk2012
187	zondag 14 oktober 2012	anverstuyft	Komen er eigenlijk ook VN-Waarnemers naar hier om te kijken of de verkiezingen eerlijk verlopen? #vk2012	f	#vk2012
188	zondag 14 oktober 2012	HanneDecout ere	Ook de verkiezingen over de taalgrens volgen we. #Charleroi #vk2012 <a href="http://t.co/ZnGWXGV8">http://t.co/ZnGWXGV8</a>	f	#vk2012
189	zondag 14 oktober 2012	MelanieBellen ge	Spannend wel! #gemeenteraadsverkiezingen	f	#gemeenteraadsverkiezingen
190	zondag 14 oktober 2012	Margaux2804	Verkiezingen zijn veruit het spannendst in #brugge !! #vk2012	f	#vk2012
191	zondag 14 oktober 2012	AulFet	Sp.a moet winnen #vk2012 #brugge	f	#vk2012

192	zondag 14 oktober 2012	MarlinaSegers	RT @stijnbonte: Aan allen die op De Wever hebben gestemd: FUCK YOU! #antwerpen #burgemeester #vk2012 #vk12 #vrtjournaal #vtm	f	#vk2012
193	zondag 14 oktober 2012	SofieDegrendel	Stemplicht vervuld #metplezier Good luck #openvld #vk2012	f	#vk2012
194	zondag 14 oktober 2012	matthiasmaes	Grappig om #decroo te horen zeggen dat een regering normaal een termijn hoort uit te doen #een #Verkiezingen2012 #debat#stekkereruit	m	#DeCroo
195	maandag 15 oktober 2012	zeeuwsch_mei sje	Amai amai amai amai #gemeenteraadsverkiezingen	f	#gemeenteraadsverkiezingen
196	maandag 15 oktober 2012	aaerts	Zouden #Decroo #Toback en #Beke ondertussen al goede vriendjes zijn? ze dragen eenzelfde kruis #terzaketv #gemeenschappelijkvijand	m	#DeCroo
197	maandag 15 oktober 2012	Bram_VDP	N-VA vormt coalitie tegen grootste partij in Kuurne, wat nu De Wever? #kuurne2012 #vk2012	m	#vk2012
198	zondag 14 oktober 2012	kirstentweetin g	De Wever ipv Janssen? Nee. Maar goed, geen stemrecht hier. #gemeenteraadsverkiezingen	f	#gemeenteraadsverkiezingen
199	zondag 7 oktober 2012	dewildemichel	Steeds weer: 'de franstalige belastingsregering', hoe moet ik de Vlaamse dan noemen? De Vlaamse inleveringsregering? #hetgrotedebat #vk2012	m	#vk2012
200	zondag 7 oktober 2012	berreva	Is het dat voor #DeCroo? #Toback? #DiRupo? #Beke? En de rest die én partijvoorzitter én lijsttrekker is? <a href="http://t.co/0UtF9jap">http://t.co/0UtF9jap</a>	f	#DeCroo

10.3. Comparaison du résultat des élections de 2012 et de 2006 <sup>34</sup>

<b>Parti politique</b>	<b>Province</b>	<b>Résultat 2006</b>	<b>Résultat 2012</b>	<b>Evolution du résultat</b>
N-VA	Anvers	26,5 %	35,9 %	+ 35 %
	Limbourg	32,1 %	26,1 %	- 19 %
	Flandre-Orientale	27,3 %	26,1 %	- 4 %
	Brabant flamand	29,4 %	25,8 %	- 12 %
	Flandre-Occidentale	37,6 %	25,3 %	- 33 %
CD&V	Anvers	26,5 %	16,8 %	- 37 %
	Limbourg	32,1 %	27,5 %	- 14 %
	Flandre-Orientale	27,3 %	19,8 %	- 27 %
	Brabant flamand	29,4 %	19,5 %	- 34 %
	Flandre-Occidentale	37,6 %	27,6 %	- 27 %
Open Vld	Anvers	16,2 %	10,1 %	- 38 %
	Limbourg	18,3 %	14,1 %	- 23 %
	Flandre-Orientale	23,8 %	19,3 %	- 19 %
	Brabant flamand	19,2 %	16,8 %	- 12 %
	Flandre-Occidentale	16,8 %	13,4 %	-20 %
sp.a	Anvers	19,5 %	12,8 %	- 34 %
	Limbourg	25,5 %	20,1 %	- 21 %
	Flandre-Orientale	16,8 %	12,7 %	- 24 %
	Brabant flamand	16,2 %	12,1 %	- 24 %
	Flandre-Occidentale	20 %	15,8 %	- 21 %
Groen	Anvers	7,8 %	9,3 %	+ 19 %
	Limbourg	4,6 %	20,1 %	+ 337 %
	Flandre-Orientale	9,1 %	9 %	- 1 %
	Brabant flamand	8,5 %	9,6 %	+ 13 %
	Flandre-Occidentale	6,8 %	7,5 %	+ 10 %
PVDA+	Anvers	1,3 %	3,4 %	+ 170 %
	Limbourg	0,5 %	2,2 %	+ 340 %
	Flandre-Orientale	0,5 %	1,7 %	+ 240 %
	Brabant flamand	0,3 %	1,2 %	+ 300 %
	Flandre-Occidentale	0,5 %	1,3 %	+ 160 %

<sup>34</sup> Source: Vlaanderen Kiest. Base de données du résultat des élections de 2012. <http://www.vlaanderenkiest.be/verkiezingen2012/index.html> [13.04.2013].

Vlaams Belang	Anvers	28,5 %	10,9 %	- 62 %
	Limbourg	18,1 %	9,1 %	- 50 %
	Flandre-Orientale	21 %	9,3 %	- 56 %
	Brabant flamand	18,2 %	6,7 %	- 63 %
	Flandre-Occidentale	17,5 %	7,7 %	- 56 %
LDD	Anvers	Résultat non affiché	Résultat non affiché	-
	Limbourg	Résultat non affiché	Résultat non affiché	-
	Flandre-Orientale	Résultat non affiché	Résultat non affiché	-
	Brabant flamand	Résultat non affiché	Résultat non affiché	-
	Flandre-Occidentale	Résultat non affiché	Résultat non affiché	-

#### 10.4. Liste de la terminologie extraite à l'aide du logiciel TExSIS

Voir le CD-ROM.

#### 10.5. Corpus de 4 629 tweets

Voir le CD-ROM.

#### 10.6. Corpus de 15 952 tweets

Voir le CD-ROM.

#### 10.7. Fiche mémoire

## Dissertatiegegevens bij fiche

### Titel van de scriptie

L'analyse des sentiments appliquée sur des tweets politiques: une étude de corpus

### Auteur(s)

Cynthia Van Hee  
000907132  
cynthia.vanhee.q7132@student.hogent.be

### Taal van de scriptie

Frans

### Vrije trefwoorden

Taaltechnologische studie  
Automatische sentimentanalyse  
Twitter-corpus  
Nederlands

Corpusannotatie  
Manuele sentimentannotatie  
Politieke tweets

### Trefwoorden en annotatie

#### Trefwoord:

Taaltechnologische studie

#### Annotatie:

Deze masterproef is gebaseerd op drie corpora van respectievelijk 200, 4629 en 15952 politieke tweets die werden verzameld tussen 13 september en 15 oktober 2012. De tweets werden gedownload met behulp van een Python script dat de Twitter API gebruikt om tweets met een specifieke hashtag op te vragen en uit te schrijven naar een bestand. Zo werden in totaal 15952 tweets verzameld aan de hand van verschillende politieke hashtags en opgeslagen als .txt-bestand. Het eerste corpus bevat 200 tweets met verschillende politieke hashtags, gaande van Vlaamse politieke partijen en politici tot politieke praatprogramma's. Het tweede corpus bevat 4629 tweets met dezelfde hashtag '#vk2012'. Het derde corpus bestaat uit 15952 tweets met verschillende politieke hashtags.

#### Iwetocodes

H540-nederlandse-taal-en-letterkunde  
H350-linguïstiek  
H353-lexicologie  
T125-automatisatie

### Doelstelling, methode en resultaten

#### Doelstelling:

De huidige technologie voor sentimentanalyse is hoofdzakelijk gebaseerd op NLP systemen die zijn getraind aan de hand van traditionele teksten zoals krantenartikelen en recensies. Uit voorgaande studies blijkt dat annotatieschema's voor sentimentanalyse van traditionele teksten vaak niet voldoende gedetailleerd zijn om op een accurate manier sentiment te detecteren in niet traditionele teksten afkomstig uit sociale media. Het hoofddoel van deze studie bestond erin na te gaan hoe gedetailleerd een dergelijk annotatieschema moet zijn om op een nauwkeurige manier sentiment te detecteren in tweets. Aan de hand van een manuele sentimentannotatie van een Twitter-corpus werd een annotatieschema geanalyseerd en geëvalueerd. Twee bijkomende doelstellingen waren nagaan in hoeverre politieke tweets de politieke voorkeur van de Belgische kiezer op een betrouwbare wijze weergeven en onderzoeken welke terminologie vaak wordt gebruikt in politieke tweets.

#### Methode:

Voor deze studie werden drie verschillende corpora samengesteld bestaande uit respectievelijk 200, 4629 en 15952 politieke Nederlandstalige tweets. De tweets werden verzameld aan de hand van een Python script dat via de Twitter API tweets met een specifieke hashtag kan downloaden en naar een .txt-bestand omzetten. In de 200 tweets van het eerste corpus werden na normalisatie manueel sentimentannotaties aangebracht met het annotatieprogramma Brat. Op basis van de manuele sentimentanalyse werden conclusies getrokken over sentimentuitdrukking op Twitter en over de sterktes en zwaktes van het annotatieschema. In de 4629 tweets van het tweede corpus werd manueel nagegaan hoe vaak de grootste Vlaamse politieke partijen en hun (ex-)voorzitters werden vermeld en welk sentiment er doorgaans aan hen werd gekoppeld. De 15952 tweets van het derde corpus werden gebruikt voor terminologie-extractie met behulp van de software TExSIS. Via part-of-speech-tagging werden woordsoorten toegekend aan de woorden van de geëxtraheerde termen.

#### Resultaten:

Uit de sentimentanalyse blijkt dat het door ons gebruikte annotatieschema veel gedetailleerder is dan het basisschema voor sentimentanalyse dat door Wiebe (2002) werd ontwikkeld. In tegenstelling tot het schema van Wiebe (2002) biedt het huidige annotatieschema de mogelijkheid om impliciet sentiment te detecteren in tweets die enkel feitelijke informatie bevatten, woorden aan te duiden die het sentiment in een zin wijzigen en sentiment te detecteren in ingebedde zinsconstructies. Toch kan het schema verder worden verfijnd om sentimentanalyse in tweets nog nauwkeuriger te maken. Daarvoor moeten dubbelzinnige categorieën zoals 'andere intensiteit' en 'andere polariteit' nog duidelijker worden omschreven, extra subcategorieën worden toegevoegd aan de huidige generieke categorie 'other modifier' en ingebedde constructies nog gedetailleerder worden geannoteerd. Verder blijkt uit deze studie dat 4629 tweets niet voldoende zijn om de politieke voorkeur van de Belgische kiezers nauwkeurig weer te geven. Uit een vergelijking met de resultaten van de gemeenteraadsverkiezingen blijkt dat sommige resultaten gedeeltelijk overeenkwamen met de analyse van het Twitter-corpus, andere dan weer niet. Uit de terminologie-extractie blijkt dat Belgische kiezers op Twitter het naar aanloop van de gemeenteraadsverkiezingen in 2012 vooral hebben gehad over de partijen Open Vld, N-VA, Groen, politieke programma's zoals Het Grote Debat en Villa Politica, de stad Antwerpen en CD&V-politicus Veli Yüksel. De meest terugkerende termen waren zelfstandige naamwoorden. Verder onderzoek

met grotere datasets is nodig om algemene conclusies te kunnen trekken over de rol van Twitter als hulpmiddel om verkiezingsuitslagen te verklaren. Nieuwe onderzoeken over sentimentanalyse zouden ook moeten uitwijzen of een gedetailleerder annotatieschema effectief leidt tot een betere sentimentanalyse van niet-traditionele teksten.