



HAL
open science

L'influence des discours d'accompagnement sur le partage social : identifier et analyser les discours d'escorte sur Twitter.

Justine Simon, Bénédicte Toullec, Romain Badouard, Dario Compagno, Arnaud Mercier, Nathalie Pignard-Cheynel, Brigitte Sebbah

► To cite this version:

Justine Simon, Bénédicte Toullec, Romain Badouard, Dario Compagno, Arnaud Mercier, et al.. L'influence des discours d'accompagnement sur le partage social : identifier et analyser les discours d'escorte sur Twitter.. Journées de Recherche Internationale : "Médias sociaux et corpus de communication médiée par le réseaux. Annotation, analyse, données libres", Oct 2015, Rennes, France. hal-02637927

HAL Id: hal-02637927

<https://hal.science/hal-02637927>

Submitted on 10 Mar 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**IDENTIFIER ET ANALYSER LES DISCOURS
D'ESCORTE SUR TWITTER
L'INFLUENCE DES DISCOURS
D'ACCOMPAGNEMENT SUR LE PARTAGE
SOCIAL**

Simon Justine, Université de Lorraine - CREM

Toullec Bénédicte, Université de Rennes 1 - CRAPE

et

Badouard Romain, Université de Cergy-Pontoise - AGORA

Bigey Magali, Université de Franche-Comté - ELLIADD / LCP

Compagno Dario, Université Sorbonne-Nouvelle - CIM

Mercier Arnaud, Université Panthéon-Assas - CARISM

Pignard-Cheynel, Nathalie, Université de Lorraine - CREM

Sebbah Brigitte, Université Paris-Est Créteil - CEDITEC

Notre texte qui rend compte d'une recherche interdisciplinaire sur les conditions de circulation des informations d'actualité via Twitter, est l'occasion d'un retour réflexif de l'ensemble des chercheurs qui y participent, pour exposer les difficultés méthodologiques rencontrées pour traiter qualitativement des données de masse (des millions de tweets associés chacun à des dizaines de métadonnées).

Notre recherche s'inscrit dans le cadre du projet Info-RSN, financé par l'ANR, analysant les modalités de circulation et de partage d'informations numériques, et porte plus spécifiquement sur le partage sur les réseaux socionumériques de lien(s) URL (Uniform Resource Locator) pointant vers des sites d'information. L'enjeu est l'analyse des discours d'accompagnement de ces liens sur Twitter - discours jugés importants en termes d'encouragement au partage social et d'expression du point de vue des internautes sur l'actualité.

Cet article suit deux objectifs :

- expliciter les choix méthodologiques du projet de recherche global - tant dans l'approche contrastive que dans l'approche itérative - qui ont permis à la fois de construire notre base de données générale mais aussi de constituer les corpus de travail.

- expliquer notre démarche empirique qui a permis de définir et d'identifier ce que nous nommons « discours d'escorte », forme particulière de discours d'accompagnement.

Ce texte suit un développement en quatre temps. Après le cadrage général du projet ANR Info-RSN, nous exposerons la démarche méthodologique, qui se traduit par un va-et-vient constant entre un traitement quantitatif (lié aux enjeux de travailler à partir de millions de données) afin de classer les données et une approche qualitative afin d'en extraire des corpus pertinents. Dans un troisième temps, nous présenterons les différents traits définitoires de la notion de discours d'escorte, appliquée à la problématique spécifique du partage d'information d'actualité. Enfin, une synthèse d'une analyse préliminaire, réalisée à partir d'une extraction aléatoire de données, permettra de proposer une première classification des discours d'escorte.

1. CADRAGE GENERAL : L'INSCRIPTION AU SEIN DU PROJET ANR INFO-RSN

Le projet ANR Info-RSN « Circulation et partage des informations sur les réseaux sociaux numériques et transformations du journalisme » a débuté en 2013. Un axe de ce projet vise plus spécifiquement l'étude des pratiques de partage des liens URL d'actualité sur Twitter. Ce travail a pour objectif de mettre au jour les dynamiques socio-discursives à l'œuvre sur cette plateforme technologique pour exposer les convictions de tout individu et pour encourager le partage d'informations. Une quantité importante de tweets ne se limite pas à la mise en circulation d'un titre d'article accompagné de son lien URL, mais comprend des propos personnels, des marques discursives diverses qui traduisent les intentions de communication de leurs auteurs et peuvent avoir un impact sur les partages de ces messages, sur la lecture des articles associés, etc. L'approche que nous avons privilégiée est donc avant tout discursive, tout en étant empreinte d'une certaine logique conversationnelle même si celle-ci ne constitue pas le cœur de notre analyse.

La première étape de ce travail reposait sur la constitution d'un corpus, et notamment sur la sélection des titres médiatiques les plus pertinents à étudier. Se voulant *a minima* représentative du paysage médiatique, la sélection des titres à intégrer dans l'étude devait tenir

compte d'un certain nombre de critères de catégorisation et de diversité :

1. la catégorie médiatique : version numérique des médias traditionnels / médias nés en ligne (NEL) ;
2. le type de support médiatique étudié : médias télévisuels / radiophoniques / presse papier ;
3. la zone de diffusion des médias : médias nationaux et médias régionaux ;
4. le type d'informations traitées : médias d'information générale et médias d'information spécialisée (information sportive) ;
5. les publics visés : public généraliste / public jeune ;
6. le modèle socio-économique du média : médias indépendants / médias issus de conglomérats.

Cette première sélection a conduit à recenser les trente-deux sites d'information suivants :

lequipe.fr	nouvelsoobs.com	lexpress.fr	france24.com
lemonde.fr	liberation.fr	lesechos.fr	latribune.fr
lefigaro.fr	lepoint.fr	huffingtonpost.fr	franceinter.fr
leparisien.fr	20minutes.fr	rue89.com	franceinfo.fr
parismatch.com	lacroix.com	bfmtv.fr	lelab.europe1.fr
courrierinternational.com	itele.fr	metro.fr	atlantico.fr
marianne.net	francetv.info	slate.fr	lopinion.fr
lejdd.fr	melty.fr	lci.tf1.fr	mediapart.fr

1 Les trente-deux sites d'information sélectionnés

La seconde réflexion de cette étude repose sur l'identification de ce qui peut être défini comme « information ». Travaillant sur la circulation de sujets d'actualité, chaque article étant identifié sur un site médiatique par son URL, c'est donc cette même URL, identifiant unique et plutôt fiable, qui a été sélectionnée comme indicateur de base du partage d'informations. Pour effectuer le suivi des articles (et les processus de rééditorialisation), l'hypothèse est faite que l'URL est l'élément de base le plus « stable », repérable même si elle est raccourcie, et ne subissant pas d'intervention de la part des twittos - en dehors des captures d'écran et partage par photo -, considérés comme minimes. Il s'agissait donc de recouper les URL des articles issus de médias et les URL intégrées aux tweets aspirés

quotidiennement sur la période définie. Cette étape permettait non seulement de constituer deux bases de données, mais également à terme d'identifier les différences ou éventuels changements de titres (le numérique permettant aux journalistes de faire évoluer en permanence leurs articles, dont leur titre, en fonction de leur nombre de vues ou de l'évolution des événements).

Afin de pouvoir constituer ces deux bases de données (la base des articles partagés et celle des tweets publiés), il était nécessaire de circonscrire une période d'étude répondant à un double enjeu de représentativité : période suffisamment longue pour être représentative des pratiques de partage informationnel mais suffisamment courte pour autoriser des analyses quantitatives supportant nos puissances de calcul. La période choisie est celle se déroulant entre le 15 mai et le 15 octobre 2014, ce qui a permis de prendre en compte les élections européennes, et toute une série de faits d'actualité ayant fait la une.

L'étape suivante a trait au tri des métadonnées (cf. tableau n°2). Concernant les métadonnées des tweets, c'est l'API de Twitter qui les crée. Il y en a une centaine (du nom du détenteur de compte, à la couleur de son fond de profil, en passant par le nombre de ses abonnés (*followers*) ou sa géolocalisation). Concernant les métadonnées des articles, le travail est plus ardu, car il s'agissait de retrouver au sein des métadonnées de chaque site, les informations jugées utiles pour notre étude et en faire une base de données homogène, tous médias confondus. L'exercice s'avère périlleux, car si le référencement auprès des moteurs de recherche oblige chaque site à normaliser un tant soit peu certaines données, il n'en reste pas moins que chaque site et chaque CMS qui en dessine l'architecture, structure les données à sa façon. On a donc parfois dû jongler entre les sites pour retrouver et extraire des informations utiles. La captation de métadonnées liées à un article publié sur un site est donc loin d'être un long fleuve tranquille, tant il est vrai que c'est aux chercheurs de normaliser l'information pour chaque média pour en faire une base de données homogènes.

Ces métadonnées, relativement nombreuses, ne sont pas toutes pertinentes pour l'étude et alourdissent donc inutilement la base de données et les processus de traitement. Il s'agissait alors d'identifier l'ensemble des variables pouvant avoir une incidence sur ce qui figurait alors comme la question de recherche initiale, à savoir le

partage des informations sur Twitter. Parmi les nombreuses variables possibles, voici quelques éléments considérés comme importants pour traiter la problématique :

Métadonnées liées au twitto	Métadonnées liées à l'URL (à l'article)	Métadonnées liées au tweet
nom du twitto	URL de l'article	texte du tweet
identifiant du twitto	titre initial de l'article	date et heure du tweet
profil du twitto	nom du média	identifiant du tweet généré par Twitter
date de création du compte du twitto	nom de l'auteur de l'article	identifiant unique du tweet créé par SemioCast
nombre de tweets émis au jour de la capture	date de mise en ligne de l'article	nombre de retweets
nombre de followés	identifiant unique de l'article créé par SemioCast ¹	source du retweet
nombre de followers		authenticité du retweet

2 Aperçu de certaines métadonnées significatives dans le cadre de notre étude

Deux principales bases de données ont donc été constituées puis recoupées :

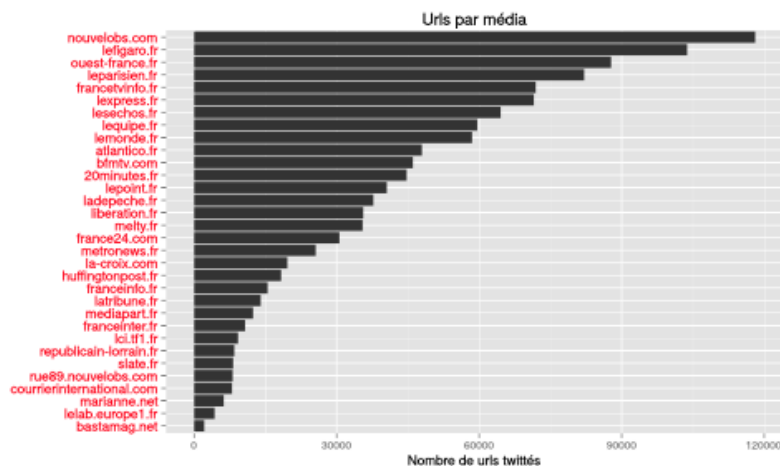
1. une base de données (BD1) de tous les tweets contenant une URL pointant vers un des sites d'information retenus pour l'étude (et comprenant la centaine de métadonnées construites par Twitter, dont on n'a retenu finalement qu'une cinquantaine, avec une quinzaine utilisées régulièrement comme facteurs explicatifs ou interprétatifs) ;
2. une base de données (BD2) recensant l'ensemble des articles de presse qui ont été ainsi mis en circulation (les sites périphériques tels que *Maville* par exemple pour *Ouest-France* ou les sites participatifs tels que *L'Obs Le Plus* n'ont pas été pris en compte).

L'obtention de ces premières données a conduit à des ajustements du processus de collecte des informations et de vérification de la procédure, mais également à identifier les bruits ayant contribué à induire d'éventuels biais dans la base de données. Ces vérifications ont reposé dans un premier temps sur l'identification quantitative des données recueillies. Deux principaux critères ont conduit à l'identification d'erreurs :

¹ L'institut d'études marketing et d'opinion travaillant sur les médias sociaux SemioCast est un partenaire du projet ANR.

1. la surreprésentation de certains articles voire de certains médias (on a ainsi dû écarter *Atlantico.fr*) car pour plusieurs sites d'information, un même article pouvait soit être débité en plusieurs "sous url", soit correspondre à des dizaines voire des centaines d'URL différentes à cause d'un jeu avec la page d'atterrissage ou la rubrique ;
2. l'identification de doublons (à savoir la comptabilisation à deux reprises d'une même URL dont les métadonnées sont identiques). Cette question des doublons pose problème car il s'agit d'identifier en amont le critère permettant de discriminer les tweets originaux des tweets générés artificiellement (par exemple un tweet produit par un twitto ou produit de façon automatique).

Sur l'ensemble de la période étudiée, 22 millions de tweets ont été récoltés. Après tri, on en conserve 17,3 millions dans notre base (soit 21% mis à l'écart). Cela représente 1.149.529 usagers uniques de Twitter et permet de constituer la seconde base forte de 997.002 articles uniques. Voici le tableau récapitulatif du nombre d'URL twittées pour chaque média étudié.



- 3 Graphique récapitulatif du nombre d'URL twittées pour chaque média sur la période du 15 mai au 15 octobre 2014

Si cette première étape de travail a représenté une partie assez importante du temps de recherche lié au projet, elle a par ailleurs contribué à établir une définition stabilisée de l'objet étudié.

2. CONSIDERATIONS METHODOLOGIQUES

D'un point de vue méthodologique, le travail effectué a reposé sur la convergence et la combinaison de différents types d'expertise nécessaires à appréhender un objet d'étude sollicitant l'association d'une approche quantitative et qualitative. Des chercheurs de différents domaines scientifiques ont ainsi collaboré, renforçant la notion d'« interdisciplinarité focalisée » proposée par Patrick Charaudeau (2008) - cela à plusieurs niveaux :

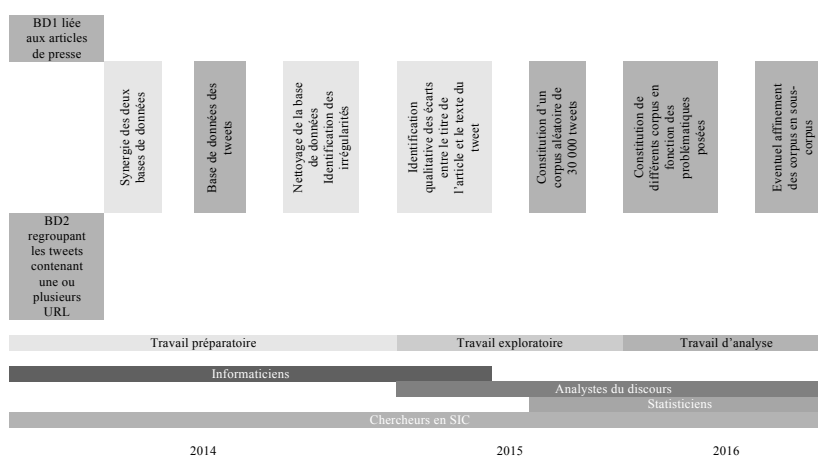
1. dans la collecte et l'analyse des tweets (SemioCast présentant un partenariat historique avec Twitter et une expertise en informatique et en intelligence artificielle) ;
2. dans le développement d'algorithmes informationnels et dans une approche quantitative des données (SemioCast et informaticiens du laboratoire Lcoms) ;
3. dans l'approche statistique (statisticiens du LETA) ;
4. et enfin - par rapport à ce qui est plus spécifiquement abordé au sein de cet article - dans une approche qualitative des discours (chercheurs en Sciences de l'Information et de la Communication et en Sciences du langage affiliés à différents laboratoires, dont le CREM).

La démarche proposée par le projet en général dépasse alors la seule approche discursive en mobilisant des méthodes et pratiques liées au traitement de millions de données, tout en conservant un regard critique sur ses opportunités et ses limites (Boyd & Crawford, 2011). Ainsi, au terme de la constitution de la première version de la base de données, différentes limites ont été posées :

1. la dépendance à la production d'URL produites par les médias (exemple d'Atlantico.fr) ;
2. la dépendance aux métadonnées produites par le dispositif technologique Twitter ;

3. la prise en compte d'autres dispositifs liés à la publication de tweets, tels que Bitly, Tweetdeck, Hootsuite, Buffer, etc. ;
4. la présence de redondances (liées aux limites des algorithmes de récupération de l'information).

Un « nettoyage » central et nécessaire de la première base de données constituée a ainsi été effectué par les différents chercheurs mobilisés avant toute création d'un corpus exploitable. Le schéma ci-dessous récapitule les différentes étapes ayant permis la constitution de corpus exploitables et parallèlement la mise à plat des concepts définis dans les deux parties suivantes.



4 Histogramme des étapes de constitution des bases de données (BD1 et BD2) et des corpus d'analyse

Ces différentes étapes de constitution de corpus ont également contribué à affiner les hypothèses initialement formulées ; certaines ont ainsi pu être écartées et d'autres intégrées :

1. contrairement à ce qui était anticipé, les émoticônes et emojis ne constituent pas des facteurs de partage pertinents, leur nombre étant relativement faible (en moyenne de 10 sur 10 000 tweets) ;
2. initialement, peu d'intérêt à la production automatique de tweets n'était consacré mais cette dimension s'est avérée incontournable

puisque 22% des tweets relèvent de publications automatiques (Pignard-Cheynel & Compagno, 2015).

Parallèlement, les conditions d'exploitabilité propres à la question des corpus ont donc été fixées, à savoir (Pincemin, 1999) :

1. la temporalité : identifier les dynamiques temporelles existantes ;
2. les thématiques : mettre en évidence les thématiques et rubriques des partages générés ;
3. les formats et formulations : identifier les corrélations entre la nature et les phénomènes de partage observés et le contenu des articles, leur mode de rédaction, leur format, leur longueur, le nombre et la nature des commentaires ;
4. l'augmentation sémiotique des tweets (ajout de vidéos, diaporamas, photos, etc.) : en étudiant les différentes modalités de valorisation de ces contenus à partir de la plateforme Twitter (liens hypertextuels, insertions d'images, etc.).

Le travail de construction de corpus de discours d'escorte s'est donc caractérisé par différentes phases s'inscrivant dans une méthodologie hypothético-déductive. Les questionnements d'origine et hypothèses qui avaient été dans un premier temps établis ont été soumis à un travail de déduction de prédictions issu des phases de test et de prélèvement notamment aléatoire de données. La dernière étape reposant alors sur la validation d'hypothèses émises lors de travaux réalisés à partir de corpus problématisés et interrogeant la capacité à confirmer (ou non) les hypothèses premières. Cette méthodologie a été validée après discussion par l'ensemble des partenaires dans un souci d'évitement de tout biais et d'optimisation des corpus alors constitués. Ces phases ont ainsi combiné approche quantitative et approche qualitative renvoyant à chaque fois aux limites de chacune dans une logique heuristique - chaque difficulté ayant fait l'objet de discussions collectives pluridisciplinaires contribuant à la résolution du problème.

Les conséquences de ce processus itératif de travail sont la co-construction des problématiques et de l'objet de la catégorisation des discours d'escorte. Elles se concrétisent par constitution de corpus de travail d'analyse de discours d'escorte orientés chacun vers une

problématique qui sera par la suite confrontée aux résultats d'analyses statistique effectuées par ailleurs.

3. LE DISCOURS D'ESCORTE

3.1. DEFINITION

Afin de pouvoir analyser les degrés d'influence des discours d'accompagnement sur le partage social, nous avons réinvesti la notion de discours d'escorte.

C'est dans la tradition littéraire qu'apparaît la notion de « discours d'escorte » à partir des travaux de Gérard Genette (1982). Dans la continuité des recherches en analyse du discours qui mettent l'accent sur son caractère dialogique et interdiscursif (Volochnov 1977/1929 ; Moirand 2007 ; Garric & Longhi 2013), cette notion a par la suite été reprise pour analyser des discours apparaissant sur des dispositifs numériques (Rabatel 2011 ; Simon 2011, 2015). Dans une problématique interdiscursive, le discours d'escorte relève d'un acte d'énonciation.

Appliquée à l'étude de notre corpus de tweets contenant des liens URL d'articles d'information, nous proposons la définition suivante : le discours d'escorte est un discours original qui entoure une URL d'un article ou une unité discursive composée de l'URL et de sa formule de partage telle que générée par les médias et qui tente d'influencer la lecture et l'interprétation du contenu mis en lien.

Le discours d'escorte du lien URL présent dans le tweet permet ainsi, d'une part, de mettre au jour sa dimension pragmatique dans le fait d'inciter le lecteur à cliquer sur le lien, et, d'autre part, d'influencer la lecture et l'interprétation du discours lié. Dans cette manière de poser le problème, la production d'un discours d'escorte est perçue comme une pratique discursive « stratégique » où l'énonciateur principal a un rôle central. Qui plus est, ce dernier peut implicitement chercher à faire adhérer le lecteur à sa propre appréciation du discours qu'il met en lien (constitution d'une preuve, doute relatif à la valeur de vérité du contenu, commentaire décalé, ironique ou humoristique, etc.).

Deux types d'énonciateurs sont alors différenciés, en fonction du degré d'originalité de la publication : l'énonciateur qui crée un discours personnel unique autour du lien et l'énonciateur dont

l'investissement personnel est pratiquement nul puisque son tweet est le résultat d'une écriture assistée (dans le cas de la reprise de la formule de partage telle que générée par les médias). De plus, du fait des propriétés du dispositif Twitter, la dimension pragmatique du discours d'escorte ne se résume pas à l'incitation à cliquer sur le lien mais elle prend en compte toutes les possibilités de partage social, telles que le retweet, la mention, etc.

Dans un premier temps, le réinvestissement de cette notion a pour objectif d'identifier ce qui correspond à un discours d'escorte afin de proposer une première classification. C'est l'objet de notre quatrième partie, qui synthétise le travail fait à partir d'une extraction aléatoire de données de 30 000 tweets. Dans un deuxième temps, la notion de discours d'escorte devra se confronter aux limites théoriques et méthodologiques introduites par la problématique du numérique à travers plusieurs analyses plus poussées de corpus spécifiques. Nous sommes en effet conscients, comme plusieurs chercheurs travaillant en analyse du discours numérique (Calabrese 2011 ; Longhi 2014 ; Paveau 2013), que l'application d'une notion - qui au départ n'a pas été pensée pour l'étude de corpus numériques - peut soulever de nouvelles difficultés. Comme le précise Marie-Anne Paveau (2013), analyser le discours numérique suppose de modifier un peu les théories et méthodes - les théories du discours n'étant pas forcément outillées pour appréhender le numérique. L'analyse du discours d'escorte sur Twitter pose ainsi la question du renouvellement des pratiques discursives appréhendées dans leur matérialité spécifique.

3.2. LE DISCOURS D'ESCORTE FACE A LA VARIETE DES DISCOURS D'ACCOMPAGNEMENT

Avant de pouvoir proposer une classification des discours d'escorte jugés pertinents pour l'analyse future, nous avons dû identifier toutes les formes de discours d'accompagnement, à partir de l'extraction aléatoire de données.

Une première opération de tri de cet échantillon aléatoire a été menée afin de distinguer ce qui relève du texte produit automatiquement (notamment par des « bots », des boutons de partage ou encore des retweets) qui méritent toutefois d'être considérés comme des actes d'énonciations (Magri-Mourgues & Rabatel, 2015), mais que nous qualifierons de « frustrés », c'est-à-

dire limités à la stricte répétition d'une énonciation première, de ce qui relève d'un acte d'énonciation donnant lieu à la production d'un discours complémentaire.

Grâce à la métadonnée appelée « source », nous avons pu écarter les tweets générés automatiquement à travers des outils de publication assistés par des algorithmes. Les publications uniquement constituées d'un lien URL ont aussi été mises de côté. Les retweets serviles, correspondant au simple clic sur le bouton retweet et ne donnant pas la possibilité à l'internaute de modifier le contenu initial, n'ont pas été jugés pertinents pour l'analyse. Il en est de même pour les tweets serviles (générés grâce au *tweet button* présent par exemple sur le site d'un média, produisant un discours se composant de l'URL et reproduisant le début de l'article partagé). Bien que ceux-ci fassent partie de la définition proposée plus haut (unité textuelle composée de l'URL et de sa formule de partage telle que générée par les médias), le tweet servile n'engage pratiquement pas l'internaute du point de vue de son investissement personnel. Il sera en revanche utile afin d'analyser les formulations médiatiques.

Nous avons ensuite identifié plusieurs formes de discours d'accompagnement d'indexation, qui ont été mises de côté également. Cet ensemble correspond :

1. aux formes de tweets générés de manière semi-automatique; (ajoutant au tweet servile une formule spécifique comme par exemple une signature préprogrammée) ;
2. aux tweets présentant un hashtagage de(s) mot(s) du titre de l'article ;
3. aux tweets où des termes de rubriquage ont été ajoutés.

Dernière précision : afin de mettre l'accent sur le partage social, c'est-à-dire sur la circulation de l'information via des internautes ordinaires, nous avons écarté les publications émanant de comptes de médias officiels ou de personnes se présentant comme professionnels de l'information. Elles font l'objet d'une étude parallèle.

4. CLASSIFICATION DES DISCOURS D'ESCORTE

L'étape qui a suivi l'opération de tri présentée précédemment nous a permis de mettre au jour trois classes de discours d'escorte : 4.1. Les discours d'escorte d'opinion, 4.2. Les discours d'escorte d'éditorialisation et 4.3. Les discours d'escorte de communication. De manière opératoire, nous avons distingué ces trois classes mais elles ne sont pas exclusives.

4.1. LES DISCOURS D'ESCORTE D'OPINION

Les discours d'escorte d'opinion correspondent à des tweets qui présentent un fort degré de subjectivation. Ils mettent en avant un avis, une opinion, que ce soit par rapport à l'événement, aux modalités de son traitement journalistique ou en contraste avec l'avis d'un autre internaute.

On relève tout d'abord un assez grand nombre d'expressions de sentiments, de mises en avant d'un point de vue critique sur l'événement et sur les acteurs concernés évoqués dans l'article. Cela peut se faire de manière positive ou négative et à travers différents marqueurs discursifs. L'ajout d'émoticônes et d'emoji vient souvent renforcer le texte. On remarque aussi l'utilisation récurrente de formes oralisées. La critique peut tout autant passer par la syllabisation, le renforcement grâce à l'usage de majuscules, comme c'est le cas dans l'exemple suivant :

1- « MA GNI FI QUE !!! Est ce qu'il existe plus con comme raisonnement ? ift.tt/1xAyL1b »²

Nous avons ensuite rencontré plusieurs discours d'escorte discutant le traitement journalistique du média mis en lien. Les formes de médiacritiques peuvent constituer une recommandation à lire, grâce à une présentation positive, ou au contraire, être de véritables critiques négatives. Dans l'exemple 2, le twitto va jusqu'à interpellier le média grâce à l'utilisation d'une mention afin de partager sa critique.

2- « Quand [@Melly](#) fait de la merde pudibondo-démago qui donne mauvaise haleine melty.fr/jennifer-lawre... via [@Tris_Acatrinei](#) »

² Afin de respecter l'anonymisation des tweets, nous ne reproduisons que le texte du tweet, sans modifications orthographiques. De plus, nous avons modifié la mise en forme des liens pour homogénéiser leur présentation sur l'ensemble de l'article.

La mise en avant d'une opinion peut enfin correspondre au partage d'expérience de lecture. Le twitto vise clairement à orienter la réception de l'article. Il peut le commenter (exemple 3), le reformuler au second degré, le résumer de manière subjective, le présenter comme l'essentiel à retenir à propos d'un fait d'actualité.

3- « Prendre un peu de recul et lire : liberation.fr/politiques/201... »

4.2. LES DISCOURS D'ESCORTE D'EDITORIALISATION

Bruno Bachimont parlait en 2007 d'éditorialisation pour évoquer la pratique d'indexation pour la publication numérique. Il précisait que cette notion permettait de « souligner le fait que les segments indexés sont enrôlés dans des processus éditoriaux en vue de nouvelles publications ». Nous étendons cette définition à un usage par des individus lambda qui, dans leur publication et dans une intention objectivante, s'attachent au contenu informationnel en le valorisant, en le mettant en relief ou en le contextualisant.

La formulation d'une opinion est ici absente. On ne cherche pas ouvertement à influencer le lecteur. La rhétorique est proche de celle des professionnels de l'information. Il s'agit souvent de discours impersonnels, caractérisés par l'effacement énonciatif. La formulation du discours d'escorte est de plus souvent construite à partir d'une syntaxe objectivante : juxtapositions de phrases averbales (exemple 4), infinitivations, passivations, etc.

4- « Assassinat de Hervé Gourdel : tristesse et indignation chez les politiques [#dlvr.it/70ICXb](http://dlvr.it/70ICXb) [#Les verbatims](https://twitter.com/hashtag/Les_verbatims) »

La mise en avant d'une logique de catégorisation de l'information est enfin caractéristique de ce mode de traitement. Le twitto peut ajouter un nom de rubrique (avec ou sans hashtag) et créer des hashtags indexateurs à partir de noms de lieux, de personnalités (exemple 5), de dates clés, etc.

5- « La rébellion de [#Bartolone](https://twitter.com/Bartolone) : Il va publier son ouvrage "Je ne me tairai plus - Plaidoyer pour un socialisme populaire" <http://lelab.europe1.fr/Claude-Bartolone-va-publier-son-ouvrage-Je-ne-me-tairai-plus-Plaidoyer-pour-un-socialisme-populaire-16712> »

4.3. LES DISCOURS D'ESCORTE DE COMMUNICATION

Sont classés dans cette catégorie les discours d'escorte qui reposent sur une intention de communication, c'est-à-dire dont la dimension interlocutive est explicite. Ce qui importe n'est pas tant l'avis du twitto sur le sujet évoqué ou encore l'éditorialisation du contenu informationnel, mais l'accent est mis sur l'interpellation directe cherchant à accrocher l'attention du lecteur, à créer un fil conversationnel, voire même à l'inciter à l'action.

Le signe arobase offre tout d'abord cette fonction d'adresse, cela à plusieurs niveaux d'interaction : l'interpellation peut se faire de manière interpersonnelle (exemple 6), l'adressage peut être multiple, mais encore, l'interpellation peut se faire de manière publique et visible dans la timeline des followers, grâce à l'ajout d'un signe généralement non invasif (souvent le point) devant l'arobase ([.@xxx](#)).

6- « [@NemoOm](#) Enfoiré c toi qui avais raison ;) la il y a plus de doute... [rmcsport.bfmtv.com/football/om-bi...](#) »

D'autres moyens sont à la portée du twitto qui souhaite attirer l'attention de son public à propos de l'information qu'il souhaite partager. Dans l'étude de l'ensemble des données aléatoires, nous avons pu rencontrer différentes pratiques discursives considérées comme des modes d'interpellations : la question, la question rhétorique, le remerciement (exemple 7), le conseil, l'avertissement, la menace, etc. L'usage de l'impératif et l'emploi des pronoms de la deuxième personne (« tu », « vous ») accompagnent souvent ces pratiques discursives.

7- « [@lifrancky](#) Merci m'sieur ! Voilà le bon lien : [metronews.fr/paris/paris-un...](#) »

L'interpellation peut directement être en lien avec le sujet évoqué dans l'article partagé ou au contraire porter sur autre chose (sur l'acte de partager lui-même par exemple). Elle peut de plus chercher à toucher un public spécifique : les motards, les parisiens (exemple 8), les parents mais encore les journalistes (exemple 9).

8- « PARISIENS, ET SI ON PARLAIT DE VOS IMPOTS ? [lepoint.fr/politique/pari...](#) »

9- « On s'en fiche, vous savez. “[@LePoint](#) : Valérie [#Trierweiler](#) est à Madagascar pour une semaine : [lepoint.fr/people/valerie...](#)” » (tweet accompagné d'une carte photo de Valérie Trierweiler)

Enfin, les discours d'escorte de communication peuvent avoir une dimension pragmatique. Les incitations à l'action sont variées :

incitations à donner un avis, à débattre sur un sujet polémique, incitations à retweeter, incitations à devenir follower d'un compte (exemple 10), incitations à voter, incitations conatives qui poussent à l'action en dehors du dispositif numérique (exemple 11).

10- « Divorce : les bijoux de famille, ça se rend... à la famille [po.st/KCschy](https://www.po.st/KCschy) suivez moi : [#stefakir](https://twitter.com/stefakir) »

11- « Les [#SansDents](https://twitter.com/SansDents) enflamment les réseaux sociaux + RDV 19h à l' [@Elysee](https://twitter.com/Elysee) : lefigaro.fr/politique/2014... »

4. EN GUISE DE CONCLUSION

D'un point de vue méthodologique, ce travail témoigne des difficultés que peuvent éprouver les chercheurs de différentes disciplines à travailler sur de grandes bases de données et confirme la nécessité de combiner des compétences complémentaires, en conciliant notamment une approche macro (quantitative) et micro (qualitative). Cette étude permet également de souligner la difficulté de travailler sur des éléments décontextualisés (tweets sous forme de tableau), conduisant à adopter une posture de vigilance constante afin d'éviter les biais analytiques (absence de prise en compte des images). Enfin les méthodes *top-down* et *bottom-up* s'avèrent complémentaires et permettent autant d'affiner la définition de concepts nécessaires à la poursuite du travail (comme celui du discours d'escorte), que d'établir des critères pertinents de constitution de corpus.

Un autre constat s'impose du point de vue de la problématique envisagée : celui de l'intérêt d'approfondir la question des conditions de circulation de l'information via les réseaux socionumériques. En termes de pratiques, il a tout d'abord été essentiel de pouvoir distinguer les publications manuelles individuelles des publications automatiques et semi-automatiques, générées en grande partie par les médias eux-mêmes. Les rhétoriques professionnelles et individuelles ont également présenté quelques différences dont l'analyse reste à approfondir. Une autre piste de recherche encore importante à creuser à propos de la question de la circulation de l'information repose sur les différences de pratiques entre publications manuelles et serviles (sous formes de tweets ou de retweets).

Cette étude préliminaire a enfin pointé une richesse du point de vue de l'élaboration des discours d'escorte, mettant en œuvre différents types de stratégies, mais elle nécessite des prolongements

de recherche à partir de corpus spécifiques : analyse de tweets à dimension polémique, usage des hashtags, analyse de la dimension interlocutive des tweets - notamment grâce à l'usage de mentions, étude de tweets multimodaux (tweets composés de texte et d'images / vidéos / diaporamas) - soulignant l'importance de la place de l'image au sein du discours d'escorte et l'impact qu'elle peut avoir sur la construction du sens (Simon & Toullec in Mercier & Cheynel, à paraître), etc.

REFERENCES BIBLIOGRAPHIQUES

- Bachimont, Bruno (2007). « Nouvelles tendances applicatives : de l'indexation à l'éditorialisation », in Gros Patrick (ed.), *L'indexation multimédia*. Paris : Hermès, 313-326.
- Boyd, Danah & Crawford, Kate (2011), « Six Provocations for Big Data. A Decade in Internet Time », Symposium on the Dynamics of the Internet and Society, accessible en ligne sur : <http://ssrn.com> abstract=1926431 ou <http://dx.doi.org/10.2139/ssrn.1926431>.
- Calabrese, Laura (dir.), (2011) : « L'Internet, corpus sauvage. Nouvelles ressources, nouveaux problèmes ? », *Le Discours et la Langue, Revue de linguistique française et d'analyse du discours*, 2.1.
- Charaudeau, Patrick (2008) : « La justification d'une approche interdisciplinaire de l'étude des médias », *Communication, L'analyse linguistique des discours des médias : apports, limites et enjeux*, Éditions Nota Bene, Québec, accessible en ligne sur : <http://www.patrick-charaudeau.com/La-justification-d-une-approche.html>.
- Compagno, Dario & Pignard-Cheynel, Nathalie (2015) : « L'information automatisée : le rôle des bots dans le partage de l'information d'actualité sur *Twitter* ». Colloque « Diversité culturelle et pluralisme de l'information à l'ère numérique : mêmes combats ? », Université Sorbonne Nouvelle - Paris 3.
- Garric, Nathalie & Longhi, Julien (2013) : « Atteindre l'interdiscours par la circulation des discours et du sens », *Langages et Société*, 144, 65-83.
- Genette, Gérard (1982) : *Palimpsestes*, Paris : Seuil.

- Longhi, Julien (2014) : « L'hybridation du discours institutionnel à l'épreuve du numérique. Renouveau et reconfiguration de la parole institutionnelle », in J. Longhi & G.-E. Sarfati (dir.) : *Contribution à l'analyse des discours institutionnels et politiques*. Paris : L'Harmattan, 167-188.
- Magri-Mourgues, Véronique & Rabatel, Alain (2015) : « Quand la répétition se fait figure », *Semen*, 38, accessible en ligne sur : <http://semen.revues.org/10285>.
- Moirand, Sophie (2007), *Les discours de la presse quotidienne, observer, analyser, comprendre*, Paris : PUF.
- Paveau, Marie-Anne (2013) : « Ce qui s'écrit dans les univers numériques », *Itinéraires*, 3. Accessible en ligne sur <http://itineraires.revues.org.bases-doc.univ-lorraine.fr/2313>.
- Pincemin, Bénédicte, 1999 : « Construire et utiliser un corpus : le point de vue d'une sémantique textuelle interprétative », in A. Condamines, M.-P. Pery-Woodley & C. Fabre (coord.) : Actes de la Conférence TALN'99, 26-36. Accessible en ligne sur http://icar.univ-lyon2.fr/membres/bpincemin/biblio/pincemin_taln99.pdf.
- Rabatel, Alain (2011) : « Analyse énonciative des s/citations du site d'Arrêt sur images », in A. Jaubert & alii (dir.) : *Citations II, Citer pour quoi faire ? Pragmatique de la citation*. Louvain-La-Neuve : Harmattan-Academia, 13-36.
- Simon, Justine (2011) : « Analyse des liens hypertextes sur le site Mediapart dans l'affaire Woerth/Bettencourt/Mediapart », in I. Saleh & alii (dir.) : *H²PTM 2011, Hypermédias et pratiques numériques*. Paris: Hermès-Lavoisier, 153-164.
- Simon, Justine (2015) : « Le discours hypertextualisé : Une notion essentielle pour l'analyse du web », in I. Saleh & alii (dir.) : *H²PTM 2015, Le numérique à l'ère des objets, de l'hypertexte à l'hyper-objet*. Paris: Hermès-Lavoisier, 3-20.
- Simon, Justine & Toullec, Bénédicte, à paraître : « Quand le tweet multimodal renouvelle la pratique de partage de l'information : approches sémio-discursive et argumentative », in A. Mercier & N. Pignard-Cheynel (éds.) : *Commentaires et partages d'informations sur les réseaux socionumériques*. Paris : Éditions de la MSH.
- Volochinov, Valentin Nikolaevič (1977/1929) : *Le Marxisme et la philosophie du langage*, Paris : Minuit.