Projet TAL sur le sexisme ordinaire

« Il faut beaucoup aimer les hommes. Beaucoup, beaucoup. Beaucoup les aimer pour les aimer. Sans cela, ce n'est pas possible, on ne peut pas les supporter. »

Marguerite Duras, *La Vie matérielle*, 1987

Avertissement

Le texte des tweets est inscrit dans un cadre sur fond gris. Par la nature de notre sujet d’étude, leur contenu peut être très cru.

Certains tweets diffament directement une personne nommée. Nous avons choisi de les reproduire ici tel quel mais cette reproduction, dans un but scientifique, ne constitue en rien une infirmation ou une confirmation des propos tenus dont l’auteur du tweet reste seul responsable.

Table des matières

[Introduction 2](#_Toc514017699)

[1. Constitution de notre corpus 2](#_Toc514017700)

[I.A. Les spécificités Tweeter : 2](#_Toc514017701)

[1) Un service gratuit mais privé 2](#_Toc514017702)

[2) Les textes 3](#_Toc514017703)

[3) Avantages 3](#_Toc514017704)

[I.B Moissonnage et sélection des tweets 3](#_Toc514017705)

[1) Solution pour moissonner les vieux tweets 3](#_Toc514017706)

[2) Sélection manuelle des tweets de notre corpus 3](#_Toc514017707)

[2. Adaptation des catégories et analyse du corpus 3](#_Toc514017708)

[2.A Modifications apportées aux catégories 3](#_Toc514017709)

[1) Nouvelles catégories : 3](#_Toc514017710)

[2) Modification des catégories existantes : 4](#_Toc514017711)

[2.B Analyse du corpus 5](#_Toc514017712)

[1) Statut discours 5](#_Toc514017713)

[2) Type discours 5](#_Toc514017714)

[3) Acte verbal 5](#_Toc514017715)

[4) Acte physique 6](#_Toc514017716)

[5) Thème 6](#_Toc514017717)

[6) Nature de relation entre les personnes impliquées 6](#_Toc514017718)

[7) Acte de langage 7](#_Toc514017719)

[8) Milieu 7](#_Toc514017720)

[9) Sexe de l’agissant 7](#_Toc514017721)

[10) Sexe du subissant 7](#_Toc514017722)

[11) Mots-dièse et références 8](#_Toc514017723)

[3. Essai d’annotateur automatique 10](#_Toc514017724)

[3.A Ensemble de mots 10](#_Toc514017725)

[3.B Structures remarquables 11](#_Toc514017726)

[Annexes : 11](#_Toc514017727)

[A1) Chaîne de traitement 11](#_Toc514017728)

[A2) Phénomènes remarqués 11](#_Toc514017729)

[Références bibliographiques 12](#_Toc514017730)

# Introduction

* Contexte
  + Tweets comme matériel de base
  + Sélection sur le mot-dièse #balancetonporc
    - Décrivant une interaction entre une victime et un agresseur qui est cité
    - On a choisi de ne pas aller vers des comptes comme @chaircollab qui reprennent des tumblrs pour se distinguer du travail de nos camarades.
  + Phénomène qui éclate en octobre 2017 suite à #metoo.
    - Répercussions mondiales du phénomène qui est décliné dans de nombreuses langues : #Aufschrei : cri, tollé, #Nopiwouma au Sénégal)
  + Un hashtag généraliste (ce n'est pas comme les tumblrs spécifiques à un métier).
  + Spécificité du mot-dièse #balancetonporc : l’appel à nommer
  + Nous allons nous restreindre à un contexte pro et ne pas prendre en compte les tweets utilisant ce mot-dièse pour le commenter.
* Problématique
  + Catégoriser les tweets récupérés (et récolter des informations pertinentes)
* Plan

1. Constitution de notre corpus
2. Adaptation des catégories et analyse du corpus
3. Essai d'annotateur automatique

# 1. Constitution de notre corpus

## I.A. Les spécificités Tweeter :

### 1) Un service gratuit mais privé

Il monétise ses données (produites gratuitement par utilisateurs) => limitations de l'API gratuite :

* Il faut donc payer pour chercher dans les tweets vieux de + de 7 jours
* Fenêtre de 15 minutes avec un max de 450 requêtes
* Mais possibilité de chercher via le site web https://twitter.com/search-advanced

### 2) Les textes

* Très courts (140 caractères)
* Pleins d'abréviations, d'écriture phonétique, de fautes, d’émojis
* Utilisent parfois du texte dans des images

### 3) Avantages

* Réseau social :
  + Les textes sont dits par des utilisateurs. Cela donne de l’information contextuelle.
  + Citation des utilisateurs par des références @référence
  + Les tweets sont reliés dans une discussion. Les tweets successifs permettent parfois d’avoir des précisions sur le tweet initial (pour savoir si c’état de l’ironie par exemple).
* Présence de mots-dièse #hashtag dans le texte des tweets.

## I.B Moissonnage et sélection des tweets

### 1) Solution pour moissonner les vieux tweets

API non officielle « GetOldTweets » pour moissonnage automatique :  
 <https://github.com/Jefferson-Henrique/GetOldTweets-python>

Produit des fichiers CSV retravaillés dans Excel et Google Spreadsheet.

### 2) Sélection manuelle des tweets de notre corpus

* Essayer de retrouver la structure et le contenu des posts de tumblr :
  + Une interaction entre une victime et un agresseur dans un contexte pro
* On regarde seulement la catégorisation. On considère que le tweet a été correctement identifié comme correspondant à notre critère auparavant (ici, manuellement).

4213 tweets lus, **50** retenus : 1.2%

L’appel a nommé explicitement n’est pas retenu car très peu de tweets respecte ce critère (cela doit être mis en relation avec la polémique sur l’aspect délation/dénonciation du mouvement). Seul 11 tweets mentionnent le nom de l’agissant.

Les tweets sont sélectionnés sur la période de novembre 2017, du 3 au 29.

# 2. Adaptation des catégories et analyse du corpus

## 2.A Modifications apportées aux catégories

### 1) Nouvelles catégories :

* Compte auteur
  + Très important pour information contextuelle = sexe, autres tweets, type de compte (humoristique/parodie).
* Nombre de retweets (de 886 à 0)
* Nombre de favoris (de 596 à 0)
  + Retweets et favoris dépendent beaucoup du réseau de l’auteur. Son tweet sera plus vu et donc potentiellement retweeté et/ou favorisé s’il est très connecté.
* Milieu pour détecter le milieu professionnel
* Agit pour détecter qui est l’agresseur
* ASX pour détecter son sexe (ASX)
  + M = Masculin
  + F = Féminin
  + U = Inconnu
  + G = Groupe de sexe indéterminé
* Subit pour détecter qui est la victime
* SSX pour détecter son sexe (SSX)
  + Mêmes valeurs que pour ASX
  + Il peut être déterminé, si la victime est l’énonciateur, en regardant le compte auteur et en essayant de déterminer le sexe en fonction de celui ex : @PrénomFéminin.
* URL pour aller voir le tweet directement

On aurait pu également avoir une catégorie pour déterminer l’acte physique (baiser forcé, main forcé…) ainsi que les parties du corps mise en jeu.

### 2) Modification des catégories existantes :

* Acte verbal :
  + Ajout de « Demande / Proposition »
  + Ajout de « Aucun » (il n’y a pas d’acte verbal)
  + Généralisation de « Compliment » en « Remarque »
  + Possibilité du double Remarque + Demande

La généralisation de compliment en remarque permet de ne pas avoir à se prononcer sur le contenu mélioratif ou dépréciatif d’un énoncé, parfois assez difficile à saisir :

Un collègue vient de me dire tu es désagréable, mortuaire, un peu vilaine et pas pro. Le seul truc qui te sauve c'est ton IMC" #balancetonporc"

Il s’agit à la fois d’un contenu dépréciatif (« désagréable », « mortuaire », « un peu vilaine », « pas pro ») suivi d’un contenu mélioratif (« ton IMC », indice de masse corporelle, qui implique la personne qui subit correspond aux canons populaires / publicitaires actuelles en ce qui concerne cette variable biologique).

* Acte physique :
  + Contact (il est décrit dans le texte)
  + « Incertain » plutôt que « Non contact » car c’est souvent ambigu et non précisé
  + De plus ambiguïté de l’acte auto-physique (la victime se touche elle-même, par exemple à la demande de l’agissant)
* Acte de langage :
  + Ajout de « Aucun »
  + Ajout de « Inconnu »
  + Ajout de « Interrogatif »

Exemple d’acte de langage interrogatif :

Un mec qui m'harcelait dans ma classe lache Tu veux manger ma bite" alors que je suis au téléphone avec ma mère. J'avais 13 ans. #balancetonporc"

Il y a dans cet énoncé aucune promesse / menace, ordre ou déclaration plus ou moins subjective. Il s’agit d’une proposition / demande faite par l’agissant à la personne qui subit.

## 2.B Analyse du corpus

Dans cette partie, nous analysons les statistiques catégorie par catégorie et nous comparons avec les résultats avec mes camarades qui ont étudié les posts. Leurs résultats sont en pourcentage en gris à la fin de chaque tableau quand ils étaient disponibles.

Une dernière sous-partie est dédiée à l’analyse des mots-dièse et des références.

### 1) Statut discours

Cette catégorie prend deux valeurs. Si l’énonciateur est la personne qui parle, il s’agit d’un « vécu » personnel qu’elle exprime sur les réseaux sociaux. Sinon c’est un fait rapporté arrivé à une autre personne.

|  |  |
| --- | --- |
| **Vécu** | **Rapporté** |
| 24 | 26 |
| 48% | 52% |
| 80% | 20% |

Le corpus se divise équitablement. De nombreux tweets reprennent des informations parues sur des sites d’information pour la relayer. À noter des tweets complexes comme ce dernier :

#Harcèlementsexuel au #MJS par #ThierryMarchallBeck (militant... #féministe et #lgbt , ,membre cabinet #Hamon ): «J’ai dû le masturber pour m’en débarrasser» - #PS #BalanceTonPorc #MeToo #Politique . Libé - 14.11.17http://www.liberation.fr/france/2017/11/14/harcelement-sexuel-au-mjs-j-ai-du-le-masturber-pour-m-en-debarrasser\_1610103?utm\_medium=social&utm\_source=twitter …

Il rapporte une information publiée dans un article de *Libération* mais à l’aide d’une citation issue directement du vécu de la victime. Nous avons privilégié la valeur « vécu » mais dans ce cas c’est très discutable.

### 2) Type discours

Cette catégorie note s’il y a une « citation », indiquée à l’aide de guillemets, dans le tweet. Sinon, on considère qu’il s’agit d’un « récit ». La présence d’une citation montre à voir une parole captée de la situation sexiste ayant eu lieu et exposée directement, renforçant l’impact du tweet.

|  |  |
| --- | --- |
| **Récit** | **Citation** |
| 36 | 14 |
| 72% | 28% |
| 9% | 91% |

Il semble y avoir beaucoup plus de récit que de citation, les auteurs préférant décrire la situation et éventuellement les paroles prononcées.

### 3) Acte verbal

Cette catégorie essaye de catégoriser l’échange verbal.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Aucun** | **Remarque** | **Demande** | **Menace** | **Insulte** | **Remarque + Demande** | **Autre** |
| 29 | 8 | 8 | 2 | 2 | 1 | 0 |
| 58% | 16% | 16% | 4% | 4% | 2% | 0% |
| 0% | 3% | 0% | 17% | 10% | 0% | 70% |

Pour 58% des situations rapportées, il n’y a pas eu d’échange verbal. Lorsqu’il y en a un, c’est en en premier lieu autant de remarques (16%) que de demandes (16%). La menace et les insultes ne venant que bien après (4% chacune).

### 4) Acte physique

Cette catégorie indique s’il y a un acte physique clairement mentionné dans le tweet ou pas. Dans ce dernier cas, l’acte physique est incertain : certains tweets peuvent le sous-entendre fortement mais contenu explicite, nous nous contentons de mettre « Incertain ».

|  |  |
| --- | --- |
| **Incertain** | **Contact** |
| 32 | 18 |
| 64% | 36% |

Un contact explicite ne concerne que 38% des tweets. On aurait pu introduire une troisième valeur, « probable » pour indiquer que si aucun contact n’explicite, il a une forte présomption qu’il se soit réalisé, comme dans ce tweet :

#balancetonporc JEAN LUC UGUEN, ancien maire de CLEDER dans le Finistère. et magistrat! J'étais jeune il m'avait demandé de venir un samedi à la mairie pour étudier mon CV pour un poste en petite enfance, à peine fermé la porte ....... il me dégoute et le pire c'est qu'il a

### 5) Thème

Cette catégorie aborde le thème du tweet, son idée principale qui fait qu’il est sexiste.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Sexuel** | **Autre** | **Apparence** | **Rôle traditionnel** | **Apparence et sexuel** | **Critique du féminisme** | **Médecine** | **Reproduction** |
| 23 | 21 | 3 | 2 | 1 | 0 | 0 | 0 |
| 46% | 42% | 3% | 2% | 1% | 0% | 0% | 0 |
| 20% | 14% | 22% | 15% | 0% | 4% | 13% | 12% |

Sans surprise, le thème est en premier lieu sexuel (46%). Agression sexuelle proprement dite, ou commentaires subjectifs, propositions / demandes et promesses / menaces en ce sens consituent le thème le plus présent dans les tweets.

### 6) Nature de relation entre les personnes impliquées

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Vertical personne d’autorité** | **Vertical professionnel** | **Horizontal** | **Inconnue** | **Relation clientèle** |
| 14 | 14 | 10 | 6 | 6 |
| 28% | 28% | 20% | 12% | 12% |
| 48% | 29% | 11% | 12% | 0% |

On notera qu’on n’a pas pris la peine de coder le sens de la relation vertical, c’est-à-dire qui, de celui qui subit et de celui qui agit, est en position supérieure dans la relation. Dans nos tweets, ce n’est jamais le subordonné qui agit, mais il aurait pu être intéressant d’ajouter cette information aux catégories.

Par personne d’autorité on entend des policiers, des politiciens, des professeurs qui sont dans leur champ professionnel et agissent de façon sexiste envers des citoyens. Ces derniers peuvent être amener à avoir cette interaction dans le cadre professionnel (une présentatrice de télé), scolaire (des étudiants en cours) ou pas (un citoyen arrêté par un policier). Dans tous les cas, celui qui subit et celui qui agit ne sont pas dans le même champ professionnel.

Le vertical professionnel ne s’applique que si celui qui subit et celui qui agit sont dans le même champ professionnel : un rédacteur en chef avec une journaliste par exemple.

On a préféré qualifier les interactions mêlant professeurs et étudiants dans la catégorie « personne d’autorité » car même s’il s’agit du même milieu, ils ne sont pas dans le même champ professionnel. Il n’existe pas de lien de subordination contractuel direct entre eux, issu d’un contrat de travail, mais un lien d’autorité du fait de la fonction.

### 7) Acte de langage

Cette catégorie essaye de catégoriser l’acte de langage.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Aucun** | **Expressif** | **Promissif** | **Interrogatif** | **Inconnu** | **Injonctif** | **Déclaratif** |
| 29 | 5 | 4 | 4 | 3 | 3 | 2 |
| 58% | 10% | 8% | 8% | 6% | 6% | 4% |
| 0% | 24% | 11% | 0% | 0% | 11% | 54% |

Beaucoup d’actes rapportés dans ces tweets ne font pas état d’un échange verbal (58%). Le premier acte qui se dégage est l’expressif : il s’agit d’une information énoncée avec une forte valeur subjective. Les deux suivants sont les promesses ou menaces et les demandes ou propositions.

### 8) Milieu

Cette catégorie indique le milieu professionnel impliqué.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Autre** | **Politique** | **Études** | **Santé** | **Police** | **Cinéma** | **Média** | **Syndicat** | **Sport** |
| 19 | 9 | 8 | 4 | 3 | 3 | 2 | 1 | 1 |
| 38% | 18% | 16% | 8% | 6% | 6% | 4% | 2% | 2% |

Beaucoup d’indétermination dans cette rubrique. On remarque néanmoins que la politique (18%) et les études (16%) arrivent en tête.

Cette dernière catégorie agrège les environnements universitaires et scolaires (lycée 🡺 tweet sur la fête du bac, collège, école élémentaire, maternelle 🡺 tweet sur l’instituteur).

La catégorie Média ne regroupe que la presse et la radio (le cinéma a sa propre rubrique).

### 9) Sexe de l’agissant

|  |  |  |  |
| --- | --- | --- | --- |
| **Masculin** | **Inconnu** | **Groupe de sexe indéterminé** | **Féminin** |
| 40 | 8 | 1 | 1 |
| 80% | 16% | 2% | 2% |

Les agissants sont en grande majorité (80%) de sexe masculin.

Le groupe de sexe indéterminé est la police, or la proportion de femmes policiers (titulaires dans la police active) était de 16 % en 2007 en excluant le personnel administratif et les adjoints de sécurité (ADS, « emploi jeunes »), beaucoup plus féminisé (76% et 42,4% respectivement en 2005) (Pruvost, 2008).

Le seul agissant déterminé comme étant de sexe féminin semble en fait relever d’une ironie critique :

« **La coiffeuse** qui me masse le crâne pendant 5 minutes au lieu de me faire un shampoing, je me suis senti harcelé #BalanceTonPorc ».

### 10) Sexe du subissant

|  |  |  |  |
| --- | --- | --- | --- |
| **Féminin** | **Inconnu** | **Masculin** | **Groupe de sexe indéterminé** |
| 27 | 16 | 5 | 2 |
| 54% | 32% | 10% | 4% |

Ceux qui subissent sont en premier lieu les femmes (54%). Les hommes ne comptent que pour 10%.

Une forte proportion d’indétermination du sexe (32%) est présente.

3 valeurs pour le sexe de celui qui subit ont été déterminées en regardant que celui qui subit était bien l’énonciateur et en regardant le nom de compte :

@MarineTaxil, @sylvielegall1 @helo\_maindiaux (helo comme Héloïse)

On peut remarquer la difficulté que consiste ce tweet en particulier :

Ouais, aujourd'hui **j'ai été livré** par le même livreur dans un intervalle de 2j. Il m'a même appelé avant d'arriver pour me demander si j'étais chez moi et en partant il m'a souri en me regardant. J'ai peur, il a mon numéro et mon adresse... #BalanceTonPorc

*« J’ai été livré »*, sans e, fait pencher pour que la victime énonciatrice soit un homme. Après analyse du compte énonciateur, @Shiwie\_, il s’avère que c’est une utilisatrice ayant commise une faute d’accord. De plus, la discussion que génère ce tweet nous apprend que cette déclaration est ironique, faite pour se moquer du phénomène autour du mot-dièse. Nous n’avons néanmoins pas utilisé ces informations supplémentaires, nous concentrant sur le tweet en lui-même, enregistré comme ayant une victime masculine.

### 11) Mots-dièse et références

#### a) Mots-dièse

Nous avons compté la fréquence des mots-dièse utilisés et nous avons essayé de dresser une typologie de ceux-ci :

|  |  |  |  |
| --- | --- | --- | --- |
| **N°** | **Mot-dièse** | **Fréquence** | **Classification** |
| 1 | #balancetonporc | 50 | Thème central |
| 2 | #metoo | 8 | Thème central |
| 3 | #directan | 2 | Politique (en direct de l’Assemblée nationale) |
| 4 | #fn | 1 | Politique (parti) |
| 5 | #cpolitique | 1 | Politique (émission) |
| 6 | #draguignan | 1 | Lieu |
| 7 | #education | 1 | Milieu professionnel |
| 8 | #gaboury | 1 | Personne |
| 9 | #noël | 1 | Autre |
| 10 | #québec | 1 | Lieu |
| 11 | #buade | 1 | Lieu (rue) |
| 12 | #vieuxquébec | 1 | Lieu |
| 13 | #yqb | 1 | Lieu (code de l’aéroport de Québec) |
| 14 | #disney | 1 | Société |
| 15 | #stopvaw | 1 | [Thème central](http://www.stopvaw.org/) |
| 16 | #orangetheworld | 1 | [Thème central](http://www.undp.org/content/undp/fr/home/ourwork/ourstories/orange-the-world.html) |
| 17 | #violcencesfaitesauxfemmes | 1 | Thème central |
| 18 | #harcelement | 1 | Thème central |
| 19 | #porc | 1 | Thème central |
| 20 | #cauet | 1 | Personne |
| 21 | #montreuil | 1 | Lieu |
| 22 | #marchalbeck | 1 | Personne |
| 23 | #alfranken | 1 | Personne |
| 24 | #rosearmy | 1 | [Thème central](https://www.rosearmy.com/) |
| 25 | #chefdescale | 1 | Fonction |
| 26 | #martinique | 1 | Lieu |
| 27 | #airfrance | 1 | Société |
| 28 | #harcèlementsexuel | 1 | Thème central |
| 29 | #mjs | 1 | Parti politique |
| 30 | #thierrymarchallbeck | 1 | Personne |
| 31 | #féministe | 1 | Thème central |
| 32 | #lgbt | 1 | Autre |
| 33 | #hamon | 1 | Personne |
| 34 | #ps | 1 | Parti politique |
| 35 | #politique | 1 | Politique |
| 36 | #lauriannedeniaud | 1 | Personne |
| 37 | #hollande | 1 | Personne |
| 38 | #cagnes | 1 | Lieu (Cagnes sur mer) |
| 39 | #lasiesta | 1 | Lieu (boîte à Cagnes) |

Les mots-dièse permettent d’insister sur un point, de raccrocher ce tweet à d’éventuelles recherches dessus dans twitter.

On peut compter les mots-dièse ensuite par catégorie :

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Thème central** | **Politique** | **Parti Politique** | **Lieu** | **Milieu pro.** | **Personne** | **Autre** | **Société** | **Fonction** |
| 10 | 4 | 2 | 9 | 1 | 8 | 2 | 2 | 1 |
| 26% | 10% | 5% | 23% | 3% | 20% | 5% | 5% | 3% |

Si des mots-dièse reliés à notre thème central forment ¼ de ceux-ci. Viennent ensuite les lieux (23%) puis mes personnes (20%). Les personnes citées ne sont pas forcément celles ayant agi. On peut citer ce tweet :

#Harcèlementsexuel au #MJS par #ThierryMarchallBeck (militant... #féministe et #lgbt , ,membre cabinet **#Hamon** ): «J’ai dû le masturber pour m’en débarrasser» - #PS #BalanceTonPorc #MeToo #Politique . Libé - 14.11.17http://www.liberation.fr/france/2017/11/14/harcelement-sexuel-au-mjs-j-ai-du-le-masturber-pour-m-en-debarrasser\_1610103?utm\_medium=social&utm\_source=twitter …

Il utilise le mot clé #Hamon pour dire que Thierry Marchall Beck était membre de son cabinet. Volonté de montrer l’importance de celui-ci par sa proximité avec les cercles du pouvoir ou volonté de salir Benoît Hamon ? D’un autre côté, le mot-dièse #ThierryMarchallBeck désigne bien un supposé « porc à balancé ».

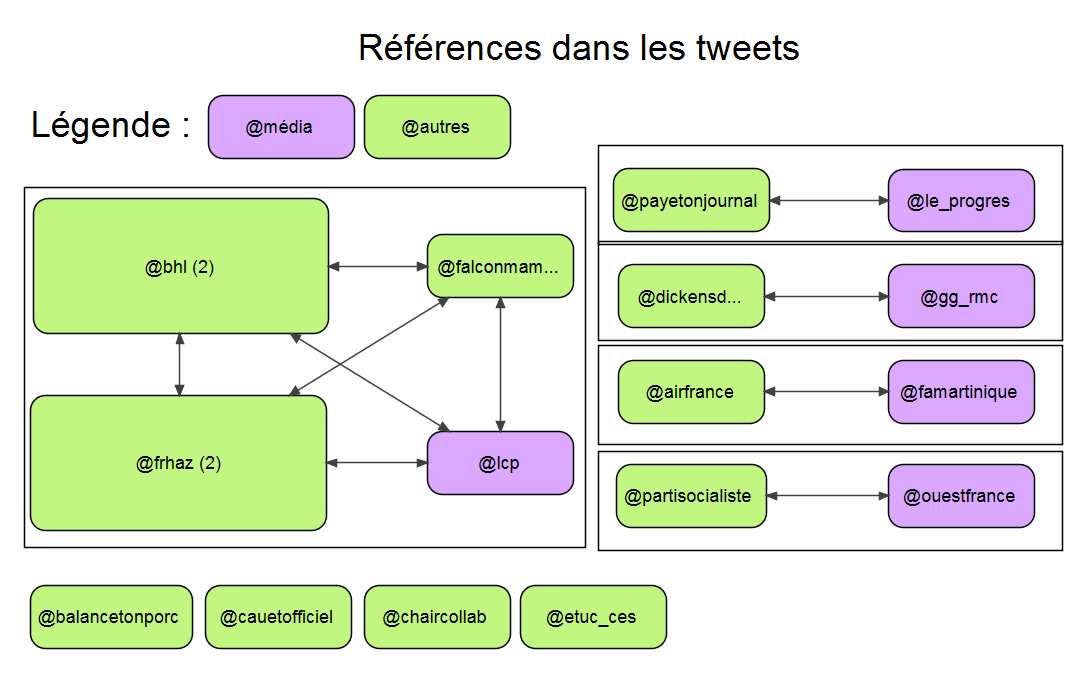
Puis enfin les politiques et partis politiques (15% à eux deux). Le corpus est prisonnier de l’actualité à laquelle il sert de caisse de résonnance : l’affaire Thierry Marchall Beck, ancien président du MJS, parti politique connu pour ses positions féministes, est suspecté d’avoir agressé sexuellement 8 femmes. Leurs témoignages sont apparus dans la presse durant le mois dont ce corpus a été extrait.

#### b) Références

Pour les références à des comptes, on retrouve ceux-ci :

* @frhaz 02
* @bhl 02
* @etuc\_ces 01
* @cauetofficiel 01
* @le\_progres, 01
* @payetonjournal 01
* @airfrance 01
* @famartinique 01
* @partisocialiste 01
* @ouestfrance 01
* @balancetonporc 01
* @lcp 01
* @falconmammouth 01
* @assasparis2 01
* @chaircollab 01
* @gg\_rmc 01
* @dickensdavid1 01

Ceux fait écho à une autre affaire touchant le journaliste Haziza Frédéric, accusé de gestes déplacés envers des collègues. Bernard-Henri Lévy l’a défendu publiquement. Cela permet de voir une notion importante : la cooccurrence des références, exprimée dans le schéma ci-dessous :



On remarque souvent un média, presse ou radio (@gg\_rmc, une émission sur RMC), en violet, est impliqué dans le tweet. Il s’agit d’une utilisation de tweeter comme relaie d’un article. Le tweet annonce son contenu et donne une URL pour aller le lire. Il s’agit de capter l’audience sur Tweeter, la même chose fonctionne sur Facebook.

# 3. Essai d’annotateur automatique

## 3.A Ensemble de mots

Parler de Mallet

Présence de mots :

- parties du corps : fesse, sein, main, cheveux

- locutions verbales ou verbes de contact : embrasse, caresse, mettre la main

- verbes de parole : dire, hurler

- mots indiquant un métier (journaliste)

- mots indiquant une position hiérarchique

- Présence de ponctuation :

- « et », :, !

- Présence de construction syntaxique :

- pronom objet à la 1ère personne + locutions verbales ou verbes de contact (m'embrasse)

- pronom possessif à la 1ère personne + mots indiquant une position hiérarchique (mon chef)

## 3.B Structures remarquables

# Annexes :

## A1) Chaîne de traitement

Pour obtenir le fichier corpus :

* Ouvrir dans **Google Spreadsheet**
  + Vérifier qu'il n'y a pas de \n et de ; dans les cellules
  + Exporter au format XLSX (Google Spreadsheet n'offre que l'export en CSV avec ',' comme séparateur)
* Ouvrir avec **Microsoft** **Excel**
  + Exporter au format CSV séparé par des ; (Excel offre soit UTF-8 avec ',' soit ASCII avec ';')
* Ouvrir avec **Notepad++**
  + Enlever la première ligne qui est le nom des colonnes
  + Supprimer la dernière ligne qui est vide
  + Convertir en UTF-8 (pour avoir enfin ';' comme séparateur en UTF-8
* Ouvrir avec mon **script Python** en mode "1"
  + Il envoit les tweets à Talismane
  + Talismane les analyse
  + On sauvegarde les résultats dans un modèle objet Python
  + Sauvegarder le modèle objet au format JSON
* Ouvrir avec mon **script Python** en mode "2"
  + Désérialisé le JSON de la sauvegarde pour retrouver le modèle objet
  + Plus besoin de Talismane, on peut travailler directement sur le modèle objet rechargé.

## A2) Phénomènes remarqués

* #balancetatruie
* Agisme
* Dénonciation / délation (le nom ou pas ?)
* Qui ?
  + Relai par les hommes
* Finkielkraut et l’islam (noyer le poisson)
* Spécisme
* Détournements (#balancetonpédo, #balancetapub, #balancetamafia, #balancetoncorps, #balancetaporcherie = le système)
* Émojis
* Ironie (s’il n’y a rien à détecter)
* Sexisme ou violence dans la famille

# Références bibliographiques

Pruvost, G. (2008). Le cas de la féminisation de la Police nationale. Idées économiques et sociales, 153,(3), 9-19. doi:10.3917/idee.153.0009. <https://www.cairn.info/revue-idees-economiques-et-sociales-2008-3-page-9.htm>