



**HAL**  
open science

## Mémoires des attentats terroristes du 13-Novembre 2015 : ce que peut nous apprendre l'analyse de discours

Denis Peschanski, Bénédicte Pincemin, Serge Heiden, Charlotte Lacoste

### ► To cite this version:

Denis Peschanski, Bénédicte Pincemin, Serge Heiden, Charlotte Lacoste. Mémoires des attentats terroristes du 13-Novembre 2015 : ce que peut nous apprendre l'analyse de discours. *Biologie Aujourd'hui*, 2023, 217 (1-2), pp.113-121. 10.1051/jbio/2023017 . hal-04154303

**HAL Id: hal-04154303**

**<https://hal.science/hal-04154303v1>**

Submitted on 9 Sep 2024

**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.

# Mémoires des attentats terroristes du 13-Novembre 2015 : ce que peut nous apprendre l'analyse de discours

## Memories of the 13-November 2015 terrorist attacks: what discourse analysis can teach us

Denis Peschanski<sup>15</sup>, Bénédicte Pincemin<sup>25</sup>, Serge Heiden<sup>35</sup>, et Charlotte Lacoste<sup>45</sup>

1. Université Paris 1 Panthéon Sorbonne, EHESS, CNRS, Centre Européen de Sociologie et de Science Politique (CESSP) UMR8209

2. Univ. Lyon, CNRS, IHRIM UMR 5317, ENS de Lyon

3. Univ. Lyon, ENS de Lyon, IHRIM UMR 5317

4. Centre de Recherches sur les Médiations (CREM) UR3476, Université de Lorraine

5. Programme 13-Novembre, équipex MATRICE

Auteur correspondant : denis.peschanski@univ-paris1.fr

### *Résumé*

Le Programme 13-Novembre vise à analyser la mémoire individuelle et collective des attentats terroristes du 13 novembre 2015. Au cœur se trouve l'Etude 1000 qui doit recueillir, à 4 reprises en 10 ans, les témoignages des mêmes 1000 personnes dans des entretiens audiovisuels. Disposant des transcriptions, nous choisissons ici de montrer l'importance de l'analyse de discours en rappelant ses fondements théoriques, de présenter l'un des outils permettant cette analyse statistique, l'Analyse Factorielle des Correspondances, et de l'utiliser pour analyser le sous-corpus des entretiens réalisés à distance des événements parisiens, auprès de 76 habitants de la région de Metz. Croisant ces volontaires avec les mots qu'ils utilisent, on constate que deux variables ressortent clairement qui opposent les vocabulaires, la variable du genre et la variable de l'âge.

**Mots-clés** : sciences de la mémoire, mémoire individuelle, mémoire collective, textométrie, Analyse Factorielle des Correspondances

### *Abstract*

The *Programme 13-Novembre* aims to analyze the individual and collective memory of the terrorist attacks of November 13, 2015. At its heart is the *Etude 1000*, which is to gather the same 1000 people in audiovisual interviews four times in 10 years. Having the transcripts at our disposal, we choose here to show the importance of discourse analysis by recalling its theoretical foundations, to present one of the tools allowing this statistical analysis, the Correspondence Factor Analysis, and to use it to analyze the sub-corpus of interviews conducted at a distance from the Paris events, with 76 inhabitants of the Metz region. Crossing these volunteers with the words they use, we see that two variables clearly stand out that oppose the vocabularies, the gender variable and the age variable.

**Keywords** : memory sciences, individual memory, collective memory, textometry, Correspondence Analysis

## Introduction

Depuis 2016, le Programme 13-Novembre vise à recueillir quelque 1000 entretiens de volontaires, du plus proche au plus lointain des lieux des attentats terroristes islamistes perpétrés en France le 13 novembre 2015 (Stade de France et Paris). Le protocole prévoit de retrouver le maximum de ces 1000 personnes à quatre reprises en dix ans. L'objectif scientifique est clair : il s'agit d'étudier non l'événement traumatique, mais la mémoire qu'en ont conservée des témoins directs ou indirects. Il n'en est pas plus simple pour autant dans la mesure où, au-delà du double défi que constitue le recueil même de ces témoignages, qui plus est sous forme audiovisuelle, et la capacité à retrouver un maximum d'entre eux à quatre reprises en dix ans (2016, 2018, 2021 et 2026), la démarche se veut résolument transdisciplinaire (Eustache & Peschanski, 2022, 2023). Partant du postulat qu'il est impossible d'appréhender pleinement à la fois (i) la mémoire collective sans prendre en compte les dynamiques cérébrales de la mémoire et (ii) ces dernières sans prendre en compte l'impact du social, le questionnaire lui-même, qui structure les entretiens de l'Étude 1000, a été élaboré en commun par des sociologues et historiens mais aussi des psychologues, psychopathologues et neuroscientifiques. On soulignera par ailleurs la dimension patrimoniale de ces fonds audiovisuels : il s'agit bien de constituer un patrimoine audiovisuel sur les attentats islamistes de novembre 2015, qui ont tant marqué la société française.

Mais c'est de la dimension strictement scientifique dont nous parlerons ici. Arrêtons-nous à la première phase de captation, celle de 2016, réalisée avec les équipes techniques de l'Institut National de l'Audiovisuel et de l'Établissement de Communication et de Production Audiovisuelle de la Défense. Pour cette seule première série de captations, on ne compte pas moins de 1431 heures d'entretiens audiovisuels pour 934 volontaires. Ces entretiens audiovisuels comprennent deux parties principales. La première consiste en un entretien semi-directif, trois grandes questions étant posées avec pour consigne stricte donnée aux intervieweurs de ne pas interrompre le volontaire et donc de n'intervenir, par une relance ou la question suivante, qu'à partir du moment où le volontaire a terminé son développement. La deuxième partie tient dans un questionnaire de mémoire émotionnelle qui appelle des réponses par « oui » ou par « non » et des précisions complémentaires de valence. Cette partie est également filmée dans la mesure où les questions peuvent susciter des développements de l'initiative du volontaire et une réaction qui sera analysée par ceux qui travailleront sur l'image et le son.

Nous avons regroupé les volontaires en 4 cercles, du plus proche au plus lointain de l'événement. Le cercle 1 réunit les personnes exposées, donc présentes sur les lieux ce soir du 13 novembre 2015 (rescapés ; intervenants : policiers, personnels de santé, magistrats, politiques ; témoins ; endeuillés) ; dans le cercle 2 se trouvent ceux qui habitent les quartiers concernés : Xe et XIe arrondissements et Saint-Denis, mais aussi ceux qui y travaillent mais qui n'ont ni vu ni entendu ; le cercle 3 concerne les habitants d'autres arrondissements de Paris ou de villes de la métropole parisienne ; pour le cercle 4 nous avons choisi trois villes en région, Caen, Montpellier et Metz.

Dans un premier temps, nous proposons une forme de mise au point sur les enjeux théoriques de l'analyse de discours. Nous appuyant sur un logiciel intégré TXM, nous privilégierons l'Analyse Factorielle des Correspondances (AFC) avant de proposer, dans un troisième temps, une forme d'étude de cas, celle de 76 personnes interviewées en septembre 2016 à Metz, à savoir l'une des trois villes dites de cercle 4.

## Une affaire de théorie et de méthodes

Mais avant de présenter ces résultats, il faut parler choix méthodologiques et théoriques. Il faut remonter aux années 1970 pour trouver les premiers travaux d'envergure de l'école française d'analyse des données appliqués au vocabulaire. Avec une méthodologie différente, Maurice Tournier (1975) d'un côté, Etienne Brunet (1981) de l'autre en étaient des acteurs majeurs. On parlait alors de lexicométrie. On parle aujourd'hui de textométrie ou de logométrie. Pour ceux qui se confrontent à ces approches, se pose d'emblée le défi de la transdisciplinarité. Inutile de se lancer si l'on ne peut s'appuyer sur des bases en linguistique et en statistiques, sans oublier bien entendu... sa discipline d'origine. Pour autant, à quoi sert de s'intéresser à l'analyse de discours si l'on n'est pas convaincu de l'autonomie relative du discours. Tel est le cas pour l'Histoire et ce n'est pas un hasard si les premiers travaux d'historien en lexicométrie se sont développés dans une période où l'école des Annales prônait l'autonomie relative de chacune des composantes de l'Histoire (l'économique, le social, les mentalités, le discours etc.), récusant l'idée selon laquelle il y aurait des lois en Histoire et l'autre idée qu'en dernière instance, l'une de ces composantes primerait, telle l'économie (Peschanski, 1988). A quoi servent ordinateurs et linguistique si le texte n'est qu'un vecteur transparent de la pensée et qu'il ne mérite donc pas d'être étudié comme tel ? Si le postulat est inverse, il faut donc constituer le vocabulaire comme objet d'histoire et, au même titre que l'économique, l'art, le social ou les mentalités, le discours a son autonomie relative avec son rythme propre d'évolution. Mais, qui dit relatif, dit bien sûr qu'il ne peut, dans le même temps, être pensé isolément. Ainsi, au cœur de ces démarches, se trouve l'analyse des rapports entre le discours, le monde dont il parle et le sujet énonciateur.

S'il s'agit de penser les rapports entre mot et sens, deux approches sont possibles et doivent être menées de conserve. On doit considérer le mot dans son unité, en dehors de son contexte syntagmatique. Nous posons ainsi que la répétition du mot « terroriste » dans la bouche de tel ou tel volontaire lors de son entretien est porteuse de sens. Comme celle de « nous », de « je » ou même de « le » ou de « on ». On doit aussi considérer le mot dans son environnement syntagmatique, le mot tirant une partie de son sens de la position qu'il occupe dans le texte, là dans le témoignage. L'association de « terrorisme » et « islamique » dans une forme de lexie figée est porteuse de sens, comme au demeurant toute association systématique de mots grammaticaux. Mais nous parlons de lexie figée ; or la proximité anormalement élevée à quelques mots de distance est également porteuse de sens.

Nous parlons ici de « mots » ou de « vocabulaire » au fil du discours. On imagine sans peine qu'on peut aussi parler de la syntaxe. On parle de « textométrie », mais on peut aussi privilégier des approches lexicologiques non quantitatives ou, à l'inverse, profiter des outils qui se sont développés depuis dans la révolution du traitement automatique du langage, de l'Intelligence Artificielle et ainsi du *deep learning* ou, à la rencontre d'autres disciplines, dans la psycholinguistique et, plus généralement, dans la psychologie, telle celle des émotions ou de la mémoire autobiographique. Et encore ne parlons-nous là que du texte. Rappelons, dans le cas qui nous occupe, que le matériau de base est formé d'entretiens audiovisuels et que le défi à moyen terme sera d'analyser, de conserve, texte, image et son avec les outils des Humanités numériques dont la textométrie elle-même en un sens élargi, Intelligence Artificielle ou autres.

## Des outils pour comprendre

Nous avons à peine esquissé l'ampleur des outils disponibles pour l'analyse de discours. Nous resserrons ici doublement la focale. Nous partons de la batterie d'outils regroupés dans le logiciel open source TXM développé sous la coordination de Serge Heiden (Heiden, Magué, Pincemin, 2010). On connaît d'autant plus ses principaux outils qu'ils avaient déjà été

développés dès les années 1970, mais à moindre puissance, par le laboratoire de lexicologie politique de l'ENS Saint-Cloud dirigé par Maurice Tournier, et singulièrement grâce à Pierre Lafon et à André Salem (Lafon, 1984 ; Lebart & Salem, 1988 ; Salem, 1993) On parlait déjà des « concordances » alignant les mots triés comme on le souhaite avec environnements à droite et à gauche ; de « l'analyse des spécificités » permettant de calculer le sous-emploi ou le suremploi d'un mot dans un sous-corpus, compte tenu de son emploi dans l'ensemble d'un corpus donné ; des « cooccurrences » calculant, elles, la proximité ou l'éloignement anormalement important entre deux mots ; de « l'Analyse factorielle des correspondances » (AFC), croisant, dans notre cas, chacun des témoins avec les mots qu'ils utilisent, ou de la « Classification Ascendante Hiérarchique », regroupant les témoins ou les mots, en particulier dans les mêmes espaces que l'AFC.

L'AFC part, comme les autres calculs, d'un principe de base : il n'y pas de fréquence en langue. En conséquence, on ne peut se référer qu'à des variables internes au corpus. Elles sont en fait au nombre de 4 : un mot (occurrence du corpus repérée par sa forme graphique, dit item formel, ou par son lemme, à savoir l'entrée du dictionnaire correspondant au mot observé) a une fréquence  $F$  dans l'ensemble du corpus formé de  $T$  mots, et une sous-fréquence  $f$  dans le sous-corpus de taille  $t$ . Le reste est affaire de calculs de probabilités.

Nous mettons ici la focale sur cette méthode, l'AFC. Nous avons choisi les lemmes (qu'on appellera indifféremment « mots »), plutôt que les items formels. Autrement dit, le mot « entrer » regroupe toutes les flexions du verbe. Nous avons aussi choisi de sélectionner les 200 lemmes ou mots les plus employés dans ce sous-corpus de Metz. On aboutit à un tableau croisé : les volontaires sont en colonnes et les lemmes en lignes. Au croisement de chaque volontaire et de chaque lemme, on a la fréquence dudit lemme dans le témoignage de ce volontaire. Pour résumer, l'AFC permet de visualiser les correspondances que deux ensembles de données entretiennent l'un avec l'autre. Mais le nuage ainsi formé par ce croisement de données est impossible à appréhender directement. Il faut donc se doter d'une méthode qui permette non seulement de faire apparaître les écarts, mais aussi de hiérarchiser les informations du nuage en question.

On doit la solution au mathématicien français Benzecri (1973) : il détermine d'abord ce qu'on appelle l'axe principal d'inertie sur lequel se projettent les points. C'est cet axe linéaire qui contient le plus d'informations sur le nuage. Le reste de l'information se répartit sur les autres axes d'inertie qui, on l'imagine bien, sont classés hiérarchiquement. Pour compléter à destination des férus de statistiques et de géométrie, tous les axes sont orthogonaux et passent par le centre de gravité du nuage. L'importance de chaque axe est donnée très simplement par le pourcentage d'information qu'il contient du nuage, un pourcentage qu'on compare au pourcentage théorique d'une équi-répartition.

Cela conduit à nous poser deux questions générales : comment les mots se situent-ils les uns par rapport aux autres dans l'espace des volontaires ? Comment les volontaires se situent-ils les uns par rapport aux autres dans l'espace des mots qu'ils emploient ?

### **Une étude de cas : les volontaires de Metz**

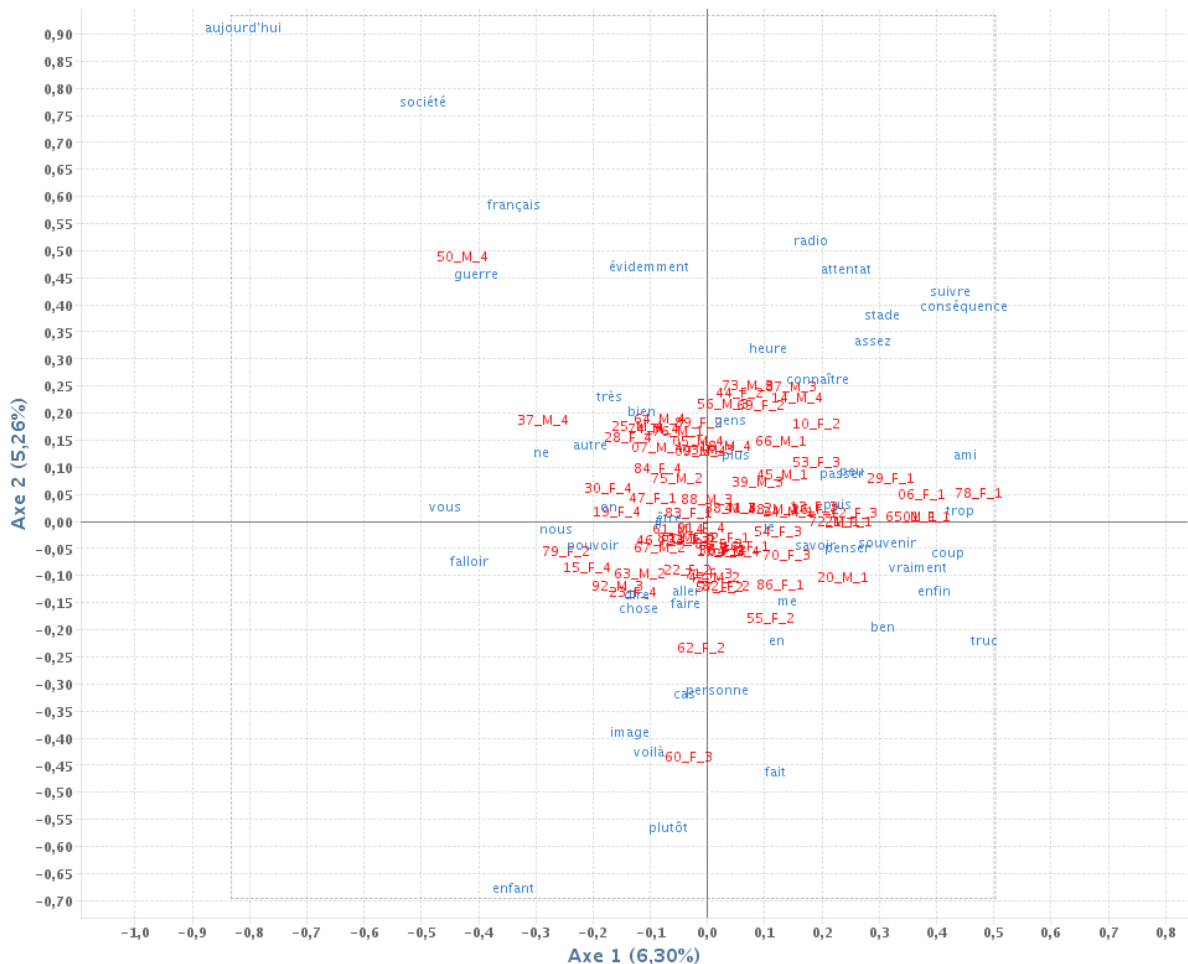
Telles sont les deux questions que nous avons posées, comme étude de cas, à un sous-corpus formé par les 76 témoins indirects et lointains des attentats qui ont accepté d'être interviewés dans le cadre du Programme 13-Novembre en juin et juillet 2016 à Metz. Ces 76 volontaires que nous disons du cercle 4 n'étaient donc pas présents à Paris ou Saint-Denis au moment des attentats et n'ont pas de liens amicaux ou familiaux avec les victimes.

Cela donne un tableau de 76x200 ; ce qui s'illustre par un nuage à 75 dimensions (on prend le plus petit côté et l'on enlève 1).

Dans le cas théorique d'indépendance et de poids identique des différents textes, les 75 axes seraient équivalents et contiendraient donc 100%/75 de l'information, soit 1,33%.

Dans les faits, l'axe 1 concentre 6,30% de l'information et le 2<sup>e</sup> axe 5,26%. Ils sont donc du même ordre de grandeur et à eux deux ils pèsent 11,56% du nuage, soit 4,4 fois plus que les 2,66% qu'ils devraient peser si la distribution était totalement équilibrée. C'est dire que ces deux axes sont significatifs aux seuils choisis. C'est maintenant cette information qu'il faut aller chercher. La représentation des deux axes est simple dans un plan, et par convention on place l'axe 1 horizontal et l'axe 2 vertical, comme le montre la figure 1.

Plan factoriel de l'analyse des correspondances de P1-METZ-P1-ET-P2-2021-07-02/cercle4/id\_genre\_agequartile/ < [frpos="NOM|VER.\*|ADV|ADJ|PRO:PER" & \_div\_n="1"]>@frlemma ≤637 792 /200/@frlemma /200



**Fig. 1** : AFC 200 mots x 76 témoins. Pour la qualité de lecture, ne sont pas représentés les mots dont la position est très déformée par la projection factorielle ( $\cos^2$  au plan  $< 0,20$ ) et dont la contribution à l'un comme à l'autre des deux axes est faible (moins de 1 %).

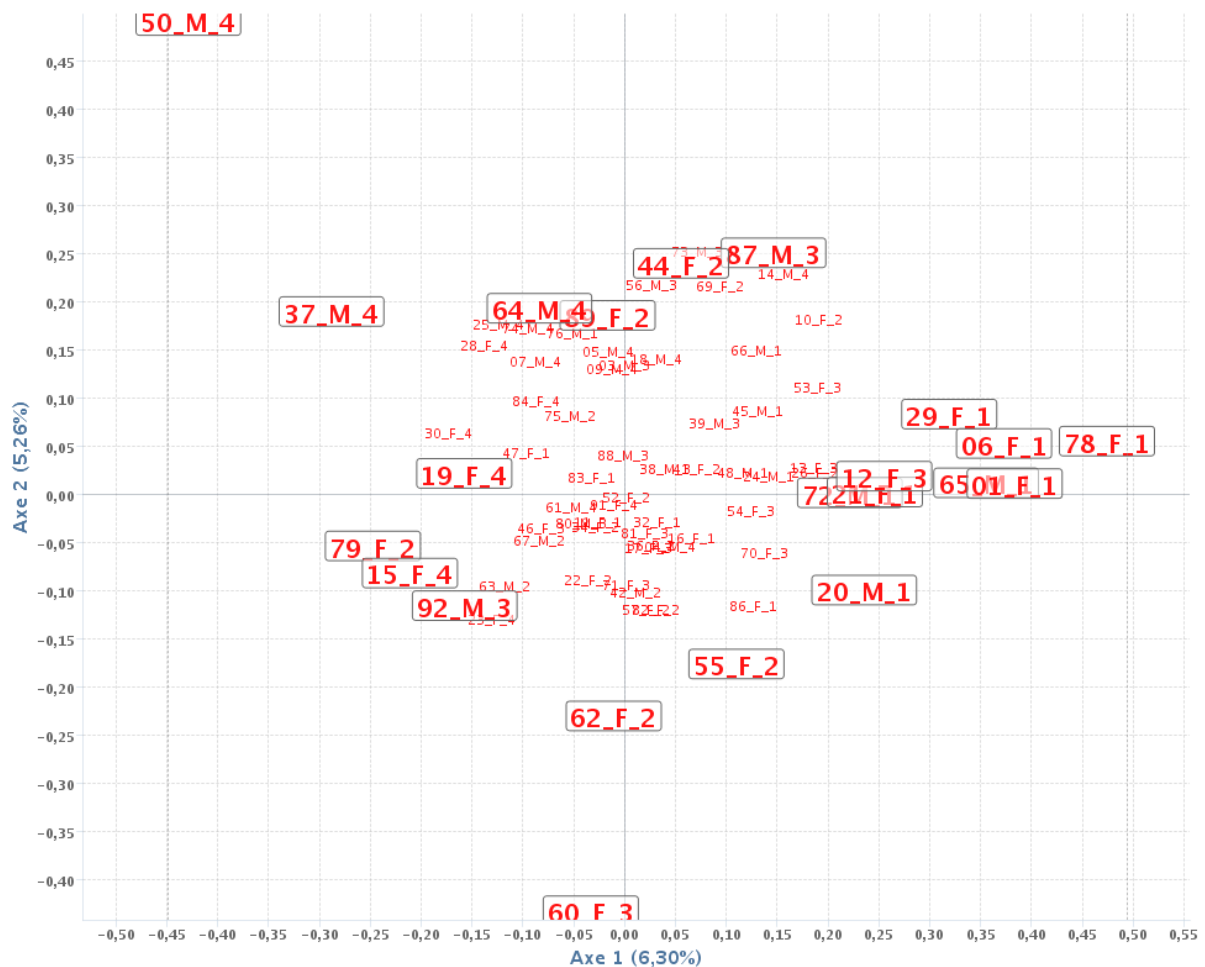
Le texte considéré ici pour cette analyse correspond à la seule première partie de l'entretien, à savoir l'entretien semi-directif au cours duquel le volontaire doit répondre à des questions très générales ; suit un questionnaire de mémoire émotionnelle relevant de l'entretien directif qui n'est pas retenu pour l'analyse de vocabulaire compte tenu du système de questions-réponses. Nous avons choisi par ailleurs de ne pas prendre en compte les mots outils car ils ne

correspondent pas à nos objectifs d'analyse. Nous conservons donc les NOMs communs, les VERbes ; les ADVerbes, les ADJectifs, les PRONoms PERsonnels et, comme indiqué plus haut, nous ne considérons que les 200 mots les plus utilisés.

Compte tenu du nombre de valeurs, la lecture est compliquée mais le retour sur les documents du Programme, en particulier la fiche sociodémographique et les résumés rédigés par les intervieweurs, mais aussi une base de données qui réunit toutes les variables, permet d'esquisser une première conclusion d'ensemble : les deux variables qui ressortent, c'est-à-dire qui apparaissent les plus associées aux différences dans le choix des mots, la façon de s'exprimer et le contenu du propos, sont l'âge et le genre.

La figure suivante (Fig. 2) efface les mots du graphique pour mieux voir la position de ces 76 volontaires sur les axes et confirme cette première impression.

Plan factoriel de l'analyse des correspondances de P1-METZ-P1-ET-P2-2021-07-02/cercle4/id\_genre\_agequartile / < [frpos="NOM|VER.\*|ADV|ADJ|PRO:PER" & "\_div\_n="1"]>@frlemma ≤637 792 /200/@frlemma /200



**Fig. 2 :** AFC 200 mots x 76 témoins. Les mots ont bien participé au calcul mais ne sont pas représentés, pour mieux examiner les contrastes entre volontaires. Les points mis en évidence correspondent aux individus pour lesquels la contribution à l'un ou l'autre axe est d'au moins 2 %.

Pour en avoir une représentation plus claire et une analyse plus fine, nous avons réparti les 76 volontaires dans 8 groupes, en distinguant femmes et hommes et en regroupant en 4 quartiles les âges (par tranche de 25%) (Tableau 1). Dans la désignation de chaque groupe, le premier caractère donne le quartile et le dernier le genre. Exemple : 1∩F regroupe les femmes (F) du quartile (25 %) le plus jeune (1). Ce ne sont donc plus 76 volontaires mais 8 groupes que nous

croisons avec les 200 mots les plus employés, soit un espace mathématique à 7 dimensions (on enlève 1 au côté le plus petit du tableau de données).

**Tableau 1.** Tableau des aides à l'interprétation de l'AFC sur les 8 groupes Âge (quartiles) × Sexe sur les deux premiers axes.

Colonne	Masse	Dist.	Cont1	Cos <sup>2</sup> 1	Cont2	Cos <sup>2</sup> 2	c1	c2
1∩F	14,54	0,03	13,16	0,35	0,25	0,00	-0,11	-0,01
1∩M	9,75	0,08	23,42	0,42	9,58	0,12	-0,18	0,10
2∩F	20,34	0,01	0,87	0,03	1,55	0,04	-0,02	-0,03
2∩M	6,35	0,06	0,05	0,00	2,07	0,04	-0,01	0,06
3∩F	15,56	0,04	2,00	0,04	32,53	0,51	-0,04	-0,14
3∩M	8,77	0,03	0,09	0,00	15,63	0,34	0,01	0,13
4∩F	13,41	0,07	38,75	0,62	11,46	0,13	0,20	-0,09
4∩M	11,28	0,06	21,65	0,41	26,92	0,35	0,16	0,15

*Colonne* : les différents groupes. Ex. : 1 ∩ F regroupe les femmes du premier quartile (le plus jeune)

*Masse* : le pourcentage du nombre total de mots (limités aux occurrences des 200 lemmes sélectionnés) pour la catégorie considérée. (La somme de la colonne vaut 100.)

*Dist.* : la distance à l'origine (qui correspond au profil lexical moyen de toutes les catégories).

*Cont1* et *2* : contribution absolue à l'axe 1 et à l'axe 2, soit la part (en %) prise par le groupe dans la définition de l'axe factoriel. (La somme de la colonne vaut 100.)

*Cos<sup>2</sup>1* et *2* : contribution relative (mesurée par le cosinus au carré), ou proportion dans laquelle un groupe est expliquée par un axe. (Pour une catégorie, la somme des cosinus carrés sur tous les axes (7 au total) est égale à 1 – somme de toute la ligne lorsqu'on a les cosinus carrés pour toutes les dimensions.)

*c1* et *c2* : coordonnées sur l'axe 1 et sur l'axe 2, le signe indiquant si c'est à gauche ou à droite, ou en haut ou en bas.

La lecture du tableau 1 qui rassemble les données de l'AFC sur les 8 groupes est sans appel. Reportons-nous à la colonne « CONT 1 » qui indique les contributions à l'axe 1. Quatre groupes ressortent : les quartiles F et H les plus jeunes et les quartiles F et H les plus âgés. La colonne « c1 » indique de quel côté de l'axe on se trouve. On voit que les deux quartiles de jeunes sont d'un côté et les deux plus âgés de l'autre. A eux quatre, ils contribuent à 97% de l'information de l'axe !

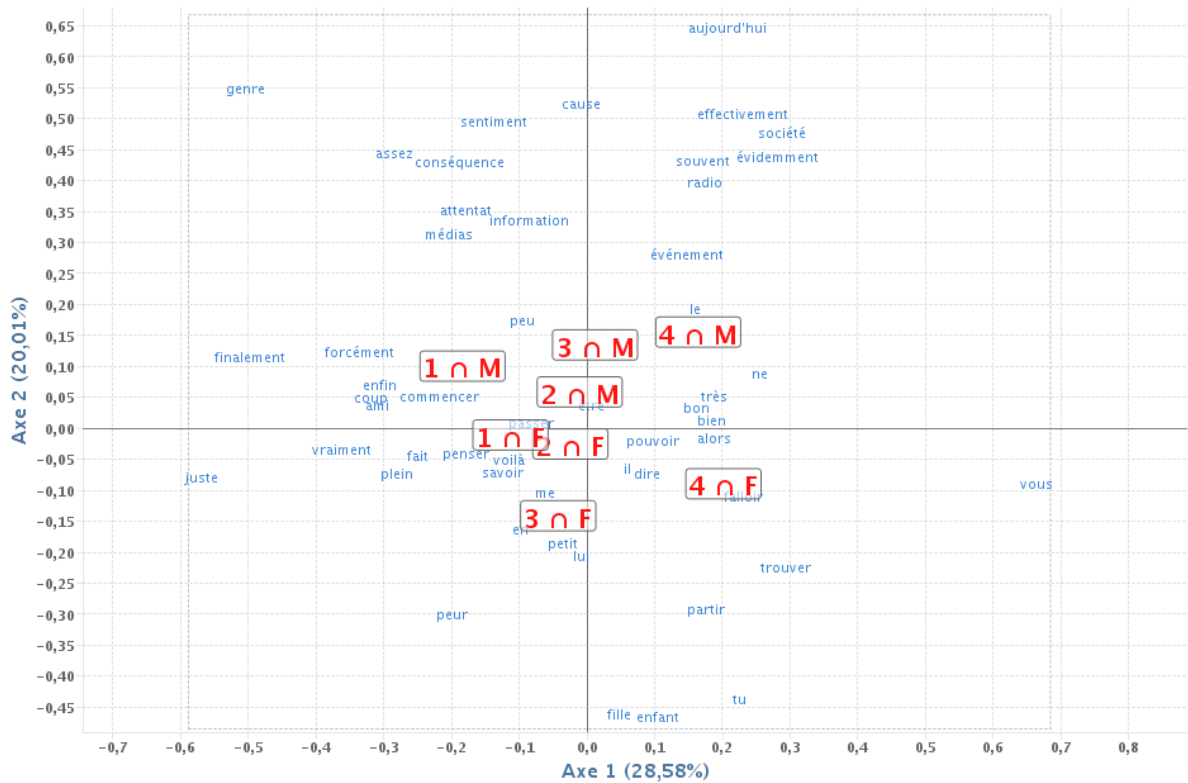
Reportons-nous à la colonne « CONT 2 » qui indique donc les contributions à l'axe 2. La lecture de la colonne « c2 » montre que le rangement est parfait : les 4 groupes d'âges de femmes se retrouvent d'un côté de l'axe, les 4 groupes d'âges d'hommes de l'autre côté.

C'est donc peu dire que le regroupement par quartiles a synthétisé la première exploration globale du corpus, comme le montrent les figures suivantes.

La figure 3 regroupe les deux axes en identifiant l'ensemble des 8 groupes définis et les mots qui contribuent le plus aux deux axes parmi les 200 les plus employés. Les figures 4 et 5 précisent encore les résultats, les mots les plus significatifs de l'axe 1 étant entourés sur la figure 4, ce qui est le cas pour les mots les plus significatifs de l'axe 2 sur la figure 5.



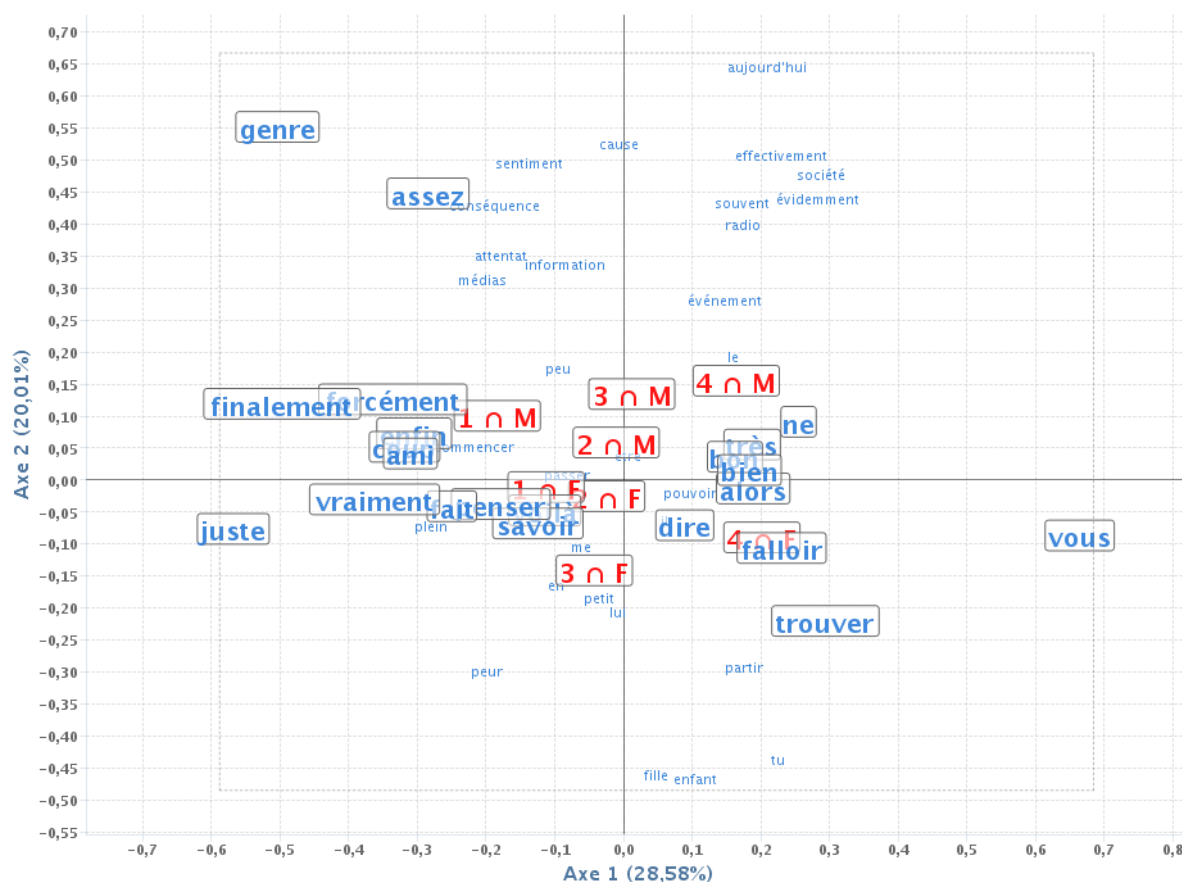
Plan factoriel de l'analyse des correspondances de P1-METZ-P1-ET-P2-2021-07-02/cercle4/text\_age-quartiles x text\_sexe / <[frpos="NOM|VER.\*|ADV|ADJ|PRO:PER" & \_div\_n="1"]>@frlemma ≤637 792 /200/@frlemma /200



**Fig. 3 :** AFC 200 mots x 8 regroupements Genre x Âge (quartiles). Pour la qualité de lecture, ne sont pas représentés les mots dont la position est très déformée par la projection factorielle ( $\cos^2$  au plan < 0,80) et dont la contribution à l'un comme à l'autre des deux axes est faible (moins de 1 %).

Car, une fois identifiées les variables principales (âge et genre), se pose alors la seconde question : quels sont les mots qui singularisent chacun de ces deux axes de contraste clairement identifiés ? On se limitera ici aux données de l'AFC. Pour information, grâce à TXM, les concordances permettent de connaître l'environnement de chaque mot, les mots qui le suivent à droite et le précèdent à gauche, et donc d'affiner l'analyse. De même, l'analyse des spécificités (Lafon, 1984) aidera à confirmer les suremplois lexicaux de chacun des groupes de témoins, comme nous allons le voir mais à titre de confirmation ou, on le verra, de précision. Restons-en donc à ces données déjà si parlantes fournies par l'AFC. Le linguiste se réjouira certainement de ce qui identifie les 25% les plus jeunes : les mots « genre », « juste », « forcément », du-« coup », « enfin », « vraiment », en-tout-« cas », « ami » sont des marqueurs dont tout un chacun peut témoigner. On repère cependant la référence à « attentat » (utilisé le plus souvent au pluriel ; rappelons qu'il s'agit du lemme). Le quartile des plus âgés s'oppose d'abord par l'emploi significativement car statistiquement plus limité de ces formes, ensuite par un vocabulaire plus diversifié qui sert, le plus souvent, à décrire et analyser ce qui s'est passé. Les quelques mots qui dépassent le seuil choisi sont, comme pour les plus jeunes, des mots qui organisent et spécifient l'oralité comme « bien », « alors », la négation « ne » avant le verbe, oubliée bien plus souvent dans les quartiles plus jeunes, mais – répétons-le – c'est la diversification du vocabulaire qui prime avec référence à l'événement lui-même et à sa qualification, à ses causes et à ses conséquences. Seul le pronom « vous », déictique par excellence, marque à la fois la condition d'énonciation et le rapport voulu avec l'intervieweur qui se trouve en quelque sorte intégré dans le récit.

Plan factoriel de l'analyse des correspondances de P1-METZ-P1-ET-P2-2021-07-02/cercle4/text\_age-quartiles x text\_sexe/ <[frpos="NOM|VER.\*|ADV|ADJ|PRO:PER" & \_div\_n="1"]>@frlemma ≤637 792 /200/@frlemma /200

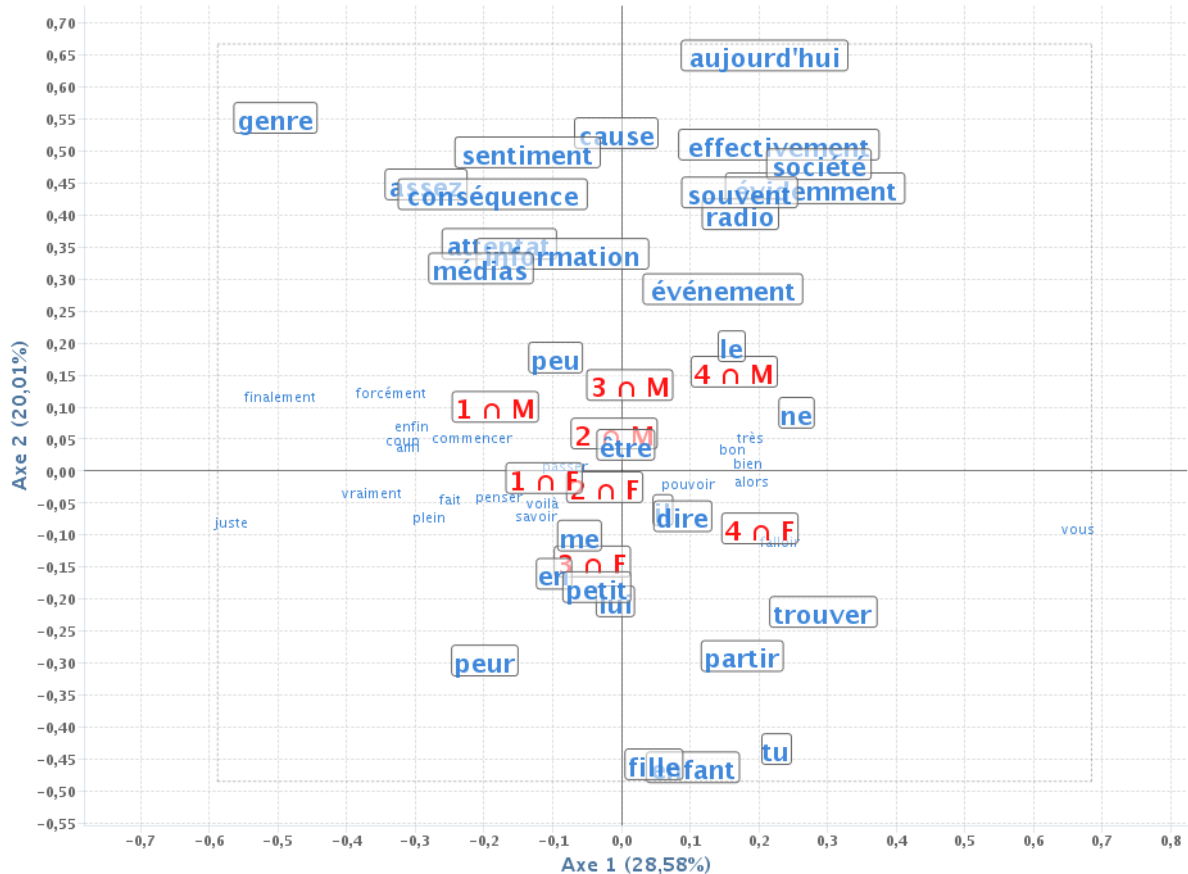


**Fig. 4 :** AFC 200 mots x 8 regroupements Genre x Âge. Les mots mis en évidence sont ceux dont la contribution à l'axe 1 est supérieure à 1%.

Mais un autre outil développé dans TXM apporte des précisions supplémentaires. Le principe en est simple et se résume à une question dédoublée : quelle était la probabilité pour qu'un mot soit utilisé autant ou plus de fois par tel volontaire, sachant qu'il a une fréquence  $F$  dans l'ensemble du corpus ? On parle alors de « spécificité positive ». Quelle était la probabilité pour qu'un mot soit utilisé si peu ou moins de fois par tel volontaire, sachant qu'il a une fréquence  $F$  dans l'ensemble du corpus ? On parle alors de « spécificité négative ». Nous n'utilisons ici cet outil que pour aider à confirmer et préciser la lecture de l'AFC. On notera par exemple que les mots « genre », du-« coup », « enfin », « assez », « finalement » mais aussi « attentat » sont tout spécialement spécifiques chez les jeunes hommes (1<sup>er</sup> quartile), avec des seuils de probabilité inférieurs à  $10^{-7}$ , et même  $10^{-9}$  pour « finalement » et  $10^{-13}$  pour « assez »,  $10^{-17}$  pour « genre » et...  $10^{-31}$  pour « enfin ». Si l'on se réfère aux deux groupes les plus âgés, soit le quartile supérieur chez les hommes et les femmes, ils se trouvent dans les spécificités négatives des suremplois du premier quartile, comme un inverse en quelque sorte. L'AFC nous disait que ce côté de l'axe comptait peu de mots. Il y avait le mot « vous » très identifié. Or, les spécificités nous le confirment avec une nuance d'importance : si les hommes le sur-emploient à  $10^{-4}$ , pour les femmes on atteint le chiffre record de  $10^{-85}$  ! Dit autrement, compte tenu de l'emploi du mot « vous » dans l'ensemble du corpus, il avait  $10^{-85}$  chances d'apparaître autant ou plus dans le sous-corpus 4nF. Revenons aux chiffres bruts tant l'écart est spectaculaire et sans tenir compte de la pondération des masses : on note 398 occurrences du mot « vous » dans ce seul groupe 4nF, et encore 161 dans le groupe 4nM, qui est le second à le sur-employer. On se retrouve

entre 50 et 100 occurrences pour chacun des 6 autres groupes, sauf 157 occurrences en 2∩F mais à rapporter à un nombre de mots plus important (de l'ordre du double des autres groupes). Si l'on examine le détail des individus, on peut préciser que cet emploi du « vous » n'est pas caractéristique de toutes les femmes plus âgées (4∩F) mais ressort du fait de certaines d'entre elle : 15\_F\_4 concentre à elle seule 183 des 1065 occurrences du corpus total, 19\_F\_4 (57 occ.), 30\_F\_4 (48 occ.), 84\_F\_4 (33 occ.), 91\_F\_4 (51 occ.), mais aussi du côté des hommes plus âgés 25\_M\_4 (42 occ.) et 61\_M\_4 (31 occ.).

Plan factoriel de l'analyse des correspondances de P1-METZ-P1-ET-P2-2021-07-02/cercle4/text\_age-quartiles x text\_sexe/ <[frpos="NOM|VER.\*|ADV|ADJ|PRO:PER" & \_div\_n="1"]>@frlemma ≤637 792 /200/@frlemma /200



**Fig. 5 :** AFC 200 mots x 8 regroupements Genre x Âge. Les mots mis en évidence sont ceux dont la contribution à l'axe 2 est supérieure à 1%.

Nous n'allons pas plus loin puisque, répétons-le, nous nous centrons sur l'AFC, mais ce détour par les spécificités conforte nos conclusions sur le premier axe. Ils le précisent dans le même temps : que ce soit pour les plus jeunes ou les plus âgés, on voit qu'au critère de l'âge s'insère déjà l'autre variable que nous révèle si clairement l'axe 2.

Que nous disent en effet les mots de l'axe 2, de poids presque équivalents à l'axe 1 ? L'opposition, de part et d'autre de cet axe 2, entre hommes et femmes était moins attendue. Pour résumer, hommes et femmes n'utilisent pas les mêmes mots pour répondre, pourtant, aux mêmes questions très générales. On remarquera déjà que les mots sont nombreux qui dépassent le seuil choisi ( $CONTR \geq 1$ ), aussi bien en haut qu'en bas de l'axe, donc pour les hommes ou pour les femmes. Les hommes surutilisent les mots qui réfèrent au contexte événementiel, qu'ils concernent la soirée du 13 novembre 2015 (« radio », « information », « attentat », « événement »), ou les causes et conséquences (« aujourd'hui », « cause », « conséquence »,

« société »). Du côté des femmes, ce n'est pas un regard de commentateur qui prédomine, mais le vécu de l'intime. Cela se retrouve dans les pronoms personnels « tu », « me », marqueurs également du rôle tenu dans cette soirée, dans l'action et l'organisation. S'ajoutent les mots qui réfèrent à la jeunesse, qu'elle soit celle de leurs propres enfants – le plus souvent – ou celle des victimes : « fille », « enfant ». Le mot « peur » renvoie notamment aux émotions ressenties au moment des attentats, même s'ils se déroulent à quelques centaines de kilomètres de là. A proximité du mot « peur », en cooccurrence, on trouve les mots « tristesse », « céder », « terroriser », « paralyser » qui éclairent la signification du mot.

## Conclusion

Telles sont quelques-unes des conclusions que nous pouvons tirer de l'AFC, l'un des outils permettant d'analyser la distribution des mots dans le discours. Le premier objectif de cette contribution était de montrer la nécessité d'étudier le discours comme tel, de préciser les enjeux théoriques et la richesse des outils statistiques mis à notre disposition. L'AFC est l'un d'entre eux, comme nous l'avons signalé, et les autres enrichiront bientôt cette analyse forte de cette boîte à outils qu'est TXM mais aussi, bien entendu, hors de cette boîte, que la démarche privilégie le quantitatif ou le qualitatif.

Pour tout dire, en privilégiant dans un premier temps le sous-corpus des 76 entretiens de Metz, nous ne pensions pas trouver des résultats aussi riches. C'était en quelque sorte ce que les biologistes appellent un groupe contrôle, sachant qu'en cette première phase de captations audiovisuelles, quelque 934 personnes ont été interviewées dont 360 avaient été directement exposées. D'autres oppositions se feront jour, mais ici les deux premiers axes structurants sont riches de conclusions, au-delà du cas du 13-Novembre. Les linguistes et sociolinguistes s'attacheront sans aucun doute à l'opposition si nette en fonction de l'âge, et spécialement entre les plus jeunes et les plus âgés du corpus. Sociologues et historiens s'intéresseront aussi à ce décalage si flagrant sept mois après les attentats. On prendra garde de ne pas oublier l'évolution longue. Nous verrons ainsi dans quelle mesure cette opposition reste de mise en phase 2 (2018), en phase 3 (2021-22) et en phase 4 (2026). Il y a 30 ou 40 ans, par exemple, constatait-on une opposition aussi structurante (au-delà des mots qui changent avec le temps et la mode) ou cette opposition n'était-elle pas si flagrante ? Nous ne disposons pas de données permettant de répondre à cette question.

Pour ce qui est de l'opposition hommes/femmes, on la repère déjà dans d'autres corpus. Dans une étude non encore publiée sur les déportés de répression pendant la Seconde Guerre mondiale, ceux et celles qui furent donc arrêtés et déportés non pour ce qu'ils étaient mais pour ce qu'ils avaient fait, une factorielle du même type montrait que le premier facteur était genre (équipex MATRICE). Autrement dit, résistantes et résistants n'utilisaient pas les mêmes mots pour raconter leur résistance et leur déportation. La conclusion n'était pas mince quand on sait qu'il y avait des différences d'âges non négligeables (une quinzaine d'années), des origines sociales très diverses comme l'étaient les appartenances aux mouvements et réseaux de résistance (communistes, gaullistes, socialistes etc.). Cette importance du genre dans le choix des mots utilisés n'est donc pas un cas spécifique à notre corpus même si – d'autres études le montreront – elle est moins évidente à observer quand on analyse le corpus des 934 entretiens de l'Etude 1000 du Programme 13-Novembre dans son ensemble, car elle interfère avec d'autres facteurs encore, comme la proximité à l'événement et le rôle dans lequel on se trouvait (survivant, témoin, intervenant professionnel, etc.). Pour autant, constatant cette opposition pour les témoins de Metz, prenons garde à ne pas utiliser le qualificatif si facile de stéréotype qui implique un jugement de valeurs et une forme de classification. Il ne s'agit pas de stéréotypes mais de représentations sociales multifactorielles. Cela vaut pour le récit qui est fait

de cette nuit du 13 novembre 2015, comme de la posture différenciée vis-à-vis de l'intervieweur quand il s'agit de raconter et d'analyser à distance. La force des outils de la textométrie est d'apporter des éléments pour la compréhension de ces différences multifactorielles, en lien avec les mécanismes mémoriels.

### **Remerciements**

Le Programme transdisciplinaire 13-Novembre coordonné par D. Peschanski et F. Eustache (pilotage : Inserm, CNRS, HESAM Université) bénéficie d'une aide de l'Agence Nationale de la Recherche au titre du programme « Investissements d'avenir » (PIA, France 2030), référence ANR-10-EQPX-21-01 Matrice\_Programme 13-Novembre [Site : [memoire13novembre.fr](http://memoire13novembre.fr)]

### REFERENCES

- Benzecri J.-P., *L'Analyse des données*. Tome 2 : *L'Analyse des correspondances*, Dunod, 1973, 619 p (ISBN 2-04-007225-X).
- Brunet E., *Vocabulaire français de 1789 à nos jours*, 3 tomes, 1824 p., Genève-Paris, Slatkine-Champion, 1981
- Eustache F., Peschanski D., “Toward new memory sciences : The *Programme 13-Novembre*”, in *Progress in Brain Research (PBR)*, volume 274, “Collective Memory, edited by O’Mara S., 2022, p.177-202.
- Eustache F., Peschanski D., « Le Programme 13-Novembre entre mémoire individuelle et mémoire collective », *Biologie Aujourd’hui*, 216, 2023. <https://doi.org/10.1051/jbio/2023016>.
- Heiden S., Magué J.-P., Pincemin B., « TXM : Une plateforme logicielle open-source pour la textométrie – conception et développement », in Bolasco S., Chiari I., Giuliano L. (eds), *Proceedings of 10th International Conference on the Statistical Analysis of Textual Data - JADT 2010*, Rome, Edizioni Universitarie di Lettere Economia Diritto, 2010, vol. 2, p. 1021-1031.
- Lafon, Pierre, *Dépouillements statistiques en lexicométrie*, Slatkine/Champion, 1984.
- Lebart L., Salem A., *Analyse statistique des données textuelles. Questions ouvertes et lexicométrie*, Paris, Dunod, 1988, 210p.
- Peschanski, D., *Et pourtant ils tournent. Vocabulaire et stratégie du PCF (1934-1936)*, Paris, Klincksiek et INaLF, 1988, 252 p.
- Salem A., *Méthodes de la statistique textuelle*, thèse de Doctorat d’Etat, Université Paris 3, 1993, 819p. et 178p. d’annexes.
- Tournier M., *Un Vocabulaire ouvrier en 1848. Essai de lexicométrie*, Thèse de Doctorat d’Etat, ENS-St Cloud, 1975.