



HAL
open science

L'évolution de la mémoire de la Shoah au prisme de la statistique textuelle

Damon Mayaffre, Bénédicte Pincemin, Serge Heiden, Philippe Weyl

► **To cite this version:**

Damon Mayaffre, Bénédicte Pincemin, Serge Heiden, Philippe Weyl. L'évolution de la mémoire de la Shoah au prisme de la statistique textuelle. Denis Peschanski; Brigitte Sion. La vérité du témoin, 2, Hermann Éditeurs; Institut National de l'Audiovisuel, pp.93-124, 2018, Mémoire et mémorialisation, 9782705697365. hal-01890536

HAL Id: hal-01890536

<https://hal.science/hal-01890536>

Submitted on 8 Oct 2018

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.

L'évolution de la mémoire de la Shoah au prisme de la statistique textuelle

Damon MAYAFFRE¹, Bénédicte PINCEMIN², Serge HEIDEN³, Philippe WEYL⁴

¹ CNRS, BCL UMR7320, Université Nice-Sophia Antipolis (Nice – France)

² Univ. Lyon, CNRS, IHRIM UMR5317 (Lyon – France)

³ Univ. Lyon, ENS de Lyon, IHRIM UMR5317 (Lyon – France)

⁴ FMS – Fondation pour la Mémoire de la Shoah (Paris – France)

1. Introduction

Entre mémoire individuelle et mémoire collective, les récits de vie autour d'événements historiques majeurs comme la Grande Guerre ou la Shoah sont depuis plusieurs décennies des objets privilégiés d'étude pour les Sciences Humaines et Sociales. S'ils interrogent ainsi les chercheurs comme la société, c'est sans doute au premier chef parce qu'ils articulent de manière complexe, dans une narration qu'il reste à décrire, Mémoire et Histoire [Ricoeur 2000], témoignage intime et grand récit, souvenir individuel et mémoire sociale.

Dans le cadre de l'Équipex Matrice (<http://www.matricememory.fr>), nous avons montré que les témoignages de la Shoah, recueillis par la Fondation pour la Mémoire de la Shoah (FMS), recélaient dans leur matérialité textuelle deux types de lexique entremêlés : un lexique sans doute directement issu de l'événement traumatique – lexique primaire pourrait-on dire ou intime du témoin – et un lexique reconstruit ou réélaboré au fil des années – lexique secondaire pourrait-on dire, social ou collectif, partagé par le grand récit historique [Mayaffre et Ben Hamed 2014].

Cette contribution met à l'épreuve cette conclusion sur cette mémoire palimpseste, ce souvenir originel constamment réélaboré, ce double lexique entrelacé qui permet écriture et réécriture de l'événement c'est-à-dire aussi mémorisation, mémorialisation et commémoration [Peschanski 2013]. Nous étudions en effet ici un corpus de textes doté de sa dimension chronologique en recueillant 20 témoignages majeurs d'Auschwitz selon trois moments d'écriture des témoins : l'immédiat après-guerre, les années 1970-1990 et les années 2000-2010 (fig. 1). L'hypothèse de travail est que, sur l'évolution linguistique de fond, sous-jacente, du genre textuel du témoignage, se superpose et se co-construit l'évolution du vécu personnel, intime et social, affectif et cognitif, et qu'ainsi les moments d'écriture déterminent la forme et le fond des témoignages. Particulièrement, le matériel lexical (les mots utilisés pour témoigner) pourrait varier selon que l'on se souvienne 5 ans, 20 ans ou 50 ans après l'événement. Plus précisément encore, le lexique qualifié de primaire que nous avons identifié devrait se trouver surreprésenté dans les témoignages proches de l'événement, là où le lexique qualifié de secondaire devrait sur-apparaître, plus tard, au détour des années 2000 : la primauté ou la secondarité de ces lexiques s'en trouveraient ainsi confirmées.

Méthodologiquement enfin, cette introspection de la sédimentation mémorielle se fera grâce aux outils de l'analyse de données textuelles (ADT) [Lebart et Salem 1994]. Nous posons en effet que la mémoire aussi, et en tout cas la narration, est pour partie un phénomène statistique (*ie.* qui répond à des règles de récurrences non dues au hasard) ; et l'approche textométrique (ou logométrique) donne les moyens d'une analyse à la fois qualitative et quantitative au plus

près du texte, sur un corpus d'un million de mots. Nous avons eu recours à deux logiciels open-source : TXM (<http://textometrie.ens-lyon.fr>) [Heiden, Magué, Pincemin 2010] dont le développement a reçu le soutien de l'Équipex Matrice, ainsi qu'IRaMuTeQ [Ratinaud & Déjean 2009, Ratinaud & Marchand 2012], pour la classification lexicale.

<i>Id.¹</i>	<i>Auteur</i>	<i>Titre</i>	<i>Nb. mots</i>
1945a	Mayer, Alex	Auschwitz, le 16 mars 1945	40 477
1945b	Kohen, Guy	Retour d'Auschwitz	22 015
1945c	Oppenheimer, Jean	Journal de route, 14 mars – 9 mai 1945	58 516
1945d	Unger, Julien	Le Sang et l'Or	59 827
1945e	Holstein, Denise	Le Manuscrit de Cayeux-sur-Mer	12 724
1946a	Altmann, Erich	Face à la mort	39 597
<i>Total période 1945-1946 :</i>			<i>233 156</i>
1978a	Klein, Eugène	Les Loups. Témoignage d'un déporté, matricule 126026	32 252
1986a	Grinbaud, Simon	XIe commandement : « Tu n'oublieras point »	88 367
1990a	Roth, Nicolas	Avoir 16 ans à Auschwitz. Mémoire d'un juif hongrois	185 267
1995a	Goltman, Pierre	Six mois en enfer	11 753
1998a	Grossman, Adèle	La Mémoire dans la chair	119 675
<i>Total période 1978-1999 :</i>			<i>437 314</i>
2002a	Hirsch, Claude	Matricule A-16689	23 230
2008a	Toros-Marter, Denise	J'avais seize ans à Pitchipoï	25 753
2008b	Skorka-Jacubert, Régine	Fringale de vie contre usine à mort	30 964
2008c	Palant, Charles	Je crois au matin	42 281
2008d	Mitzner, Charles	Seuls au monde	15 454
2008e	Rosenbaum, Isidore	Je suis né le 8 mai 1945	23 428
2009a	Golgevit, Eva	Ne pleurez pas, les fils...	19 149
2009b	Lichtszejn-Montard, Sarah	Chassez les papillons noirs.	69 235
2010a	Spingarn, Odette	J'ai sauté du train. Fragments	24 925
<i>Total période 2002-2010 :</i>			<i>274 419</i>
<i>TOTAL (nb. de mots du corpus) :</i>			<i>944 889</i>

Figure 1. Corpus – 20 témoignages (issus de la Collection Témoignages de la Shoah – FMS)

2. Première approche globale du corpus : thèmes et temps de la mémoire

2.1. Thématiques des témoignages de la Shoah : stabilité des mondes lexicaux

Le corpus étudié dans ce travail, équivalent à un million de mots, issu de la collection *Témoignages de la Shoah* de la FMS et centré sur le camp d'extermination d'Auschwitz, augmente et améliore le corpus étudié par Mayaffre et Ben Hamed [2014] : 9 textes sont communs aux deux corpus, 7 textes ne figurent que dans le corpus de 2014 et 11 nouveaux textes ont été introduits en 2016 dans cette présente étude (fig. 1).

Est-ce que ce renouvellement pour moitié des témoignages modifie les principales thématiques qui en ressortent globalement ?

Dans un premier temps nous calculons sur le corpus 2016 une classification lexicale selon la méthode [Reinert 1993], avec un paramétrage permettant des résultats détaillés² (fig. 2).

¹ L'identifiant correspond à l'année de rédaction ; la lettre distingue les textes d'une même année.

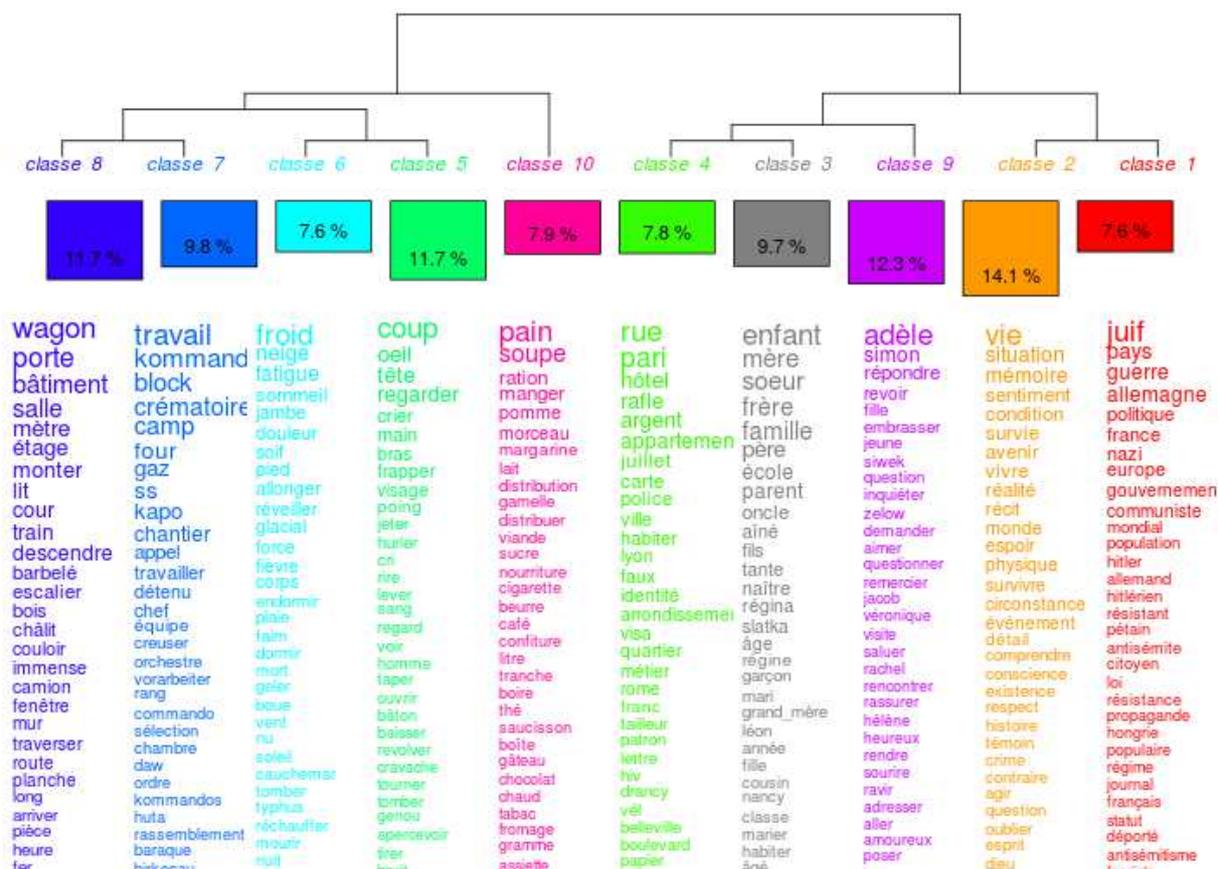


Figure 2. Classes lexicales du corpus (sortie Iramuteq, méthode Reinert)

Il ressort de l'analyse deux conclusions principales.

Premièrement, la comparaison des classes lexicales obtenues ici avec celles calculées en 2014 (fig. 3) nous permet de conclure à la stabilité des thématiques représentées dans ces deux ensembles de témoignages, ce qui conforte l'hypothèse que ces thématiques sont des thématiques de fond pour ce type de textes. Bien sûr, nous pourrions désormais envisager d'élargir encore le recueil de témoignages sur la Shoah mais il est fort à parier que le lexique essentiel, structuré thématiquement comme indiqué, se trouve identifié par nos études.

Deuxièmement, comme en 2014, la classification distingue bien, globalement, deux lexiques (classes 5, 6, 7, 8 et 10 d'un côté, *versus* classes 1, 2, 3, 4 et 9 de l'autre³). Le premier lexique pourrait être qualifié de sensitif ou encore de concret (« coup », « pain », « froid », « travail », « wagon », « porte », etc.) : il nous semble relever de la « trace » affective ou de « l'impression » – pour reprendre les termes-clefs de Ricoeur [2000] – proche du traumatisme vécu. Le deuxième lexique, lui, est plus intellectualisé, historicisé ou généalogique (« juif », « mémoire », « guerre », « Paris », « vie », « famille », « enfant », etc.) : il nous semble

² Nous soumettons à Iramuteq les textes intégraux du corpus sans passer par une étape de filtrage aux 300 substantifs les plus fréquents, et nous poussons un peu plus avant la classification descendante afin de détailler davantage de classes (16 classes demandées à l'issue de la phase 1, soit 10 classes finalement produites). Ce réglage s'avère très satisfaisant en termes de couverture du corpus, car chaque segment est décrit par en moyenne 35 mots, et 90 % des segments sont classés (89.07 % exactement).

³ Nous verrons que seule la classe 9 est un peu particulière car héritée d'un seul témoignage.

relever de la mise en récit ou, au sens strict, de la re-présentation sociale, familiale et historique de l'événement.

<i>Les 6 classes de 2014</i>	<i>Les 10 classes de 2016</i>	<i>Interprétation</i>	<i>Analyse</i>
4 : guerre, année, déportation, France, mémoire, Paris, quartier, rafle, juif, Allemagne...	1 : juif, pays, guerre, Allemagne, politique, France...	socio-politique et religieux	§3.6, §3.2
	2 : vie, situation, mémoire, survie, avenir...	recul, réflexion, mémoire	§3.1-2
	4 : rue, Paris, hôtel, rafle, appartement...	cadre élargi	§3.2
2 : mère, père, enfant, parent, frère...	3 : enfant, mère, sœur, frère, famille, père, école, parent...	famille	§3.7
3 : coup, œil, homme, neige, souffrance...	5 : coup, œil ⁴ , tête, regarder, crier, frapper...	violence et souffrance	§3.4
	6 : froid, neige, fatigue, sommeil, douleur...	quotidien des camps	§3.2-4
1 : camp, prisonnier, SS, travail, kommando, bâtiment, barbelé, mère...	8 : wagon, porte, bâtiment, salle, mètre, lit, train, barbelé...	cadre des camps	§3.2-3
	7 : travail, kommando, block, crématoire, camp, SS, détenu...	cadre et quotidien des camps (dont personnes)	§3.3
5 : pain, soupe, ration, pomme, soir, block, salle, détenu...	10 : pain, soupe, ration, manger, pomme, morceau...	quotidien des camps : alimentation	§3.3
	9 : Adèle, Simon, répondre, revoir, fille, embrasser, jeune, Siwek...	type de narration (3 ^e pers.) propre au texte (1998a)	

Figure 3. Comparaison entre les classes obtenues sur les corpus de 2014 et de 2016, et liens (en dernière colonne) vers les sections de ce chapitre approfondissant l'analyse. La ligne épaisse délimite la classification principale. La classe 9 est grisée, étant non pertinente pour la suite de l'analyse car générée par un seul texte.

C'est la distribution de ces deux lexiques dans le corpus structuré chronologiquement, en fonction de la date d'écriture des témoignages, qui est mise à l'épreuve dans ce chapitre.

2.2. Cartographie lexicale : la structuration chronologique du corpus ou le temps de la mémoire

Depuis plusieurs années maintenant, sinon depuis l'origine, l'Analyse des données textuelles (ADT) a appris à fonctionner sur des « corpus maquettes » [Metwally 2017]. Pour des raisons linguistiques et méthodologiques, nous réduisons ici le corpus à ses 400 noms communs les plus fréquents dans les témoignages sans perdre l'ordonnement ou la linéarité du texte. Cette réduction peut se concevoir linguistiquement par l'importance du substantif pour dire le monde, les choses et l'idée. Et, de fait, expérimentalement, sur notre corpus et pour notre problématique de contraste entre les périodes d'écriture, nous avons observé que les mots des autres catégories grammaticales (déterminants, adverbes, etc.) étaient moins pertinents pour décrire et caractériser les différentes périodes⁵. Cette réduction se justifie aussi méthodologiquement, particulièrement pour les traitements statistiques qui franchissent le pas

⁴ Les deux classifications Iramuteq ont été opérées sur des mots lemmatisés, c'est-à-dire dont les variations de terminaison (dues au genre, au nombre, à la conjugaison) ont été neutralisées : c'est ainsi qu'avec l'intermédiaire du lemme « coup », le « coup d'œil » a été rapproché des « coups » infligés avec violence (cf. « crier », « frapper », « poing », etc.). Dans l'étude détaillée sur les spécificités, nous avons ainsi préféré distinguer singulier et pluriel pour les noms.

⁵ Les verbes, les adjectifs, les adverbes ont été étudiés mais les indices sont moins marqués et beaucoup d'entre eux se sont avérés spécifiques d'un texte plutôt que d'une période. Quant aux pronoms personnels ou aux déterminants, nous avons considéré qu'ils traduiraient ici des choix rédactionnels et stylistiques plus difficiles à cerner dans le cadre de notre problématique.

de la visualisation : représenter simultanément des centaines de milliers de mots rend impossible la lisibilité et la perception de la structuration globale des données, là où donner à voir 400 items produit une vue d'ensemble réaliste et synthétique révélant des « lignes de force » structurant le corpus.

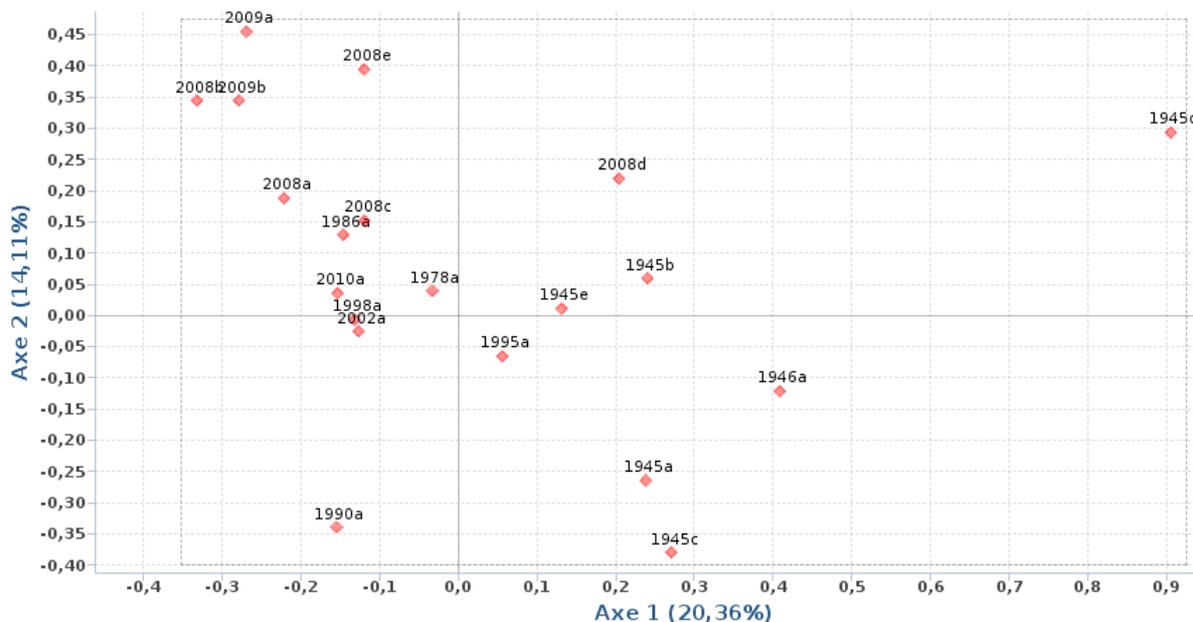


Figure 4. Plan factoriel des textes caractérisés par les 400 formes nominales les plus fréquentes du corpus

L'analyse factorielle des correspondances des 400 substantifs les plus utilisés⁶, distribués dans les 20 témoignages donne une carte (fig. 4) qui peut se lire selon une logique visiblement chronologique ou historique, avec les témoignages les plus anciens d'un côté de l'axe (à droite) et les témoignages plus récents de l'autre côté de l'axe ; seul un point (2008d) faisant exception à cette logique diachronique. Le calcul statistique fait donc très nettement ressortir le temps comme première dimension de structuration lexicale du corpus et de contraste entre les textes⁷.

Dès lors, la projection des mots sur la carte rend évidente l'interprétation (fig. 5). De manière spectaculaire en effet, le vocabulaire à droite de l'axe, c'est-à-dire utilisé de manière privilégiée par les témoignages les plus anciens, relève du lexique primaire que nous avons identifié comme contemporain de la Shoah et de la vie des camps. C'est d'abord un vocabulaire sensoriel qui exprime la douleur des corps et des esprits : les « cris », le « silence », le « sang », la « mort », le « feu », les « coups ». C'est également, très présent, le

⁶ N'ont été lemmatisés qu'une vingtaine de noms, tous ceux dont la variation singulier / pluriel n'est pas significative dans le plan factoriel 1x2 étudié (le singulier et le pluriel du mot sont groupés et dans le même quadrant, ou angulairement proches et ayant des profils de contributions comparables) ; et ont été neutralisées deux variations graphiques qui sinon accaparent en bonne partie l'axe 1 (au vu des contributions) : fusion des lignes « SS » et « S.S. », « bloc » et « Block ». Par ailleurs ont été élagués (« peeling ») trois mots quasi-exclusivement liés à un seul (gros) texte (1990a), qui perturbent une vision d'ensemble équilibrée (à savoir « comportement », « Häftlinge », « hongrois »).

⁷ Soulignons bien que le calcul statistique travaille uniquement sur la répartition des mots entre les différents textes, et ne connaît pas / n'utilise pas l'ordre chronologique des textes ou leur situation dans le temps. Cette information se lit *a posteriori* pour le chercheur sur le graphique à travers les noms donnés aux textes, mais ces noms sont ici pour la machine de simples étiquettes qui ne sont pas interprétées lors du calcul de l'AFC.

vocabulaire du besoin vital et principalement celui de la nourriture : le « pain », la « soupe », mais aussi les « couvertures », la « veste » ou la « chemise ».

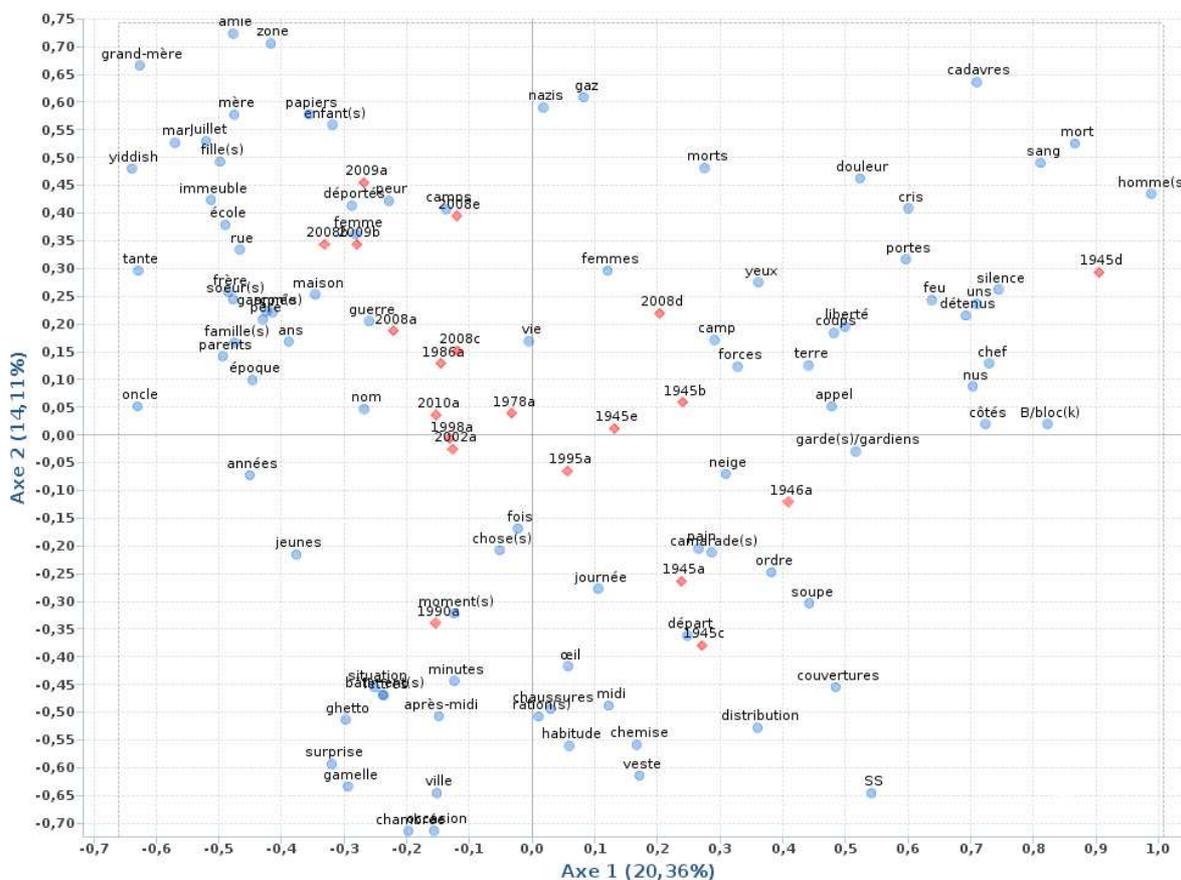


Figure 5. Plan factoriel des textes avec affichage des 97 substantifs structurant le plan 1 x 2⁸

Symétriquement, le vocabulaire à gauche de l'axe, c'est-à-dire celui utilisé principalement dans les témoignages tardifs ou postdatés, répond à notre description d'un vocabulaire ré-élaboré, c'est-à-dire appartenant plus à la re-présentation sociale de la Shoah qu'à la vie des camps elle-même. L'ensemble du vocabulaire familial s'y trouve (« grand-mère », « mère », « enfant(s) », « parent », « tante »...). Dans les années 2000 notamment, la Shoah ne semble pouvoir se dire et se comprendre par les témoins sans la reconstruction du roman familial précisément brisé par la déportation et l'extermination ; ce lexique fait écho au choix éditorial effectué par la Fondation pour la Mémoire de la Shoah d'éditer dans sa collection les photos et la généalogie des témoins. Comme le lexique généalogique ne concerne pas uniquement le couple mère-enfant éventuellement présent dans les premiers mois de la déportation, mais qu'il s'étend à l'« oncle », à la « tante », au futur « mari » rencontré après-guerre, ou aux « enfants » qui ont vu le jour depuis, ce lexique, sans doute étranger à la vie concentrationnaire elle-même puisque les familles sont le plus souvent décomposées, traduit

⁸ Pour la lisibilité, la fiabilité et la pertinence de la carte, nous avons limité la projection aux 97 noms (sur 399) vérifiant les critères suivants : (i) contribution sur l'axe 1 supérieure ou égale à 0,5 (ce qui sélectionne un ensemble de 43 points contribuant à 68 % de l'inertie de l'axe), ou (ii) contribution sur l'axe 2 supérieure ou égale à 0,5 (ce qui sélectionne un ensemble de 56 points contribuant à 60 % de l'inertie de l'axe), ou (iii) qualité dans le plan 1x2 (somme des cosinus carrés sur les axes 1 et 2) supérieure ou égale à 0,5 (notamment pour pouvoir visualiser des points moins contributeurs du fait de leur moindre fréquence, mais pourtant bien illustratifs des axes et enrichissant leur compréhension).

une réélaboration ou une mise en récit *a posteriori* du souvenir. De la même manière, le vocabulaire politico-religieux se trouve à proximité des textes tardifs : « Yiddish », « déportés », « zone » ou simplement « guerre » attestent d'une grille de lecture et de compréhension, politique ou religieuse, surplombante aux souvenirs intimes du déporté.

Nous admettons donc que dans l'architecture de leur mémoire, les témoins disposent de deux grands types de lexique qu'ils mobilisent de manière contrastée pour se remémorer la Shoah : un lexique proche de l'événement, fait de sang⁹ et de larmes, mobilisé à la sortie des camps et de la guerre ; un lexique plus (re)élaboré, fait souvent de notions plus abstraites, qui se trouve mobilisé à distance de l'événement dans les années postérieures.

2.3. Mesure et observation des évolutions lexicales. Éléments méthodologiques

Divisé naturellement en 20 textes datés, le corpus (formé de l'ensemble des 20 textes) a été structuré en trois périodes (l'immédiat après-guerre, les années 1970-1990 et les années 2000-2010) dont nous venons de constater l'influence sur la forme et le fond lexical du récit.

Pour la caractérisation précise et contrastive de ces divisions, l'outil textométrique le plus approprié est le calcul des spécificités [Lafon 1980], qui dresse un inventaire hiérarchisé des mots statistiquement surreprésentés ou sous-représentés dans les parties au regard du corpus dans son ensemble. Ce calcul mesure statistiquement l'écart entre la fréquence observée d'un mot dans un texte et celle attendue si les mots du corpus avaient été répartis au hasard.

Prenons deux exemples. Le mot « pied » est utilisé 193 fois dans l'ensemble du corpus dont 84 fois dans la première des trois périodes. Le calcul de probabilité permet de relever qu'il avait 1 chance sur 10 millions (!) d'apparaître autant de fois (ou plus) par rapport à une distribution égale relativement entre les trois périodes. On dit aussi qu'il a un indice de +7, soit le nombre de zéros de la probabilité. Par l'indice élevé (+7), le calcul pointe ainsi une concentration exceptionnelle du mot « pieds » en première période.

Le mot « Libération » présente à l'inverse une spécificité négative de -4. Traduisons : apparaissant une seule fois dans la première période, alors qu'il apparaît 39 au total dans l'ensemble du corpus, il est à l'évidence sous-employé. Soyons précis : il avait une chance sur 10.000 d'apparaître si peu, voire moins, par rapport à une distribution égale entre les trois périodes.

« L'égalité » évoquée est évidemment relative car elle tient compte de la longueur respective des sous-corpus ou périodes ici distingués. On remarquera donc qu'on a besoin de quatre variables pour calculer une spécificité, toutes internes au corpus étudié : la longueur du corpus d'ensemble, la longueur du sous-corpus étudié, la fréquence du mot dans le corpus dans son ensemble et la fréquence du même mot dans le sous-corpus.

Nous avons procédé au calcul catégorie grammaticale par catégorie grammaticale afin de neutraliser les effets de style liés à des écritures plus nominales vs verbales [Mayaffre 2006] : ainsi, le suremploi d'un nom commun est mesuré par rapport à l'emploi général des noms communs dans le corpus, et non par rapport à l'ensemble des mots.

De plus, la géométrie de notre corpus, dont les parties (les périodes) comptent chacune moins d'une dizaine de textes, impose pour chaque mot calculé comme spécifique de la partie de s'assurer que cette spécificité n'est pas due à une surreprésentation dans un seul texte, mais est un point commun partagé par différents textes de la partie. Cette vérification s'opère en

⁹ Cf. par exemple dans notre corpus l'ouvrage de 1945 de Julien Unger, *Le Sang et l'Or*.

consultant parallèlement les mêmes données, mais cette fois-ci partitionnées sur les textes (cf. en fig. 6, la zone d'écran en haut à droite). Ainsi on observe par exemple que la surreprésentation de « mort » en 1945-46 n'est pas représentative de l'ensemble des textes de la période, mais due au suremploi de ce mot dans le seul texte 1945d.

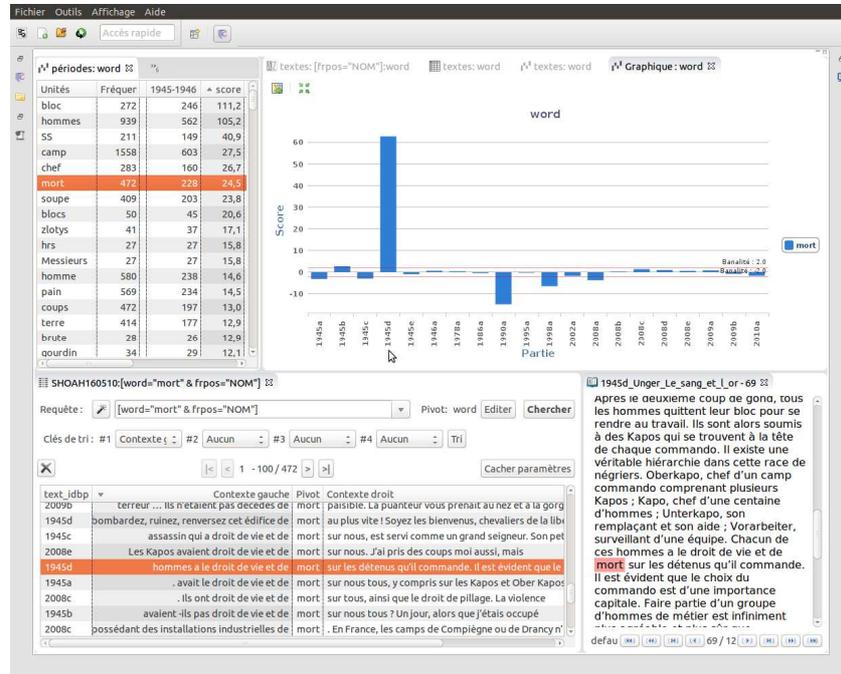


Figure 6. Interface pour le dépouillement des spécificités

Par ailleurs la consultation des contextes d'usage (par la concordance, cf. en fig. 6 la zone d'écran en bas à gauche) reste un préalable indispensable pour l'interprétation. Elle permet, par exemple, de s'apercevoir de l'hétérogénéité sémantique d'un mot (ex. « classe » : scolaire, sociale, tarifaire...), ou de repérer des emplois différents selon les périodes, ce qu'un calcul de cooccurrences dans chaque période permet éventuellement de confirmer.

Ainsi, pour le « départ » : en première période c'est souvent un moment attendu ou redouté, un événement qui a sa consistance propre (ex. « le jour tant redouté du départ », « Ils n'étaient pas du prochain départ. », « on reparla de départ », « nouvelle rumeur de départ » (1945a), « Toujours rien de nouveau concernant le départ. », « Cette fois-ci, cela se précise et prend des formes tangibles. Je parle du départ, bien entendu. », « La fièvre d'un départ éventuel s'est calmée beaucoup, on n'en parle même plus du tout ce matin. », « La vraie nouvelle du vrai départ n'a pas provoqué l'éclat, l'effet de bombe qu'on aurait pu attendre » « le signe du départ tant désiré » (1945c), « Fort de la rumeur relative à notre départ en Allemagne » (1946a)). Alors que dans les témoignages ultérieurs domine un usage plus phraséologique (« au départ ») ou ponctuel (avant/après/pour/depuis... le/leur/notre... départ).

Dernière remarque enfin, certaines unités lexicales sont de l'ordre du syntagme (mot composé, locution) plutôt que du mot. Les syntagmes les plus fréquents peuvent être recensés systématiquement ; puis un calcul de spécificité peut être appliqué individuellement à tout syntagme pour confirmer l'irrégularité de sa distribution (par ex. « zone libre » a une spécificité de +5,6 en 2002-2010).

Dès lors, nous nous proposons de décrire dans une troisième partie les mémoires en œuvre dans le corpus en insistant particulièrement sur l'immédiat après-guerre et, à l'opposé, les périodes les plus récentes (plus de 30 ans – soit après l'arrivée d'une ou deux nouvelles générations – après la guerre). Les 7 courtes sections présentées ne prétendent pas épuiser le

corpus mais mettent en valeur ses plus fortes spécificités chronologiques afin de repérer l'influence du temps dans le choix du lexique mobilisé par les témoins à trois moments distincts d'écriture. Dans tous les développements qui suivent une spécificité positive s'exprime donc de la façon suivante : « spécificité de 5 », « indice de 5 » ou « +5 » pour exprimer la même chose, dans ce cas une surreprésentation à hauteur d'une chance sur 100 000 (le mot en question avait une chance sur 100 000 d'apparaître autant ou plus dans le sous-corpus évoqué).

3. Mémoires. Distributions chronologiques et observation des évolutions lexicales dominantes

3.1. Au sortir d'Auschwitz, l'humanité en question

Le mot « homme(s) » ressort comme l'un des plus fortement caractéristiques des témoignages écrits en 1945-46 (spécificité de 110)¹⁰. L'observation des contextes en corpus révèle que ce mot concentre en première période davantage d'usages génériques au sens d'être humain (fig. 7). Dans cette période, le plus fort cooccurrent d'« homme(s) » est l'article défini (fig. 8), alors que par la suite, les usages sont davantage anaphoriques ou désignent concrètement un individu ou un groupe. On peut également noter la récurrence de l'expression « la vie d'un homme », dont les 6 occurrences dans le corpus se retrouvent toutes en première période, dans trois textes différents (fig. 9).

1945a	s'adressant pour une fois non à des numéros, à des esclaves, mais à des hommes.
1945b	la misère ravale l'individu au rang de la bête et lui fait perdre toute dignité. L'homme vraiment mis à nu, sans nul masque, n'est pas beau.
1945b	C'est à des moments semblables que l'homme se doit de résister à la tentation de suicide
1945c	Et voilà la preuve que l'homme est un drôle d'animal
1945c	Mais combien les hommes se laissent mener, impressionner facilement !
1945d	Il n'est pas permis à l'homme véritable de se taire lorsqu'il est témoin d'un pareil massacre !
1945d	Ô, hommes, qu'avez-vous fait des hommes, vos frères !
1945d	– Comment des hommes peuvent-ils faire cela ? – Des hommes, non ! Mais les SS sont-ils des hommes ?
1945d	Quel homme peut prévoir la parcelle de vie qui lui sera attribuée ?
1946a	L'homme dans une telle situation est une créature bien particulière.
1978a	Son cœur candide ne savait pas encore jusqu'où peut aller la méchanceté de l'homme
1978a	le régime concentrationnaire se chargeait de réduire l'homme à l'état d'animal ne pensant qu'à sa propre survie.
1978a	j'avais cru avoir atteint l'extrême limite de ce qu'un homme peut endurer, et pourtant ...
1995a	La faim. Ceux qui n'ont pas subi cet enfer ignorent le sens de ce mot. La faim dégrade. Elle devient une obsession. Elle détruit le léger vernis de civilisation acquis au cours des siècles et ramène l'homme à la bête de l'état de nature.
2002a	Ce comportement est de tous les temps et de tous les hommes et aurait pu être le nôtre si nous n'avions pas été juifs
2002a	il me paraît très important de ne pas oublier ces tragédies et de rester vigilant. L'homme ne change pas et ce qui s'est fait se refera.
2008c	Les hommes sont la seule espèce vivante qui, au cœur de la nuit, ne cesse de croire au matin.
2008c	Entre eux et nous, c'est nous qui étions les hommes. La déshumanisation qu'ils voulaient nous infliger, ce sont eux qui l'ont subie
2008d	Je suis resté un homme. Ils n'ont pas réussi à nous déshumaniser.

Figure 7. Exemples d'usages du mot « homme(s) » au sens d'« être humain », plus présents dans les premiers écrits.

¹⁰ Dans une autre langue, *Se questo è un uomo* est effectivement exactement contemporain à nos témoignages de l'immédiat après-guerre ; le titre en général et le mot *uomo* de Primo Levi ont suscité beaucoup d'interrogations auxquelles ces lignes peuvent apporter des éléments de réponse. Pour le français, et dans notre corpus, cf. particulièrement *Le Sang et l'Or* de Julien Unger où la question de l'humanité/déshumanisation est posée.

<i>Cooc.</i>	1945-1946			<i>Cooc.</i>	1978-1999			<i>Cooc.</i>	2002-2010		
	<i>Fréq.</i>	<i>Cofrq</i>	<i>S+</i>		<i>Fréq.</i>	<i>Cofrq</i>	<i>S+</i>		<i>Fréq.</i>	<i>Cofrq</i>	<i>S+</i>
le	15204	389	60	jeune	503	37	40	du	4792	73	21
du ¹¹	4148	120	22	un	9168	95	29	un	5900	80	20
cent	71	18	19	ce	6042	70	24	ce	3388	52	15
deux	379	32	18	brave	29	7	12	jeune	274	16	13
.	11027	195	12	du	6649	47	9	deux	588	15	7
@card@ ¹²	1362	45	10	deux	812	14	7	@card@	1457	22	6

Figure 8. Déterminants et adjectifs antéposés cooccurrents avec « homme(s) » selon les périodes (en lemmes¹³). Les colonnes chiffrent successivement la fréquence du mot cooccurrent globalement dans la période, sa fréquence au voisinage de « homme(s) », et la mesure statistique de sa spécificité dans ce voisinage (cf. § 2.3) qui évalue la force de la cooccurrence.

1945b	En Allemagne, la vie d'un homme est meilleur marché que celle d'un lapin.
1945c	une de ces corrections qui comptent dans la vie d'un homme.
1945d	La vie d'un homme vaut moins ici qu'une cigarette ou une assiette de soupe.
1945d	– Qu'est-ce qu'est la vie d'un homme ? Qui osera dire : – La vie d'un homme n'est absolument rien.
1945d	Quelle importance avait pour lui la vie d'un homme ? Bien moins certainement qu'un petit trou au milieu de la rue.

Figure 9. Ensemble des passages du corpus contenant « la vie d'un homme ».

Si l'on définit plus globalement un champ lexical lié à la déshumanisation et à la bestialité¹⁴, avec 249 occurrences en 1945-46 ce champ est caractérisé par une spécificité de +9 : « Nous n'étions plus que des épaves à face vaguement humaine » (1945b). Certains noms d'animaux accompagnent ce mouvement de déshumanisation : la surreprésentation du « chien » (+3,6) tient à « l'existence de chien » menée par les victimes (« comme un chien ») et à des désignations injurieuses (« Sale chien ! crie-t-il à l'homme. » (1945d)). Les « cochons » (+3,3) s'avèrent mieux nourris que les hommes, ou relèvent également de désignations injurieuses.

¹¹ Le lemme « du » se réalise quasiment toujours en « des » (indéfini ou défini contracté), dans tout le corpus on ne compte que 3 occurrences de cette construction au singulier (« du jeune homme » (2), « du saint homme »).

¹² Dans le modèle de langue utilisé par TreeTagger ici, @card@ est le lemme des nombres écrits en chiffres arabes. En d'autres termes les mots « 1 », « 2 », « 3 », « 4 », etc. sont rassemblés dans un seul lemme qu'on pourrait appeler « chiffres arabes », qu'on nomme ici par convention @card@, dont la fréquence est la somme des fréquences de tous ces chiffres arabes. Imaginons (pure fiction) que seuls les chiffres arabes « 1 », « 3 » et « 9 » soient utilisés dans votre corpus, que « 1 » soit utilisé 10 fois, « 3 » 25 fois et « 9 » 15 fois, le lemme @card@ aurait pour fréquence 10+25+15=50.

¹³ Le lemme est une forme normalisée conventionnelle du mot, tel le mot vedette d'une entrée de dictionnaire, à laquelle on rapporte l'ensemble des formes fléchies et variations typographiques en contexte du mot : ainsi les déterminants et adjectifs sont ici désignés par leur forme au masculin singulier, mais les fréquences décomptent aussi les occurrences au pluriel.

¹⁴ Lemmes (formes non fléchies) des attestations de ce champ lexical dans le corpus (avec indication de leur fréquence) : abattoir (10), abattre (17), animal (46), animalité (2), avilir (1), avilissant (1), avilissement (6), bestial (9), bestialité (5), bestiaux (42), bestiole (3), bétail (17), bétailière (1), bête (74), déshumanisation (5), déshumaniser (6), digne (27), dignement (5), dignité (30), dompteur (5), humain (221), humainement (4), humanisation (1), humaniser (2), humanisme (3), humaniste (3), humanitaire (8), humanité (39), inhumain (28), inhumanité (3), instinct (26), parquer (11), sauvage (24), surhumain (6), troupeau (23).

La nudité est souvent imposée, avilissante, fragilisante : « nus » présente une spécificité de +7 en 1945-46, dans la construction « tout/complètement/entièrement nus » pour un tiers des occurrences, et « pieds nus » pour un peu moins d'un quart. Du côté des verbes, « déshabiller » est également surreprésenté (+3).

Les « cadavres » (indice de 4,4) sont très présents, s'agissant cependant aussi bien de « monceaux » entassés (9 cooccurrences, indice de 18, indiquant rappelons-le le nombre de zéros, soit une change sur un million de milliards !), que de cadavres « vivants » (6 cooccurrences, indice de 5), « ambulants » (3 occurrences, indice de 6). Quant aux « brutes » (+3,2), elles sont ambiguës : il s'agit tantôt des bourreaux, tantôt des victimes (dormir, travailler « comme des brutes »).

L'homme, dépouillé de tout, aspire à la « liberté » (+5), tout en sombrant dans l'arbitraire, l'indifférence (« quelconque » (+5)). Ce n'est qu'avec le recul du temps que réapparaît la « décision » (-6, 51 occurrences, aucune en 1945-46), l'« idée » (-4), le fait d'être « conscient » (-4), et que la « peur » (-4) et le « choc » (-3) peuvent être nommés ; c'est plus tard aussi que le « ton » (-3,9) de la voix ouvre d'autres modes d'expression que les « cris » échappés sous les « coups » (premier cooccurrent de « cri » avec une co-fréquence de 20 et un indice de 14), et que la « figure » (+3,7), agressée, blessée, cède davantage la place au « visage » (-7) expressif et se prêtant à la description (cooccurrents : « sourire » (9), « expression » (8), « barbe » (7), « nez » (7), « rond » (6), « encadré » (5), « méconnaissable » (5), « tuméfié » (3)). C'est à distance de l'expérience d'Auschwitz que la survie devient une dimension centrale du témoignage (« survivre », « survivant », « survie » : -16 en 1945-46).

3.2. Cadrages temporels et spatiaux : du local et immédiat à des repères élargis

En ce qui concerne les cadres temporels et spatiaux du témoignage, c'est aussi les éléments les plus concrets, tangibles, proches, qui dominent dans les récits rédigés immédiatement après-guerre (fig. 10).

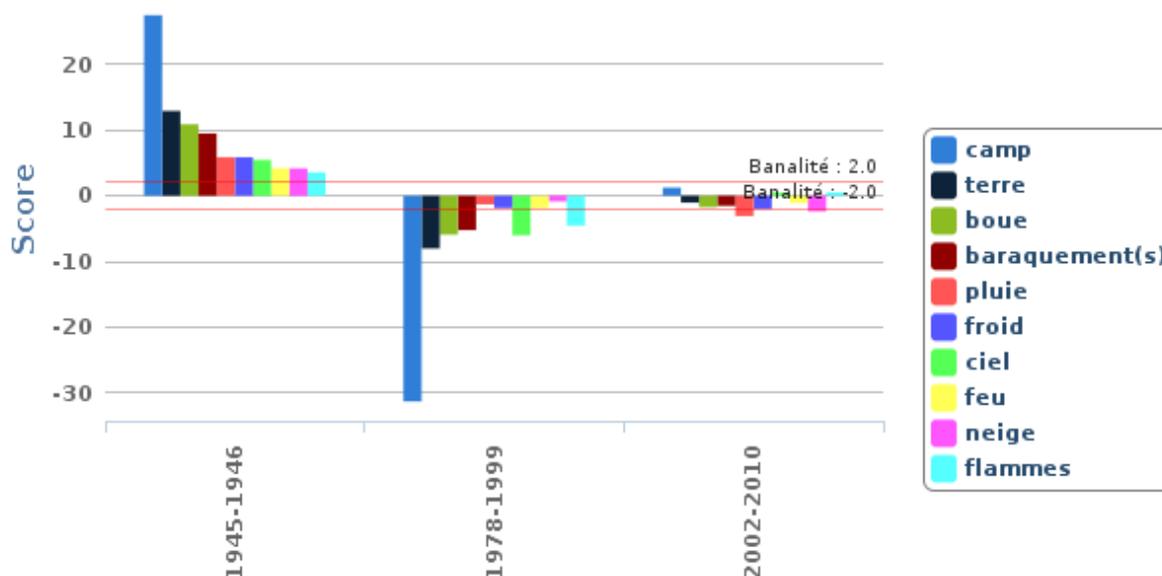


Figure 10. Distribution chronologique des éléments concrets du cadre

Les lieux sont d'abord et surtout le « camp » (+27) avec ses « bloc(s) » (+131) ou « baraquement(s) » (+9,5). Les conditions extérieures sont prégnantes : le « froid » (+16), la « pluie » (+5) (et « pleuvoir » (+4) concernant pour moitié la météo mais aussi pour moitié des « coups »), la « neige » (+4), « glacé » (+4), « geler » (+3) (le plus souvent des membres ou des personnes, 16 occurrences sur 23 en 1945-46), les « courant(s) d'air » (+3), la « boue »

(+10), « dehors » (+7). La « terre » (+12) cumule (i) le lieu ou l'on gît (coucher/écrouler/tomber/jeter/amonceler à/par terre), (ii) l'endroit ou le pays (terre de liberté/maudite/étrangère), (iii) le monde (les damnés de la terre, disparaître de la surface de la terre), (iv) le matériau (cailloux et terre, terre glaise, poignée de terre), et tout bonnement (v) les pommes de terre. Le « ciel » (+5) peut être une ouverture vers la liberté (ciel bleu, ciel étoilé : « On se console en se disant qu'en France les mêmes étoiles brillent encore de tout leur éclat, les mêmes constellations éclairent le ciel magnifique. » (1945d)) ou le théâtre de l'horreur (ciel enflammé des cheminées, des combats aériens, ciel vers lequel sont tournés les yeux des cadavres). Le « feu » (+4,2) aussi est présent, avec ses « flammes » (+3,5) et le verbe « brûler ».

Dans tout le corpus on parle de l'ouverture des portes (du wagon, ou des bâtiments qui laissent entrer le froid, ou ouverture brusque), mais en première période plus qu'ailleurs on évoque aussi des portes lourdes, que l'on ferme (fig. 11).

1945a	nous conduit à notre prison ; énormes portes	, verrous impressionnants, couloir immense et glacé,
1945a	des miradors. Nous passâmes sous des portes	entièrement cernées par des cordons de lampes
1945a	. Sans s'attarder à enfoncer une des portes	, ils pénétrèrent dans l'enceinte du camp,
1945b	Le soir, au verrouillage et fermeture des portes	, je posai la question à un des soldats de la Wehrmacht
1945b	leur donnait savon et serviettes, puis les portes	se fermaient hermétiquement ; par des trous grillagés
1945b	les fantassins défoncèrent les portes	grillagées. Dans les baraques, il y avait des morts et
1945d	sa visite matinale et quotidienne. Les portes	se refermèrent de nouveau et tous reprirent leur souffle,
1945d	laisse pas entrer. Il nous renvoie aux portes	arrière. Mais là, c'est un spectacle horrible.
1945d	En effet nous sommes devant les portes	de son bloc. Mais, comme on ne laisse pas pénétrer
1945d	sueur froide, ils attendaient derrière les portes	qui ne s'ouvraient pas. Ici on ne voulait pas d'
1945d	noire. La pluie tombait toujours et les portes	étaient toujours fermées. Lorsqu'on les ouvrit enfin pour
1945d	deux hommes à l'aspect farouche. Les portes	arrière ont été barricadées par de grosses poutres,
1945d	N'est-ce pas derrière ces portes	, condamnées maintenant, que s'entassaient des monceaux
1945d	regardent Le camp est vide. Toutes les portes	sont consignées, les camions et les SS, l'arme à
1945d	avaient même pas de bat-flanc, ni de portes	, ni de fenêtres. Les pauvres malheureuses n'avaient
1945d	appel du commando disciplinaire. Les portes	de ce lieu de réclusion s'ouvrent dès le second coup de
1945d	la mort. Les alertes nous ferment les portes	de fer et nous volent le peu d'oxygène nécessaire à la
1945d	crimes innombrables. Faites sauter les portes	de nos cavernes pour que nous puissions au moins mourir
1945e	Les scellés sont collés sur nos portes	, heureusement Monsieur Ferry reste dans la maison
1945e	voyant la panique vinrent clouer les portes	pour que personne ne tente de s'échapper. Il faisait déjà
1986a	une savonnette, s'y engouffraient. Les portes	bouclées hermétiquement, les douches débitaient un gaz
1990a	le ghetto Le 15 mai 1944, les portes	du ghetto se refermaient sur nous. Nous étions installés
1990a	Avant que les cadenas ne ferment les portes	, on nous distribua un seau d'eau et un autre pour
1990a	le boulanger de ma ville natale, des portes	ou des guichets métalliques fermés, groupés par deux.
1995a	père dans le garage, dont les doubles portes	coulissantes ont été fermées. Tout à coup, je l'entends
1998a	, nous avons ordre d'attendre. Les portes	sont plombées, nous n'avons aucune liste et nous ignorons
1998a	suffisamment plein qu'ils ferment les portes	. Compressée contre son père, Adèle distingue la seule
2002a	halle était fermée la nuit par de lourdes portes	métalliques. Dans la journée, ces portes restaient ouvertes
2002a	usage de son parachute. Les lourdes portes	d'acier de ce qui restait de notre palais se dégonflèrent et
2002a	à l'écrasement lors de la chute des portes	d'acier du garage à Nordhausen, j'avais été pris dans
2008a	parents, soit bénie ! Enfin, les portes	du wagon qui ont été scellées depuis trois nuits – Confiez
2008b	remontés dans nos wagons et les portes	se sont refermées derrière nous. Puis vint la troisième nuit
2008c	au quatrième étage d'un bâtiment sans portes	ni fenêtres, sans murs non plus d'ailleurs car l'immeuble
2008e	du coucher, les gardes fermaient les portes	de nos cellules individuelles avec une manivelle. On nous
2008e	trois jours et trois nuits sans que les portes	soient jamais ouvertes. Les femmes et les enfants pleuraient
2009b	tiède, mais de l'air ! Les portes	des wagons étaient verrouillées. Un peu de paille était

Figure 11. Sélection des contextes où les « portes » sont liées à l'enfermement.

Dans les témoignages ultérieurs (1970-2010), le vocabulaire englobe davantage de lieux du quotidien fréquentés avant ou après la déportation : le logement (« maison » (-8 en 1945-46), « appartement » (-7), « hôtel » (-4) dont l'hôtel Lutetia, « immeuble » (-4)) et son environnement (« rue » (-21), « arrondissement » (-5), « métro » (-4), « quartier » (-4),

« boulevard » (-3), « mairie » (-3), « ghetto » (-6)). Le « bâtiment » (-4) et la « baraque » (-3) semblent notamment remplacer un certain nombre d'occurrences du « bloc » dont l'usage était particulièrement massif en 1945-46.

Ces témoignages plus tardifs peuvent aussi davantage recourir à un vocabulaire géopolitique, resituant la réalité des camps dans un contexte international. On relève en effet dans le corpus 32 noms de continent, de pays ou de capitale de fréquence supérieure ou égale à 10. 13 d'entre eux sont en déficit (au seuil spécificité de -2) dans les témoignages rédigés en 1945-1946 (jusqu'à « Roumanie » sur la fig. 12), tous les autres ont une répartition plus équilibrée au fil du temps, aucun n'étant surreprésenté en 1945-46. Dans cette première période, quand les pays sont évoqués, c'est plus particulièrement à travers les noms de nationalité, donc à travers des groupes de personnes, plutôt que des entités géopolitiques (« Français » (spécificité de +7 en 1945-46), « Polonais » (+6), « Russes » (+4), « Hollandais » (+3), « Tchèques » (+3), etc.).

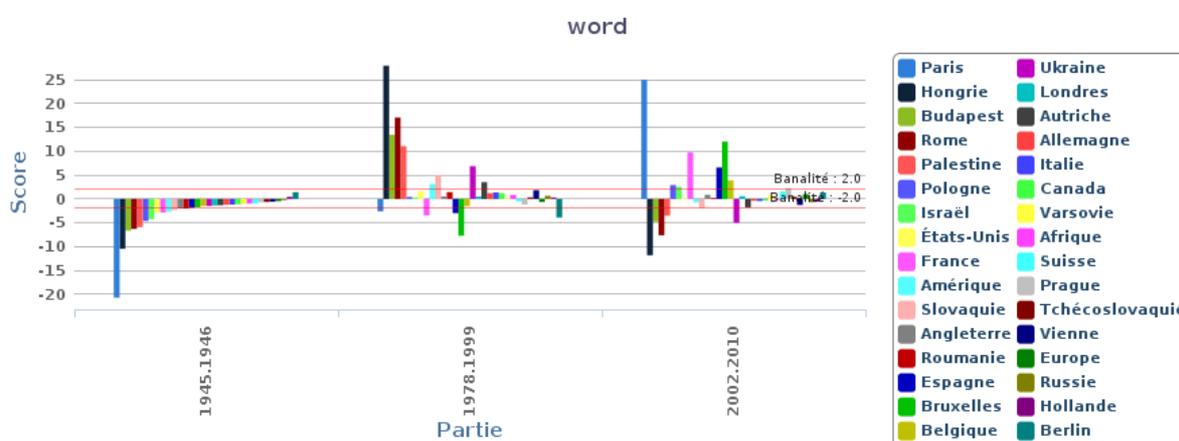


Figure 12. Distribution chronologique des 32 noms de continents, pays ou capitales mentionnés plus de 10 fois dans le corpus

On observe, pour le cadre temporel, le même mouvement de rétrécissement/élargissement que pour le cadre spatial (fig. 13). Dans les témoignages de 1945-46, l'échelle temporelle est avant tout celle de la « journée » (+3,5), de ce qui revient « quotidiennement » (+2) : le « matin » (+11) et le « soir » (+6) (avec l'« appel » (+10) du matin ou du soir qui rythme la journée), le « midi » (+4) (dont la « soupe de midi » (+3)), la « nuit » (+4,7). On peut entrer dans le détail des « heure(s) » (+7,7), quart(s) d'heure, demi heure, heure(s) et « demie » (+3), « minutes » (+4,9). On est dans le temps vécu, le déroulement, la durée, avec le suremploi des adverbes « vite » (+5), « lentement » (+5), « encore » (+25) (dans les différents sens de répétition, de durée, d'attente et déception), « enfin » (+4). Les repères sont immédiats : « hier » (+10), « maintenant » (+7), « ici » (+7).

En contrepoint, le vocabulaire sous-employé en 1945-46 par contraste avec les années 1978-2010, sont les « année(s) » (-30) (et les « ans » (-32), à rapprocher du vocabulaire de la description de vie : « âge » (-10), « originaire » (-6), « aîné » (-5), « maternel » (-5)), l'« époque » (-15), les « mois » (-10), les « semaines » (-5), et dans une moindre mesure les saisons. Les témoignages plus tardifs se repèrent par rapport au « début » (-5) ou à la « fin » (-6) de la guerre, d'un mois, d'une saison, d'une année, ils embrassent le « présent » (-3) (« à présent » pour deux tiers des occurrences), l'« avenir » (-5) et l'« origine » (-4). Le mot « temps » (-6) participe à des expressions qui montrent que le cadre temporel s'est élargi : en ce temps(-là), (depuis/pendant) tout ce temps, dans/en/par ces temps, peu de/quelque temps après/avant, dans peu de temps, avec le temps, en son temps. Les noms de mois suivis d'une

année (ex. « janvier 1945 », « juillet 1942 »), donnant un ton plus historique au récit, sont significativement surreprésentés en dernière période (spécificités de -17 en 1945-1946, -7 en 1978-1999, et +36 en 2002-2010).

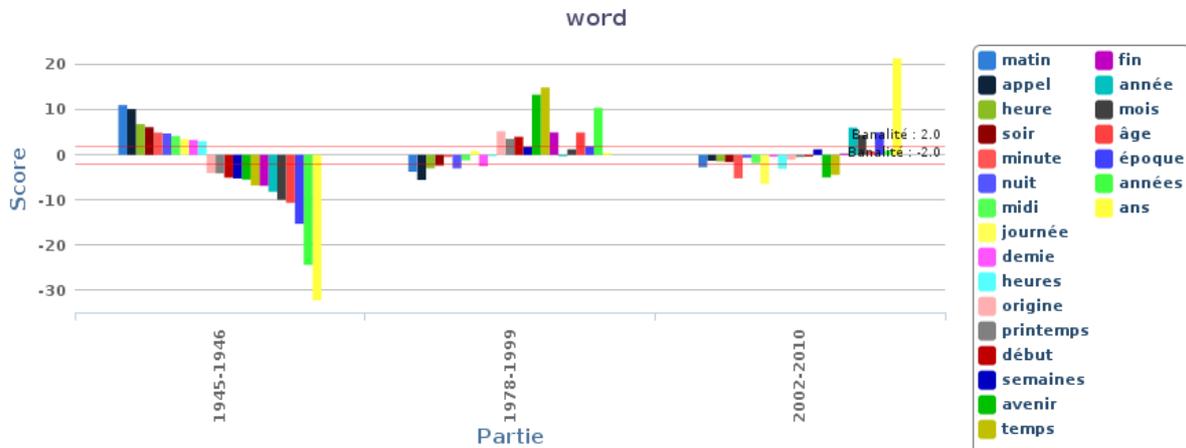


Figure 13. Distribution chronologique de noms liés à l'expression du temps

3.3. L'organisation et le quotidien des camps au cœur des premiers témoignages

En 1945-1946, les personnes évoquées reflètent l'organisation des camps : « SS » (+40), toute une gamme de « chef(s) » (+35) (fig. 14), « commando(s) » (+10) dont « commando spécial » (+5), « détenus » (+8), « sentinelles » (+3,5), « doyen d'âge » (+3), mais aussi « coiffeurs » (+4,9), « spécialistes » (+4,8), « docteur(s) » (+6), « médecin(s) » (+4,7), « malades » (+3,5), « mourants » (+3,2). Le témoin parle de ses « camarade(s) » (+8) ou énumère par groupes : « hommes, femmes¹⁵, enfants et vieillards ».

<i>expression</i>	<i>Fréquence (corpus entier)</i>	<i>Spécificité en 1945-1946</i>
chef(s) de/du B/bloc(k)(s)	65	17
chef de/du camp	44	7
chef(s) de groupe	13	4
chef de chambre/chambree	12	3
chef(s) d'équipe	11	-
chef(s) de chantier	7	-
chef de section	7	4
chef d'atelier	6	-
chef de famille	5	-
chef de/du Kommando	5	-
chef de production	4	-

Figure 14. Syntagmes « chef(s) de + NOM » attestés dans le corpus (avec indication de la spécificité quand elle est significative)

Même chose pour les activités quotidiennes : surreprésentation dans les premiers récits de plusieurs séries de termes :

¹⁵ Ainsi le pluriel « femmes » est surreprésenté dans les écrits de 1945-46 (+6), alors que le singulier « femme », se prêtant au sens de « épouse » et rejoignant ainsi le vocabulaire familial, est caractéristique des écrits plus tardifs (+8,3 pour 1978-2010).

- « désinfection » (+6) (dans des contextes comme conduit/emmené/envoyé/passer à la désinfection, bâtiment/bloc de désinfection), « visite médicale » (+4), « contrôle » (+4) (notamment « contrôle de poux » (+3)), « nettoyage » (+3,6), « propreté » (+3) (« travaux de propreté »), « douche » (+3,3), « ranger » (+7) (dont le sens de mettre en rang, et de se ranger) ;
- « distribuer » (+6) et « distribution » (+3,9), principalement « de (la) soupe » et « de pain », mais aussi « manger » (+6), « voler » (+5) et « vendre » (+4) ;
- « coucher » (+5), « dormir » (+4), « fatiguer » (+4, essentiellement au participe passé (30 occurrences sur 34) en 1945-46), et le « repos » (+3,2) que l'on cherche ou dont on peut être plus ou moins privé ; on peut aussi rencontrer l'« ennui » (+3,7)
- « falloir » (+20) (avec comme constructions les plus fréquentes « falloir faire » et « falloir attendre »), « devoir » (+8)
- « conduire » (+10), en 1945-46 dans des tournures de sens passif pour deux tiers des occurrences (être conduit(s), (on) nous/les conduit)

Pour ce qui concerne l'alimentation, « soupe » (+23), « pain » (+14,5) et « margarine » (+8) sont cooccurrents les uns des autres et attirent « morceau », « ration(s) », « tranche », « grammes » (surtout de pain), « litre » (surtout de soupe), « distribution », « midi », « soir », « supplément », « gamelle ». Avec « beurre » (+6) (mais aussi « margarine ») se déploie en cooccurrence le paradigme des gourmandises et des extras : « pain » (blanc), « confiture » (+3,2 en 1945-46), « lait », « saucisson », « gâteaux » (+4,2), « thé », « sucre », « fromage », « œufs », « jambon », « corned-beef », « porridge », « viande », « cigarettes » (+4,2), « volailles », « chocolat », « poisson », « lard ». Les « pommes » (+4,2) ressortent également de la première période, majoritairement comme « pommes de terre » (+5). Le qualificatif d'« infect » (+8) s'applique certes à la nourriture (soupe, tisane, eau, pain, etc.) mais s'étend à toutes sortes d'objets : lieu, personnes, objets (couchage, rasoir), mœurs, plaies, odeur, etc.

Au titre des « effets » personnels (+4,4), le calcul met en évidence les « bagages » (+6) et « paquets » (+4), les « couvertures » (+10) et les « lits » (+7), les « souliers » (+5), le « savon » (+5), les « lunettes » (+3,9). Plutôt que de parler d'habits, le suremploi de « couvrir » (12) montre qu'on évoque les hommes couverts de « chiffons » (indice de cooccurrence de +8), de « sueur » (+8), de « boue » (+5). Les « poches » (+3,6) sont fouillées et vidées, le moindre avoir devient « butin » (+3,8). Quant à l'« or », ses premiers cooccurrents dans le corpus sont « dent(s) » (indice de cooccurrence +16), « bijoux » (+11), « montres » (+3), soit ce dont on dépouvoit les arrivants ou les cadavres, et qui devient monnaie comme les « dollars » (+5) d'un « trafic » (+6).

3.4. L'expression initiale du ressenti, dominé par la violence et la souffrance

De façon saisissante, les verbes les plus spécifiques des témoignages écrits au sortir de la guerre multiplient et concentrent l'expression d'actes de violence ou de douleur : « brûler » (+19), « tuer » (+16), « crier » (+14), « chasser » (+9), « arracher » (+9), « mourir » (+7), « frapper » (+7), « trembler » (+5), « déchirer » (+5), « tomber » (+5) (mais avec des usages plus hétérogènes), « tordre » (+5), « enfler » (+4), « pleuvoir » (+4) (26 occurrences en 1945-46, pour moitié concernant des coups), « achever » (+4) (32 occurrences en 1945-46, pour moitié au sens de « achever un homme » ou « achever sa vie, ses souffrances »), « traîner » (+4) (dont se traîner (un quart des occurrences), traîner la/le(s) jambe(s)/pied(s)/patte, personnes qui sont traînées), « renverser » (+4), « écraser » (+4), « fatiguer » (+4),

« condamner » (+3), « briser » (+3) (19 des 25 occurrences de 1945-46 portent sur l'homme qui est brisé dans son corps/cœur/courage/espoir etc.).

<i>Syntagme</i>	<i>Fréquence (corpus entier)</i>
coups de bâton(s)	31
coups de cravache(s)	21
coups de pied(s)	21
coups de feu	18
coups de trique(s)	16
coups de crosse	13
coups de poing(s)	14
coups de gourdin	9
coups de botte(s)	9
coups de matraque	7
coups de canon	5
coups de revolver	5
coups sur le derrière	5

Figure 15. Les syntagmes construits autour de « coups » dans le corpus (de fréquence 5 ou plus)

Du côté des noms ce sont les « coups » (+13) qui se déploient en un large paradigme (fig. 15) et dominant, associés au « gourdin » (+12) et au « poing » (+3), et premier cooccurrent de « cris » (+4,6). Cris de « souffrance(s) » (+10), appels au « secours » (+3), « plaies » (+3), « martyre » (+3), « misère » (+3). Si un « traitement » (+4,8) peut être médical, il s'agit aussi souvent de mauvais traitement : traitement « inhumain » (indice de cooccurrence de +3), les trois premiers cooccurrents de « traitement » étant les verbes « infliger » (+12), « subir » (+7), « supporter » (+5). Cela va jusqu'à un droit « de vie et de mort » (+3), l'expression étant attestée dans quasiment tous les textes de 1945-46 (fig. 16). Dans ces témoignages, les cadres nazis sont dénoncés comme « assassins » (+4,4), ils commettent des « crimes » (+3,3) et doivent être punis.

<i>Id.</i>	<i>contexte gauche</i>	<i>pivot</i>	<i>contexte droit</i>
1945a	N'importe quel S. S. avait le droit de	vie et de mort	sur nous tous
1945b	entière merci : n'avaient-ils pas droit de	vie et de mort	sur nous tous ?
1945c	décidait d'un geste négligent de la	vie et de la mort	de milliers et de milliers d'individus.
1945c	Lui, l'assassin qui a droit de	vie et de mort	sur nous, est servi comme un grand
1945d	bloc sept et je suis le maître de votre	vie et de votre mort	! Finis vos bavardages
1945d	Chacun de ces hommes a le droit de	vie et de mort	sur les détenus qu'il commande.
1946a	[Le chef de camp] était le maître « à la	vie et à la mort	» des occupants du camp.
1946a	Le jeune homme avait décidé de la	vie et de la mort	. Nos camarades ont directement été conduits
2008c	Ils ont droit de	vie et de mort	sur tous, ainsi que le droit de pillage.
2008e	Les Kapos avaient droit de	vie et de mort	sur nous. J'ai pris des coups moi aussi

Figure 16. Le « droit de vie et de mort » et ses variantes dans le corpus

Le corps transparait d'abord au prisme de cette violence et des souffrances endurées. Si « pied(s) » est la partie du corps quantitativement la plus saillante (+8,8), c'est certes dans une locution comme « à pied », mais aussi « coup(s) de pied(s) », « pied(s) nu(s) », « pieds enflés ». Les corps semblent réduits à quelques parties et démembrés, les « têtes » (+4,1) que l'on compte, les « jambes » (+3). Les « dents » (+3,8) sont les dents en or (qui sont arrachées), mais participent aussi à des expressions marquant l'épreuve comme « serrer les dents », « claquer des dents », « grincer/grincements de(s) dents ».

Évolution chronologique déjà signalée par la carte factorielle (§2.2), le vocabulaire du sens (*vs* du raisonnement), ici centré sur la souffrance physique (celle du quotidien éprouvant des camps comme celle des épisodes de violence), apparaît très marqué au commencement pour

s'effacer au fil du temps. Pour le continuum – c'est-à-dire la continuité mais aussi la démarcation – entre mémoire et histoire [Joutard 2013], ce constat est non obvie, au regard des récits de vie. Sans conteste, le souvenir d'abord sensoriel des témoins semble condamné à se rationaliser à mesure de l'éloignement du traumatisme. L'historicisation de la Shoah passe donc par un déclin de l'expression de la souffrance physique, par définition individuelle, pour aller vers une intelligibilité d'ordre social et rationnel. Une telle évolution pourrait être un schéma général qui marque la transformation d'un épisode en événement historique. Du reste, les psychologues et neuroscientifiques soulignent que le trauma est l'envahissement du présent par un passé qui s'impose donc comme présent. On pourrait se dire, avec eux, que le changement de registre de vocabulaire est la marque d'une résilience en cours.

3.5. Premiers mots pour dire l'indicible : un vocabulaire de l'extrême et de l'intensité

Beaucoup d'adjectifs et adverbessurreprésentés dans les témoignages de 1945-1946 marquent l'intensité :

- de façon générique : « véritable » (+8), ainsi que le superlatif relatif (« le plus » et ses variations, +10), les usages de « pareil » (+6) et « combien » (+3), cf. fig. 17,
- mais aussi par rapport au volume : « grand » (+3), « immense » (+2), « innombrable » (+2),
- à la force : « formidable » (+2), « fort » (spécificité de +4 pour l'adverbe et +2 pour l'adjectif), « vif » (+2),
- à l'intégralité : « complet » (+5), « partout » (+9), « tout » (+5), « rien » (+4), « entier » (+3), « entièrement » (+3), « strictement » (+3), « exclusivement » (+2),
- à l'excès : « si » (adverbe) (+4), « trop » (+3), « absolument » (+4), « infiniment » (+2).

1945b face vaguement humaine. Qui pourrait oublier de **pareils** moments ? Aujourd'hui encore, il m'est pénible
 1945c nous étions victimes d'un cynisme, d'une hypocrisie sans **pareils**.
 1945d combien de temps une **pareil** existence pourra-t-elle durer ? Jusqu'où la honte, l'avilissement
 1945a **Combien** de malheureux ont été retrouvés, atrocement ployés sur cette barrière électrique
 1945c Mouillés [...] dès le matin, **combien** de fois gardions-nous les vêtements trempés sur le dos jusqu'au soir
 1945c c'est tout à fait normal. Mais **combien** les hommes se laissent mener, impressionner facilement !
 1945d Mais aussi, **combien** extraordinaire de voir cette chose impossible : chacun de ces hommes, dans son immense détresse, gardait encore au cœur une toute petite lueur d'espoir.
 1945d Vous pouvez vous rendre compte, **combien** peu il faut à l'homme pour vivre et pour espérer.
 1945d Ah, hommes, que la vie est belle ! **Combien** voudrions-nous la vivre encore !

Figure 17. Exemples d'usages intenses de « pareil » et de « combien »

On relève aussi des adjectifs négatifs au sens fort (« invraisemblable » (+2), « impossible » (+2)), des adjectifs intenses péjoratifs (« terrible » (+7), « épouvantable » (+6), « atroce » (+4) et « atrocement » (+2), « violent » (+4), « affreux » (+3), « misérable » (+3), « affamé » (+3), « terrifiant » (+2), « sévère » (+2), « lugubre » (+2)) mais aussi quelques mélioratifs (« beau » (+5), « superbe » (+3), « excellent » (+3), « bon » (+2), « charmant » (+2), « meilleur » (+2)).

3.6. La recontextualisation politique, sociale, religieuse, et l'ascension remarquable du terme « juif »

À l'inverse du vocabulaire sensoriel, le vocabulaire politico-religieux augmente au fil du temps. L'exemple le plus spectaculaire concerne le terme « juif », fortement sous-représenté

dans les premiers écrits (spécificité de -31 en 1945-46) et par contraste surreprésenté ensuite (spécificités de +6,6 et +5,8 dans les deux périodes suivantes) (fig. 18). De fait, les témoignages du corpus écrits dans l'immédiat après-guerre (1945-1946) n'utilisent que très peu la référence identitaire ou religieuse à la judaïté : à la Libération, le mot « juif » semble chargé négativement – presque un tabou lexical – ; c'est le mot de la stigmatisation, le mot imposé par l'Autre. C'est dans les sous-corpus des années 1970 et suivantes que les « juifs » deviennent les acteurs (en l'occurrence les victimes) de la Shoah : « juif » devient alors le mot de la revendication identitaire – il n'est plus péjoratif. De tabou, il peut devenir étendard. L'évolution du mot dans le corpus reflète du reste l'évolution de la société française : au sortir de la guerre, les déportés sont seulement reconnus par l'Etat comme « déportés politiques » (et non religieux), il s'agit à la fois de se (re)fondre dans une société laïque et humaniste pour mieux se reconstruire, revenir à la vie, et tirer un trait sur la stigmatisation ; dans les années 2000, en revanche, on affirme l'identité du groupe.

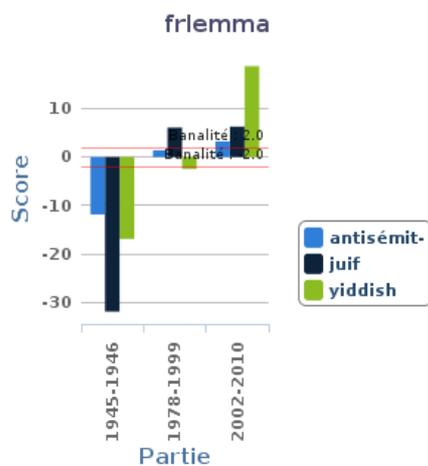


Figure 18. Distribution chronologique du mot « juif » et de termes apparentés

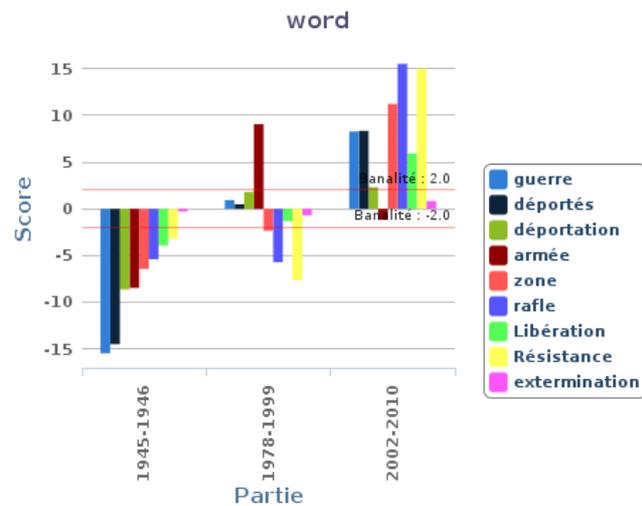


Figure 19. Distribution chronologique de termes historico-politiques dans le corpus

Les témoins utilisent au fil du temps un vocabulaire de plus en plus général, social, historique ou politique au sens étymologique (fig. 19). Replaçant son témoignage dans le cadre général de la « guerre » et de la mobilisation de l'« armée », le récit se repère en évoquant les lieux (« zone » libre ou occupée), les acteurs (« Résistance », « déportés »), les moments (« rafle », « Libération ») d'un vécu partagé, national ou mondial : l'histoire individuelle embrasse désormais l'histoire collective.

Mais ici encore l'interprétation se doit de rester prudente, et cette analyse serait à confirmer sur un corpus plus étendu. En effet, pour nos mesures sur le présent corpus, certains termes analogues semblent échapper à l'influence temporelle : « extermination », « peuple », « paix », qu'on aurait pu attribuer à ce vocabulaire plus conceptuel et historique, sont repérés comme banals¹⁶ avec une spécificité inférieure à 1 dans les trois périodes (en dépit d'une fréquence globale supérieure à 70 pour les deux premiers et de 34 pour le troisième, qui auraient permis de capter statistiquement leur éventuelle irrégularité de distribution).

¹⁶ Le repérage des termes banals dans un tableau de spécificités peut se faire dans TXM avec la macro BasicVocabulary.

3.7. La reconstruction tardive du roman familial

Nous avons noté l'importance remarquable du roman familial dans les récits de vie ou « récits de morts » des rescapés d'Auschwitz [Mayaffre et Ben Hamed 2014]. Comme catharsis à l'extermination, la famille abimée par l'univers des camps prend, au fil du temps et *a posteriori* pour le témoin, une valeur cardinale dans la narration des événements. On est frappé par la concentration des noms de parenté en tête des spécificités les plus négatives des témoignages écrits en 1945-46 : les cinq premiers noms communs étant « père » (-52), « mère » (-43), « frère » (-37), « parents » (-36), « famille » (-36). Et, naturellement, ces mots et bien d'autres du genre apparaissent en tête des spécificités les plus positives pour la période globale ultérieure, celle de 1978-2010.

L'examen chronologique du vocabulaire généalogique peut s'organiser en plusieurs groupes, et s'avère pour tous éloquent. Le premier groupe concerne le cercle le plus étroit de la cellule familiale (fig. 20). Les deux principales exceptions ne remettent pas en cause le mouvement de fond. En 1945-46, 26 des 33 occurrences de « frères » sont au sens de la fraternité humaine (« frères de misère »). Si l'on refait le calcul en ne gardant que les 7 occurrences relevant du sens familial, l'indice fortement négatif (-8,9) rejoint celui des autres termes généalogiques. Quant à « papa », 13 de ses 14 occurrences de 1945-46 sont concentrées dans un seul texte (1945e). Si l'on écarte ce texte comme non-représentatif de la partie de ce point de vue, l'indice recalculé s'accorde également au mouvement négatif d'ensemble (-1,7).

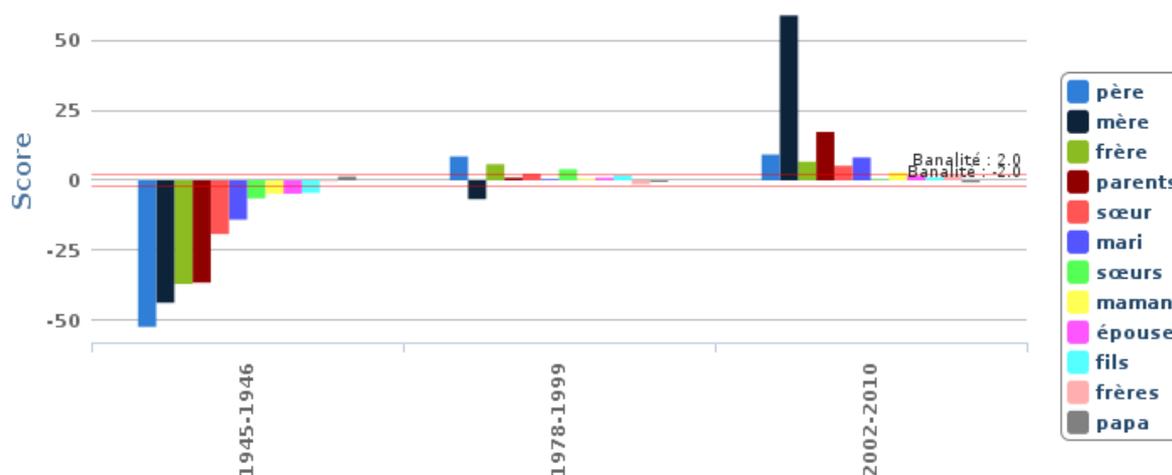


Figure 20. Distribution chronologique du vocabulaire familial : groupe 1, la cellule familiale.

Dans le second groupe, en étendant les liens familiaux aux relations plus indirectes et transversales, le même sous-emploi massif caractérise les premiers témoignages (fig. 21). Les indices les plus faibles s'expliquent statistiquement par la fréquence globale relativement basse des termes concernés (« petits-enfants » n'a que 17 occurrences en tout dans le corpus, dont une seule en 1945-46 ; pour « neveu », c'est 1 occurrence sur seulement 11 au total).

Un troisième groupe peut concerner non plus les membres mais les structures et événements familiaux (comme « famille » (-36) et « mariage » (-9)) (fig. 22), et un quatrième groupe se spécialise sur les mots plus ambigus, ne concernant pas toujours nécessairement un lien de parenté (comme « fille » (-21), « enfants » (-11) ou « femme » (-8)) (fig. 23) : pour ces deux derniers groupes encore, la sous-représentation en première période est fortement marquée.

C'est bien dans une réélaboration tardive que le vocabulaire de la famille intervient dans des témoignages autobiographiques, avec la reconstruction d'une saga familiale le plus souvent détruite. Explicitement souvent, les récits des années 2000 sont ceux de grands-parents qui, ayant dans un premier temps protégé leurs propres enfants de leur passé douloureux et

traumatisant, acceptent d'écrire pour leurs petits-enfants, soucieux de leur transmettre l'histoire d'une tragédie universelle, bien sûr, mais aussi par concrétion d'une tragédie familiale. Parce que la Shoah est désormais comprise comme un génocide, la mémoire du rescapé des années 2000 convoque concrètement sa généalogie. Dans les camps, l'horizon d'attente, c'était la mort et l'extermination. A la Libération, ou plutôt dans les années ultérieures, c'est la lente reconstruction et le retour de la chaîne de la vie. Le recul des années permet progressivement non seulement de reconstruire une généalogie ascendante mais de poser la question du lien familial en devenir vers les descendants. Davantage que par la référence intime à la souffrance, au corps, à la déshumanisation des récits des premiers témoignages, les témoignages plus tardifs permettent à l'émotionnel de rejaillir dans les textes par l'irruption de la famille, aval et amont. Âgés, les témoins racontent davantage l'enfance heureuse (brisée brutalement par la guerre) et le contexte familial d'alors. Ils évoquent, souvent à l'usage direct des petits-enfants, les aïeux d'avant la Shoah pour les origines, et les disparus pendant, sans sépultures.

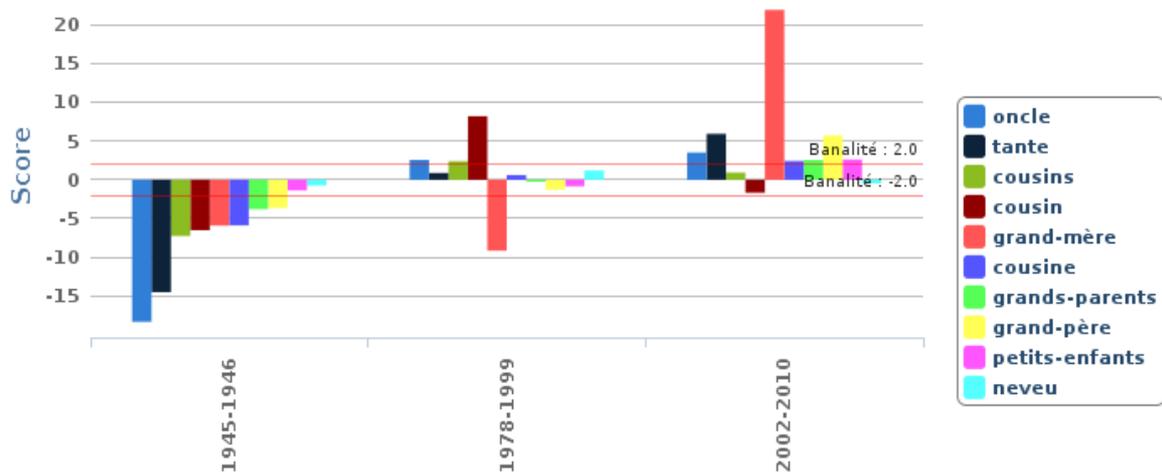


Figure 21. Distribution chronologique du vocabulaire familial : groupe 2, la famille élargie.

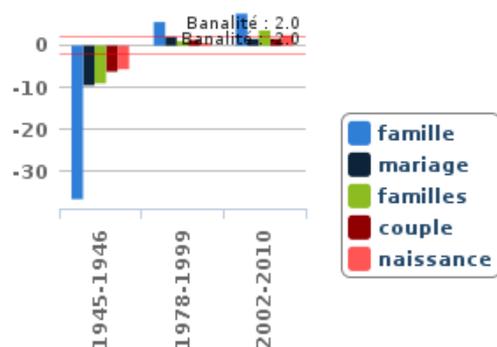


Figure 22. Distribution chronologique du vocabulaire familial : groupe 3, structures et événements familiaux.

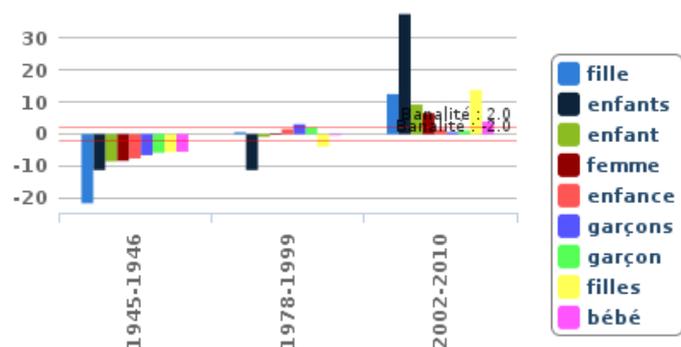


Figure 23. Distribution chronologique du vocabulaire familial : groupe 4, termes ambigus.

4. Conclusion

À la « concurrence des mémoires » qui se joue peut-être dans nos cerveaux et dans nos manuels scolaires, cette contribution a essayé d'opposer la complémentarité des lexiques. Si le souvenir se construit et se reconstruit dans et par la matérialité langagière – chaque mot étant à la fois trace et média, quasi physiques, de l'événement reconvoqué –, l'étude des lexiques mise en œuvre pour faire advenir le témoignage doit être envisagée avec le plus grand soin méthodologique.

Les récits de vie à n années de distance semblent devoir répondre au double impératif de (faire) revivre l'événement et de le rendre intelligible : nous sommes là au cœur à la fois de *l'identité narrative* du témoin [Ricoeur 2000] et de la narration historique.

Ce double impératif passe selon nous par deux lexiques complémentaires. Et cette étude a montré que la portion respective de ces lexiques variait au fil des ans pour décrire une historicisation originale, ici de la Shoah ; cette historicisation se repère ainsi au cœur même des récits de vie.

D'abord sensoriel, et par là sans doute égocentré, le témoignage se désincarne au fil du temps pour devenir plus rationnel, plus collectif. L'intelligibilité de l'événement passe par une rationalisation, et la mémoire sensitive, comme évanescence, laisse la place à une mémoire policée et réélaborée.

Reste alors l'essentiel dans les témoignages de la Fondation pour la Mémoire de la Shoah, toujours changeant mais identique à lui-même au long du corpus, présent dès le départ mais chaque année plus impérieux : la peur de l'oubli – personnel¹⁷ et collectif – et le devoir, sinon la nécessité vitale, de témoigner – pour les camarades morts, pour et à la demande des enfants et petits-enfants, des frères en humanité et des générations nouvelles (fig. 24)¹⁸.

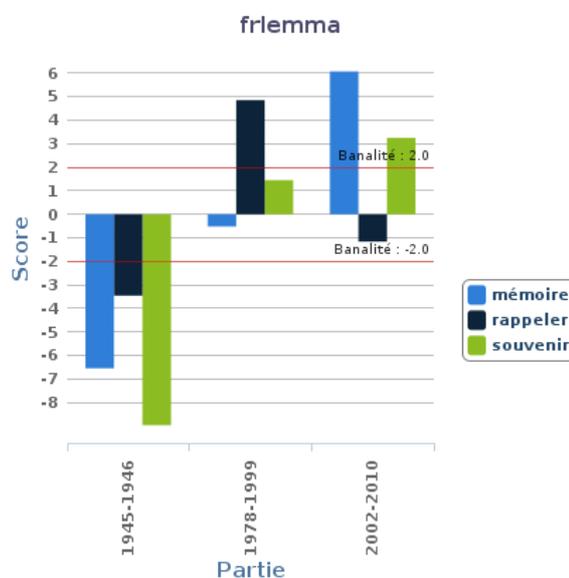


Figure 24. Distribution chronologique des termes de la mémoire

¹⁷ Avec l'âge grandit l'inclination bien naturelle à écarter de sa mémoire les moments les plus atroces ou douloureux de son existence, pour ne se souvenir que des moments heureux. Le recueil de tels témoignages demande ainsi une particulièrement grande délicatesse.

¹⁸ Dans le corpus c'est Simon Grinbaud qui exprime en 1986 le plus vivement l'impératif par le titre de son témoignage : *XIe commandement* : « Tu n'oublieras point »

Références

- Heiden S., Magué J.-Ph., Pincemin B. (2010). « TXM : Une plateforme logicielle open-source pour la textométrie – conception et développement », in Sergio Bolasco, Isabella Chiari, Luca Giuliano (eds), *JADT 2010*, Edizioni Universitarie di Lettere Economia Diritto, p. 1021-1031, http://lexicometrica.univ-paris3.fr/jadt/jadt2010/allegati/JADT-2010-1021-1032_025-Heiden.pdf.
- Joutard Ph. (2013). *Histoire et mémoires, conflits et alliance*. Paris : La Découverte.
- Lafon P. (1980). « Sur la variabilité de la fréquence des formes dans un corpus », *Mots*, 1, p. 127-165, http://www.persee.fr/web/revues/home/prescript/article/mots_0243-6450_1980_num_1_1_1008.
- Lebart L., Salem A. (1994). *Statistique Textuelle*. Paris : Dunod <http://ses-perso.telecom-paristech.fr/lebart/ST.html>.
- Mayaffre D. (2006). « Faut-il prendre en compte la composition grammaticale des textes dans le calcul des spécificités lexicales ? Tests logométriques appliqués au discours présidentiel sous la V^{ème} République », in J.-M. Viprey (éd.), *JADT'06*, Besançon : PU de Franche-Comté, p. 677-685, <http://lexicometrica.univ-paris3.fr/jadt/jadt2006/PDF/II-060.pdf>.
- Mayaffre D. et Ben Hamed M. (2014). « Récits de mort et souvenir traumatique. Trames et traces lexicales des témoignages sur la Shoah », *Argumentation et Analyse du Discours*, 13, <http://aad.revues.org/1836>.
- Metwally H. (2017). *Analyse thématique du Monde diplomatique. Parcours logométriques et traitement logiciel d'un grand corpus*. Thèse de doctorat – Université de Nice.
- Peschanski, D. (éd.). (2013). *Mémoire et mémorialisation – Volume 1 : De l'absence à la représentation*. Paris : Hermann
- Ratinaud P., Déjean S. (2009). « IRaMuTeQ : implémentation de la méthode ALCESTE d'analyse de texte dans un logiciel libre », Colloque *Modélisation Appliquée aux Sciences Humaines et Sociales* (MASHS2009), Toulouse, 10 juin 2009, http://reperer.no-ip.org/Members/pratinaud/mes-documents/articles-et-presentations/presentation_mashs2009.pdf/view.
- Ratinaud P., Marchand P. (2012). « Application de la méthode ALCESTE à de "gros" corpus et stabilité des "mondes lexicaux" : analyse du "CableGate" avec IRaMuTeQ ». In Anne Dister, Dominique Longrée, Gérald Purnelle (éds), *Actes des 11eme Journées internationales d'Analyse statistique des Données Textuelles* (JADT 2012), Liège, 13-15 juin 2012, 835-844, <http://lexicometrica.univ-paris3.fr/jadt/jadt2012/Communications/Ratinaud,%20Pierre%20et%20al.%20-%20Application%20de%20la%20methode%20Alceste.pdf>.
- Reinert M. (1993). « Les "mondes lexicaux" et leur "logique" à travers l'analyse statistique d'un corpus de récits de cauchemars », *Langage et société*, 66, 5-39.
- Ricoeur P. (2000 rééd.). *La mémoire, l'histoire, l'oubli*. Paris : Seuil.