



HAL
open science

L'Histoire mot à mot

Damon Mayaffre

► **To cite this version:**

Damon Mayaffre. L'Histoire mot à mot : Les mots des migrations dans le discours politique des années trente. Cahiers de la Méditerranée, 1997, 54, pp.159-171. hal-00916730

HAL Id: hal-00916730

<https://hal.science/hal-00916730>

Submitted on 10 Dec 2013

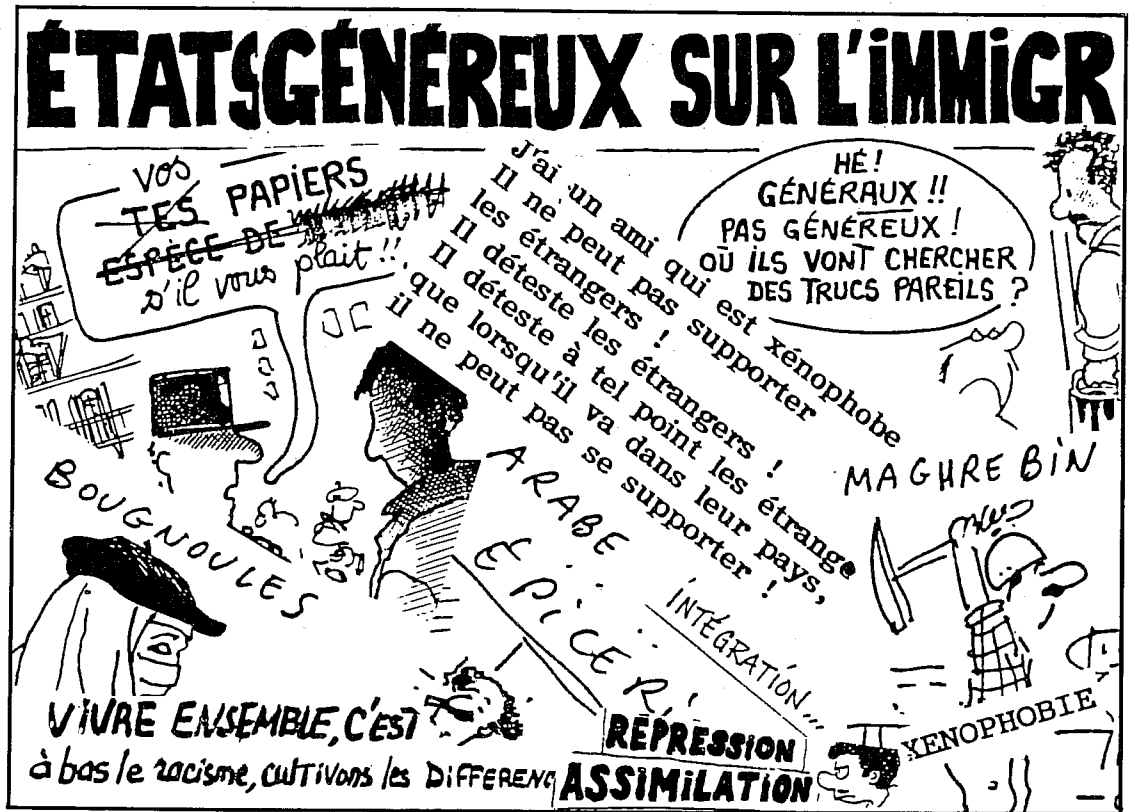
HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

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.

M Cahiers de la MEDITERRANEE

Fac-similé

MOTS ET MIGRATIONS



L'HISTOIRE MOT À MOT. LES MOTS DES MIGRATIONS DANS LE DISCOURS POLITIQUE TRADITIONNEL DES ANNÉES 30.

Fac-similé

CONTRIBUTION MÉTHODOLOGIQUE.

Damon MAYAFFRE
C.M.M.C. Université de Nice.

L'historien a toujours affaire aux mots. Sa discipline commence précisément lorsqu'apparaissent l'écriture, le récit, l'archive. Aussi est-il à jamais condamné à reconstituer les événements passés grâce à des vecteurs qu'il lui faut savoir déchiffrer et une matière linguistique dont l'analyse sort de ses compétences immédiates.

La relation originelle de l'histoire au langage ou à l'écriture prend un tour particulier lorsque l'objet historique étudié est lui-même les mots ; les mots du discours, les mots des journaux, les mots des romans...

Et si, en analysant de front le vocabulaire, le problème de la distorsion entre objet d'étude et archive, qui tout à coup se confondent, est éludé, c'est pour faire surgir plus cruellement encore le rapport entre Histoire et Linguistique dans la recherche d'une herméneutique scientifique.

Le développement de l'analyse du discours, des études de presse et de l'opinion publique insiste particulièrement sur l'importance des mots dans le champ social et politique. Si les mots médiatisent l'histoire pour l'historien, en amont, ils médiatisent la réalité pour les contemporains.

Dans l'imaginaire collectif le signe construit parfois le réfèrent. Nommer c'est souvent édifier, dénoncer souvent abolir ; souvent, "*dire c'est faire*"¹. Cette "performativité" du langage, l'importance que prend le mot dans la construction sociale font du vocabulaire un objet riche et complexe pour l'historien qui l'oblige à réfléchir sur une méthode pour le traiter.

Une lecture intuitive ou "*une lecture sauvage*"² des textes peut-elle tenir lieu de méthode scientifique ? Ou faut-il, pour garantir des résultats, une lecture contrôlée ?

1 J. L. Austin invente les "performatifs" : *How to do things with Words*, 1962.

Trad. française, *Quand dire, c'est faire*, Ed du Seuil, 1970.

2 L. Bardin, *Analyse de contenu*, PUF, 1977.

I - LA LEXICOLOGIE QUANTITATIVE : PRINCIPES ET RÉFLEXIONS

La lexicologie quantitative - appelée aussi, de manière trop réductrice pour ce qu'elle implique aujourd'hui, lexicométrie- assistée par ordinateur traite le texte et son vocabulaire de manière systématique en associant à chacun des mots d'un corpus, grâce à une indexation et une statistique lexicale de plus en plus performantes, plusieurs valeurs chiffrées (fréquence absolue, fréquence relative, écart réduit, etc.) qui correspondent au nombre de ses occurrences dans le corpus, la distribution de ces occurrences à l'intérieur d'un corpus fractionné, l'écart par rapport à une norme etc.

En utilisant cette méthode d'analyse du texte, le chercheur veut montrer que les conclusions qu'il pourra tirer sur l'importance d'un thème dans le corpus sont légitimes, et lorsqu'il voudra illustrer ses propos par des citations, il pourra affirmer le caractère représentatif des ces dernières³. Et ceci est remarquable lorsque le corpus embrassé est important. Dans des macro-corpus qui représentent parfois des centaines de livres de poche, seuls l'ordinateur et des logiciels appropriés permettent une description précise, objective, nuancée du vocabulaire sur laquelle l'historien pourra réfléchir.

Depuis le début des années 1960, sous l'impulsion des statisticiens du vocabulaire, P. Guiraud, puis Ch. Muller, et les travaux pionniers du centre lexicométrique de l'ENS de Saint-Cloud constitué en 1965 autour de R.-L. Wagner, la méthode lexicométrique, sa critique, tout comme ses techniques sont connues et font l'objet d'une littérature abondante⁴. Les présenter dans le détail nous amènerait trop loin. Néanmoins, il convient de revenir sur trois questions qui sont au fondement de la lexicométrie, et qui posé, en arrière plan, la question de l'usage du mot en Histoire.

Premièrement -et même si la réponse semble aller de soi- l'analyse lexicale est-elle recevable ? Est-il légitime d'aborder les discours ou les articles de presse *via* leur lexique, ou d'une manière plus réductrice encore, par les "mots"⁵ qui les composent ?

³ "Dès avant la guerre de 14, (François Simiand) avait dénoncé l'usage incontrôlé de la citation comme preuve, en lui opposant comme seul principe scientifique légitime, la preuve statistique, l'histoire quantitative...", G. Noiriél, préface de S. Bonnafous, *L'immigration prise aux mots*, Kimé, 1991.

⁴ Voir particulièrement les premiers numéros de la revue *Mots/ Ordinateur/ Textes*, publiée par le Laboratoire de lexicométrie et textes politiques, INALF - ENS de Fontenay- Saint-Cloud. Voir aussi les ouvrages et articles de R. Robin, D. Peschanski, A. Prost, M. Tournier, etc. L'aspect technique (statistique et informatique) sera étudié dans Ch. Muller, É. Brunet, P. Lafon, A. Salem.

⁵ Sur la distinction entre "unité lexicale", "vocabulaire", et "mot" cf. avant tout les travaux de Ch. Muller : "Le mot, unité de texte et unité de lexique", *Travaux de linguistique et de littérature*, Fac. de Stasbourg, Klincksieck, 1963, T1, pp.165-175; *Principes et méthodes de statistique lexicale*, ch. 3 et 4 "La norme

En partie contestable linguistiquement -les analyses structurales s'intéressent plus à la syntaxe et à la distribution des mots qu'aux mots eux-mêmes- l'étude du discours par son vocabulaire semble incontournable pour l'historien, à la recherche du sens des textes. Le vocable peut être considéré comme l'atome de signification du discours, même si une linguistique nucléaire distingue encore le sémantème et le sème. Toute idée passe, en effet, par l'énonciation de vocables, tout thème développe un réseau sémantique tissé de mots clefs.

Si aborder un discours par ces mots paraît donc naturel, cette approche devient objective et rigoureuse lorsqu'elle répond au critère d'exhaustivité.

"Pour la lexicométrie, l'objectivité commence quand les projections tendent à faire place à des formalisations. On ne fixe d'avance ni mots-témoins, ni grilles, ni schèmes, ni hiérarchies. On prend tout (exhaustivité des relevés), de la même manière (uniformité du dépouillement) et selon un critère unique (invariance d'une unité de base anonyme)"⁶.

Cette exigence d'exhaustivité garantit la neutralité de l'analyste et la scientificité de l'analyse. Le chercheur n'impose pas ses choix, ses *a priori*, il ne dicte pas l'analyse. Tous les mots du texte sont traités de manière identique par l'ordinateur, selon leur forme graphique et les mêmes critères quantitatifs; et seules entrent en ligne de compte dans un premier temps ces valorisations quantitatives.

Deuxièmement, les critères quantitatifs sont-ils pertinents ? Si l'on a pu dire qu'un locuteur a besoin de certains mots pour exprimer certaines idées, on peut penser que la multiplication d'emploi d'un mot est révélatrice du développement d'un thème donné dans le discours. Antoine Prost, après une pratique fructueuse d'analyse quantitative du vocabulaire, et après la compilation des travaux lexicométriques en Histoire, n'hésite pas à l'affirmer.

"La fréquence des termes, des expressions... paraît un indicateur sûr de leur importance objective " ou encore "les fréquences lexicales constitueraient ... un guide fiable dans l'exploration des textes." ⁷

Ces affirmations, sans doute recevables pour la plupart des productions langagières, sont particulièrement valables pour le discours politique dans son acception la plus élargie. Les textes politiques exigent une lisibilité, une clarté, une efficacité, une performance, un aspect persuasif qui forcent le locuteur

lexicologique. Le mot." et *"La norme lexicologique. Le vocable."* Champion, rééd. 1992. Le mot n'a pas de définition linguistique précise. Il est défini simplement comme une unité graphique (suite de lettres comprises entre deux blancs ou/et deux ponctuations).

⁶ M. Demonet, A. Geffroy, P. Lafon, M. Tournier...*Des tracts en Mai 68*, Champ Libre, 1978, p. 21.

⁷ A. Prost, *"Les mots"*, dans R. Rémond (s.d.), *Pour une histoire politique*, Seuil, 1988 p. 259.

politique aux répétitions lexicales. Les répétitions et les redondances, que l'on sait être au fondement du discours pédagogique ou du message publicitaire, semblent l'être du discours politique. Celui-ci, en effet, est un discours prosélyte, un discours de propagande. Il a besoin de marteler certaines idées et certains mots pour être facilement reconnu, retenu et reproduit, pour remporter l'adhésion de tous et inciter à l'action.

Il faut cependant rajouter que les critères quantitatifs deviennent réellement efficaces lorsqu'ils permettent des comparaisons. Pas toujours parlante en soi, la fréquence d'un terme devient signifiante lorsqu'on la compare à d'autres fréquences; celle d'un autre mot pour le même corpus, celle du même mot dans un corpus similaire mais différent ou à l'intérieur d'une sous-partie du corpus si celui-ci est subdivisé. Cela exige d'une part que d'autres études soient réalisées (et la base *Frantext* de laquelle a été issu le *Trésor de la Langue Française* ⁸, peut offrir une norme et un corpus de référence), et d'autre part que le corpus étudié soit assez vaste, diachronique et contrastif pour fournir des possibilités de comparaisons internes; comparaisons chronologiques, c'est une nécessité pour l'historien, ou comparaisons qualitatives entre plusieurs locuteurs.

Enfin notons que la description quantitative du vocabulaire gagne encore en efficacité depuis qu'elle jouit d'une finesse et d'un raffinement remarquables. La lexicométrie ne révèle plus seulement, comme à ses balbutiements, la valeur absolue des occurrences des mots du corpus dans un index alphabétique ou hiérarchique, elle indique aussi grâce à un traitement statistique poussé à son terme, le poids relatif des mots par rapport à l'ensemble des occurrences du corpus, leur répartition ou leur distribution dans un corpus divisé en sous-parties, les spécificités lexicales des sous-parties en question, etc., jusqu'aux possibilités d'analyses factorielles des correspondances.

Possibilités de comparaisons internes et externes au corpus, raffinement de la statistique lexicale, l'objection de Roland Barthes, qui pouvait jeter le discrédit sur une lexicométrie primitive et sur l'approche quantitative du vocabulaire ne tient plus. En effet, s'il s'avère patent, même dans un discours politique, c'est à dire un discours de propagande, que

"le sens ne naît point par répétition mais par différence, de sorte qu'un terme rare dès lors qu'il est saisi dans un

⁸ La base *Frantext* est la plus formidable base de données de textes français. Elle compte quelques 3000 oeuvres et 100 millions d'occurrences. Sa constitution a exigé un quart de siècle de travail. Le *Trésor de la Langue Française*, dont le projet remonte à 1957 et qui vient d'aboutir à la publication du 16^{ème} et dernier volume, est un Dictionnaire de la langue du XIX^º et du XX^º siècle. Il a été constitué sous la direction de P.Imbs et B Quemada dans le cadre du CNRS par l'Institut National de la Langue Française à partir du corpus de *Frantext*.

système d'exclusion et de relation signifie tout autant qu'un terme fréquent ⁹.

alors la lexicométrie mesurera cette "différence", ce "système d'exclusion" grâce à une statistique lexicale appropriée (spécificités chronologiques d'un terme, comparaison avec la norme du corpus...) et grâce à des mises en relief comparatives.

Troisième question -la plus grave-, l'approche désincarnée et lexicale, comme les critères quantitatifs qui paraissent donc nécessaires sont-ils pour autant suffisants, notamment lorsque surgit la nécessité d'interpréter les chiffres et les données et de "comprendre" le texte ? Derrière cette question se profilent les limites de la lexicométrie. Il convient de les rappeler sans pour autant les exagérer. Les lexicologues le savent pertinemment, le mot -plus encore que le vocable- n'a de valeur que dans et par la phrase qui le contient.

La polysémie des mots, les phrases négatives ou dénégatives, les formules antiphrastiques, le discours rapporté, etc. montrent la nécessité du retour constant au texte et au contexte. Les phrases caricaturales qui suivent, affichent toutes une occurrence d'"immigration":

"Je suis pour l'immigration". "Je suis contre l'immigration". "Il n'est pas vrai que je suis contre l'immigration". "Il paraît que je suis pour l'immigration". "Mon adversaire dit : "je suis contre l'immigration"."

Pourtant on ne peut conclure à l'élan xénophile du locuteur que dans la première des propositions, et encore, faut-il être sûr que cette formulation n'est pas ironique. Toute conclusion passe donc par une étude plus ou moins élargie du contexte; étude de la phrase -ceci est rarement suffisant-, étude du paragraphe, étude du discours, étude, parfois, de manière générale, de l'idéologie du locuteur et des circonstances historiques dans lesquelles le mot et le discours ont été prononcés. Les chercheurs en lexicométrie, conscients des limites de l'approche paradigmatique qu'ils faisaient de l'énoncé, n'ont pas renoncé à contrôler cette mise en contexte. Et à l'étude systématique du vocabulaire se combine aujourd'hui une lexicométrie syntagmatique qui fait ressortir, d'une manière tout aussi systématique, le contexte immédiat des mots.

Le logiciel d'Étienne Brunet, *Hyperbase* dont nous voulons ici illustrer les possibilités, permet, parmi d'autres choses, l'analyse méthodique des "concordances" (environnement syntagmatique proche d'un terme) et du "contexte" (environnement syntagmatique élargi).

Aussi fait-il apparaître, à volonté, sous forme de listes facilement consultables, l'ensemble des "lignes" (concordance) ou des paragraphes (contexte) qui comportent un mot donné, permettant ainsi de vérifier le sens du mot et la signification de son emploi. D'une manière plus générale *Hyperbase*

⁹ Roland Barthes cité par D. Labbé, *Le discours communiste*, Presse de la FNSP, 1977.

permet aussi un dialogue incessant entre les mots présentés en index et le texte dans sa globalité et sa continuité linéaire¹⁰.

Pour cette raison, par ce dépassement, nous préférons le terme de lexicologie quantitative, à celui, plus réducteur, de lexicométrie. La lexicologie quantitative permet avant tout une description contrôlée et chiffrée du vocabulaire du texte par la mise en exergue de critères quantitatifs; mais elle ne se contente pas de ces critères. Elle n'exclut pas -et demande au contraire- un retour au texte et au contexte, elle implique une réflexion syntagmatique. Elle nécessite de dialoguer avec l'ensemble du discours.

L'approche du texte est certes avant tout quantitative, mais elle ouvre la voie à l'analyse qualitative, implique une "*science du lexique*", une réflexion sur l'utilisation des mots, un discours sur le discours et son vocabulaire.

II - LES MOTS DES MIGRATIONS CHEZ THOREZ, BLUM, TARDIEU ET FLANDIN DANS LES ANNÉES 30.

Fort de ces réflexions, nous nous proposons de faire connaissance avec *Hyperbase* qui est sur Macintosh sans doute le logiciel de lexicométrie le plus élaboré et le plus convivial. Ces quelques premiers pas devront inciter les chercheurs à faire plus ample connaissance avec le logiciel, et non pas masquer l'ensemble de ses possibilités.

La démonstration portera sur le corpus que nous traitons dans le cadre d'un doctorat d'histoire. Ce corpus compte quelques 1.600.000 mots ou occurrences (une cinquantaine de livre de poche) et regroupe la plupart des discours et articles que Thorez, Blum, Flandin et Tardieu ont prononcés ou écrits entre 1929 et 1939.

A - L'Index des formes :

A la base de toute démarche lexicométrique il y a l'index des formes. C'est-à-dire le catalogue ou le dictionnaire de l'ensemble des mots du discours avec leur nombre d'occurrences, auxquelles *Hyperbase* ajoute la distribution de celles-ci dans les différentes sous-parties du corpus.

¹⁰ Dans le même esprit, le logiciel pionnier en lexicométrie, *Lexico1*, mis au point par le laboratoire de lexicologie de Saint-Cloud, analyse non seulement les mots mais aussi les "*segments répétés*" (suites de plusieurs mots (allant parfois jusqu'à la phrase) qui apparaissent plusieurs fois dans un corpus) et dépasse ainsi lui aussi les limites paradigmatiques de la lexicométrie de la première génération. Les inconvénients majeurs de *lexico 1* sont : les limites de sa capacité de traitement (corpus limité à 700.000 mots) et l'absence d'interface active qui interdit le dialogue entre le texte et le traitement statistique.

Cet index peut être présenté sous une forme alphabétique ou sous une forme hiérarchique, du mot le plus employé jusqu'aux hapax (les mots employés une seul fois.) Quel chercheur n'a pas, à défaut de pouvoir lire un ouvrage dans sa totalité, consulté l'index des noms propres pour être renvoyé directement à des thèmes particuliers ? Cette pratique est ici "tout simplement" systématisée.

C'est l'ensemble des mots du corpus qui est indexé, et ce critère d'exhaustivité n'est pas anecdotique, il est, cela a été dit, gage de la scientificité de la démarche. Il n'y a pas de choix *a priori* du chercheur qui imposerait sa lecture et ses centres d'intérêt, tous les mots sont, dans un premier temps, pris en compte.

L'index constitué par la machine, il reste à le consulter. Dans notre corpus, les entrées des mots traitant directement des migrations ont été consignées dans le Tableau 1.

Tableau 1 : Nombres d'occurrences des mots ayant trait aux migrations dans le corpus

	Thorez	Blum	Flandin	Tardieu	Total
migration	0	4	1	0	5
migrations	0	2	1	0	3
immigration	0	0	2	3	6
immigrés	21	0	1	0	22
immigrée	5	0	0	0	5
immigrants	0	0	0	2	2
émigration	2	6	0	2	10
émigrés	17	4	2	2	25
émigrants	0	0	0	2	2
Total	45	16	7	12	79

Pour être grossier, ce constat, qui ne repose que sur des valeurs absolues et non sur un traitement statistique, est néanmoins déjà intéressant. 79 mots traitant directement du phénomène migratoire sur les 1.570.868 mots de notre corpus, le nombre est dérisoire. Et dès lors deux conclusions sont possibles :

- le thème est pratiquement absent des discours de la représentation politique classique ou parlementaire dans les années trente.

- le thème est développé avec d'autres mots qu'il devient intéressant de chercher.

Une étude plus poussée, notamment grâce aux mots "*étranger(e)(s)*" nous invite plutôt à conclure en faveur de la première proposition. Pourtant, nous savons avec les études de Ralph Schor que le problème de l'immigration

est important dans la société française des années trente¹¹. Il apparaît donc que ce thème est en grande partie confisqué par l'extrême droite¹², par la presse et peut être dans une moindre mesure par le discours syndical. En tout cas les représentants des grands partis parlementaires s'en privent.

Au regard du nombre d'occurrences des mots "*immigrés*" ou "*immigrée*" nous pouvons même parler de véritable tabou lexical chez Blum (2 occurrences en 10 ans), chez Flandin ou chez Tardieu (aucune occurrence !). Ces trois personnages parmi les plus influents de l'entre-deux-guerres semblent délibérément refuser de porter ces mots à leur bouche durant les années trente. Sans aucun doute cela nous invite à réfléchir, en période de crise, sur la distorsion entre les préoccupations majeures des Français (ou en tout cas les préoccupations telles qu'elles peuvent apparaître dans la presse), et celles de la classe politique dirigeante. Dans les années trente la presse résonne de propos sur l'immigration ; Blum, Flandin et Tardieu, restent muets.

B - Concordance et Contexte :

Malgré la rareté des termes liés aux migrations dans le corpus, il est intéressant d'étudier leurs conditions d'utilisation. Plusieurs possibilités s'offrent au chercheur avec *Hyperbase*. D'abord, il suffit de "cliquer" sur les mots de l'index alphabétique pour voir défiler les portions des discours dans lesquels ils se trouvent. Et pour faciliter la lecture, le mot est alors encadré à l'écran. Mais plus rigoureuse est l'étude systématique des "*concordances*" et du "*contexte*". En effet pour un mot donné il est possible de demander à *Hyperbase* de faire apparaître l'ensemble des lignes ou des paragraphes dans lesquels il figure.

Par exemple l'étude des concordances des 21 occurrences d'"*immigrés*" chez Thorez donne :

Forme(s) recherchée(s) : "*immigrés*".

Partie(s) du corpus étudié(s) : Thorez.

= 21 occurrences

1 ... *mais si les travailleurs français et immigrés, surtout à la veille du chômage...*

2 ...*des milliers d'ouvriers français et immigrés sont condamnés au chômage partiel ...*

11 Voir par exemple, la recrudescence des articles consacrés par la presse à la question des étrangers dans l'économie française en 1930, 1931, 1932, dans R. Schor, *Histoire de l'immigration en France*, 1996, A. Colin, tableau, p. 122.

12 Cf R. Schor, "L'extrême-droite française et les immigrés en temps de crise : années trente - années quatre vingt", *REMI*, 1996, vol. 12, p 241-261.

- 3 ...des ouvriers français et des ouvriers immigrés, subissant la même misère, courbés...
- 4 ...les luttes des ouvriers français et immigrés en soutenant la bataille de nos frères...
- 5 ...persécution contre les ouvriers immigrés, etc.
- 6 ...sociales, les ouvriers français et immigrés portent un coup sérieux au patronat et...
- 7 ...les jeunes et surtout les ouvriers immigrés. La conférence du Parti se...
- 8 ...l'égalité absolue entre les ouvriers immigrés et les ouvriers français,...
- 9 ...union étroite des ouvriers français et immigrés et, en outre, gros danger de voir se...
- 10 ... convaincre , une partie des ouvriers immigrés pourraient se trouver devant nous,...
- 11 ...l'égalité absolue entre les ouvriers immigrés et français, salaires identiques,...
- 12 ...prolétaires de la métropole français et immigrés et tente de briser, par les moyens...
- 13 ... jeunes , des femmes et des ouvriers immigrés. En d'autres régions on se trouvera...
- 14 ...(1100000 salariés dont 180000 immigrés). Les 180000 mineurs, les 200000...
- 15 ...le refoulement des travailleurs immigrés en poursuivant avec l'aide des...
- 16 ...du jour, jeunes et vieux. Français et immigrés, ont répondu à l'appel des militants...
- 17 ...jeunes, soldats, femmes, ouvriers immigrés et coloniaux. Un effort particulier...
- 18 ...les jeunes, les soldats, les ouvriers immigrés, tous ceux qui ne votent pas, il y a...
- 19 ...anciens combattants, jeunes, femmes, immigrés, coloniaux à la victoire contre le...
- 20 ...et les marins, pour les ouvriers immigrés et coloniaux. Puis, nous proposons...
- 21 ...alors que l'on expulse les travailleurs immigrés coupables de se défendre, aux côtés...

Cette étude des concordances permet une lecture rapide, contrôlée et systématique. Elle ouvre la voix à une analyse qualitative et permet de

déboucher sur une interprétation historique. Sans même aller plus loin nous pourrions ici conclure que les "immigrés" apparaissent toujours, dans le discours communiste des années trente, dans le cadre d'une solidarité de classe entre travailleurs.

Cependant, l'étude des concordances n'offre pas toujours un environnement assez élargi pour tirer des conclusions sûres. Il est alors nécessaire de demander le *contexte* (paragraphe) du mot, quitte à rendre la lecture systématique plus longue. Ainsi le *contexte* de la forme "immigration" chez Tardieu se présente ainsi :

Forme(s) recherchée(s) : immigration

Partie(s) du corpus étudiée(s) : Tardieu

= 3 occurrences

"Problème de l'outillage, problème de la race aussi, dans son sens le plus large. J'entends par là pour un peuple qui n'est plus le nombre, la nécessité de développer au maximum les oeuvres d'hygiène, la nécessité de régulariser, de systématiser une politique d'IMMIGRATION qui a été abandonnée au hasard depuis la guerre."

"Il y avait en 1918, en Palestine, 500000 Juifs. En 1938, on en comptait 500000, venus de quarante-deux pays. Cette IMMIGRATION a transformé la Judée. Le rendement des terres a doublé ou triplé."

"Car, en Palestine, il y avait, en face des Arabes, les Juifs, qui, malgré leur dispersion séculaire, y apportaient un nationalisme aigu; qui entendaient que la promesse sur le Foyer juif fût intégralement tenue; que l'IMMIGRATION ne fût ni arrêtée, ni limitée et qui avaient, pour cela l'appui moral des États-Unis."

Il est confirmé que Tardieu ne s'intéresse pas à l'immigration en France. Une seule des utilisations concerne l'immigration française. Elle date du 3 septembre 1929, c'est-à-dire avant que la crise économique n'éclate aux États-Unis et touche la France, et Tardieu se montre plutôt favorable à une politique en faveur de l'immigration qu'il faut certes contrôler mais systématiser.

Évidemment il est possible de pousser l'analyse à l'ensemble du corpus et de ne pas se contenter, comme nous l'avons fait, d'un seul locuteur (Thorez ou Tardieu). Il est même particulièrement intéressant d'établir une comparaison sur l'utilisation d'un même mot chez deux, trois ou quatre hommes politiques d'horizons différents.

Et les possibilités du programme des concordances et du contexte ne s'arrêtent pas là ! Il permet aussi de réaliser des études non pas seulement sur le mot mais aussi sur le vocable¹³. En demandant à *Hyperbase* d'étudier les

13 A condition que les flexions du vocable aient la même racine.

contextes du vocable "immigré", il sortira l'ensemble des phrases comportant "immigré", "immigrés", "immigrées" et "immigrées". Plus globalement encore le programme permet d'étudier des chaînes de caractères. En demandant à *Hyperbase* d'étudier la chaîne 'migr' c'est le contexte de l'ensemble des mots qui nous intéresse qui sera traité (imMIGRation, éMIGRès, MIGRations...).¹⁴

Enfin, suprême raffinement, le programme des contextes étudie les co-occurrences des mots. Si le chercheur pressent l'imbrication de deux thèmes, il peut demander la recherche systématique de tous les paragraphes dans lesquels deux mots ont une présence simultanée.

Chez Thorez "immigrés" est-il co-occurent de "chômage" ? *Hyperbase* répond de manière systématique :

Forme(s) recherchée(s) : "immigrés" croisé avec "chômage"

Partie du corpus : Thorez

Occurrences : 2

"Or plus que jamais nous avons besoin en France de syndicats puissants. groupant la masse des ouvriers industriels les plus mal payés et rassemblant sans distinction les travailleurs français et IMMIGRÉS, surtout à la veille du CHOMAGE."

"Déjà vingt millions d'ouvriers chôment dans les grands pays capitalistes. Et, alors que Tardieu nie toujours l'existence du CHOMAGE en France, des milliers et des dizaines de milliers d'ouvriers français et IMMIGRÉS sont condamnés au chômage partiel."

III - DISTRIBUTION ET SPÉCIFICITÉ CHRONOLOGIQUES :

Un corpus gagne à être contrastif, c'est à dire regroupant des locuteurs différents (ici Thorez, Blum, Flandin et Tardieu) afin de nourrir l'analyse de comparaisons qui donnent souvent sens aux chiffres. Pour l'historien il est aussi avantageux qu'il soit diachronique c'est-à-dire divisé en plusieurs tranches chronologiques (ici 1929, 1930, 1931...1939).

Ainsi, de même qu'il est possible pour un mot donné d'étudier sa distribution entre les orateurs, il est possible de voir comment ses occurrences sont distribuées chronologiquement. Pour l'historien cela est fondamental. Cela lui permet de distinguer les termes dont l'usage est constant et ceux dont la fréquence d'usage évolue (disparition progressive ou brutale d'un terme, a

¹⁴ *Hyperbase* permet ainsi de réaliser au cas par cas un début de lemmatisation. Pour les vocables dont les flexions n'ont pas de racine unique, il est possible d'étudier alors non plus une chaîne, un vocable ou un mot mais une "liste" de mots. Ainsi en établissant la liste "allé", "vient", "irons"... on peut étudier l'utilisation du verbe "aller". Ce système de liste, permet d'étudier des champs sémantiques : "immigrés" + "étrangers" + "Italiens", ...

contrario apparition et progression) Cela permet de renouer avec le premier geste de la démarche historique, de démêler dans le vocabulaire, ce qui relève du structurel et du conjoncturel, du permanent et de l'évolutif; de repérer, selon les termes de Jacques Le Goff "*la dialectique de la continuité et du changement*"¹⁵. Cela permet surtout d'identifier et de dater les grands tournants discursifs, les changements de discours et de politique des locuteurs.

En valeur absolue, le terme "*immigrés*" par exemple se distribue dans le corpus Thorez ainsi :

	1930	1931	1932	1933	1934	1935	1936	1937	1938	1939
immigrés	6	7	2	1	5	0	0	0	0	0

Tableau 2 : distribution chronologique du terme "*immigrés*" chez Thorez.

Le tableau est limpide. Il n'y a pas de préoccupation constante, mais une préoccupation conjoncturelle. Le thème de la solidarité de classe avec les immigrés appartient à la première partie des années trente, celle du "*front unique à la base*". Avec la politique de Front populaire le thème disparaît complètement. La problématique politique du Parti communiste se déplace¹⁶ et dans le cadre d'un discours devenu patriotique, les "*immigrés*" n'avaient plus, momentanément, leur place.

Ici et pour conclure, une idée fondamentale doit être soulevée. Les constatations sur des fréquences absolues sont trop aléatoires (elle dépendent notamment de la longueur du corpus et des sous-parties chronologiques) pour être vraiment significatives et le traitement lexicométrique n'aurait guère de sens s'il s'en contentait. La lexicométrie implique une statistique lexicale, dont nous avons la place d'évoquer que le principe.

L'emploi d'un mot peut être qualifié d'important ou de faible uniquement par rapport à une norme. Cette norme est une norme endogène, elle est constituée par le corpus lui même dans son ensemble. Ainsi si le corpus est segmenté (politiquement ou chronologiquement) on pourra comparer l'utilisation du mot dans une sous-partie par rapport à l'utilisation "moyenne" ou générale dans le corpus. Il s'en suit une série de calculs qui aboutissent à l'établissement d'un écart réduit¹⁷ et qui vise à mesurer le degré de "*spécificité*" d'un terme dans une sous-partie.

15 J. Le Goff, *Saint Louis*, Gallimard, 1996, p. 14.

16 Cf à ce propos la plus fructueuse étude historique avec une méthode lexicométrique: D. Peschanski, *Et pourtant ils tournent. Vocabulaire et stratégie du PCF (1934-36)*, Klincksieck, 1988.

17 L'écart réduit que l'on note z est l'écart absolu divisé par l'écart type. Pour son calcul voir Ch. Muller, *Initiation aux méthodes de la statistique linguistique*, Hachette, 1973, p 68 et ss.

Ce calcul se fait en prenant en compte la fréquence globale du terme (21 occurrences d'immigrés), la sous fréquence du terme dans une partie (par exemple 7 occurrences en 1931), la longueur globale du corpus (Thorez = 394.661 occurrences) et la longueur de la sous-partie (Thorez en 1931 = 40.991 occurrences). D'autre part ce calcul prend en compte la part de hasard dans la distribution et mesure la probabilité que telle ou telle fréquence apparaisse dans telle ou telle partie : l'écart par rapport à un tirage aléatoire.

Pour notre exemple "*immigrés*" apparaît statistiquement chez Thorez comme une spécificité chronologique de 1931. Son écart réduit par rapport au corpus s'élève à +3,4¹⁸.

Lorsqu'un corpus est segmenté *Hyperbase* automatise cette démarche et l'effectue pour tous les mots. Le chercheur peut alors convoquer un mot pour voir dans quelle partie il se trouve "*sur-employé*". Il peut aussi consulter une partie (par exemple 1930) et demander à *Hyperbase* l'ensemble des mots spécifiques de cette partie. L'ordinateur indique alors tous les mots qui qualifient statistiquement l'année 1930 par rapport aux autres années. C'est alors tout l'univers lexical de l'année, dans ce qu'il a de spécifique, qui apparaît, aux yeux de l'historien, sur l'écran. Selon une formule consacrée, la lexicologie apparaît avoir deux valeurs : une probatoire, l'autre heuristique.

Sa valeur probatoire est celle que nous venons avant tout d'illustrer. Le chercheur arrive avec des questions précises (est-ce que le thème de l'immigration est important dans le discours politique traditionnel des années trente ? Le mot de "*migration*" fait-il partie de l'univers lexical de l'époque, le thème de l'immigration est-il constant chez Thorez ou limité à une certaine période ?) et l'ordinateur donne réponse et administre la preuve.

Mais la lexicométrie a surtout une valeur heuristique. Lorsque le chercheur a fini d'interroger les chiffres, ce sont souvent les chiffres qui à leur tour l'interrogent. La lexicométrie, en proposant une lecture différente du texte, lecture désincarnée et non linéaire, lecture paradigmatique et non syntagmatique, permet de voir autre chose, ou de voir différemment.

Un sens caché ou inconnu apparaît alors parfois sous la forme de questionnement. Pourquoi la fréquence de tel mot est-elle si importante, pourquoi tel autre mot disparaît-il, au fil des années du corpus, après avoir été si présent, pourquoi "*immigrés*" est-il spécifique de l'année 1931 chez Thorez? Mille et une questions apparaissent en effet devant un index hiérarchique, des spécificités chronologiques ou une analyse factorielle des correspondances, pour stimuler inestimablement la réflexion historique.

Réponse scientifique à des questions *a priori* grâce une lecture contrôlée, interrogation nouvelle grâce à une lecture différente des discours, tels sont les deux espoirs que l'on fonde sur la lexicologie quantitative assistée par ordinateur. Si elle sait ne pas décevoir ces espoirs, comment ne satisferait-elle pas l'historien dans son corps à corps avec les mots.

18 De même *immigrés* est spécifiquement négatif (-3,1) en 1936.