

DOI: <http://dx.doi.org/10.15446/male.n9.54913>

Les nouvelles technologies face à l'historiographie positiviste de l'histoire de France : étude sur le 19^{ème} siècle en français : enjeux et perspectives¹

Las nuevas tecnologías se enfrentan a la historiografía positivista de la historia de Francia: estudio del siglo 19 en francés: retos y perspectivas

New technologies Confront the Positivist Historiography of the History of France: A study of the 19th Century in French: Challenges and Perspectives

Dr. Olivier Bailblé

olivier.bailble@hotmail.fr

Assistant Professor in Chinese Linguistics, Université des langues étrangères de Corée (HUFS), Séoul, Corea del Sur

Dr. Eric Bailblé

eric.bailble@gmail.com

Profesor, certificado d'histoire et de géographie au lycée français de Budapest, Hungría

Recibido: 5 de julio de 2015

Aprobado: 20 de diciembre de 2015

Résumé

Cet article a donné la possibilité de rendre compte d'une analyse lexicale du 19^{ème} siècle français dans l'édition mondiale produite durant les 19^{ème} et 20^{ème} siècles à l'aide d'un nouvel outil lexicométrique nommé « *N-gram-Viewer* » (*visualisateur de fréquence de mots*). Cet instrument numérique en ligne sur l'Internet nous a permis de réaliser une approche assez précise des occurrences lexicales des deux derniers siècles. Ce logiciel qui manipule un corpus de 5 millions de livres numérisés est à la fois linguistique et historique. Il offre en effet des possibilités nouvelles pour apprécier et expliquer l'histoire du 19^{ème} siècle français entre mythe et réalité.

¹ This paper was supported by Hankuk University of Foreign Studies research fund of 2015.

La problématique centrale de cet article a donc été la suivante : dans quelle mesure peut-on dire que les nouvelles technologies linguistiques permettent d’obtenir un autre regard sur l’histoire dite traditionnelle d’un pays ? Et plus avant encore : dans quelle mesure peut-on dire aussi qu’une analyse structurale macro-lexicale peut débusquer des représentations inédites et constitutives gravitant autour des « Grands Hommes » liés naturellement à l’histoire événementielle ou des positivistes d’un pays ?

Les données obtenues ont permis de construire des courbes d’occurrences lexicales (nommées *N*-grammes) qui nous ont donné accès à un monde numérique jusqu’ici caché. Cet outil nous a permis surtout de voir un double immatériel de la mémoire de l’humanité car il offre l’accès à des traitements algorithmiques accélérés et à des enregistrements massifs d’information connectés à l’édition mondiale.

La science de fouilles de données (data mining), la science de réseaux (les graphes obtenus par l’Internet) et leur mise à jour par des « datas centers », tels ont été les trois paramètres essentiels de cette recherche. Ces trois aspects techniques ont constitué la toile de fond de cette étude qui dispose au final de plus de 500 milliards de mots, soit environ 5 téraoctets. La création numérique de ce méga-texte réalisé par les laboratoires « Google » et l’université d’Harvard en 2010 a donc ouvert des perspectives inédites entre culture et numérique, d’où le néologisme récent de « Culturomique » pour nommer ce microscope d’analyse linguistique au service de l’histoire.

Mots-clés : *représentation, histoire, culturomique, mythes en histoire, N-Gram-Viewer, histoire de France.*

Resumen

El presente artículo dio la posibilidad de dar cuenta de un análisis lexical del siglo XIX francés en la edición mundial francesa producida durante los siglos XIX y XX con la ayuda de una nueva herramienta lexicométrica llamada *N-gram-viewer* (visualizador de frecuencia de palabras). Dicho instrumento virtual en línea permitió realizar una aproximación bastante precisa de las ocurrencias lexicales de los dos últimos siglos. Este programa, que maneja un corpus de 5 millones de libros virtuales, es a la vez lingüístico e histórico. En efecto, ofrece nuevas posibilidades para apreciar y explicar la historia francesa del siglo XIX entre mitos y realidades.

La problemática central de este artículo fue la siguiente: ¿En qué medida se puede afirmar que las nuevas tecnologías lingüísticas permiten obtener otra mirada acerca de la historia llamada “tradicional” de un país? Y más aún: ¿En qué medida se podría afirmar que un análisis estructural macro-lexicoal puede encontrar representaciones inéditas y constitutivas que gravitan alrededor de los “Grandes Hombres” ligados naturalmente a la historia cronológica o positivista de un país?

Los datos obtenidos permitieron construir curvas de ocurrencias lexicales llamadas *N-grammes* que nos dieron acceso a un mundo virtual hasta ahora oculto. Esta herramienta

Revista Electrónica Matices en Lenguas Extranjeras, número 9. ISSN 2011-1177. Páginas 68 – 90.
Universidad Nacional de Colombia - Facultad de Ciencias Humanas - Departamento de Lenguas Extranjeras. Bogotá. <http://revistas.unal.edu.co/index.php/male>

nos permitió sobre todo ver un doble inmaterial de la memoria de la humanidad ya que dio acceso a tratamientos algorítmicos acelerados y a grabaciones masivas de información conectadas a la edición mundial.

La ciencia de búsqueda de datos (*data mining*), la ciencia de redes (grafos obtenidos por Internet) y su actualización por *data centers*, fueron los tres parámetros esenciales de esta investigación. Estos tres aspectos técnicos constituyeron el telón de fondo de este estudio que dispone de más de 500 mil millones de palabras, es decir, cerca de 5 *terabytes*. La creación virtual de este mega-texto, realizada por los laboratorios de Google y la Universidad de Harvard en 2010, ha abierto nuevas perspectivas entre la cultura y lo virtual, de ahí el neologismo reciente de la “*culturomique*” para nombrar este microscópico de análisis lingüístico al servicio de la historia.

Palabras clave: *representación, historia, culturomique, mitos en la historia, N-Gram-Viewer, historia de Francia.*

Abstract

This article presents a lexical analysis of the 19th century French edition produced in the world during the 19th and 20th centuries with a new lexicometric tool called "N-gram-Viewer" (word frequency browser). This online digital instrument allowed to do a fairly accurate approach to the lexical occurrences of the last two centuries. This program, which handles a corpus of 5 million numbered books, is at the same time linguistic and historical. Indeed, it offers new possibilities to appreciate and explain the French story of the XIX century amongst myths and realities.

Actually, the burning issue of this article could be therefore summarized as follows: to what extent can one say that the new linguistic technologies allow to get another look at the so-called traditional history of a country? And further still. To what extent can we also say that a macro-structural lexical analysis can uncover unprecedented and constitutive representations gravitating around the "Great Men", who are naturally linked to the chronological or positivist history of a country?

The data obtained allowed to construct the curves from lexical occurrences (named “*N-grams*”). These curves gave us access to a digital world up here hidden. This tool, thus, gave us the possibility to see an intangible double of the memory of humanity as it offered access to algorithmic and accelerated treatments and massive information records connected to the world edition.

The science of searching of data (*data mining*), the science of networks (the graphs obtained by Internet), and their update by "data centers", are three essential parameters in our research. These three technical aspects formed the backdrop for this study that also has 500 billion words, that is to say, approximately 5 terabytes. The numeric creation of this mega-text performed by the "Google laboratories" and the University of Harvard in 2010, has

Revista Electrónica Matices en Lenguas Extranjeras, número 9. ISSN 2011-1177. Páginas 68 – 90.
Universidad Nacional de Colombia - Facultad de Ciencias Humanas - Departamento de Lenguas Extranjeras. Bogotá. <http://revistas.unal.edu.co/index.php/male>

opened fresh perspectives between culture and numeric aspects, from there we get the recent neologism “culturomics” to crystallize this study. A new linguistic tool connected with History.

Key words: *representations, history, culturomics, cherished myths in history, N-Gram-Viewer, French History*

S’il était possible de condenser l’histoire du 19^{ème} siècle français, quels seraient les items à retenir, quelles seraient les représentations entre mythes et réalités du 19^{ème} siècle français et surtout, pourquoi ont-ils évolué durant les 19^{ème} et 20^{ème} siècles à travers des publications de l’édition mondiale ? Grâce à la parution de nouveaux outils technologiques à l’ère d’Internet, nous avons entrepris de mener une étude, qui pourrait d’ailleurs paraître irréaliste, en utilisant un nouvel outil lexicométrique nommé « N-Gram Viewer »². Il s’agit d’un instrument numérique en ligne sur l’Internet qui permet d’analyser les occurrences lexicales de l’édition mondiale sur un sujet choisi en utilisant des mots clés.

Pour cela, il nous a fallu aborder l’école historique que l’on qualifia longtemps de « positiviste », car elle est constituée d’« objets-histoire » souvent matérialisés par « les Grands Hommes », c’est-à-dire des acteurs dits classiques de l’histoire récit et de l’histoire romanesque. En effet, pour aborder les productions de l’édition mondiale des 19^{ème} et 20^{ème} siècles, l’histoire-populaire met en relief l’état historiographique au niveau mondial et reconnaît l’héritage structurel de Braudel F. cristallisé dans son ouvrage majeur en 3 volumes et intitulé « Civilisation matérielle et capitalisme » publié à partir de 1967.

Mais, ce retour à « l’histoire-récit » n’est pour nous qu’une stratégie pour mieux utiliser notre outil numérique qui reste inédit dans l’univers des historiens, si l’on considère la puissance d’analyse quantitative de cet instrument.

En effet, les noms propres ont pour qualité intrinsèque d’éviter plus facilement l’écueil des homonymes. Un nom propre dispose de sa propre nature, de sa topographie, de sa morphologie. Ces trois traits permettent ainsi de repérer plus facilement les occurrences lexicales dans le programme *N-Gram-Viewer*. Avec deux ou trois noms propres, on peut synthétiser un thème en histoire pour l’intégrer à ce programme informatique. Une fois les courbes d’occurrences lexicales obtenues, nous pouvons alors dégager des hypothèses d’explication, sans que cela nous garantisse que les noms propres sélectionnés soient les seuls éléments d’une histoire pouvant constituer des représentations collectives. C’est ainsi que le programme peut suggérer des pistes d’explication et rien de plus.

² *traduction littérale* : visualisateur de fréquence de mots - <https://books.google.com/ngrams>

Pour illustrer l'histoire des représentations de ce siècle, nous avons choisi les items suivants : Dreyfus, Pasteur, Napoléon III, Waterloo, Jules Ferry, Austerlitz, La Commune, Napoléon 1^{er}, Tour Eiffel, Bérézina et les Frères Lumière. Encore une fois cette liste est un choix empirique et d'autres choix auraient été possibles. Ces noms propres ainsi choisis dans l'univers dit, « positiviste », ne sont donc que des outils pour faire émerger les représentations autour du thème choisi dans cet article et, selon les langues sélectionnées, nous proposerons une lecture thématique des courbes obtenues.

À partir de ce constat, nous nous sommes demandés pourquoi existe-il une croissance durant une année particulière dans telle langue sélectionnée ? Pourquoi une chute brutale une autre année dans une autre langue ? Quelles hypothèses explicatives peut-on apporter pour interpréter des courbes d'occurrences lexicales sur deux siècles et quelles hypothèses explicatives peut-on apporter pour interpréter des courbes d'occurrences lexicales sur les deux siècles ? Pour y répondre, il est impératif tout d'abord de faire une mise au point méthodologique et épistémologique. Ensuite, nous traiterons le concept de ce logiciel informatique, connu tout particulièrement sous l'appellation de la culturomique. Finalement, nous tenterons d'analyser la représentation des courbes lexicales du français de l'anglais américain, de l'anglais britannique, de l'allemand, de l'italien et de l'espagnol.

Tout d'abord, il serait important de réfléchir au cadre méthodologique et épistémologique concernant le logiciel *N-Gram viewer*, mais une première difficulté s'impose. En effet, on peut d'emblée nous reprocher de piocher dans le vieux musée « lavissien » : une approche romantique des récits contés par ces « Grands hommes ».

Cependant, cette critique est de moins en moins pertinente depuis la rupture épistémologique créée par l'américain White H (1990) dans son concept clé de « discours narratif ». Cette théorie littéraire et linguistique a en effet remis en question toutes les écoles historiques au début des années 1990, jusqu'au travail même des historiens. Un doute profond s'est alors installé, mais quelques conclusions en furent tirées. Selon White H. (ibid), tout discours ou écrit en histoire ne serait que représentation et mise en littérature quelle que soit l'école historique ciblée.

Tout ne serait que mise en scène, une sorte de « Making fiction operation ». Face à cette rupture, l'historien s'est mis à douter de lui-même, car toute histoire et son écriture devient ainsi subjective. Pourtant, l'idée de représentation ne fut pas toujours un danger. Le concept de représentation n'est pas nouveau chez les universitaires, notamment français. Ces derniers ont commencé à utiliser plus régulièrement ce mot depuis les années 1960 lorsque des sociologues tels que Moscovici S. (1961) l'ont introduit dans le champ des sciences humaines.

Au début des années 1990, l'idée de représentation renvoie à l'imaginaire social des individus. Des historiens français tels que Duby G., Mandrou R. ou Ariès P, Vovelle M. ont participé à la construction de ce concept. Ces démarches, alors parcimonieuses, ont permis

peu à peu de se diriger vers la conclusion épistémologique soulignée par l'historien Prost A. (1996), à savoir que l'Histoire ne serait qu'une mise en intrigue, ce qui faisait un écho direct à White H. Cette « mise en intrigue » confirme ainsi la logique romanesque du travail d'écriture de l'historien, ce dernier se retrouvant dans une posture d'écrivain-romancier.

C'est ainsi que l'on découvre qu'analyser l'histoire des représentations sur un sujet, à partir de noms propres liés aux « Grands Hommes » n'est plus aussi positiviste, puisque désormais tout ne serait que mise en intrigue chez les historiens. White H. et Prost A. ont, pour ainsi dire, relativisé les frontières entre les écoles historiques, ce qui nous permet alors, avec des outils informatiques, d'aller plus loin et de dépasser des débats idéologiques devenus moins présents et moins essentiels.

Ensuite, il est nécessaire d'analyser le logiciel informatique et un néologisme formulé par le mot « culturomique ». Le logiciel informatique « *N-Gram Viewer* » accepte sans idéologie tout type d'items. Il intègre les mots sélectionnés dans une logique mathématique qui nous fait découvrir un monde et une discipline nouvelle. Quelles en seraient les racines morphologiques ?

Ce néologisme cristallise culture et numérique. Cette formule vient étymologiquement de la discipline de la biologie moderne connue sous le nom de génomique : aller au plus profond du système humain, comme aller au plus profond de la culture des hommes en numérisant tous leurs écrits. C'est le mariage improbable entre le numérique et la culture selon Delahaye J.P et Gauvrit N (2013).

En effet, la « culturomique » fut un projet piloté surtout par l'université d'Harvard à partir de 2009. Cette prestigieuse université a décidé de numériser un nombre gigantesque d'ouvrages en collaboration avec les laboratoires Google. Cette numérisation a été opérée sur une grande échelle puisque cela concerne plus de cinq millions d'ouvrages. Selon Delahaye J.P. (2011), chercheur au laboratoire d'informatique fondamentale de Lille (LIFL), analyser les représentations globales issues de millions d'ouvrages sur un sujet de manière exhaustive était jusqu'ici impossible mais, désormais, cet outil novateur peut débloquer la situation. Il précise ainsi :

L'informatique élargit le champ des variables évaluables. Elle offre des capacités de stockage massif d'informations, elle réalise des collectes de faits, de chiffres et d'images à grande échelle et les automatise. Elle donne aussi les moyens de traitement numérique ultra-rapide pour l'exploitation de gigantesques bases de données (p. 82).

Une nouvelle question surgit alors : comment fonctionne l'ensemble de ce capital de cinq millions de livres numérisés ? Le « data mining », la fouille des données, est sans cesse renouvelée par des milliers de « data-centers », ce qui est un avantage par rapport à l'exercice (plus classique et analogique) de l'enquête qui restait liée à un moment « T ».

La culturomique suit donc une logique active de mise à jour permanente. Cette approche est aussi quantitative. Ces livres ciblés ont été choisis dans les bibliothèques de plus de 40 universités dans le monde. Ces millions de titres ainsi numérisés représentent 4 % des livres publiés dans l'Histoire des Hommes. Les critères de sélection de cette gigantesque numérisation furent choisis pour la qualité technique de leur numérisation ainsi que pour le caractère de leurs « métadonnées » (Evans J., et Foster J. 2011), c'est-à-dire qu'ils présentent des informations générales en grand nombre. L'ensemble de ces 5 millions d'ouvrages est constitué de cinq cents milliards de mots répartis entre plusieurs langues. 361 Mm (*Mm = milliard de mots*) de mots pour la langue anglaise (la majorité du corpus numérisé), 45 Mm pour la langue française, quarante-cinq milliards pour la langue espagnole, trente-sept milliards pour l'allemand, et quelques milliards d'autres pour les autres langues (comme par exemple la langue russe : 35 Mm, la langue chinoise : 13 Mm, et la langue hébraïque : 2 Mm).

Ce projet de numérisation fut coordonné par l'universitaire Michel J-B. (2010), au sein de ce que ce dernier nomme alors « *L'observatoire Culturel* », un observatoire situé à Harvard. L'outil « *N-gram-Viewer* » fut mis en ligne, officiellement, en décembre 2010. Toutefois, le corpus numérisé n'est accessible que partiellement. En effet, on ne peut pas consulter directement les livres qui constituent ce corpus numérisé, d'une part parce que cela poserait un problème relevant de la propriété intellectuelle et, d'autre part, parce que ce corpus est d'une telle taille (taille qui atteint les 5 téraoctets) qu'il ne peut être traité par un ordinateur courant. Les 5 millions d'ouvrages numérisés n'incluent pas les journaux, les revues, les tracts, les publicités et d'autres imprimés. La date de publication retenue pour les livres sélectionnés est celle donnée par l'éditeur et donc, dans les courbes obtenues, le livre peut apparaître plusieurs fois avec des dates différentes, s'il a été réédité. Cet accès gratuit à l'outil « *N-gram-Viewer* » n'est donc que partiel, mais cette approche nous a tout de même permis d'obtenir des occurrences significatives sur l'histoire des représentations du 19^{ème} siècle français.

Finalement, quant à la représentation des courbes lexicales, il est nécessaire de remarquer qu'elles ne peuvent que suggérer des pistes de réflexion, des hypothèses, mais celles-ci ont l'avantage de faire « résonner » les personnages sélectionnés dans une structure éditoriale inédite. Le défi essentiel est de trouver un lien entre le discours (les tendances définies par les courbes) et le corps « muet » qui le soutient (les ouvrages numérisés, mais non accessibles). En d'autres termes, il nous faudra ici construire une « pragmatique » d'historien face aux graphiques obtenus. La démarche s'inscrit donc dans une « ontologie relative » dans le traitement des données proposées par nos courbes. Il y a dans ces graphiques une sorte d'auto-organisation du discours, les tendances éditoriales observées faisant « vibrer » des signes entre eux avec une certaine saveur. Ces « signes-indices » traduisent une perception spécifique et quantitative du passé, par le prisme d'une langue et donc d'une culture spécifique.

Pour terminer, nous avons fait le choix de travailler sur six langues (par rapport aux neuf langues disponibles de l'outil *N-gram-Viewer*). Nous avons voulu pour cela suivre le découpage culturel de l'UNESCO (Organisation des Nations Unies pour l'éducation, la science et la culture) où l'Europe et l'Amérique du Nord se retrouvaient ensemble. Ce découpage permettait de ne pas trop élargir le rayon de notre compas au niveau géographique afin d'obtenir une photographie de graphiques de courbes qui nous semblait déjà assez ambitieuse. Nous commencerons ainsi notre analyse de l'histoire des représentations du 19^{ème} siècle français par l'usage de la langue française, puis nous poursuivrons avec les autres langues proposées par notre outil, à savoir, l'anglais américain, l'anglais britannique, l'allemand, l'italien et l'espagnol.

L'analyse des N-grammes en langue française

Pour commencer notre analyse, nous aborderons en premier lieu les N-grammes spécifiques à la langue française (voir Figure 1). En effet, d'après nos résultats, la prédominance dans l'édition française de l'innovation technologique avec un grand gagnant est claire : Pasteur L. et ses découvertes. L'autre surprise, c'est le pic lié à « l'affaire Dreyfus » (surtout de 1894 à 1906) qui finalement fait émerger l'existence d'une opinion populaire en France sur le conflit social et politique le plus important de la Troisième République. Pourtant, une fois l'affaire terminée en 1906, la prédominance de l'affaire Dreyfus résiste devant Napoléon III, et conserve finalement une solide troisième place. (Doat F., 1988).

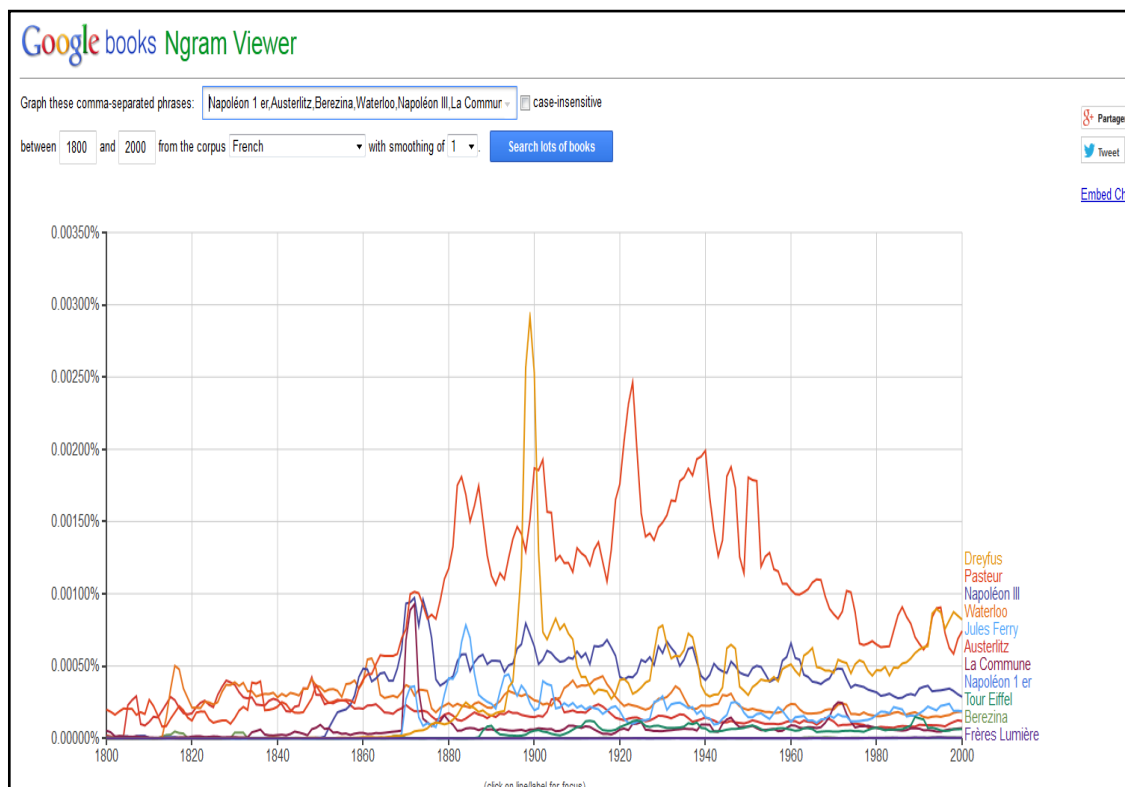


Figure 1. N-grammes en langue française.

En quatrième position, on trouve l'importance sociétale de l'œuvre de Ferry J. avec ses lois de 1881-1882 ce qui donnera les bases de l'éducation laïque en France. Cela s'exprime par des « pics-impacts » autour de 1880-1885, puis on perçoit une stabilisation des courbes Ferry (Andler Ch., 1918), à partir de 1910 jusqu'en 2000. Dans une position assez voisine, on remarque la présence de « Waterloo » (Largeaud J-M., 2000) qui suggère donc un Napoléon 1^{er} dans la défaite et non en position dominante, comme cela aurait pu être le cas avec Austerlitz (Gallo M., 1997). Ce pic est ainsi piégé dans sa topographie temporelle, quasi enracinée. Il ne peut être l'événement qui voyage dans le temps. Austerlitz résiste toutefois timidement dans les deux siècles et ne disparaît jamais totalement, mais il semble condamné à rester occulté par Waterloo. La défaite l'emporte sur les victoires concernant l'ogre corse de sorte que la légende noire est ici perceptible.

On peut enfin souligner la présence au 20^{ème} siècle de la « Tour Eiffel » (Dumont J-B., 2013) et de « la Commune » (Grenet J., 2003). Cette dernière a connu son pic naturel en 1871. Les « Frères Lumière » (Puthod M-F., 1994) sont condamnés à l'axe de l'abscisse, en toile de fond, comme tendus sur un écran plat. Donc, on voit ici un siècle à la fois de rupture technologique (effet Pasteur), mais un siècle aussi de rupture politique et sociétale qui émerge à la fin du 19^{ème} siècle : une opinion publique matérialisée par la presse papier grand public et les joutes politiques qui se multiplient et dont le porte-drapeau fut le « *J'accuse* » de Zola E. (1965).

L'analyse des N-grammes en anglais des éditions nord-américaines

Concernant les *N-grammes* en anglais (voir Figure 2) on distingue trois grands ensembles sur ce graphique représentant les éditions américaines. On identifie une première domination, celle de Waterloo, qui s'étend de 1810 à 1910. De 1910 à 1960, c'est « Pasteur » qui domine le graphique, Waterloo passant en deuxième position pendant le 20^{ème} siècle. On distingue ensuite un pic dominateur sur Dreyfus de 1960 à 1965. Quelle analogie pourrait-on faire sur cette année pic de 1960 ?

Le 11 mai 1960, le président Eisenhower reconnaît un mensonge d'Etat : c'est l'affaire des vols d'avions espions qui ont eu lieu les quatre dernières années au-dessus du sol soviétique. Le 15 mai 1960, les vols sont stoppés. Washington reconnaît ses torts. Une question se pose : l'atmosphère trouble entourant ces vols aurait-elle généré un « effet Dreyfus » (White G.R., 1996) au cœur des éditions américaines ?

Un autre pic notoire pour Dreyfus en 1965 : cette année-là, le Président Johnson reçoit Martin Luther King M. en juin 1965, et on vote le « *Voting-Rights-Act* » en août 1965, ce qui donne - enfin - un ensemble de droits civiques non limités à la communauté noire, longtemps après la guerre de Sécession. Une autre page de l'histoire de la communauté

noire commence à s'écrire (Branch T., 2010). L'idée de reconnaissance juridique après un long combat politique pourrait rapprocher le militaire français et le pasteur protestant.

Ce qui est curieux au 20^{ème} siècle, c'est le retour inattendu de Waterloo de 1970 à 2000. Nous nous sommes là aussi interrogés sur l'histoire américaine et les tentatives analogiques possibles. Ainsi, nous avons pu mettre en parallèle la réduction des troupes américaines au Vietnam qui n'est plus constituée que de 350000 hommes en 1970 (Du Berrier H., et Major J. R., 1965). Serait-ce un écotype d'un Waterloo américain au Vietnam ? Le « Dien Bien Phû » des Français ou la fuite de Saïgon par les Américains alimentent-ils alors la rhétorique Waterloo ? La défaite de Napoléon 1^{er} n'est plus si nette vers 2000. Les attentats de 2001 ont-ils changé la donne ? Le programme « N-Gram-Viewer » ne nous donne pas accès à cette hypothèse possible (le corpus s'arrête en 2000), mais on peut le supposer, au regard des discours politiques entrepris bien avant le 11 septembre 2001 par le président républicain et ses « faucons ». La culture de la défaite n'est plus à l'ordre du jour. Waterloo devient ainsi un contre-modèle.

Que dire des autres items ? Napoléon III (Anceau E., 2008) et Austerlitz semblent se partager la quatrième place sur les deux siècles étudiés. Napoléon 1^{er} et Ferry J. sont loin derrière. *La Tour Eiffel* s'aligne quasiment sur l'axe de l'abscisse, car non traduite en anglais. Ce fut notre parti pris de garder au maximum les expressions et métonymies de l'histoire de France dans leur morphologie d'origine afin d'éviter le crible de la traduction.

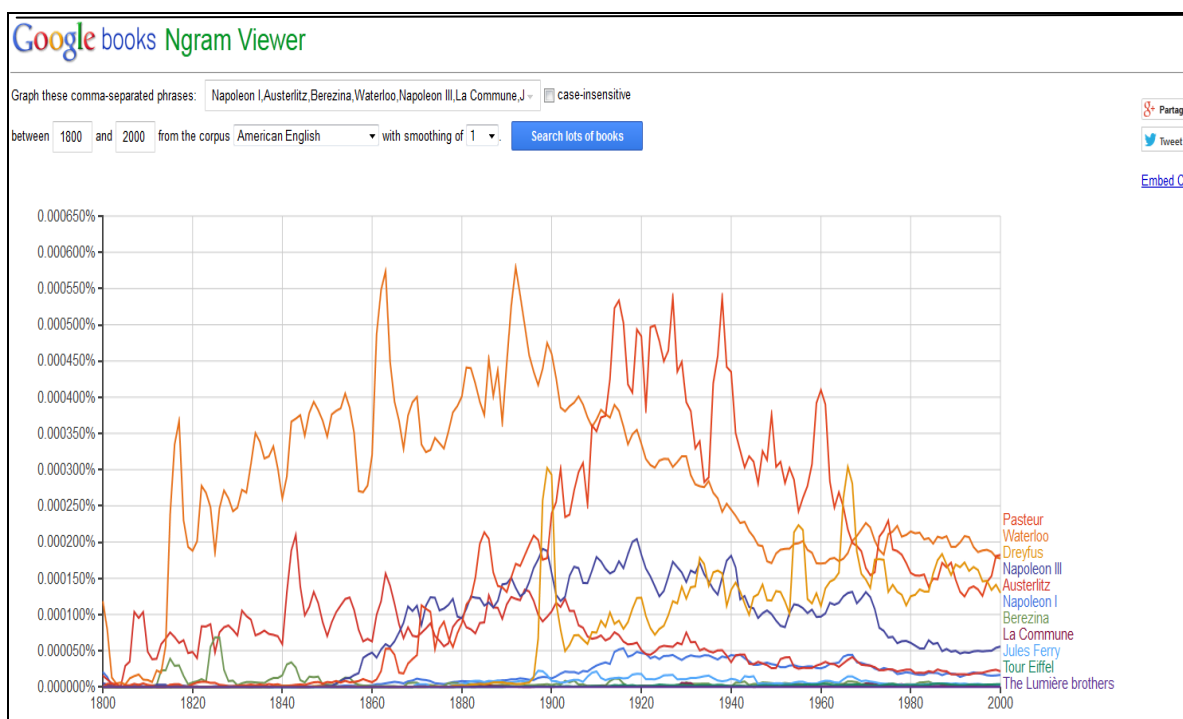


Figure 2. N-grammes en anglais des éditions nord-américaines.

L'analyse des N-grammes en anglais des éditions de Grande-Bretagne

Quant à l'analyse des *N-grammes* en anglais (voir Figure 3), Waterloo comme item obtient une place d'honneur. Ce succès s'étend de 1810 à 1910, mais « l'Entente Cordiale » (Boyd W. et al., 2004) obligeant et face à l'ennemi allemand, la grande défaite de Napoléon 1^{er} perd « discrètement » de son charme. Londres semble mettre en sourdine l'événement. À la mort du Roi d'Angleterre, Edouard VII, le 6 mai 1910, les relations entre la France et le Royaume-Uni sont donc devenues solides ; on nommait d'ailleurs Edouard VII, le Roi « Pacificateur » (Flourens E., 1906). Alors pourquoi encore évoquer Waterloo ? Edouard VII a - par ailleurs - donné son nom à un square à Paris, à une place et à un théâtre. Il est l'homme de « l'Entente Cordiale » par définition, celui qui a fait oublier - pour un temps - Waterloo...ou Fachoda.

Après 1918, c'est l'item « Pasteur » qui dominera jusqu'en 1965. Nous sommes dans une période qui veut de nouveau croire aux progrès et à la reconstruction, notamment après les deux guerres mondiales. Pasteur symbolise la guérison après la maladie, notamment celle de la guerre et de ses violences. Néanmoins, le graphique est perturbé à la fin du 20^{ème} siècle.

Des fantômes reviennent au début des années 1970 : Waterloo revient sur la scène; est-ce l'expression d'un retour de l'inimitié de la France vis-à-vis de la Grande-Bretagne ? La Grande-Bretagne intègre enfin, après de multiples résistances gaulliennes, la CEE (Communauté Européenne Economique). En fait, c'est en 1972 que le Président Pompidou avait fait lever le veto du général de Gaulle à l'entrée du Royaume-Uni dans le projet européen. Cette intégration est-elle perçue comme une victoire par la Grande-Bretagne ? Pour le premier ministre, Edward Heath, c'est quoi qu'il en soit une réussite qui s'accomplit le 1er janvier 1973.

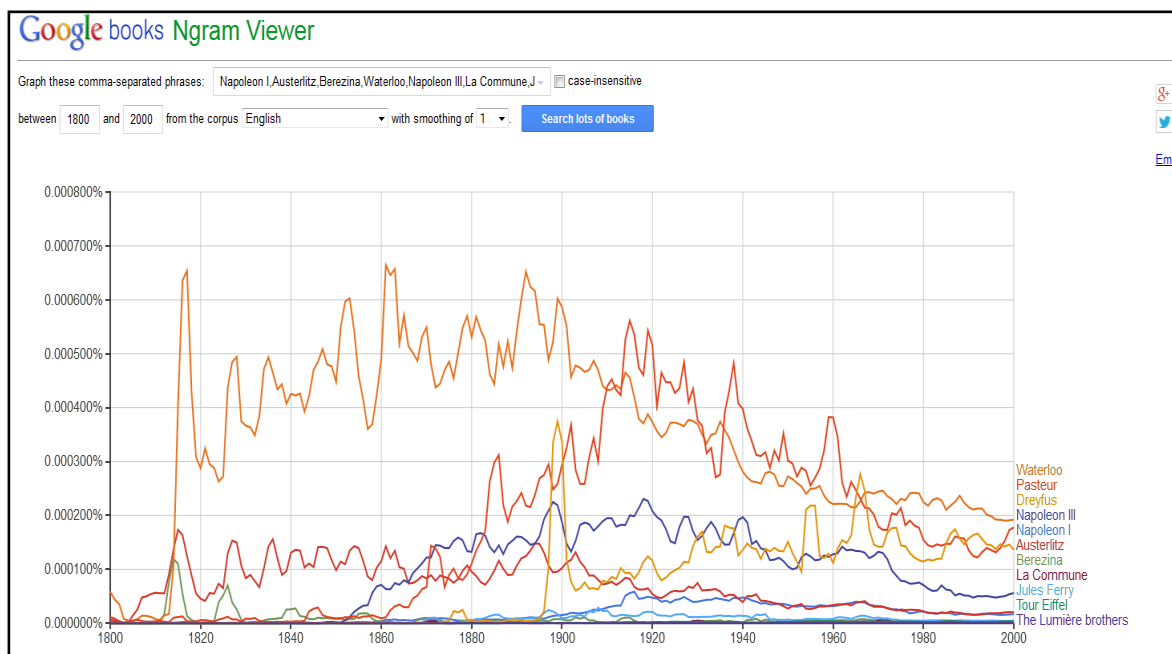


Figure 3. N-grammes en anglais des éditions de Grande-Bretagne.

Que dire des autres items ? Austerlitz et Napoléon III occupent la quatrième place, avec Dreyfus qui s'écroule après 1906 (fin de l'affaire) pour revenir autour de 1970, mais de manière ponctuelle notamment à une troisième position en 2000. Est-ce l'alignement politique qui se prépare discrètement entre les Etats-Unis et le Royaume-Uni de Tony Blair ? (Lee P., 2012). L'esprit de l'intrigue politique soutient l'univers-item de Dreyfus. Ici, c'est la préparation vers la guerre en Irak. Les Frères Lumière restent alignés sur l'abscisse.

L'analyse des N-grammes en allemand

Austerlitz est le premier item actif dans notre graphique (voir Figure 4), en première position, puis Waterloo prend le relais, mais sur un temps moins long, car à partir de 1850, Napoléon III amorce une hausse vertigineuse qui se concentre sur deux pics dont l'un en 1860. Serait-il lié à sa victoire de « *Solférino* » de 1859 ? (Pellissier P., 2012)

L'autre pic de Napoléon III se situe en 1870, année de la victoire de Bismarck sur la France et sur l'Europe. Grâce à sa victoire, l'Allemagne bénéficie d'un boom économique : c'est le « *Gründerzeit* », notamment alimenté par les 7 milliards versés par la France suite à sa défaite (Bainville J., 1918). De 1880 à 1920, c'est Pasteur qui prend le relais, dépassé curieusement par Napoléon III vers 1925. Comment interpréter ce retour de Napoléon III ? Serait-ce un retour symbolique (Costa A., 1928) des 7 milliards exigés à la France après 1870 ? Le Traité de Versailles serait-il un « *Gründerzeit* » à l'envers ? Puis, de 1925 à 1935, ressurgit la thématique Dreyfus. Peut-on supposer que cette thématique surgit avec la montée structurelle de l'antisémitisme des années 1930 en France avec les « *Ligues* » (Nicosia F. et al. 1990) et en Allemagne, plus explicitement, avec les Nazis ?

De 1935 à 1950, Napoléon III domine, pour être de nouveau dépassé par un très haut pic de 1960 à 1965 sur Dreyfus, un pic très largement supérieur aux années 1930. Le pic Dreyfus de 1960 correspond au retour des juifs allemands qui n'ont pas pu s'adapter à la vie en Israël. Une polémique va naître à ce sujet, venant des organisations juives américaines. Comment peut-on revenir dans un pays qui extermina une communauté ? Telle est la question que se posaient ces organisations outre-Atlantique. La thématique tournait autour de la culpabilité et de l'absurdité de ce retour.

Autre explication possible de l'item « Dreyfus » ; le chômage touchant les juifs venus de l'Europe de l'Est et notamment de l'ex-Urss et vivant en Allemagne, expliquerait-il aussi une remontée de l'antisémitisme ? On les qualifie alors d'« indisciplinés » et les préjugés se multiplient à leur égard, notamment dans les Länder de l'Est (Bensimon D., 2004).

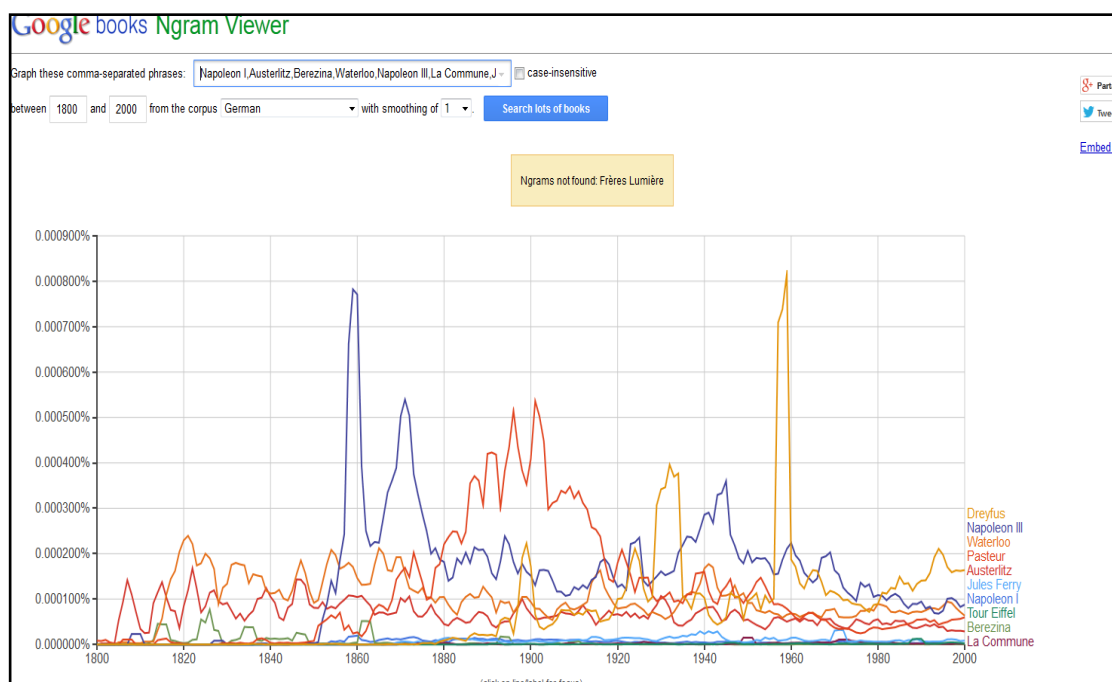


Figure 4. N-grammes en allemand.

L’item d’ « Austerlitz » constitue une bande moyenne basse sur les deux siècles et on peut percevoir, proche de l’abscisse, une vie non négligeable de Jules Ferry qui connaît une envolée dans les années 1935-1940. La réforme scolaire au pas de l’oie ou l’impérialisme de Jules Ferry auraient-ils séduit les éditions allemandes durant la période nazie ? Le volontarisme anti-humaniste des Nazis faisait sans doute feu de tout bois (Heydet X., 1937).

L’analyse des N-grammes en italien

Austerlitz amorce logiquement la période, puis s’impose Waterloo (voir Figure 5). Le 19^{ème} siècle est une période de batailles dans une logique toute lavissienne. Mais, dès 1860, c’est « Pasteur » qui prend le relais avec un pic très important en 1890 et 1900, pour ensuite retomber et laisser une place très ponctuelle à Dreyfus pendant la période dite de « l’affaire » prenant fin en 1906.

Ce pic de 1890 pourrait faire écho à l'Encyclique du Pape qui date du 15 mai 1891, qui était un texte à valeur sociale et qui s'intéresse au sort du monde ouvrier et à ses conditions de vie, en prenant soin d'écartier la pensée socialiste.

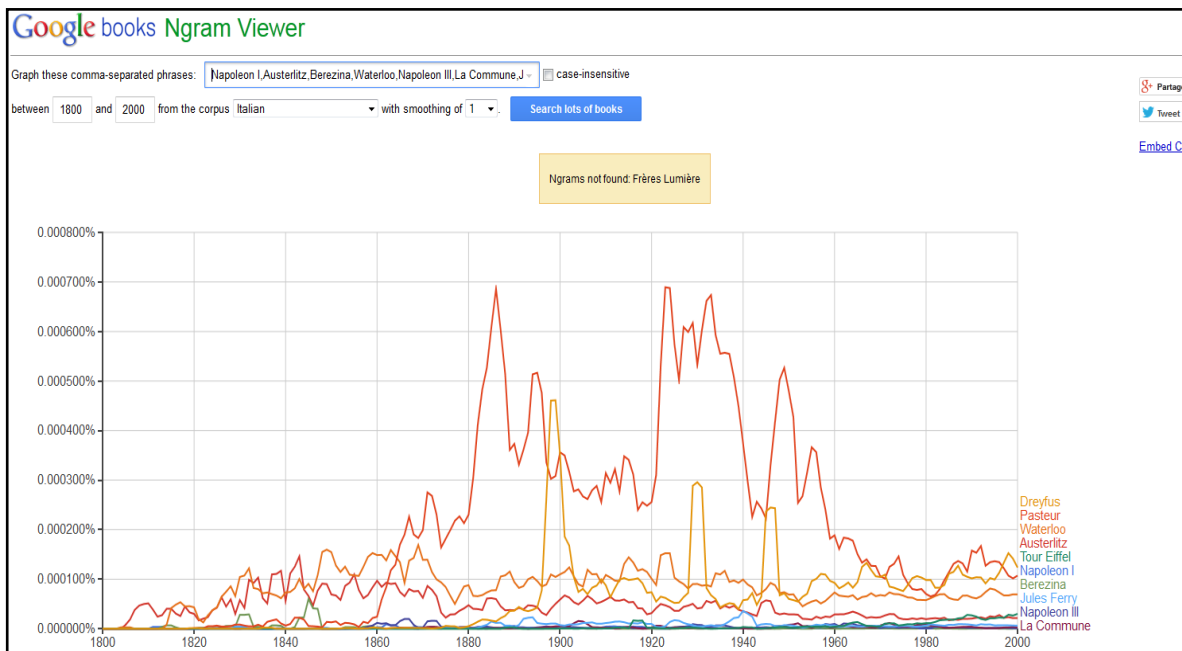


Figure 5. N-grammes en italien.

La pensée hygiéniste pour le monde du travail sans la pensée socialiste est l'objectif politique de ce texte sociétal du Pape (Richard J., 1993). Pasteur serait donc implicitement l'outil parfait de progrès social sans le socialisme, suivant la logique théologique du Pape d'alors.

Pasteur bénéficie ensuite d'une nouvelle série de pics assez sensibles entre 1920 à 1940, puis il chute lourdement au point d'atteindre durant la seconde guerre mondiale le niveau de Dreyfus, cristallisant l'antisémitisme d'Etat du « Duce ». La vérité n'est plus à chercher dans les sciences (Pasteur) ni dans la justice (Dreyfus), mais dans la seule volonté du Duce. Ce dernier veut assécher les « Marais Pontins » sans l'aide de Pasteur ou de toute logique scientifique ou écologique. Seul compte l'objectif : la « Grande bataille du blé » lancée en grande pompe en 1925 (Ministère de l'agriculture italien, 1930). Une fois l'égaré passé, l'après-guerre ouvre une nouvelle époque pour Pasteur. Ce dernier repart à la hausse de 1945 à 1960, mais il finit par chuter et il est de nouveau dépassé par Dreyfus dans les années 1980. Comment pourrait-on expliquer ce retour ? Rappelons que c'est une année où l'idée du complot politique revient, réminiscence de l'antisémitisme, et qui serait ici symbolisée par le scandale de la « loge P2 » (Marhic (R., 1995), affaire qui date du 17 mars 1981.

Waterloo a gardé en revanche une troisième place assez stable dans les deux siècles étudiés, ainsi qu'Austerlitz, mais à un niveau plus bas. Napoléon III existe après « *Solférino* » autour de 1865 et 1870, mais sa position reste bien timide et il est vite remplacé sur l'axe de l'abscisse par Ferry J. qui s'agit pour des réformes scolaires ou pour ses entreprises coloniales. D'ailleurs, c'est autour des années 1880 que l'Italie obtient, au détriment de la France, la concession ferroviaire « *Tunis - La Goulette*. » (Ben-Giat R., 2005). L'Italie a des objectifs coloniaux. Ferry semble encourager - par sa présence éditoriale - la péninsule italienne dans cette aventure.

L'analyse des N-grammes en espagnol

L'édition espagnole donne une large place de 1805 à 1855 à Austerlitz (voir Figure 6), puis la première place est prise par Napoléon III de 1852 à 1870, entre temps dépassé par Dreyfus durant l'affaire elle-même. Fin 19^{ème}, l'affaire Dreyfus génère une onde de choc de 1890 à 1910, puis une deuxième vague moins importante de 1925 à 2000. Il est intéressant ici de remarquer que pour le 20^{ème} siècle, c'est la domination quasi sans partage de Pasteur qui bénéficie d'un haut pic en 1920 puis en 1930, avant de diminuer progressivement jusqu'en 2000. La date de 1920 pour Pasteur n'est pas anodine, car l'Espagne, neutre pendant la première guerre, est vivement touchée par la « grippe dite espagnole » qui fait des ravages. (Hochart M., 2007)

Il est curieux de remarquer la « renaissance » de Napoléon 1er entre 1860 à 1910, comme réanimé par Napoléon III et son coup d'Etat de 1852. C'est l'un des rares graphiques à réagir ainsi. Peut-on parler de modélisation des Napoléon (Ménager B., 1998). Cette culture des coups d'Etat napoléoniens renvoie à la vie politique espagnole d'alors. Ainsi, ne pourrait-on pas songer au coup d'État du Général Joan Prim de 1868 qui force Isabelle II d'Espagne à quitter le pouvoir et à s'exiler en France ? Finalement, la « Tour Eiffel » existe mais elle reste écrasée sur l'abscisse. Elle reste condamnée à son rôle de toile de fond, comme pour tous les autres graphiques précédents.

Pour conclure, une ouverture vers la pensée lexicométrique reste nécessaire. Même si cette étude comparée de l'histoire des représentations du 19^{ème} siècle français nous a permis de mieux comprendre les images mythiques ou réelles véhiculées à travers l'édition mondiale appartenant à la période des 19^{ème} et 20^{ème} siècles. « *Penser, c'est comparer* » rappelait Malraux sans ses « *Antimémoires* » Malraux A, (1967). L'approche lexicométrique (donc des occurrences lexicales) que nous avons proposée suit cette logique de comparaison et évite les déterminismes politiques de pensées : le combat entre école positiviste contre école des Annales semble bien loin.

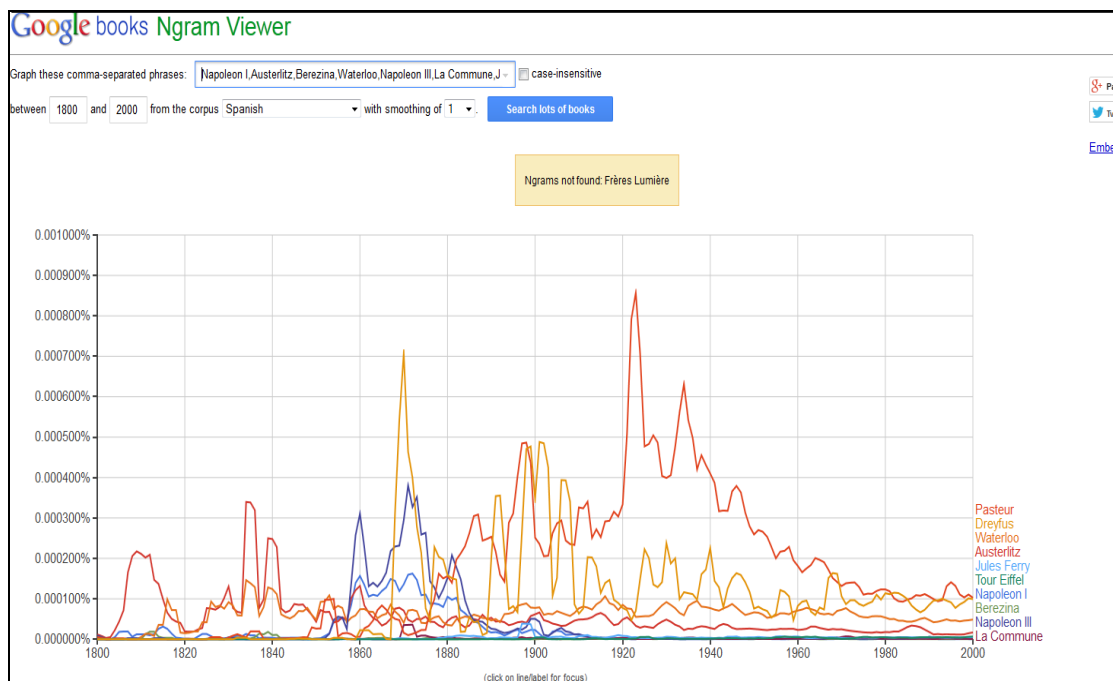


Figure 6. N-grammes en espagnol.

Cette approche quantitative de l'histoire nous a ainsi permis d'ouvrir de nouvelles « historicités ». Ainsi, l'historien français spécialiste des histoires internationales, Patrick Boucheron P. (2012), historien ouvert à une vision globale de l'histoire du monde, nous conseillait déjà de quitter une vision trop eurocentrée. La révolution numérique qui s'accélère encourage cette ouverture et la « culturomique » participe à ce mouvement. Les révolutions arabes de 2011, la divulgation de fichiers diplomatiques sur « Wikileaks » ou les actions d'« Anonymous » illustrent, à leur façon aussi, cet emballement d'un nouvel espace-temps virtuel pour la politique et les sciences humaines en général.

Ainsi, l'Internet participe à cette mutation plus singulière des représentations de l'histoire et de ses mythes. L'espace numérique est donc bien la nouvelle zone de combat où chacun invente sa vérité.

L'outil « *N-Gram Viewer* » permet donc une étude réflexive sur le temps par des techniques innovantes, mais l'entreprise reste encore fragile et sans doute trop empirique. En effet, les « *N-grammes* » présentent une histoire panoramique liée à une production éditoriale sur deux siècles. Ces courbes présentent une polymorphie, un autre rivage lié à une nouvelle histoire, totalement séculière et éloignée de toute théologie.

Ces tendances obtenues dans les graphiques sont, comme aurait pu nous le formuler l'historien comparatiste, Detienne M. (2009) :

Joyeusement utopiques et prêtes à faire voir les civilisations scintiller comme au-tant d'étoiles sur les vagues de la nuit », (...) un « temps inaugural de l'imagination comparée. (p. 19)

Les graphiques analysés sont - d'une certaine façon - des expériences physiques plus que l'expression d'une histoire accomplie. Celle-ci n'est pas ordonnée selon un axe allant du simple au complexe, le tout encadré dans un modèle trop longtemps évolutionniste. Bien au contraire. Ces courbes lexicométriques permettent surtout à l'histoire de redevenir une science de l'existence et non une science de l'essence. Ceci dit, les limites de l'outil N-gram-Viewer restent pourtant importantes. Le corpus composé de 72% de mots en langue anglaise ne doit-il pas être souligné comme perspective d'analyse limitée ? Dans l'histoire mondiale de l'édition, ne trouve-t-on pas ainsi une sur-représentation de l'anglais, surtout lorsque l'on sait que cela fut produit par les laboratoires « Google » ? Reste un autre problème méthodologique majeur. N'ayant pas accès direct aux ouvrages numérisés, nous ne pouvons qu'avancer des hypothèses à partir du contexte politique lié à chaque langue. Nous sommes dans le stade de l'indice. Il n'empêche que les pistes qui sont « ouvertes » sont passionnantes. Elles sont improbables mais elles ont de la valeur. La difficulté reste localisée sur la valeur d'interprétation : les variations de la fréquence éditoriale d'un personnage sont très difficiles à expliquer.

La « culturomique » reste donc une discipline en devenir dans sa méthode et ses modes de recherches. Il nous reste encore pour demain à mieux identifier les liens possibles entre les courbes obtenues et les hypothèses explicatives. Notre démarche avait ici pour objectif de se poser la question des entrées (datas) sur des analyses possibles. Le quantitatif de ce travail est l'objet même de cette étude. Cette présente étude avait donc pour objectif implicite d'être un appel à contribution face à un champ de recherche immense et encore novateur. Nous restons donc au stade de la curiosité et dans l'attente d'une perspective stimulante à vouloir encore « faire de l'histoire », toujours plus mais autrement.

Références

- Anceau E., (2008). *Napoléon III : un Saint-Simon à cheval*, 2008, Paris : Tallandier.
- Andler Ch., (1918). *Les lois Jules Ferry*, Paris : Berger-Leverault.
- Bainville J., (1918). *Bismarck et la France ; d'après les Mémoires du prince de Hohenlohe*, Paris : Librairie Nationale.
- Andall J., (2014). *Italian colonialism, legacy and memory*, Oxford, Berne, Berlin: P. Lang.
- Bensimon D., (2004). *Juifs en Allemagne aujourd'hui*, Paris : L'Harmattan.
- Boucheron P., (2012). *Entretiens : conversations sur l'histoire*, Lagrasse : Verdier.
- Boyd W., et al, (2004). *France / Grande-Bretagne : l'Entente cordiale = Great-Britain-France*, Paris : ADPF.
- Branch T., (2013). *The King year: historic moments in the civil rights movement*, N.Y: Simon and Schuster.
- Braudel F., (1967). *Civilisation matérielle et capitalisme*, Paris ; A. Colin.
- Costa A., (1928). *Des indemnités dues par l'Allemagne en vertu du traité de Versailles*, Thèse Université de Paris, Faculté de droit, Paris : PUF.
- Delahaye J.-P., (2011). La culturomique, in, *Pour la science*, n°406, p. 82.
- Detienne M., (2009). *Comparer l'incomparable*, Coll. Points Essais, Paris : Ed. Du Seuil.
- Doat F., (1988). *Histoire de France à travers les journaux passés : De Napoléon III à l'Affaire Dreyfus : 1851-1898 /Montreuil ; L'arbre verdoyant*.
- Du Berrier H., et Major J. R., (1965). *L'échec américain au Vietnam : vu par un Américain*, Paris : La Table Ronde.
- Dumont J.-B., (2013). *Les grands travaux du siècle : la Tour Eiffel*, Paris : Ed. Manucius.
- Evans J., et Foster J., (2011). Metaknowledge, in. *Science*, vol. n.331, pp. 721-724.
- Flourens E., (1906). *La France conquise : Edouard VII et Clemenceau*, Paris : Garnier Frères.
- Revista Electrónica Matices en Lenguas Extranjeras, número 9. ISSN 2011-1177. Páginas 68 – 90.*
Universidad Nacional de Colombia - Facultad de Ciencias Humanas - Departamento de Lenguas Extranjeras. Bogotá. <http://revistas.unal.edu.co/index.php/male>

- Gallo M., (1997). *Napoléon - partie 2 : Le soleil d'Austerlitz*, Paris : Laffont.
- Grenet J., (2003). *La mémoire de la Commune de Paris*, sous la direction de Serge Bertein, Master 2 d'histoire, Paris : IEP.
- Heydet X. (1937). *L'école au service du Troisième Reich*, Paris : H. Didier.
- Hochart M., (2007). *Le virus de la grippe : de l'antiquité à nos jours*, mémoire de droit et santé, Université de Lille.
- Largeaud J.-M. (2000). *Waterloo dans la mémoire des Français (1815-1914)*, thèse d'histoire sous la dir. De Isabelle Brelot, France: Université de Lyon 2.
- Lee P., (2012). *Blair's just war: Iraq and the illusion of morality*, New York: Palgrave MacMillan.
- Malraux A., *Antimémoires*, Paris, 1967 : Gallimard.
- Marhic R., (1995). *Enquête sur les extrémistes de l'occulte ; de la loge P2 à l'ordre du temple solaire*, Bordeaux : L'horizon chimérique.
- Ménager B., (1998). *Les Napoléon du peuple*, Paris : Aubier.
- Michel J.-B., et al., (2010). *Quantitative analysis of culture using millions of digitized books*, in. Science, vol. 331, pp. 176- 182.
- Millman R., (1990). *Les ligues catholiques et patriotiques face à la question juive en France de 1924 à 1939*, thèse sous la direction de S. Berstein, Paris, 1990 : IEP.
- Ministère de l'Agriculture en Italie (1930) *La bataille du blé en Italie*, Rome.
- Moscovici S., (2004). *La psychanalyse, son image et son public*, Paris : P.U.F.
- Nicosia F., et al., (2010). *Jewish Life in Nazi Germany, dilemmas and responses*, N.Y: Berghahn Books.
- Pasteur L., (1940-1951). *Correspondances de Pasteur : 1840-1895 / réunie et annotée par Pasteur Vallery-Radot , (4 volumes)*. Paris : Flammarion.
- Pellissier P., *Solférino ; 24 juin 1859*, Paris : Perrin.

Puthod M-F., (1994). *Le cinéma des Frères Lumière*, Paris : Publication de l'école moderne française.

Richard J., (1993). *La question sociale et aujourd'hui ; Colloque du centenaire de Rerum Novarum, 12 au 17 mai 1991*, Québec : Université Laval.

White G. R., (1996). *The accused: the Dreyfus trilogy* / George R. Whyte; [preface by Dieter W. Benecke], Bonn.

Zola E., (1965). *La vérité en marche : J'accuse*, 1965, Paris : Pauvert.

Ouvrages généraux sur les représentations en histoire

Amalvi Ch., (1988). *De l'art et la manière d'accommoder les héros de l'Histoire de France. De Vercingétorix à la révolution*, Paris : Albin Michel.

Bradley F.H., (1965). *Les présupposés de l'histoire critique*, [Trad. De P. Fruchon, (Première édition Oxford, 1874)], Paris : Les belles Lettres.

Brutier A., (1995). Enseignement de la représentation et représentation de l'enseignement : Lavis et pédagogie de l'Histoire, in. *Histoire de l'éducation*, n° 65, janvier, p.27-50.

Bruguière A., (1979). Histoire d'une histoire : la naissance des Annales, in *Annales, ESC*, novembre-décembre, pp. 1347-1359.

Burrin P., (1992). L'historien et l'historisation, in R. Frank (dir.) *Ecrire l'histoire du temps présent. En hommage à François Bédarida*, Editions du Cnrs, Paris, pp. 77-82.

Burke P., (1991). *New perspectives on Historical writing*, Cambridge: Polity Press.

Campbell M., (1999). *Wonder and Science: imagining Worlds in Early Modern Europe*, Ithaca: Cornell University Press.

Cavaillé J.P., (1991). *La fable du monde*, Paris : Vrin / Editions de l'Ehess.

Carr E.- H., (1988). *Qu'est-ce que l'Histoire ?*, Paris : La découverte, [1ère édition en anglais, 1961].

Certeau M. de., (1975). *L'écriture de l'Histoire*, Paris : Gallimard.

Certeau M. de., (1974). L'opération historique, in. Jacques Le Goff et Pierre Nora, in. *Faire de l'Histoire, I, Nouveaux problèmes*, pp. 3-41.

Chaudonneret M.-C., (1987). *La figure de la République. Le concours de 1848*, Paris : Rmn.

Chartier R., (1989). Le monde comme représentation, in. *Annales ESC*, novembre-décembre, pp. 1505-1520.

Chartier R., (2009). *Au bord de la falaise. L'Histoire entre certitudes*, Paris, Albin Michel.

Christie J., Shuttleworth S., (dir.) (1989). *Nature transfigured Science and literature, 1700- 1900*, Manchester University Press.

Corbin A., (1992). Le vertige des foisonnements : esquisse panoramique d'une Histoire sans nom, in *Revue d'histoire moderne et contemporaine*, janvier-mars, pp. 103-126.

Corbin A., Gérôme N., et al., (1994). *Les usages politiques des fêtes aux XIX et XXème siècles*, Paris : Publications de la Sorbonne.

Deleuze G., (1985). *L'image temps*, Paris : Minuit.

Delorme J., (1970). *Chronologie des civilisations*, Paris : Puf, troisième éd.

Desimone R., (1992). Les fonctions de la métaphore du mariage politique du Roi et de la République. France XV –XVIIIème siècles, in. *Annales, ESC*, pp.1127-1147.

Dosse F., (1987). *L'histoire en miettes. Des Annales à la nouvelle Histoire*, Paris : La Découverte.

Dubois C.G., (1977). *La conception de l'Histoire en France XVIème siècle*, Paris, A. Nizet.

Farge A., (1992). *Dire et mal dire. L'opinion publique au XVIIIème siècle*, Paris : Seuil.

Fierro-Domenech A., (1986). *Le pré-carré. Géographie historique de la France*, Paris : Laffont.

François M., (1972). *La France et les Français*, Paris : Gallimard [coll. Bibl. de la Pléiade].

Friedlander S., (1975). *Histoire et psychanalyse. Essai sur les possibilités et les limites de la psycho-histoire*, Paris : Ed. Du Seuil.

Revista Electrónica Matices en Lenguas Extranjeras, número 9. ISSN 2011-1177. Páginas 68 – 90.
Universidad Nacional de Colombia - Facultad de Ciencias Humanas - Departamento de Lenguas Extranjeras. Bogotá. <http://revistas.unal.edu.co/index.php/male>

- Gehlen A., (1990). *Anthropologie et psychologie sociale*, Paris : Puf.
- Ginzburg C., (1989). *Mythes, emblèmes et races. Morphologie et histoire*, Paris : Flammarion.
- Hartog F., (1980). *Le miroir d'Hérodote*, Paris : Gallimard.
- Joutard J., (1993). Une passion française : l'histoire, in Burguière A., et Revel J., (dir.), *Histoire de la France : les formes de la culture*, Paris : Ed. du seuil, pp. 507-570.
- Marin L., (1979). *Le récit est un piège*, Paris : Ed. de Minuit.
- Marrou H.-I., (1954). *De la connaissance historique*, Paris : Ed. du seuil.
- Milo D. S., (1991). *Trahir le temps*, Paris : les belles lettres.
- Moraze Ch., (1948). *Trois essais sur histoire et culture*, Paris : Armand Colin.
- Noiriel G., (1989). Pour une approche subjective du social, in *Annales Esc, nov.déc.*, pp. 1435-1459.
- Nora P., (1986). L'histoire de France de Lavissee, in. *Les lieux de mémoire : tome 2, La nation*, Paris : Gallimard, pp. 317- 375.
- Prost A., (1996). Histoire, vérités et méthodes. Des structures argumentatives de l'histoire, in *le débat, n° 92, nov. Décembre*, pp.127-140.
- Prost A., (2000). Comment l'histoire fait elle l'historien ?, in. *Vingtième siècle. Revue d'histoire, n°65, janv. Mars*, pp.3 -12.
- Tutiaux–Guillon N., et Mousseaux M. –J., (1998). *Les jeunes et l'Histoire : identités, valeurs, conscience historique : enquête européenne « Youth and history »*, Paris : Inrp.
- Vendryes P., (1952). *De la probabilité en histoire*, Paris : Albin Michel.
- Wieviorka A., (1998). *L'Ere du témoin*, Paris : Plon.
- White H., *The content of the form: narrative discourse and historical representation*, Baltimore, 1990, John Hopkins University Press, 244 pages.

Ouvrages sur la culturomique

Candless Mc. D., (2011). *Datavision*, Paris : Robert Laffont.

Cellier J., et Cocaud M., (2012). *Le traitement des données en histoire et sciences sociales*, Université de Rennes, PUR.

Delahaye J.-P., (2010). « Jusqu'où l'ordinateur calculera-t-il ? », *Pour la science*, N° 283, pp.100-105.

Delahaye J.-P., (2011). « La culturomique », *Pour la science*, N°406, pp. 82 à 87.

Delahaye J.-P., (2003). « Les chiffres de la complexité informatique », *Pour la science*, N°314, décembre, pp.162-165.

Gauvrit N., et Van Rillaer J. (2010). *Les psychanalyses. Des mythologies du 20ème siècle?*, Sofia-Antipolis : Book e Book.

Henno J., (2012). *Silicon Valley ou Prédateurs Vallée*, Paris : Télémaque.

Michel J.-B., et al., (2011). Quantitative analysis of culture using millions of digitized books, in. *Science*, 14 331 (6014), pp. 176 - 182.

Pommier F., (2010). *Mots en tocs et formules en tics. Petite maladie du parler d'aujourd'hui*, Paris : Seuil.