

La mémoire kaléidoscopique: l'histoire au prisme des algorithmes d'autocomplétion

Anna Jobin, Olivier Glassey, Stéfanie Prezioso,
Frederic Kaplan

When looking up information online, people encounter algorithmic output not only on the results pages of web search engines: already during the process of typing a query, autocompletion algorithms suggest what to look for. By analyzing such suggestions for search queries related to the First World War this contribution shows how historical information online is constantly recalculated and reconfigured according to different algorithmic logics. Ultimately, autocompletion produces a hybrid form of collective memory that is subtly shaped by both technology and social processes.

83

Durant les dernières décennies, le web est devenu un lieu fort de la diffusion et de la recherche d'information. Au sein de ces usages du web, les moteurs de recherche occupent une place cardinale. Ils fonctionnent à la fois comme des facilitateurs de recherche mais aussi comme des filtres qui reconfigurent sans cesse, à l'aide d'un système algorithmique puissant, la hiérarchie de ce qui est accessible en ligne. Ce travail des algorithmes qui pondèrent en temps réel des pertinences, agrègent les intérêts des internautes et calculent des proximités sémantiques constitue une part omniprésente mais peu visible du web, influençant la visibilité des contenus en ligne. Structurants mais sans cesse changeants, les algorithmes suscitent un intérêt grandissant, bien que les données empiriques liées à ce qu'ils «produisent» fassent encore largement défaut. Quelles opérations effectuent-ils? De quelle façon l'accès à l'information relative à l'histoire est-elle façonnée par ces dispositifs? Comment ces opérations s'articulent-elles avec les enjeux de mémoire? La présente contribution propose de questionner le regard contemporain porté sur les événements historiques à travers la structuration algorithmique de l'information en ligne d'un moteur de recherche. Il s'agit plus précisément de rendre compte des diverses manières dont les algorithmes d'autocomplétion structurent les contenus à caractère historique en ligne en prenant l'exemple des sujets liés à la Grande Guerre.

L'autocomplétion: algorithmes entre passé et présent

Pour la plupart des gens disposant d'une connexion internet, les moteurs de recherche occupent aujourd'hui un rôle central dans l'accès à l'information en ligne. Loin du modèle du web comme un ensemble de sites interconnectés par des liens hypertextes formant une structure ressemblant au rhizome,¹ le passage prépondérant des internautes par un moteur de recherche pour accéder à l'information confère à ces moteurs le statut important de «gardien» (ou *gatekeeper* en anglais).² Ainsi, les grands acteurs du web, tel Google, occupent aujourd'hui un rôle semblable aux médias traditionnels en ceci qu'ils façonnent le rapport du grand public à l'information, que ce façonnage s'opère au niveau des sujets, de leur cadrage ou du mode de consommation de l'information.³ Puisque la majorité des internautes utilise Google pour rechercher du contenu web, il est pertinent d'étudier comment ce moteur de recherche participe à structurer l'interaction des internautes avec les informations liées au passé et, par conséquent, de s'interroger sur la place des algorithmes de Google dans le rapport du grand public au passé.

L'analyse des résultats associés à une recherche en ligne (listes de sites, collections d'images, de vidéos, etc.) offre une manière d'appréhender le type d'opération à l'œuvre dans les moteurs de recherche.⁴ Cependant, les algorithmes de Google agissent non seulement dans la production des résultats de recherche, mais également en amont, lorsqu'il s'agit de définir les mots-clés qui donneront lieu à une recherche. En quelques fractions de seconde, et alors que l'internaute est encore en train de formuler sa demande, le moteur lui fait déjà des suggestions, non seulement sur la manière de terminer le mot qu'il est en train de composer, mais également en lui proposant des termes à associer à sa requête pour l'affiner. La fonction nommée «autocomplétion» lie des mots-clés supplémentaires

1 A noter que la forme même de ce modèle a, lui aussi, évolué de l'idée du web comme un réseau plutôt homogène à la représentation d'un ensemble de connexions inégales, formant une structure ressemblant davantage au rhizome, pour reprendre la métaphore de Deleuze et Guattari. Voir DELEUZE Gilles, GUATTARI Félix, *Mille plateaux*, Paris 1980.

2 HESS Aaron, «Reconsidering the Rhizome. A Textual Analysis of Web Search Engines as Gatekeepers of the Internet», in: SPINK Amanda, ZIMMER Michael (éds.), *Web Search Multidisciplinary Perspectives*, Berlin 2010.

3 BOZDAG Engin, «Bias in Algorithmic Filtering and Personalization», *Ethics and Information Technology* 15/3 (2013), 209-227.

4 Les exemples de ce type de recherche sont nombreux, c'est pourquoi nous ne citons ici qu'une étude de recherche empirique (signalée par un des relecteurs) qui va plus loin, en examinant non seulement les résultats de recherche mais également les effets d'une manipulation de ces derniers: EPSTEIN Robert, ROBERTSON Ronald E., «The Search Engine Manipulation Effect (SEME) and its Possible Impact on the Outcomes of Elections», *Proceedings of the National Academy of Sciences*, 112/33 (2015), E4512-E4521.

aux mots-clés initiaux et fait ainsi la médiation entre l'intention de recherche d'un utilisateur Google et son expression à travers un ensemble de mots-clés.⁵

Lancée en 2004 sous le nom interne de *Google Suggest* pour les utilisateurs intéressés, l'autocomplétion est intégrée comme mode par défaut du champ de recherche quatre ans plus tard. Sur le blog de l'entreprise, Google liste «l'aide pour formuler des recherches» comme première utilité de cette fonction, suivie par «la réduction des fautes d'orthographe» et «l'économie du nombre de touches tapées».⁶ Bien qu'ayant évolué depuis, la logique d'utilisation de cette fonction reste la même et le processus d'association entre mots-clés est en grande partie automatisé, ce qui lui vaut régulièrement d'être objet de débats sociaux lorsque des résultats inattendus ou questionnables font surface.⁷

Ces suggestions automatisées s'apparentent donc à un fléchage qui indique des chemins à suivre sur la base de recherches effectuées par d'autres et évaluées comme similaires par les algorithmes. Recalculées en permanence, ces recommandations dessinent donc un paysage singulier de proximités sémantiques, reflets des questions des autres utilisateurs et mises au diapason de ce qui constitue «l'actualité». Elles se trouvent alors érigées en hiérarchie de probabilités. Traduit en langage commun, le dispositif déclare à l'utilisateur que «si vous avez commencé à formuler cette question, sur la base de multiples variables dont les requêtes formulées récemment par d'autres usagers, il y a de très bonnes chances que ce qui vous intéresse est d'associer votre requête au terme suivant...». Dans un parcours de recherche d'information, ces propositions d'autocomplétion, bien avant les résultats de la recherche Google, constituent donc déjà un premier «résultat» de l'enquête informationnelle, car elles suggèrent un champ sémantique à l'utilisateur, favorisant certaines expressions au détriment d'autres.

Les commémorations du centenaire de la Première Guerre mondiale de 2014 nous offrent l'occasion d'aborder cette thématique en questionnant la manière dont le moteur de recherche Google tente de préempter les interrogations des internautes en leur suggérant des pistes de recherche sur la base des premiers termes introduits, l'algorithme anticipant des

5 JOBIN Anna, KAPLAN Frederic, «Are Google's linguistic prosthesis biased towards commercially more interesting expressions? A preliminary study on the linguistic effects of auto-completion algorithms», in: Conference Abstracts, Center for Digital Research in the Humanities, Lincoln (NE) 2013, 245-248.

6 LIU Jennifer, «At a Loss for Words?», Official Google Blog, 25. 8. 2008, <https://googleblog.blogspot.com/2008/08/at-loss-for-words.html> (consulté le 26. 7. 2015).

7 Pour un tour de la question, cf. JOBIN Anna, «Google's Autocompletion: Algorithms, Stereotypes and Accountability», *Sociostrategy*, 21. 10. 2013, <https://sociostrategy.com/2013/googles-autocompletion-algorithms-stereotypes-accountability/> (consulté le 24. 1. 2017).

propositions.⁸ Ce sont donc précisément ces suggestions liées à différents concepts et notions faisant référence à la Grande Guerre qui nous permettent d'explorer de quelle manière le passé se manifeste pour un utilisateur de Google au moment de sa recherche en ligne.

Le passé en accès gardé

Les commémorations officielles du centenaire du premier conflit mondial dans le monde semblent soumettre l'histoire aux exigences du présent. Les commémorations qui ont eu lieu, et celles encore à venir, en particulier en 2018, vont charrier leur lot d'usages publics.⁹ «Sans se pencher sur le présent, il est impossible de comprendre le passé», écrivait Marc Bloch.¹⁰ Le questionnement sur nos sociétés contemporaines ne peut aujourd'hui faire l'impasse sur la Première Guerre mondiale, qui nous oblige à confronter «le présent comme problème historique».¹¹ Ainsi, les évocations d'homologies structurales entre 1914 et nos jours sont monnaie courante, perdant en profondeur historique ce qu'elles gagnent en présentisme. L'appel à un imaginaire historique autour de la Grande Guerre s'inscrit dès lors dans un usage officiel et/ou médiatique, mais qui au cours du temps devient opinion commune et autoréférentielle, parfois inconsciente, des traces que l'événement a laissées dans le présent.

Ces dynamiques d'«aplatissement» seraient-elles inscrites dans les dispositifs techniques contemporains utilisés par le grand public? Le fonctionnement des moteurs de recherche, qui opèrent sur la base de l'information disponible en ligne, semble en effet propice à un lissage du passé sur le présent. Ainsi, l'actualité des commémorations pourrait influencer les algorithmes d'autocomplétion d'au moins deux manières.¹² Premièrement, les commémorations font augmenter le volume des contenus consacrés à la Grande Guerre sur le web (discours, annonce d'événements, articles de presse ou scientifiques, etc.). Les données et les métadonnées de ces contenus, qui dans leur ensemble opèrent dans un certain univers sémantique, sont ensuite référencées par les algorithmes du moteur de recherche.

8 JOBIN/KAPLAN (voir note 5), 245-248.

9 A ce propos, voir: HORNE John, «The Great War at its Centenary», in: WINTER Jay (éd.), *The Cambridge History of the First World War*, vol. 3, Cambridge 2014, 618-639; JULIEN Elise, *Der Erste Weltkrieg (Kontroversen um die Geschichte)*, vol. 74, Darmstadt 2014.

10 BLOCH Marc, *L'étrange défaite*, Paris 1946, 5.

11 LUKACS Georg, *Histoire et conscience de classe. Essais de dialectique marxiste*, Paris 1960, 150.

12 Sur la logique algorithmique, voir GOOGLE, «Effectuer des recherches avec la saisie semi-automatique - Aide Recherche Web», Effectuer des recherches avec la saisie semi-automatique, s. d., <https://support.google.com/websearch/answer/106230?hl=fr> (consulté le 8. 2. 2017).

Certaines associations lexicales pourraient alors apparaître plus pertinentes, car elles reviennent plus fréquemment que d'autres dans les contenus mis en ligne pendant la période des commémorations. Deuxièmement, les mots-clés recherchés par les internautes, leur provenance et leur volume sont d'autres éléments pris en compte par les algorithmes d'autocomplétion de Google. Plus une certaine expression composée est recherchée, plus ses mots-clés seront associés. Les mêmes raisons qui sont à l'origine de l'augmentation du contenu web consacré à la Grande Guerre en 2014 expliquent également pourquoi le grand public peut être amené à effectuer des recherches web en lien avec la Première Guerre mondiale ou ses commémorations.¹³

Comprendre le présent de la Grande Guerre contribue à éclairer la conscience historique de cet événement, conçue comme le dialogue entre la construction d'une mémoire collective et celle de l'histoire savante. Ainsi, plus que jamais dans les usages du web, la «conscience du passé dépend de l'autoconscience du présent».¹⁴ La question se pose alors en termes de rapport «entre les formes de fabrication/réception des récits et la conscience de soi» de nos sociétés;¹⁵ mais aussi en fonction des liens réciproques entretenus par le passé, le présent et le futur dans les nouveaux dispositifs informationnels et communicationnels. Les extraordinaires potentialités d'Internet, en tant que vecteur de transmission, sont souvent occultées par la dénonciation pure et simple du «présentisme» que l'on suppose inhérent à ce type de média. Pourtant, la massification de l'usage des moteurs de recherche nous offre l'occasion d'aborder les «présents du passé» sous un jour résolument nouveau en nous penchant spécifiquement sur les rapports complexes et souvent tourmentés entre histoire, mémoire et sociétés.

Les médias (quels qu'ils soient) sont non seulement des «véhicules de mémoire» mais aussi des constructeurs de la mémoire historique.¹⁶ Leur importance dans la construction d'un regard sur l'histoire n'est plus à prouver. Pierre Nora ne questionnait-il pas déjà, dans les années 1980, l'«irruption» du passé dans la presse, à la télévision, à la radio...¹⁷ aujourd'hui,

13 Pour plus de détails sur l'influence notamment des actualités médiatiques sur le volume de recherche, voir GRANKA Laura Ann, «Media Agenda Setting and Online Search Traffic. Influences of Online and Traditional Media», 2010, <http://research.google.com/pubs/pub36915.html> (consulté le 10. 6. 2016).

14 BALABAN Oded, «Temps, entendement et volonté», *Diogène*, 190 (2000), 5.

15 HARTOG François, LANCLUD Gérard, «Régimes d'historicité», in: DUTU Alexandre, DODILLE Norbert (éds.), *L'état des lieux en sciences sociales*, Paris 1993, 32.

16 GLASSEY Olivier, PREZIOSO Stéphanie, «Fabrique médiatique du regard et usages publics de l'histoire. Steven Spielberg, Le Débarquement et la guerre côté Alliés», in: BADIE Marie-France, HECK Michèle-Caroline, MONBRUN Philippe (éds.), *La fabrique du regard (L'atelier des sciences humaines et sociales)*, Paris 2011, 326-335.

17 NORA Pierre, «La fin de l'histoire-mémorisée», in: id. (éd.), *Les lieux de mémoire*, Paris Paris 1984, XVII-XXV.

son «omniprésence» ne fait plus de doute. Les canaux de diffusion se sont cependant multipliés et diversifiés: film grand public, docudrama ou fiction du réel, jeux vidéo...¹⁸ Le web est devenu l'un des vecteurs principaux de la construction de points de vue sur le passé, où les internautes sont appelés à intervenir dans la construction de la mémoire, en fonction des impératifs du présent, un phénomène que Louise Merzeau résume par la notion de «plasticité de la mémoire» sur le web.¹⁹

Notre recherche s'inscrit plus généralement dans les usages publics et politiques de l'histoire dans les nouvelles technologies de l'information et de la communication.²⁰ L'autocomplétion est une façon omniprésente mais invisible dont ces mémoires sont traduites en ligne. Il convient alors d'explorer de quelles manières ces supports favorisent de nouvelles représentations de l'événement et par là même de nouveaux espaces de mémoire diffuse qui n'est ni la mémoire officielle, ni la mémoire collective, ni celle de groupe ou individuelle, mais une mémoire se situant en deçà et au-delà des autres.

En parallèle aux activités de commémorations publiques, et à la mémoire officielle qui se (re)construit au moment du centenaire, nous sommes ainsi en mesure de détecter une évolution possible de ce *regard d'aujourd'hui* et de la mémoire collective qui y est associée. L'influence des algorithmes sur le «présent du passé» se donne à voir notamment dans la manière dont l'actualité des commémorations (discours, cérémonies, productions médiatiques, etc.) se retrouve non seulement au niveau des résultats d'une recherche en ligne mais déjà dans les autocomplétions sur ces sujets.

Comment cette actualité influence-t-elle les suggestions du moteur de recherche ou, au contraire, de quelles manières se retrouve-t-elle associée à d'autres événements? Quelle est la volatilité de ces associations calculées et comment participent-elles des formes de «tyrannie du récent»²¹ qui seraient inscrites dans les logiques de fonctionnement de cette infrastructure épistémique que constitue le web?

18 Sur cette dernière question, voir GLASSEY Olivier, PREZIOSO Stéphanie, «La guerre en ligne. Analyses des pratiques liées aux simulations informatiques ludiques massivement multi-joueurs», in: ROUSTAN Mélanie (éd.), *La pratique du jeu vidéo: réalité ou virtualité?* (Sciences humaines et sociales. Consommation et société), Paris 2004, 189-202.

19 MERZEAU Louise, «Guerres de mémoires on line: un nouvel enjeu stratégique?», in: BLANCHARD Pascal, VEYRAT-MASSON Isabelle (éds.), *Les guerres de mémoires. La France et son histoire. Enjeux politiques, controverses historiques, stratégies médiatiques* (Cahiers libres), Paris 2008, 297.

20 GLASSEY/PREZIOSO (voir note 16).

21 GLASSEY Olivier, PREZIOSO Stéphanie, «Musées de la Résistance en ligne. Vers de nouveaux espaces symboliques?», *En Jeu. Histoire et mémoire vivante*, 1 (2013), 105-117.

Une méthodologie anticipatoire

Les algorithmes n'opèrent pas dans un vacuum mais s'intègrent dans le monde social et ses dynamiques, qui se traduisent en paramètres mesurables et, surtout, en mots-clés.²² En anticipant que les mots-clés et l'évaluation que Google en fait peuvent varier dans le temps en fonction des usages publics de l'histoire, nous avons collecté nos données pendant l'été des commémorations de la Grande Guerre en 2014. Ces données sont donc des listes ordinales constituées de suggestions d'autocomplétion Google pour une expression de base donnée, triées d'après leur pertinence. La récolte de ces données a posé trois défis méthodologiques majeurs que nous expliciterons ci-dessous: d'abord, il a fallu minimiser les effets de personnalisation et de profilage; ensuite, il était important de nous focaliser sur la dimension temporelle en vue du fonctionnement dynamique des algorithmes de Google qui changent au fil du temps; et finalement, il était crucial de constituer à l'avance un ensemble d'expressions de base sur lesquelles nous récolterions les suggestions d'autocomplétion, et donc d'anticiper tout actualité *potentielle* autour des commémorations de la Grande Guerre.

Aucun usage de Google n'échappe au profilage, c'est-à-dire à l'adaptation générale du moteur de recherche en fonction par exemple du lieu géographique ou du navigateur utilisé. Comme le profilage est fondé sur des données agrégées de l'ensemble des internautes, notre faible volume de recherche ne devrait pas y avoir eu d'impact. A partir de là, un focus sur l'évolution des autosuggestions a permis de retracer la volatilité de ces dernières à condition de garder les paramètres de profilage fixes. Nous avons atteint cela en créant un environnement stable, c'est-à-dire en utilisant un ordinateur exclusivement consacré à cette recherche, et en collectant nos données avec le même protocole sur le même réseau internet pendant toute la période de récolte, en ouvrant à chaque fois une nouvelle session. Pour que ces changements de mots-clés ne soient pas liés à des logiques de personnalisation, nous avons pris soin de récolter nos données en interrogeant google.fr directement depuis le terminal avec l'outil cURL, qui envoie une demande http et ne sauvegarde pas d'historique ni de témoins de connexion. Cette démarche nous a permis de neutraliser les effets de profilage et d'éviter les écueils liés à la personnalisation.

Le ciblage individuel des moteurs de recherche fait que chaque utilisateur dispose de sa propre version des propositions et des résultats de Google

22 Cf. JOBIN Anna, GLASSEY Olivier, «(I Am Not A Web Search Result! I Am A Free Word.) The Categorization and Commodification of «Switzerland» by Google», in: RASCH Miriam, KÖNIG René (éds.), *Society of the Query Reader. Reflections on Web Search*, Amsterdam 2014, 153.

à un moment donné.²³ Il n'est alors pas possible de tirer de conclusions générales à partir d'une observation des suggestions d'autocomplétion isolée (ni même des résultats de recherche par ailleurs, bien que bon nombre d'enquêtes abondent dans ce sens). Cela rend par ailleurs difficile, voire impossible, une étude qui se concentrerait uniquement sur les suggestions d'autocomplétion à un moment donné: elle permettrait d'esquisser une liste de mots-clés qui ne sera, de toute façon, pas stable dans le temps et donc anecdotique. Nous avons contourné ce problème en nous focalisant sur l'évolution des propositions, c'est-à-dire sur les dynamiques à moyen terme qui régissent les suggestions d'autocomplétion. En collectionnant nos données à une fréquence hebdomadaire pendant 22 semaines²⁴ entre avril et septembre 2014, notre récolte couvre les deux mois avant et après les dates clés du centenaire de la Grande Guerre en juin et en juillet. Ainsi, nous n'avons pas seulement étudié les simples occurrences de certaines suggestions d'autocomplétion, mais surtout leurs logiques d'évolution au fil des semaines.

Afin de pouvoir collecter des suggestions d'autocomplétion, il fallait choisir les expressions de base. Pour détecter une évolution éventuelle pendant la période celle-ci, ces expressions de base devaient impérativement être définies avant le début de la récolte, sans possibilité d'ajouter de nouvelles expressions au protocole durant la récolte même. C'est pourquoi nous avons dû définir des listes de mots-clés extrêmement larges, qui allaient couvrir autant que possible toute actualité potentielle de l'été 2014. Autrement dit, ces listes étaient marquées par l'anticipation de ce qui *aurait pu* se passer pendant l'été 2014 ayant un impact sur les recherches Google: des prises de position publiques ou des discours officiels à mots-clés marquants, des énoncés viraux par des politiciens ou lors de commémorations locales, des accidents, des gaffes ou des faux pas, et caetera. Ce qui impliquait que la plupart de nos mots-clés risquaient de devenir superflus en automne, une fois que nous connaîtrions l'existence – ou non – d'événement ou d'énoncé marquant pendant les mois précédents – une évaluation qui n'est possible qu'*ex post*.

Nous avons défini trois thesauri – constitués, respectivement, d'expressions nominatives, d'expressions spécialisées et d'expressions éventuelles –, nous étant restreints à l'étude des expressions entières. Le premier thesaurus est composé de mots qui désignent la Première Guerre mondiale, tel «Grande Guerre» ou «Der des der». Le deuxième thesaurus comprend quant à lui le lexique du Collectif de recherche international et de débat

23 BOZDAG (voir note 3).

24 Pour l'un des trois thesauri (cf. ci-dessous), nous ne disposons que de 16 semaines de données à la suite d'une erreur de sauvegarde.

sur la guerre de 1914–1918 (CRID 14–18),²⁵ allant d’«abeille» à «zeppelin». Ce thesaurus nous est apparu comme d’autant plus intéressant que le CRID 14–18 occupe une place particulière dans le champ historiographique français; tenant d’une histoire sociale du conflit, le collectif d’historiens qui a son siège à Craonne s’inscrit dans une lecture critique de la guerre moins diffusée par les grands médias. Finalement, le troisième thesaurus comprend un ensemble de mots qui auraient pu être importants au fil de l’été et des commémorations. Afin de le constituer, nous avons extrait les noms communs et les noms propres des discours présidentiels et des articles publiés par *Le Monde* consacrés au centenaire en 2013. La distinction entre les trois thesauri n’a pas eu d’impact sur la conclusion finale mais nous a permis de constituer un ensemble d’expressions de base aussi vaste qu’hétérogène.

En suivant sur plusieurs mois les propositions d’autocomplétion à partir de ces termes prédéfinis, la récolte finale a résulté en plus de 190 000 suggestions d’autocomplétion. Cependant, ce grand nombre ne se traduit pas en une analyse aussi exhaustive: d’un côté, une grande partie des expressions – surtout celles qui pourraient être qualifiées d’anticipatoires – est devenue superflue, car aucun événement ou énoncé de l’été 2014 lié à la Grande Guerre ne s’est avéré de caractère viral; d’un autre côté, nous avons constaté, pour une grande partie des mots restants, des minimales fluctuations récurrentes dans l’ordre d’apparition en tant que suggestion d’autocomplétion. Ce dernier point confirme notre prémisse d’être face à un système algorithmique en adaptation constante. Google change donc en continu, sans que l’impact de ces changements soit assez grand pour pouvoir l’attribuer à des facteurs extérieurs. Bien qu’il soit impossible d’expliquer ce qui se passe exactement à l’intérieur de la boîte noire de Google – et tel n’a pas été notre objectif²⁶ – notre analyse longitudinale nous a permis d’extraire des études de cas qui traduisent différentes représentations du lien entre passé et présent. C’est pourquoi notre analyse, ci-dessous, se focalise sur certains mots-clés qui illustrent des logiques d’autocomplétion très diverses et nous permettent de constater certains

25 BACH André et al., «Lexique des termes employés en 1914–1918», Collectif de recherche international et de débat sur la guerre de 1914–1918, 2014, http://crid1418.org/espace_pedagogique/lexique/lexique_ini.htm (consulté le 4. 1. 2014).

26 Non seulement il est impossible de connaître le fonctionnement exact des systèmes algorithmiques de Google – parce qu’il est dynamique, parce qu’il est secret et parce qu’il repose sur des logiques algorithmiques d’autoapprentissage – mais même si elles étaient réalisables, de telles connaissances n’auraient qu’un impact partiel sur la compréhension des enjeux qui sont liés. A ce propos, cf. SEAVER Nick, «On Reverse Engineering. Anthropology and Algorithms», Medium, 28. 1. 2014, <https://medium.com/anthropology-and-algorithms/on-reverse-engineering-d9f5bae87812> (consulté le 20. 8. 2015); ANANNY Mike, CRAWFORD Kate, «Seeing without Knowing. Limitations of the Transparency Ideal and its Application to Algorithmic Accountability», *New Media & Society*, 2016.

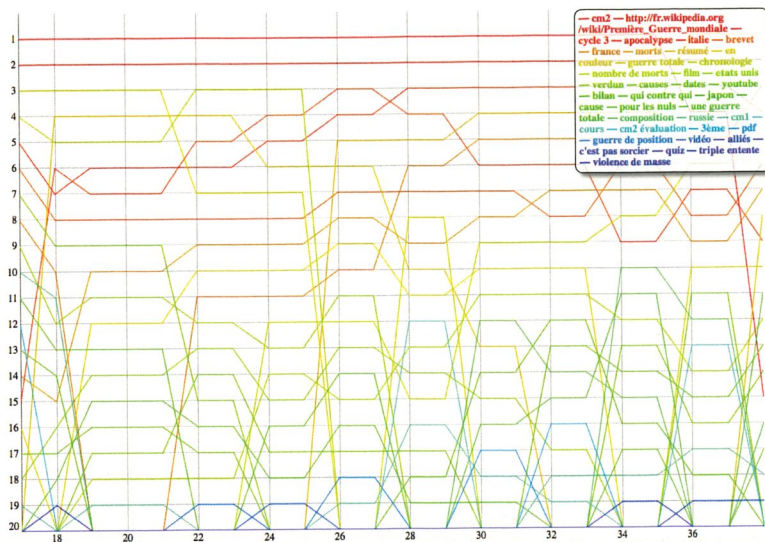


Fig. 1: Toutes les autocomplétions pour la recherche «première guerre mondiale» et leurs positions ordinales au fil des semaines

des liens avec le passé ainsi construits par les suggestions proposées automatiquement.

Notre analyse s'est faite à l'aide d'un outil de visualisation,²⁷ qui a rendu davantage lisible pour l'œil humain l'ordre des mots-clés au fil des semaines. Cet outil de visualisation nous montre toutes les suggestions d'auto-complétion pour un terme donné sous forme d'un tableau: l'axe des X représente le temps (indiqué en numéro de semaines du calendrier) et l'axe des Y montre l'ordre d'une suggestion donnée par rapport à d'autres expressions, le chiffre «1» indiquant la première suggestion d'auto-complétion, «2» la seconde, et ainsi de suite. Plus la ligne est basse, plus faible est donc la probabilité réelle que l'expression soit effectivement proposée à un utilisateur de Google lors d'une recherche (figure 1).

Au lieu d'effectuer une analyse automatique nous avons opté pour une approche privilégiant une coordination de regards humains sur les données afin d'émettre des hypothèses pertinentes et d'éviter des surinterprétations. En effet, notre recherche vise à explorer les liens du présent au passé, et donc à analyser des univers sémantiques et narratifs. Cette tâche demande

27 Pour une discussion des enjeux de la visualisation de données en histoire, voir NATALE Enrico et al., *Visualisierung von Daten in der Geschichtswissenschaft / La visualisation des données en histoire*, Zurich 2015.



Fig. 2: Deux autocomplétions particulières pour la recherche «première guerre mondiale» mises en exergue par les chercheurs et leurs positions ordinales au fil des semaines

un savoir spécialisé, contextuel, et n'est que difficilement traduisible en un processus automatisé.²⁸ Le premier cas ci-dessous illustre d'ailleurs parfaitement bien en quoi une analyse simple purement quantitative n'aurait sans doute pas été en mesure de révéler les subtilités sémantiques qui nous intéressent.

Un paysage mémoriel à l'ombre des probabilités d'occurrence

Puisque l'autocomplétion intervient en amont de la recherche sur Internet, au moment même de sa formulation par l'utilisateur, nous avons vu auparavant que les regards contemporains sur le passé (discours, cérémonies, productions médiatiques, etc.) peuvent se retrouver déjà dans les propositions qui sont faites aux internautes, et pas seulement dans les résultats de recherche.

Un premier constat est d'ordre général: la plupart des suggestions proposées automatiquement n'ont pas d'importance stable. Leur place dans

28 Pour une discussion des approches historiques qui peuvent, ou non, bénéficier des Big Data, voir BEARMAN Peter, «Big Data and Historical Social Science», *Big Data & Society*, 2/2 (2015).

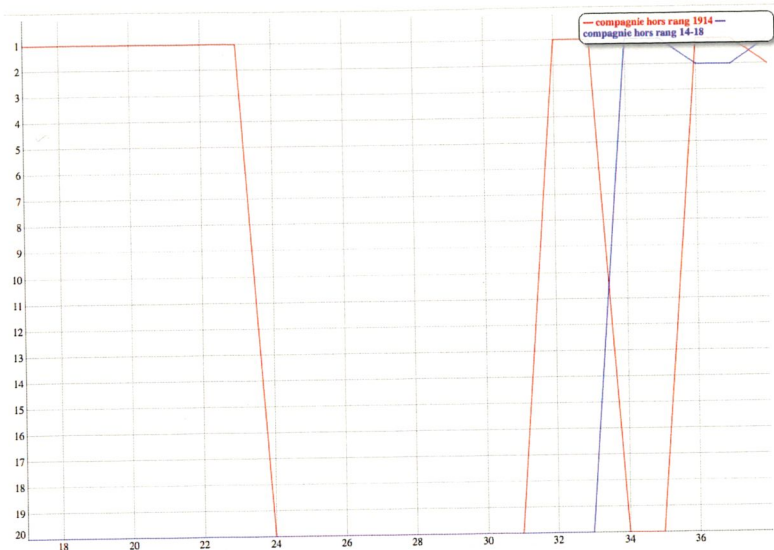


Fig. 3: Toutes les autocomplétions pour la recherche «compagnie hors rang» et leurs positions ordinales au fil des semaines

le tableau change constamment, ce qui indique que le lien algorithmique entre le présent et le passé est sans cesse renégocié. Cependant, ce lien opère à plusieurs niveaux et son impact est parfois équivoque. Un exemple qui illustre ce qui précède concerne le terme «première guerre mondiale». Pour la recherche de ce terme, par exemple, la suggestion d'autocomplétion «nombre de morts» est plutôt haut classée. D'après Google, il y a donc une certaine probabilité qu'une personne qui commence par taper «première guerre mondiale» sur son clavier souhaite finir sa recherche par «nombre de morts». Mais d'une semaine à l'autre, à la mi-juin 2014, cette suggestion d'autocomplétion ne fait plus partie des possibilités. Cependant, elle n'a pas disparu. Elle a simplement été remplacée par l'expression plus courte de «morts» (figure 2). Et alors que le changement est intéressant sous l'angle du langage, il s'avère difficile d'interpréter son impact significatif sur l'univers sémantique lié à la Première Guerre mondiale.

Dans d'autres cas, les changements des termes d'autocomplétion – ou carrément l'absence de suggestions – proposent des lectures *a priori* moins ambiguës. Prenons l'expression de «compagnie hors rang». Les suggestions d'autocomplétion associées à cette expression sont très peu nombreuses: d'abord, il n'y en a qu'une, «1914», qui disparaît pendant neuf semaines, et ensuite une deuxième suggestion, «14-18», apparaît. (figure 3). La seule

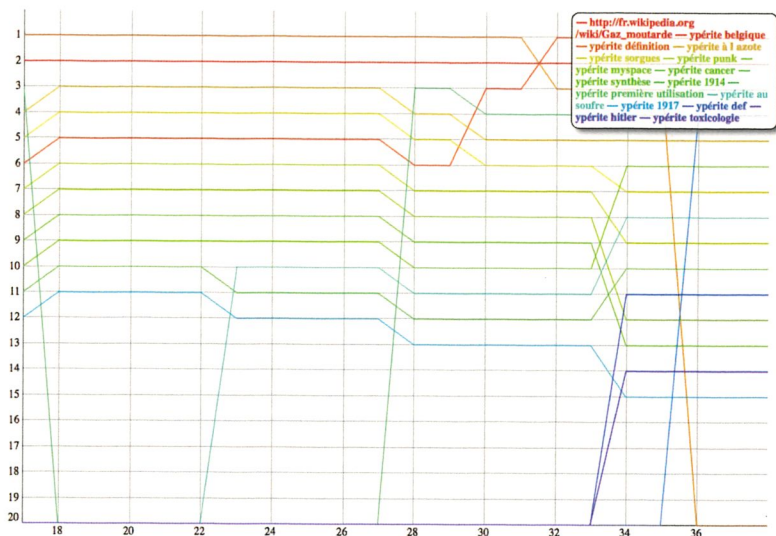


Fig. 4: Toutes les autocomplétions pour la recherche «ypérite» et leurs positions ordinales au fil des semaines

association sémantique algorithmique de l'expression «compagnie hors rang» n'est donc pas très forte mais est en revanche clairement liée au premier conflit mondial.

Une autre expression liée à la Grande Guerre est celle d'«ypérite», le gaz moutarde qui doit son surnom à sa première utilisation lors d'un combat près du village d'Ypres. Ce lien se reflète également dans les suggestions d'autocomplétion: parmi les 15 suggestions existantes, une grande partie se réfère à des lieux ou des événements de la Première Guerre mondiale, comme «ypérite belgeque», «ypérite 1914», «ypérite 1917» ou encore «ypérite hitler» (figure 4).²⁹ Cette association sémantique coexiste avec un autre champ lexical important lié à la composition chimique («ypérite à l'azote», «toxicologie», «synthèse»). Un troisième champ sémantique, composé de deux suggestions associant l'expression à un groupe de musique du même nom, semble perdre en importance à partir de la mi-août. Cet exemple nous montre que les associations entre mots-clés d'un certain champ de connaissances peuvent être, sur Google, en compétition directe avec celles d'autres champs.

Parmi les très nombreuses suggestions d'autocomplétion potentielles pour l'expression «première guerre mondiale» (figure 5), il y en a certaines

29 En effet, Adolf Hitler était parmi les soldats allemands blessés par le gaz moutarde à Ypres.



Fig. 5: Trois autoccompléments particulières pour la recherche «première guerre mondiale» mises en exergue par les chercheurs et leurs positions ordinales au fil des semaines

qui se réfèrent clairement à des produits culturels spécifiques récents (au moment de la capture des données). Nous y retrouvons notamment le film documentaire en cinq épisodes *Apocalypse: la Première Guerre mondiale*³⁰ d'Isabelle Clarke et Daniel Costelle (figure 6). Ce documentaire fait suite à *Apocalypse: la Seconde Guerre mondiale*³¹ et *Apocalypse Hitler*,³² de véritables succès d'audience. Ces documentaires ont la particularité d'avoir été réalisés avec des images d'archives colorisées, afin de les rendre plus «accessibles» pour le grand public; ils sont narrés par la voix de Matthieu Kassovitz et présentent des documents filmés d'époque montés dans un scénario filmique qui s'apparente à de la fiction et où «l'émotion» représentée mais aussi suscitée joue un rôle cardinal comme en témoignent les titres des cinq épisodes d'*Apocalypse: la Première Guerre mondiale* (Furie, la Peur, l'Enfer, la Rage, Délivrance).

Ce que dévoilent alors nos observations, c'est que les plus populaires de ces produits culturels, massivement distribués, fonctionnent à la fois comme

30 CLARKE Isabelle, COSTELLE Daniel, *Apocalypse: la Première Guerre mondiale*, 2014, 295 minutes.

31 CLARKE Isabelle, COSTELLE Daniel, *Apocalypse: la Seconde Guerre mondiale*, 2009, 320 minutes.

32 CLARKE Isabelle, COSTELLE Daniel, *Apocalypse: Hitler*, 2011, 104 minutes.

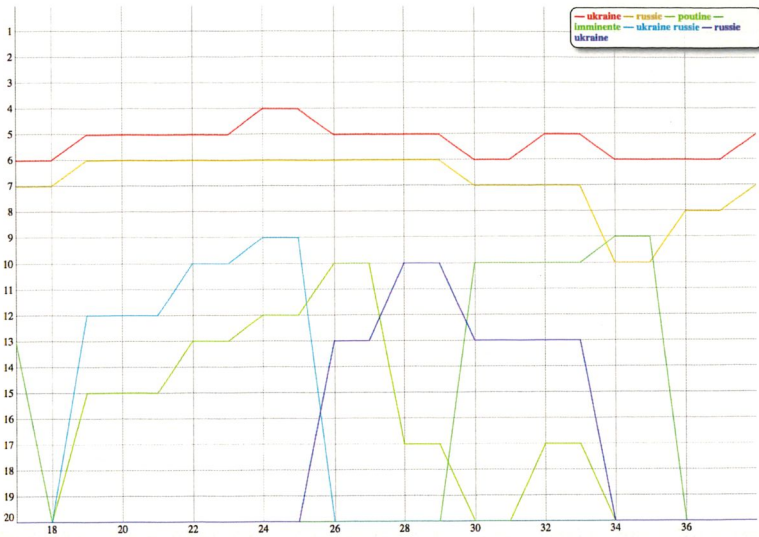


Fig. 6: Six autocomplétions particulières pour la recherche «guerre mondiale» mises en exergue par les chercheurs et leurs positions ordinales au fil des semaines

dispositifs qui mettent en avant une lecture historique spécifique et à la fois deviennent pour les internautes et pour les algorithmes une catégorie pertinente de recherche d'information sur ce contexte historique spécifique.

Un autre cas de figure nous montre que l'infrastructure numérique de Google, dans son association entre mots-clés significatifs, fait cohabiter les contextes historiques avec un futur hypothétique: en cherchant «guerre mondiale», l'une des suggestions algorithmiquement pertinentes est, comme l'on pourrait s'y attendre, le chiffre «1», pour Première Guerre mondiale. D'ailleurs, il en est de même pour le chiffre «2». Et alors que «guerre mondiale 1» et «guerre mondiale 2» font référence à des faits historiques avérés, il existe également la suggestion algorithmique «3», qui désigne donc un événement hypothétique. «guerre mondiale 3» cohabite avec les autres suggestions sans signe apparent de son caractère hypothétique.

D'autres associations algorithmiques à l'expression «guerre mondiale» montrent ce que Prezioso et Glassey ont précédemment appelé la «tyrannie du récent», à savoir les logiques de fonctionnement de cette infrastructure épistémique que constitue le web qui privilégie fortement le présent.³³

33 GLASSEY Olivier, PREZIOSO Stéfanie, «Auschwitz sur Facebook. Un livre d'or avant la visite», Culture & Musées, 20/1 (2012), 95-120.

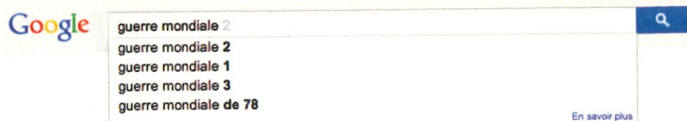


Fig. 7: Capture d'écran de Google pour la recherche «guerre mondiale»

Dans nos données, le «récent» se manifeste par exemple sous la forme de suggestions d'autocomplétion provenant de l'actualité géopolitique, comme «guerre mondiale ukraine», ou «russie», ou «poutine» (figure 5). D'ailleurs, la suggestion combinant les deux pays, «guerre mondiale ukraine russie» est inversée fin juin en «guerre mondiale russie ukraine». Ces processus sont incroyablement dynamiques: l'actualité, les centres d'intérêts des internautes, les modifications des algorithmes, d'autres modifications techniques (par exemple le design, l'interface utilisateur, etc.) contribuent tous à changer le paysage des résultats d'autocomplétion que nous avons observé.

Cependant, si la volatilité de ces associations calculées peut, dans sa forme extrême, soutenir les logiques de la «tyrannie du récent», il existe également la problématique opposée, c'est-à-dire la stabilisation erronée du passé. Le cas de la «guerre mondiale de septante-huit» en est un exemple flagrant. En effet, tapant une personne francophone qui tapait «guerre mondiale» dans le champ de recherche de Google avait de fortes chances, pendant l'été 2014, de se voir proposer non seulement les compléments «1» et «2», voire «3», mais également «de 78» (figure 6). L'origine de cette expression est une émission de télé-réalité diffusée début mars 2013, reprise dans les médias: «C'était la guerre mondiale de soixante-dix-huit»,³⁴ disait une participante décrivant une dispute. Nos données datant d'au moins un an plus tard, et s'étendant sur plusieurs mois, indiquent une stabilisation de l'autocomplétion «de 78» pour «guerre mondiale». Autrement dit, sur la base d'une «bourde» énoncée par une personne médiatisée à un moment donné, ainsi que sur la base du «buzz» qui a suivi – car ce sont bien ces répétitions qui fondent l'existence de cette «catégorie d'événement» – les algorithmes du moteur de recherche le plus utilisé en Europe suggèrent, même un an plus tard, aux internautes de s'informer

34 Voir «Clash entre Amélie et Aurélie (Les Anges 5), Téléloisir.fr, 4. 3. 2013, <http://www.programme-tv.net/videos/zapping/12939-clash-entre-amelie-et-aurelie-les-anges-5/>. A noter que «la guerre mondiale de soixante-dix-huit» est phonétiquement proche de «quatorze – dix-huit», ce qui est moins apparent dans les chiffres

sur une guerre qui n'a jamais existé. En regard de la dimension fortement dynamique des algorithmes d'autocomplétion, constater des formes de stabilisations est au moins aussi important que d'identifier des variations.

Vers des usages algorithmiques de l'histoire?

Nous l'avons vu, l'action des algorithmes des moteurs de recherche ne se situe pas uniquement au niveau des résultats d'une requête mais intervient également en amont, au moment même de sa formulation par l'utilisateur. Étudier quelles expressions sont associées à des termes liés à la Grande Guerre, et comment ces associations changent dans le temps, nous a alors permis d'analyser les différentes dynamiques de la représentation de la Grande Guerre par cette infrastructure informationnelle numérique qu'est Google. Nous avons pu identifier des formes d'écrasements épistémologiques produits par ce dispositif, qui non seulement retravaille sans cesse le passé à l'aune du récent, mais génère également de nouvelles catégories d'événements stables qui viennent peupler l'univers sémantique de la Guerre mondiale. L'examen de l'association d'expressions à des termes liés aux événements historiques nous permet ainsi de constater que les algorithmes d'autocomplétion ne font aucune différence ontologique entre un événement historique, un événement potentiel, voire une divagation. Ce constat est d'autant plus important qu'il illustre empiriquement comment les acteurs algorithmiques contemporains participent à une mémoire collective «filtrée». Ces derniers occupent alors une position stratégique et exercent, *de facto*, un pouvoir d'influence sur notre «présent du passé».

Cependant, notre recherche a également démontré qu'à l'intérieur d'un cadre bien posé, le fonctionnement algorithmique qui gouverne les représentations du passé est incroyablement dynamique, tel un kaléidoscope qui permet de voir d'innombrables motifs bariolés tout en contenant un nombre de formes et de couleurs bien défini. À l'intérieur d'une infrastructure informationnelle gérée par Google, les suggestions d'autocomplétion à caractère historique reflètent alors une mémoire collective existante puisqu'elle se fonde sur des mots-clés associés par les utilisateurs lors de précédentes recherches. En même temps, ces associations sont, à l'intérieur du cadre posé, très dynamiques puisqu'elles se recréent à l'usage, prenant en compte l'actualité ainsi que les actions des utilisateurs. Cependant, il nous semble crucial de ne pas sous-estimer le fait que les algorithmes d'autocomplétion sont un dispositif technique ayant une chance de dévier l'utilisateur de ses intentions initiales de recherche. Et alors que les résultats de recherche peuvent plus facilement être scrutés, discutés et remis en question, l'autocomplétion agit à la fois en amont et de manière subtile.

Force est de constater que différentes logiques agissent simultanément, et nous en avons montré des cas d'études pour plusieurs d'entre elles. Parmi les suggestions liées à la Première Guerre mondiale, certaines semblent clairement se rapporter à l'immédiat, comme celles concernant l'actualité géopolitique. De même, il existe des stabilisations qui relèvent elles du «récent», associant la Grande Guerre à des produits culturels grand public ou à un énoncé viral d'une vedette de télé-réalité. Nous avons également constaté que les mots-clés du champs lexical historique sont, à l'intérieur de la logique de l'autocomplétion, en compétition avec d'autres associations sémantiques. Il s'agit alors d'interroger sur la tension entre les différentes dynamiques algorithmiques et par là même la mémoire collective véhiculée.

00 Pour conclure, notre approche montre le rôle de plus en plus prépondérant que prennent les algorithmes dans nos manières d'utiliser les ressources des technologies de l'information pour appréhender le monde. Elle postule qu'il convient d'intégrer ces dispositifs techniques dans la réflexion sur les usages publics de l'histoire tels qu'ils se déploient dans leurs formes contemporaines. Ce terrain offre également l'occasion de s'interroger sur la relation changeante de nos sociétés aux récits et aux temporalités historiques. L'exemple des commémorations de la Première Guerre mondiale montre comment le passé se trouve sans cesse reconfiguré à l'aune d'un présentisme renforcé par les dispositifs techniques actuels. Cette dialectique entre l'horizon d'un présent où l'actualité dicte son agenda et la convocation d'une histoire qui oscille entre «nostalgie du passé», mémoire collective et histoire savante constitue alors l'un des espaces dans lesquels le pouvoir des algorithmes s'institue.

Anna Jobin

Chercheuse au Health Ethics & Policy Lab (ETH Zurich) et membre affilié au Laboratoire d'étude des sciences et des techniques (STS Lab) de l'Université de Lausanne. Ses recherches portent sur les logiques algorithmiques en ligne et les dynamiques de commodification du numérique.

Stéphanie Prezioso

Professeure associée à l'Institut d'études politiques, historiques et internationales (IEPHI) de l'Université de Lausanne, membre du conseil scientifique de la Mission du Centenaire de la Première Guerre mondiale et membre du Collectif de Recherche International et de Débat sur la Guerre de 14-18 (CRID).

Olivier Glassey

Maître d'enseignement et de recherche au Laboratoire d'étude des sciences et des techniques (STSLab) et à l'Observatoire Science, Politique et Société (OSPS) de l'Université de Lausanne et directeur du Musée de la main UNIL-CHUV.

101

Frederic Kaplan

Professeur assistant à l'École polytechnique fédérale de Lausanne (EPFL), où il occupe la chaire de Digital Humanities, directeur du Digital Humanities Lab (DHLab) et directeur du projet de recherche "Venice Time Machine".