



**HAL**  
open science

## De quoi le web est-il l'archive ?

Philippe Rygiel

► **To cite this version:**

Philippe Rygiel. De quoi le web est-il l'archive ? . Jean-Philippe Genet, Andrea Zorzi. Les historiens et l'informatique, un métier à réinventer, 444, École française de Rome, pp.291-310, 2011, Les historiens et l'informatique un métier à réinventer, 978-2-7283-0904-7. halshs-01856143

**HAL Id: halshs-01856143**

**<https://shs.hal.science/halshs-01856143v1>**

Submitted on 26 Jan 2023

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

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

PHILIPPE RYGIEL

## DE QUOI LE WEB EST-IL L'ARCHIVE ?

«Pas de documents, pas d'histoire».  
Langlois et Seignobos, *Introduction aux études historiques*

Je voudrais proposer ici quelques réflexions autour de l'usage d'Internet et plus particulièrement du World Wide Web comme source possible pour l'historien, parce qu'il me semble que la réflexion collective initiée par les ateliers ATHIS, qui se veut prospective, ne peut éviter de rencontrer cette question. Les pages qui suivent cependant ne peuvent proposer que quelques éléments fragiles, provisoires, témoignant d'une réflexion en cours ou plutôt d'une interrogation. Celle-ci est née d'une pratique d'enseignant, plus que de mes propres recherches. J'enseigne en effet dans un centre universitaire, le centre d'histoire sociale de l'Université Paris I, qui attire, et de plus en plus, des étudiants qui entendent travailler sur le très contemporain et parfois sur des phénomènes ou des institutions difficiles à saisir par le biais d'autres sources. Le web francophone en effet, qui a aujourd'hui une quinzaine d'années d'existence, est par excellence le lieu d'expression de mouvances culturelles ou militantes très minoritaires, dont l'étude attire certains de nos étudiants, qui se tournent donc assez logiquement vers les ressources offertes par le web afin d'amasser l'information dont ils ont besoin. Pour ne prendre qu'un exemple, j'évoquerai une étudiante qui, il y a quelques temps, a soutenu un mémoire dont l'objet était la mouvance hip hop en région parisienne<sup>1</sup>, petit milieu qui donne naissance à des productions culturelles qui expriment une identité définie, en partie, en référence à un substrat politique. Internet s'est révélé en ce cas comme un réservoir de sources et d'informations utiles, mais je me suis rendu compte qu'il m'était difficile, tant d'accompagner cette recherche, que parfois d'évaluer l'usage fait des ressources électroniques.

<sup>1</sup> M. Fouques, *Le Hip Hop face aux Zoulous. Une histoire des enfants de l'immigration*, mémoire de maîtrise, Université Paris, 2005.

Je pars donc d'un constat, celui d'une ignorance et d'une difficulté et je vise ici à en comprendre la teneur en précisant les termes de quelques questions simples. Puis je trouve sur le web des matériaux qui ont pour l'historien une utilité spécifique, c'est-à-dire auxquels je ne puis substituer d'autres matériaux? Quelles difficultés pose leur usage et sont-elles là encore spécifiques? Puis-je définir une posture, ou des stratégies, adaptées au matériau qui permettent de fabriquer des sources sans renoncer aux exigences de l'historien.

Pour éviter de donner à cette exploration un tour trop désincarné, j'ai choisi un thème de recherche qui constitue à la fois un véritable objet d'histoire sur lequel j'ai quelques lueurs et qui offre un terrain qui se prête à une enquête partiellement menée par le web. Je me suis donc demandé ce que la toile me permettait de dire des expulsions d'étrangers menées par la république française au cours des deux dernières décennies.

Cette politique, particulièrement brutale dans ses formes, n'est pas en soi une nouveauté. Durant les années trente déjà, les difficultés économiques conduisirent certains des gouvernements de l'époque à multiplier les arrêtés d'expulsion et de refoulement<sup>2</sup>. La présente crise présente cependant quelques traits spécifiques, dont l'un est qu'elle rencontre de fortes résistances au sein de la société civile qui se traduisent par la multiplication d'associations de collectifs et de réseaux ad hoc dont la principale ambition est de combattre les mesures adoptées ou d'en limiter les effets. Ces structures légères, sont originales à plusieurs titres. D'une part les résistances que rencontrent les politiques des années trente sont le fait de partis politiques, d'organisations syndicales, qui prennent en charge la question et qui défendent tant des principes que des individus dont certains sont membres, au moins théoriquement, de leur clientèle. D'autre part les structures légères que nous rencontrons aujourd'hui font un usage intensif des nouvelles technologies, nous pouvons retrouver des traces de leurs activités sur la toile. Ma question devient donc, dans un premier temps, puis je trouve, par le web, des matériaux que ne m'offriraient pas la presse ou les archives du ministère de l'intérieur?

Mon premier réflexe, puisque le web offre une multitude de discours, est d'y rechercher des tropes. Je cherche à repérer ceux qui parlent des événements qui m'intéressent, à décrire leurs discours. Mes ennuis commencent cependant lorsque je veux découper un corpus, le décrire et l'analyser.

<sup>2</sup> Voir P. Rygiel, *Le bon grain et l'ivraie. La sélection des migrants en Occident*, Paris, Publibook, 2009, troisième édition, première édition 2004.

### L'impossible corpus

La première difficulté est liée au nombre des éléments potentiellement pertinents. Une recherche sur Google, combinant les trois termes «étrangers», «expulsions» et «France», renvoie, en janvier 2009, à 583 000 documents, qu'il est bien sûr impossible de tous consulter et pour lesquelles nous ne possédons pas non plus de description pertinente au regard de nos préoccupations. Sachant que les moteurs de recherches ne parviennent jamais qu'à indexer une partie du web, sans d'ailleurs qu'il soit possible de déterminer ni la taille, ni les propriétés des parties manquantes, je ne peux que constater l'énormité de la tâche consistant à constituer un dossier documentaire complet et à en recenser les éléments.

Une autre difficulté pratique tient à la diversité des ressources qui me sont offertes. Je vais, fouillant les pages auxquelles je puis accéder, trouver des extraits d'émissions télévisées, des enregistrements sonores, des photographies, des textes de natures diverses (traités savants, articles de journaux, tracts ou pétitions électroniques, recueils de dispositions juridiques) soit un matériau dont l'exploitation n'obéit pas au mode de spécialisation habituel des historiens, qui sont souvent autant maîtres d'un ou de quelques fonds et types de sources qu'experts d'une question.

Le mode d'accès à l'information par le biais de requêtes sémantiques, ou pour mieux dire, par l'emploi de chaînes de caractères – qui est celui proposé par défaut par les principaux moteurs de recherches – a de plus pour effet, qu'en fonction des éléments retenus, nous obtenons des réponses différentes, tant par la taille du matériau auquel elles donnent accès que par le type de document qui sont proposés en premier lieu par le moteur de recherche. Soit deux requêtes formulées le même jour sur la page de garde de Google, l'une sous la forme (expulsions + France + étrangers), l'autre sous la forme («reconduite à la frontière» + France + étrangers), la première obtient 368 000 réponses, la seconde 17 600. Si nous classons, au moyen d'une typologie sommaire les requêtes proposées en page 1, nous trouvons dans le premier cas quatre sites d'information – émanation internet de médias traditionnels –, trois sites militants, un site juridique, un extrait d'encyclopédie et une page du site du Sénat. Dans le second cas, la première page nous renvoie à huit sites juridiques, à un site d'information (marocain celui-ci) et à une encyclopédie, qui n'est d'ailleurs par la même que la précédente. Ajoutons que si nous répétons la même requête à quelques minutes d'intervalle nous n'obtenons pas le même nombre de réponses (405 000 pour expulsions+ France + étrangers) et qu'intervenir

l'ordre des termes, sans modifier ni les chaînes de caractères ni les opérateurs booléens utilisés, aboutit à obtenir 627 000 réponses (expulsions + étrangers + France). Certains des éléments de ce constat ne sont pas très surprenants étant donnée la thématique choisie. Une bonne partie des luttes politiques en cours qui touchent aux questions d'immigration ont en effet pour enjeu la dénomination même des individus et des procédures et l'emploi de tel ou tel syntagme est, en soi, à la fois prise de position et intervention dans des luttes symboliques. La «rafle de sans papiers» s'oppose ainsi terme à terme à «l'arrestation de ressortissants étrangers en situation irrégulière». La conséquence en est cependant à la fois la démultiplication du nombre de pages potentiellement intéressantes et la difficulté à définir un champ sémantique pertinent, et surtout à valider les choix faits, puisque cela suppose une connaissance préalable de celui-ci.

#### Aux médiateurs inconnus

Ce simple exemple illustre d'autres difficultés liées à l'emploi du web et d'abord le fait que nous n'accédons jamais à celui-ci, mais obtenons d'une médiation les réponses aux requêtes que nous formulons. C'est bien sûr le cas aussi de l'historien qui travaille en archives ou en bibliothèques, dont le parcours est permis et guidé, voire infléchi, par des médiateurs. Deux différences importantes cependant subsistent. La première est que nous sommes familiers de ces médiations, en avons conscience et souvent les prenons explicitement en compte<sup>3</sup>, alors que l'idéologie internet, qui sacralise la transparence et l'immédiateté<sup>4</sup>, peut masquer à l'utilisateur le caractère toujours médié de ses vagabondages sur la toile. D'autre part, dans le cas de l'archive et de la bibliothèque, les opérations effectuées par les professionnels sont généralement documentées et explicitées dans des documents accessibles (plan d'inventaire, catalogue), ce qui n'est pas vrai des procédures mises en place par les opérateurs internet. Ceux-ci en effet ne sont généralement pas des opérateurs publics, tenus de rendre compte de leurs procédures, mais des sociétés commerciales dont le succès repose en partie sur leur habilité à garder secrets les algorithmes qui font la fortune de leurs actionnaires. Je suis ainsi incapable de savoir, l'algorithme de Google

<sup>3</sup> On pourra voir parmi les publications récentes en français, M., J. F. Fayet. et M. Porret, *Penser l'archive*, Lausanne, Antipodes, 2006, p. 274 et pour les bibliothèques *Matériaux pour l'histoire de notre temps*, 82, avril-juin 2006.

<sup>4</sup> Voir par exemple, P. Breton, *Le culte de l'Internet. Une menace pour le lien social?*, Paris, 2000, particulièrement p. 63 et suivantes.

n'étant pas public, pourquoi il ne me renvoie pas les mêmes réponses lorsque j'intervertis les termes d'une requête. Cette médiation a des effets sur le nombre de documents proposés, un autre moteur de recherche ainsi<sup>5</sup> renvoie 118 réponses pour «France+expulsions+étrangers», sur le choix des sites mis en valeur (parmi les 10 premiers sites proposés le même jour par cet opérateur, trois sont des sites officiels français, trois des sites juridiques, deux émanent de médias, l'un est un site de particulier, un seul est un site militant) mais aussi introduit probablement en soi un biais de sélection.

Nous ne pouvons bien sûr le montrer en comparant les milliers de pages auxquelles renvoient des moteurs de recherche généraliste, mais pouvons le vérifier à partir de sites spécialisés dans l'indexation d'une partie des ressources web, en l'occurrence les blogs. Si nous utilisons *blog dimension*, service francophone spécialisé dans l'exploration des blogs, nous trouvons, en réponse à notre requête habituelle, 1500 résultats. Parmi les premiers sites mentionnés figurent en bonne place des blogs dont les thématiques et le vocabulaire évoquent l'ultra gauche française, ainsi que deux sites pro-palestiniens et un site personnel essentiellement consacré à la défense et à l'illustration de la révolution islamiste iranienne. À l'inverse, l'utilisation d'un moteur anglo-saxon<sup>6</sup> ne me renvoie que quelques résultats, qui proviennent de pages personnelles de particuliers hébergées par le journal *Le Monde*, quand d'autres me proposent des pages m'offrant de louer une maison dans le Lot ou l'Aveyron, produites par des hôteliers amateurs. L'exemple bien sûr est malicieusement choisi. Il n'en illustre pas moins le fait, que croyant accéder au web, je ne fais que recourir aux services d'un ou de plusieurs médiateurs. Cela a des effets tant sur le volume des ressources auxquelles je puis accéder, que sur la nature de celles qui sont mises en avant et en valeur et affecte probablement la nature du corpus auquel je puis accéder, sans qu'il me soit possible ni de définir ni d'évaluer les biais introduits.

Le même exemple peut nous permettre d'illustrer une difficulté, pour l'historien, d'une toute autre nature. Je n'ai, lorsque je consulte un document internet, pas accès à un document, mais à une vue calculée selon des procédures complexes, qui sont définies les caractéristiques à la fois des machines utilisées et des codages informatiques employés, à partir de fichiers informatiques auxquels je ne peux souvent pas accéder. Tout un chacun a pu faire l'expérience de ce que, selon la machine ou l'écran utilisé, l'apparence de bien des

<sup>5</sup> <http://www.seek.fr/>.

<sup>6</sup> <http://www.blogdigger.com/>.

pages web varie. Nous sommes peut-être moins conscients de ce que, selon le moment de la consultation, ou la position de la machine que nous utilisons dans le réseau<sup>7</sup>, les éléments textuels obtenus peuvent eux aussi varier lorsque nous effectuons une même requête à partir d'un site dynamique, qui, sauf exception, ne nous permet pas d'accéder aux tables ou aux fichiers sources utilisés pour produire la réponse à une demande. Le moyen le plus simple de le vérifier est de proposer à un même moteur de recherche une requête identique à quelques heures d'intervalle, ou depuis deux postes n'appartenant pas au même sous-réseau. Il s'ensuit que, visualisant une page web, ou plutôt accédant à une Url, je ne consulte pas un document, ou fraction d'un document, au sens par exemple ou je pourrai examiner une liste de brevets, ou une liste de feux, mais obtiens la réponse à une question. De plus, ce que ce que je lis ne peut avoir le statut de document, au sens où les historiens habituellement l'entendent, que si j'enregistre, en utilisant un protocole qui en respecte l'aspect, ce que je viens moi-même de co-produire et l'archive en en garantissant ainsi l'identité.

### Les difficultés de la critique

Nous touchons là à une des caractéristiques premières des pages web, et d'autres ressources électroniques d'ailleurs, qui est la dissociation du support d'inscription de l'information et du support de visualisation de celle-ci (en l'occurrence la page web), qui n'existe que par l'instruction donnée par l'internaute, qui déclenche un processus génératif dont des pans, parfois très larges, lui sont non seulement de fait inconnus, mais aussi très souvent inaccessibles. Les conséquences pour l'historien en sont multiples. L'une d'elle en est que je ne peux mener une critique externe du document qui en établisse l'authenticité ou l'intégrité à partir de la réunion et de l'interprétation d'indices matériels. Les traces incorporées au document de sa genèse ne sont, dans le cas du document informatique, ni de même nature que celles que porte le document papier, ni surtout accessibles à celui qui visualise une page web, parce qu'éventuellement inscrites dans les entrailles de machines auxquelles il n'a généralement pas accès. La facilité avec laquelle les spécialistes du phishing produisent des copies quasi-parfaites de grands sites commerciaux, dont le caractère frauduleux n'est détectable par l'uti-

<sup>7</sup> Les grands moteurs de recherches commerciaux se sont engagés auprès des autorités d'un certain nombre de pays à bloquer l'accès à certains sites lorsque la requête reçue émane d'une machine identifiée.

lisateur qu'au moyen d'interactions directes avec des représentants de l'organisation attaquée, illustre ce point.

Ajoutons que la critique de provenance n'est, dans un certain nombre de cas, guère plus facile à mener. La première raison en est que les pages web, ou les fragments qui les composent, ne sont pas nécessairement datées. Certes il est possible, par divers moyens, dont le recours aux archives existantes du web<sup>8</sup>, de reconstituer divers états d'un site, mais d'une part cela nécessite à chaque fois enquête, d'autre part cela n'est possible ni pour tous les sites, ni pour toutes les parties d'un site. L'archivage d'un site stable et bien référencé se fait en effet généralement, c'est le cas de webarchives par exemple, par capture, à intervalles plus ou moins réguliers, de documents générés lors d'une consultation. Cela revient, un site web, quoique n'étant pas assimilable à un média de flux en partageant certains traits, à archiver un film en prélevant une image à intervalle de quelques minutes et, du fait à la fois de l'instabilité des Url et de la multiplication au cours de ces dernières années de sites dynamiques régulièrement reconfigurés, à ne proposer de ceux-ci qu'une vue très partielle. Le mode de consultation des dépôts d'archives numériques étant de plus généralement la définition et la documentation de sites, il est rare qu'ils permettent une recherche sémantique, ce qui interdit la constitution de dossiers documentaires rétrospectifs pertinents à partir de la définition thématique d'un champ ou par la définition d'expressions régulières ad hoc. En d'autres termes, je ne peux pas savoir ce que le web permettait de lire il y a deux ans ou trois ans lorsque l'on s'intéressait aux expulsions d'étrangers par l'état français, non plus que toujours dater les éléments qu'il est possible aujourd'hui de réunir.

Pour des raisons différentes, qui tiennent autant aux usages sociaux s'accompagnant de la production de ressources destinées au web qu'aux dispositifs techniques mis en œuvre, l'attribution à une instance ou à un auteur des documents web est parfois malaisée. Le fait est lié d'abord au développement des outils de production collaborative (qui ainsi est l'auteur d'une notice disponible sur un dispositif wiki? et faut-il comprendre ce qu'elle nous offre comme l'émanation d'une institution dont il serait possible de définir l'identité – une forme contemporaine de scriptorium – ou comme la résultante d'un processus socio-technique dont nous connaissons au fond mal les tenants). Bien d'autres productions pour le web, particulièrement celles issues de ce qu'il est convenu de nommer le web 2.0<sup>9</sup>,

<sup>8</sup> Par exemple <http://web.archive.org/>.

<sup>9</sup> Voir dans ce volume le texte de Serge Noiret.



conduisent à des questions similaires. Les blogs ainsi, particulièrement ceux qui évoquent des thèmes susceptible d'une politisation, voient se multiplier les contributions anonymes, fragments de textes dont nous ne connaissons pas les auteurs, sauf à considérer qu'ils sont, par le simple fait qu'il ne les a pas supprimés, élément d'un discours agencé par le propriétaire du blog. Nous voilà alors confronté à une multiplication de tracts dadaïstes dont les fragments juxtaposés seraient mobiles et parfois évanescents.

Bien sûr, je force un peu le trait, pour les besoins de l'exposé et certaines des difficultés rencontrées par l'historien qui explore le web ne sont pas sans analogies avec celles qui se présentent à des historiens découvrant, dans l'arrière salle d'un local syndicat, des piles emmêlées de tracts non signés et non datés, inventoriant patiemment les graffitis de Pompéï, ou consultant dans un fonds public une collection de lettres anonymes. Il n'en reste pas moins qu'il est des choses que le web ne permet pas à un historien qui entend l'utiliser en respectant les principes critiques qui fondent sa discipline et que ces impossibilités, ou ces difficultés, illustrent souvent des propriétés du matériau qui s'accordent mal aux pratiques spontanées des tenants d'une discipline dont les normes furent définies en référence à des documents que caractérisaient l'identité des supports d'inscription et de visualisation, la pérennité et la stabilité de ceux-ci – soit leur identité –, souvent garantie par une institution – archives ou bibliothèque – qui assurait leur conservation tout en facilitant leur identification, voire, souvent, en garantissait l'authenticité et l'attribution à un émetteur. Je ne peux ainsi pas constituer un corpus documentaire exploitable par l'historien à partir d'une recherche sémantique sur le web, parce que les données qu'il me fournit sont peu structurées, ou plutôt structurées par des opérateurs qui n'appartiennent pas à la même chaîne de savoir que moi et dont je ne peux connaître le mode opératoire, mais aussi du fait de la multiplicité des ressources et de la nécessité, pour identifier et dater nombre d'entre elles, d'une enquête dont les procédures ne peuvent être automatisées, ainsi qu'à cause de l'impossibilité d'une rétopolation des requêtes. Pour le dire dans la langue native des historiens, si le web ne me permet pas de découper un corpus thématique pertinent, c'est parce qu'il n'est ni une archive, ni une bibliothèque et qu'aucune instance identifiable n'a rassemblé, structuré et décrit les documents que je peux produire le consultant, ce qui transforme profondément les conditions de leur utilisation et en particulier les conditions de la critique, non que celle-ci soit impossible, mais elle ne peut s'appuyer ni sur les acquis d'une érudition ni sur l'usage d'indices et de systèmes de signes familiers aux historiens. Cela ne signifie ni que l'historien ne peut pas utiliser le document électronique, ni qu'il

ne le fera pas, ne serait-ce que parce que les futurs historiens de la période que nous vivons n'aurons guère le choix.

### *Y aller quand même*

#### Des instances sans papier et des pratiques sans nom

La première raison, de fait, est que de plus en plus l'activité d'un certain nombre d'institutions ne laisse de traces qu'électronique et, dans le cas par exemple d'un certain nombre de structures militantes, essentiellement sous la forme de ressources destinées au web. L'historien d'ailleurs y trouvera son compte. Le web en effet porte la trace d'échanges, de pratiques, dont nous avons pour les périodes anciennes peu de traces. La sociabilité autour d'un objet ou d'un projet, l'expression semi-publique de sentiments sur la chose publique ne sont pas ainsi choses nouvelles, mais nous n'avons pas toujours beaucoup de traces de ces pratiques, alors que le web, les blogs en particulier, regorgent de fragments de textes évoquant inmanquablement les conversations de comptoirs ou les discussions de membres de petits groupes unis par une même passion<sup>10</sup>. Je ne sais pas s'il faut considérer que nous avons là des pratiques nouvelles, ou bien une visibilité nouvelle et un changement de formes de pratiques sociales depuis longtemps vivaces, mais il me semble que s'ouvre un gisement de traces nouvelles et que l'historien, dont nous savons depuis longtemps qu'il est un ogre, ne peut que le lorgner avec gourmandise, même si ces traces seront difficiles à manier. Elles supposent, par leur volume et leur nature, à la fois la définition d'une diplomatique du document électronique, d'autant plus difficile à mettre en place que ses formes évoluent rapidement dans le temps<sup>11</sup> et des outils adaptés à la fois aux préoccupations des historiens et aux particularités du matériau, parmi lesquelles la taille des corpus à traiter. C'est là d'ailleurs une des particularités des documents pour le web, et plus généralement des documents électroniques, que d'imposer et de permettre à la fois des

<sup>10</sup> Pour une discussion de ce point voir F., *Le web 2.0 en perspective. Une analyse socio-économique de l'Internet*, Paris, ■■■ compléter.

<sup>11</sup> Les experts du W3C nous annoncent (et préparent) le successeur du web 2.0, qui prendrait la forme d'un web socio-sémantique associant les principes collaboratifs caractéristiques du web 2.0 et la multiplication d'agents intelligents permettant l'automatisation de tâches de recherche et d'interprétation en s'appuyant sur des ontologies locales, voir en français P. Laublet, *Web sémantique et ontologies*, dans C. Brossaud, B. Reber (dir.), *Humanités numériques 1, nouvelles technologies cognitives et épistémologie*, Paris, 2007.

manipulations de systèmes de signes nouveaux et des manipulations nouvelles de signes.

### De nouvelles inscriptions

Ainsi, l'une des caractéristiques des pages et des sites internet est qu'ils comportent généralement des liens à des ressources externes et sont eux-mêmes la cible de liens externes. Ces liens s'apparentent aux citations dont sont pourvues les publications universitaires, mais possèdent d'autres propriétés et d'autres fonctions. Ils sont évolutifs, ce qui permet, pour peu que l'on dispose de relevés espacés dans le temps, d'en constituer un historique. Ils constituent, du point de vue de l'utilisateur, des balises permettant l'élaboration d'un cheminement et peuvent être compris, si l'on se place du point de vue du concepteur du site, à la fois comme des appendices à son propre discours, qui contribuent à définir le destinataire attendu du site, la position du locuteur (l'impartialité peut ainsi s'exprimer par le renvoi explicite à des sites illustrant des positions opposées) ou bien l'appartenance (reconnue si les liens sont symétriques, revendiquée s'il est unidirectionnel) à une configuration sociale, le principe ou l'organisation de celle-ci pouvant être diverses. Ces dimensions ne sont bien sûr pas exclusives l'une de l'autre, non plus que notre analyse exhaustive. Nous pouvons donc à la fois reconstituer des parcours de lecture offerts aux internautes et des analyses, ou des descriptions, de réseaux sociaux, qui peuvent être conçus comme unissant des individus, des organisations ou des productions graphiques. Si nous partons d'un exemple, nous pouvons montrer assez rapidement que l'analyse peut se révéler féconde.

L'un des sites auxquels renvoie rapidement la requête (expulsions + étrangers + France) est un blog<sup>12</sup> titré «Immigration-Escoquerie-Sentimentale». Ouvert en novembre 2006, d'après la liste des posts précédents, cependant que le code source de la page donne comme date de création du document le 26 février 2006, il est encore actif en novembre 2008. L'en tête de celui-ci précise que l'auteur en est une femme, Marie-Annick Delaunay, par ailleurs auteure d'un ouvrage consacré aux mariages blancs entre Françaises et étrangers<sup>13</sup>. Le site de l'éditeur<sup>14</sup>, auquel renvoie le blog, nous apprend qu'elle est également la présidente d'une association «Non aux mariages et paternités de papiers», créée le 5 juillet 2004 et qui a

<sup>12</sup> <http://www.immigration-escoquerie-sentimentale.blogspot.com/>

<sup>13</sup> M. A. Delaunay, *L'immigration par escoquerie sentimentale*, Tatamis, 2006.

<sup>14</sup> <http://www.tatamis.fr/virtuelle.php/id/111213>, en février 2009.

elle même un site internet<sup>15</sup>. La consultation du code source de celle-ci, simple page html statique, donne comme date de création le 30/11/2004. Les textes lisibles sur la page de garde du blog et le premier post définissent la posture de l'auteure, mue par le souci de protéger des femmes (et exclusivement des femmes), de pratiques pouvant leur porter préjudice. Son action est selon elle une action «d'information et de prévention, [son] livre met en garde contre les escroqueries sentimentales qui se multiplient et appelle à un changement pour y mettre un terme»<sup>16</sup>.

La dimension politique de l'entreprise est cependant facilement repérable, au moyen d'une simple analyse des textes proposés à la lecture. Plusieurs de ceux présents en février 2009 sur la page de garde mettent directement en cause la politique menée par les gouvernements de gauche, coupables d'avoir facilité la tâche des escrocs. L'en tête du site précise ainsi que «Depuis la loi RESEDA de 1998 du gouvernement Jospin, on a assisté à une augmentation sans précédent des mariages mixtes contractés avec des citoyens français»<sup>17</sup>. Elle dénonce de plus vigoureusement l'indifférence des associations et institutions proches de la gauche française engagées dans la défense des droits des étrangers, écrivant ainsi que :

les associations anti-racistes sont bien silencieuses chaque fois qu'un crime d'honneur, une séquestration ou des repréailles à coups de couteau sont pratiquées contre une jeune fille obligée de se soumettre aux diktats de la famille, voire de la tribu. Elles sont aveugles, sourdes et muettes ... la Halde, la LDH, le Mrap, le Gisti, la Cimade, ...et tous les ayatollahs de l'antiracisme<sup>18</sup>.

La page de garde du blog comporte, fin février 2009, 6 liens externes. L'un mène au site de l'éditeur de l'ouvrage mentionné plus haut, l'autre à celui de l'association dont l'auteure est la présidente. L'un conduit au site d'une association, «l'enfant d'abord<sup>19</sup>», qui se présente elle aussi comme une association apolitique, soucieuse du bien des enfants et qui mène un combat contre les modalités de garde alternée prévue par la loi de mars 2002. Attentifs aux débats parlementaires, auxquels renvoient le site, les dirigeants de cette association ont obtenu en juin 2008 une audience auprès de la ministre de la famille. Trois autres liens renvoient à des articles parus dans la presse et disponibles sur Internet, deux sont parus

<sup>15</sup> <http://2jong.free.fr/associationnon/informations.php>, en février 2009.

<sup>16</sup> [http://immigration-escroquerie-sentimentale.blogspot.com/2006\\_11\\_01\\_archive.html](http://immigration-escroquerie-sentimentale.blogspot.com/2006_11_01_archive.html).

<sup>17</sup> <http://www.immigration-escroquerie-sentimentale.blogspot.com/>.

<sup>18</sup> *Idem*.

<sup>19</sup> <http://www.lenfantdabord.org/>.

dans le Figaro, un dans la Dépêche du Midi. Le dernier lien, est destiné à mettre en garde les Françaises se rendant en Afrique, en leur offrant de lire ce que l'on pense d'elles au Sénégal. Le texte auquel le lien renvoie<sup>20</sup>, d'une vulgarité crasse, évoque la «grosse conne européenne» qui (...) n'est généralement pas une «fille facile» [mais] ne dédaigne pas coucher avec un «Africain» qui évidemment rêve de rester au Sénégal et n'a aucune arrière-pensée. Elle s'en sort au mieux avec des mycoses sur la chatte au pire avec un DAS qu'elle aura bien mérité. Au moins à son retour, elle coûtera cher à la sécurité sociale mais pas longtemps et ne coûtera rien à la collectivité pour sa retraite.

La cible du texte est politiquement définie par un petit texte en forme de question qui placé en pied de page précise l'acception des termes employés :

Je suis un rebelle, j'ai campé dans le Larzac à l'appel de José Bové, suis-je un?

J'ai bien peur que oui. L'altermondialiste anarchiste est quasiment toujours un gros con persuadé de faire parti des gentils de cette planète. [il] s'est généralement fait taper sur la gueule «par des éléments incontrôlés du 9.3.» lors des manifs contre le CPE et ce n'est que justice<sup>21</sup>.

Le ton et les procédés linguistiques employés évoquent un anarchisme de droite à la française, qui lorgne autant vers Frédéric Dard que vers Céline. Ce texte peut être obtenu depuis un site qui se présente comme une encyclopédie en ligne du Sénégal et un site d'information utile tant aux Sénégalais de l'étranger qu'aux touristes prévoyant un séjour dans ce pays. Les descriptions des parcs nationaux y côtoient les listes d'astuces permettant d'obtenir des billets d'avion bon marché. Un article, préservé par une archive numérique sénégalaise<sup>22</sup>, nous apprend que l'auteur, un Français propriétaire du site, a été condamné par contumace pour diffamation au Sénégal et, résidant en Espagne, est à l'abri des poursuites. Plusieurs de ses textes, sont repris par les auteurs de site personnels ou de blogs qui souvent, même lorsqu'ils sont rattachés au très sérieux journal Le Monde, semblent penser qu'il s'agit là de l'expression d'une opinion répandue parmi les Sénégalais<sup>23</sup>. Un certain nombre de sites et de

<sup>20</sup> [http://www.senegalaisement.com/senegal/consconnes\\_senegal.php](http://www.senegalaisement.com/senegal/consconnes_senegal.php).

<sup>21</sup> *Idem*.

<sup>22</sup> <http://www.osiris.sn/article715.html>.

<sup>23</sup> Par exemple, «Voilà sur quel ton on parle de nous, les femmes françaises en particulier, au Sénégal», <http://coirault-neuburger.blog.lemonde.fr/2007/12/09/voila-sur-quel-ton-on-parle-de-nous-les-femmes-francaises-en-particulier-au-senegal/>, 9 décembre 2007.

forums, exprimant explicitement des thèses d'extrême-droite, reprennent également ces textes<sup>24</sup>.

Le dernier lien enfin renvoie au site personnel d'un autre journaliste et auteur, André Bercoff, auteur d'un *Précis de la décomposition française*, qui propose en ligne de courtes séances vidéo le mettant en scène commentant l'actualité du moment. Son blog comporte une dizaine de liens externes dont la plupart renvoient aux sites de média généralistes et à quelques sites au positionnement idéologique brouillé, tel celui d'une «Riposte Laïque» dont la page de garde arbore une belle cocarde tricolore et défend une république laïque et sociale dont l'un des héros est le parlementaire hollandais Geert Wilders, célèbre pour ses propos islamophobes<sup>25</sup>.

Les liens entrants (tels que repérés au moyen des outils proposés par Yahoo<sup>26</sup>) sont plus difficiles et plus longs à analyser, du fait de leur nombre (90 en février 2009<sup>27</sup>). Quelques traits peuvent cependant être dégagés. Ils émanent de sites plus ouvertement politiques que le site auquel ils renvoient, soit des sites souvent de militants actifs, d'élus ou d'organisations appartenant soit aux franges les plus conservatrices de la droite de gouvernement, soit aux diverses sensibilités de l'extrême droite. Radio-courtoisie ainsi, un blog de la droite nationaliste wallone<sup>28</sup>, ou un site estampillé MNR, dirigent l'internaute vers le site de l'association, ainsi qu'un fort intéressant répertoire de sites à l'usage des droites européennes, proposant l'accès à «1000 sites francophones de réinformation» permettant de lutter contre.

Les idéologies imposées par les médias [qui], sous différentes variantes, se résument essentiellement à deux idéologies principales : le socialisme et le mondialisme/anti-conservatisme/antichristianisme. La première idéologie veut imposer les utopies étatiques que sont l'«anti-discrimination» et le nivellement par l'égalitarisme et le collectivisme au moyen d'un état tentaculaire et omniprésent,

<sup>24</sup> Par exemple <http://www.france-echos.com/>.

<sup>25</sup> *Merci M. Wilders d'avoir lancé le débat*, <http://www.ripostelaique.com/Merci-M-Wilders-d-avoir-lance-le.html>.

<sup>26</sup> Le module de recherche d'inlinks de Yahoo est actuellement l'outil le plus efficace accessible par une interface web, l'outil similaire proposé par google étant manifestement bridé, <http://siteexplorer.search.yahoo.com/>.

<sup>27</sup> La signification de ceux-ci, et donc leur interprétation et les manipulations des liens repérés, est sujette à variations dans le temps. La multiplication des blogs ainsi a pour conséquence qu'un webmestre peut obtenir un lien vers son site en allant lui même inscrire sur le blog d'un confrère un lien vers celui-ci. Les contributions étant cependant modérées par le propriétaire du site, nous pouvons considérer que qui ne dit mot consent et que le maintien du lien vaut assentiment à l'arrimage des deux blogs.

<sup>28</sup> <http://reconquista.over-blog.com/article-7277391-6.html>.

mais souvent irresponsable avec son hyper fiscalité et ses bureaucraties pléthoriques nationales et bientôt européistes<sup>29</sup>.

Ces données nous permettent alors plusieurs constats. L'un tient aux stratégies d'énonciation de l'auteur du blog. Celui-ci met en circulation, ou relaie, un certain nombre de figures ou de récits qui confortent une idéologie qui est déniée en tant que telle, d'où une stratégie d'établissement de liens externes qui renvoient à des ressources documentaires garantes de sérieux de l'information, mais à l'affichage idéologique discret ou délibérément brouillé. À l'inverse, les webmasters des sites émanant – et se présentant comme tels – de l'extrême droite francophone renvoient à ces récits, que leur positionnement définit comme objectifs et neutres, afin de conforter par des faits des postures idéologiques. L'exemple suggère une stratégie de manipulation de l'opinion et la coordination – qui ne suppose ici ni inter-connaissance ni délibération – de groupes institués et de francs-tireurs idéologiques déposant discrètement des toxines électroniques dont la zone de diffusion correspond à un espace linguistique et non à un espace national. Le constat appelle bien sûr d'autres questions, tenant à la fréquence et à l'efficacité de ce travail idéologique, auxquelles cependant nous ne pouvons répondre sans changer d'échelle et de mode d'observation.

Si, en effet, les indices obtenus par les manipulations permises par l'observation des inscriptions que sont les liens externes des sites examinés sont pertinents, pour le sociologue ou l'historien et, combiné à d'autres, manifestent la possibilité d'enquêtes intégrant des sources électroniques. Nous nous heurtons assez vite à trois limites. La première tient au fait que l'espace ainsi défini n'intègre pas de dimension temporelle. Les sociologues ou les spécialistes d'autres sciences sociales y pallient parfois en se faisant les archivistes de leur propre recherche, effectuant des coupes à intervalles plus ou moins réguliers, ou bien enregistrant leurs données sous la forme de flux d'informations datées<sup>30</sup>. L'historien ne peut procéder ainsi qu'autant qu'il peut s'appuyer sur une archive électronique constituée. La seconde tient au nombre de manipulations à effectuer si nous voulons cartographier, non pas l'espace que définit un site, mais celui des prises de position sur une question ou des configurations définies par des renvois de site à site. Le relevé et le suivi manuel des liens repérés est matériellement impossible. Il faut alors soit renoncer, soit se doter d'une instrumentation adaptée. Celle-ci existe<sup>31</sup>, est parfois facilement accessible et gratuite, mais elle n'est

<sup>29</sup> [http://www.esprit-europeen.fr/agora\\_liens\\_1000\\_sites\\_francophones](http://www.esprit-europeen.fr/agora_liens_1000_sites_francophones).

<sup>30</sup> F. Granjon, *L'Internet militant : Mouvement social et usage des réseaux télématiques*, Rennes, 2001.

<sup>31</sup> Sur l'état de la question, pour l'étude des liens hypertextes, H. Park et M

pas conçue pour les besoins des historiens. Certains de ces outils ainsi n'intègrent pas de modules permettant de prendre en compte les évolutions dans le temps des espaces décrits<sup>32</sup>. Bien peu d'historiens de plus pourraient prendre en main ces instruments, d'abord parce qu'ils nécessitent quelques notions d'informatique, surtout parce qu'ils ne parviennent à proposer des descriptions et des représentations graphiques des configurations décrites qu'au prix d'une formalisation mathématique complexe (en l' occurrence les techniques de description des grands graphes) et constamment renouvelée. Or, de toutes les sciences sociales, l'histoire est sans doute celle qui initie le moins ses étudiants et ses futurs maîtres à la culture mathématique et informatique, pour des raisons que Jean Philippe Genet, constatant que «Former les historiens à l'informatique n'est pas facile»<sup>33</sup>, exposait il y a près de quinze ans et qui n'ont au fond guère changé depuis.

### Expressions régulières et interprétation

Ce sont là des difficultés que nous retrouvons, sous des formes plus ou moins aiguës, non seulement quand nous entendons utiliser les traces propres au web, mais encore quand nous voulons exploiter les possibilités de recherches et de manipulation de chaînes de caractère qu'il autorise, en tant qu'il est un très grand corpus textuel ouvert. La recherche de syntagme ou de tropes, sous ses différentes formes, génétique ou bien de propagation, est une pratique érudite ancienne et efficace. Ce que permet ici est le web est de la mener à partir d'un corpus de très grande taille, ouvert, c'est-à-dire non a priori défini par le chercheur et sans limites physiques pesant sur le nombre de tropes identifiables et à partir d'un matériau – citation et reproduction étant des caractéristiques de l'écriture pour le web – permettant d'assigner une signification à ces pratiques en les rapportant aux pratiques sociales et aux propriétés des acteurs. Prenons un exemple simple. Nous pouvons considérer qu'une partie de l'action militante d'un certain nombre d'organisations consiste en la mise en récits de processus sociaux ou politiques, destinés, par

Thewall., *The Network Approach to Web Hyperlink Research and its Utility for Science Communication*, dans C Hine, *Virtual Methods : Issues in Social Research on the Internet*, Berg, Oxford, 2005.

<sup>32</sup> S. Rorgers, *Mapping Web Space with the Issuecrawler*, <http://www.govcom.org>, 2006.

<sup>33</sup> J.-P. Genet, *La formation des historiens à l'informatique en France : espoir ou désespoir?* dans *Le Médiéviste et l'Ordinateur*, n° 31-32, 1995, p. 1, disponible en ligne à <http://lemo.irht.cnrs.fr/31-32/mo3115.htm> (février 2009).



leur dissémination, à provoquer une mobilisation ou à transformer des façons de nommer ou de classer certains types d'événements. Une analyse de ces récits peut conduire à mettre en évidence des chaînes de caractères les caractérisant et autoriser à évaluer leur diffusion à partir d'un point d'origine identifié. Un blog, rattaché au journal *Libération*, est ainsi utilisé par des proches ou des membres du Réseau Éducation Sans Frontières<sup>34</sup> qui y déposent de petits textes, souvent de courts récits, qui suscitent, ou non, des commentaires, généralement anonymes, ce qui bien sûr rend difficile leur interprétation. Il est possible par contre de suivre le devenir de certains de ces textes sur la toile au moyen de quelques mots clés. La petite histoire narrée par Martine Vuaillet en juin 2007, qui évoque sa visite dans un centre de rétention, a ainsi été abondamment reprise et commentée sur la toile. Rendant visite à une famille placée dans un centre de rétention elle avait «Avant d'entrer dans la cellule de visite, (...) laissé sur le comptoir un livre d'enfant qui s'appelle *Au panier* (...) en disant avec un grand sourire «ça vous fera de la lecture». (C'est une histoire d'hommes en uniforme qui mettent au panier tout ce qui est différent)»<sup>35</sup>. Les fonctionnaires en fonction ont-ils peut-être apprécié l'attention, au point que le «lieutenant Sèvre (...) m'a dit qu'il me confisquait l'ouvrage pour subversion». Retenue pour vérification d'identité, la locutrice s'est vu retirer l'autorisation de visiter le centre. Après plusieurs essais, il apparaît que la combinaison de chaînes de caractères la plus efficace pour retrouver trace de cette fable véridique est (vuaillet+rétention +visite +livre). La requête effectuée sur *Google* permet d'atteindre plusieurs centaines de pages évoquant l'épisode. Nous ne pouvons bien sûr en mener une analyse systématique dans le cadre de cette étude. Notons que l'on retrouve trace, sans surprise, de sites émanant d'associations et de forces politiques engagées dans la défense des droits des étrangers, mais aussi de nombreux blogs et sites personnels dont les auteurs expriment souvent leur rapport au récit par un retitrage de celui-ci (nous trouvons «Le lieutenant qui censurait les livres», «Attention subversion» «subversive ou courageuse» et bien d'autres) qui fait souvent de la censure d'une publication subversive par un officier (et non des conditions de vie et de séjour dans les centres de rétention) le pivot de l'historiette et l'objet du scandale. Nous pouvons ainsi là encore produire des sources permettant de qualifier une stratégie communicationnelle et réfléchir à ses effets. Pour revenir à nos préoccupations premières, Internet

<sup>34</sup> *Une lecture pour enfants jugée subversive*, [http://sanspapiers.blogs.liberation.fr/sans\\_papiers/2007/06/une\\_lecture\\_pou.html](http://sanspapiers.blogs.liberation.fr/sans_papiers/2007/06/une_lecture_pou.html).

<sup>35</sup> *Idem*.

nous offre ici les moyens d'une très classique analyse de réception, qu'il facilite et infléchit en permettant le repérage très rapide d'une masse de textes reprenant un motif, dont beaucoup sont produits par des locuteurs s'exprimant en tant qu'individus, sans pour autant que ne soit levée les interrogations tenant à la nature du corpus ainsi constitué, ni à la signification des prises de parole dont nous retrouvons la trace électronique.

De plus, là encore, si nous voulons changer d'échelle, c'est à dire produire une table ou une carte de distribution d'un certain nombre de syntagmes que nous pourrions rapprocher d'une table de liens, soit construire un dispositif savant inédit dont nous pouvons attendre des effets cognitifs, il nous faudrait passer à la fois par l'explicitation de règles de manipulations (extraction, association, tri) des expressions régulières et des symboles graphiques présents dans les pages visualisées et passer, soit par l'écriture d'un algorithme, soit par l'usage d'un outil logiciel adapté. Cela suppose à la fois des compétences et des choix lourds d'enjeux, en particulier parce ce qu'ils impliquent ou induisent sur les formes de division du travail et de coopération au sein de la communauté scientifique<sup>36</sup>.

### *Stratégies historiennes*

#### L'instrumentation de l'histoire

En l'état actuel des choses donc, le web en tant que source, du fait de ses caractéristiques et de sa faible profondeur, peut rendre à l'historien du très contemporain de signalés services et promet aux historiens futurs une masse d'information dont l'exploitation pose de sérieux problèmes à qui ne veut pas renoncer aux exigences de la discipline. Nous ne pouvons, en tant qu'historien, guère que proposer quelques pistes de réflexion et quelques éléments de stratégie. Le premier constat est que les matériaux qu'offre le web ne peuvent être maîtrisés par l'historien, c'est à dire manipulés selon les règles du métier et exploités dans toute leur richesse, qu'au prix du renouvellement de son instrumentation. L'un des intérêts de l'émergence du document électronique et de sa distribution en réseau est sans doute d'ailleurs de nous permettre de repenser le travail de l'historien et plus largement celui des spécialistes de sciences humaines, en le redéfinis-

<sup>36</sup> Le débat entre tenants d'une production locale d'outils logiciels adaptés aux besoins d'une communauté scientifique et partisans d'une appropriation d'outils génériques est ancien et récurrent et ce dans tous les champs des sciences humaines. On en trouvera un écho sociologique dans F. Chateauraynaud, *Prospero : une technologie littéraire pour les sciences humaines*, Paris, 2003.

sant comme une manipulation instrumentée et réglée de symboles connectés à des traces matérielles, ce que savaient pertinemment les fondateurs de la discipline qui, tels Langlois et Seignobos, ne passaient guère moins de temps à discuter de la bonne façon d'établir et de tenir les fiches permettant l'exploitation de leurs documents-textes, qu'à réfléchir aux exigences de la critique<sup>37</sup>. Adapter ces principes aux documents électroniques suppose, plus que l'achat de machines, que les historiens, une partie d'entre-eux du moins, aient quelques teintures d'une culture informatique et d'une culture mathématique qui leur permettent sinon de construire leur propres outils, du moins de dialoguer avec des scientifiques et des ingénieurs, sous peine de devoir utiliser des dispositifs élaborés sans tenir compte de leurs besoins propres ou de ne pouvoir exploiter une bonne partie des matériaux produits par les sociétés du très contemporain. Nous retrouvons là une thématique familière aux praticiens de l'histoire quantitative et de l'informatique historique<sup>38</sup>, à ceci près, et le changement est d'importance, que la question de l'instrumentation de l'historien n'est pas posée à une minorité de spécialistes traitant de questions souvent jugées absconces par une bonne partie de la profession, mais à tous ceux qui entendent ou entendront produire une connaissance du très contemporain exploitant le type d'inscriptions laissées par ses institutions, voire tous ceux et ils seront de plus en nombreux, qui accéderont aux sources nourrissant leur recherche par le biais de documents électroniques. La tâche est rude. Elle suppose que la formation initiale des historiens fasse une part plus grande à la culture scientifique et technique, ce qui ne pourra se faire que lentement et suppose une réflexion collective<sup>39</sup> mais aussi qu'émergent quelques filières permettant la formation de praticiens solidement spécialisés dans l'élaboration de l'instrumentation scientifique, dont le travail soit reconnu pour ce qu'il est, un travail scientifique dont la communauté historienne a collectivement besoin, si nous voulons éviter de dépendre de l'apparition aléatoire de générations spontanées composées d'individus au parcours atypiques, campant dans des positions marginales, ce qui est la règle, en France du moins, depuis plusieurs décennies. Elle le sera aussi parce que, contrairement au livre, dont les propriétés sont connues et stabilisées depuis longtemps, le document électronique demeure un devenir, le produit tran-

<sup>37</sup> C. V. Langlois et C. Seignobos, *Introduction aux études historiques*, Paris, 1992, première édition 1898.

<sup>38</sup> Voir, par exemple, *Postface*, dans *Le Médiéviste et l'ordinateur*, n° 31-32, 1995, <http://lemo.irht.cnrs.fr/31-32/mo3120.htm>.

<sup>39</sup> Voir A. Dallo, *Quelle informatique enseigner aux historiens. L'exemple de l'Université Paris I*, dans *Le médiéviste et l'ordinateur*, numéro 41, hiver 2002.

sitoire d'une évolution en cours. Les compétences qui permettent son recueil et son exploitation sont aussi coûteuses à acquérir que rapidement obsolètes, supposant, pour être constamment reconquise, plus l'accès à une culture scientifique que la maîtrise de quelques procédures rapidement abandonnées.

### L'urgence archivistique

Ces mutations rapides posent un autre problème, qui est celui des conditions de l'archivage du web et plus largement des documents sous forme électronique. La période actuelle se caractérise par la superposition d'initiatives publiques et privées destinées à préserver ceux-ci, sans cependant que la question de la pérennité à long terme des ressources électroniques, qui pose des problèmes tant matériel que logiciel, ne soit réglée. La participation des historiens à la réflexion, déjà entamée depuis plusieurs années les documentalistes et les archivistes<sup>40</sup>, apparaît nécessaire, de même que leur contribution, malheureusement trop rare, aux initiatives visant à produire, à partir de la définition d'un objet d'étude ou d'un cadre thématique, des recueils de données électroniques, tel celui porté par le Centre d'Histoire Sociale du XX<sup>e</sup> siècle afin de permettre l'étude future des « nouveaux mouvements sociaux » de la période contemporaine, dont l'action souvent s'inscrit d'abord sur des supports électroniques quoiqu'elle ne prenne pas toujours ou pas surtout la forme de pages html destinées à une diffusion par le biais du web<sup>41</sup>. Refusant le fantasme contemporain de l'archivage universel, qui méconnaît le fait que l'archive est toujours à la fois le réceptacle de vestiges ayant échappé à un processus de destruction et le produit de choix, le projet prévoit l'identification d'acteurs clés, le recueil et l'indexation de leurs productions numériques.

### Poser les questions auxquelles il est possible de répondre

La réflexion engagée, parce qu'elle est œuvre d'archivistes et de documentalistes, intègre à la fois une connaissance de l'objet appréhendé et une réflexion sur les particularités du matériau. À l'historien qui voudrait déjà, étudiant le très contemporain, s'emparer du matériau produit par l'intermédiaire d'Internet, elle offre des pistes utiles. Si l'entreprise en effet est possible, c'est parce qu'elle définit une stra-

<sup>40</sup> P. Rygiel, *Des archives numériques sans historiens?*, dans *Matériaux pour l'histoire de notre temps*, numéro 79, juillet-septembre 2005.

<sup>41</sup> F. Blum, B. Groppo R. Vaccaro et F. Veyron (dir.), *Internet et mouvements sociaux : nouvelles pratiques militantes, nouvelles sources pour l'histoire*, dans *Matériaux pour l'histoire de notre temps*, n° 79, juillet-septembre 2005.

tégie savante qui rompt avec les pratiques spontanées de l'internaute en définissant des entités dont il est pertinent d'organiser la préservation – en collaboration avec des institutions publiques, les acteurs eux-mêmes et des historiens – et dont il serait possible de poursuivre l'étude. Le projet part en effet d'une connaissance préalable d'un champ d'intervention socio-politique, qui permet le repérage et le choix d'acteurs que caractérisent à la fois leur usage intensif des outils électroniques et leur position au sein du champ. Cela fait de leur production électronique l'un des éléments d'un dossier portant trace de leur activité, soit un fonds d'archives, attaché à une entité dont les propriétés peuvent être définies. L'historien retrouve un cadre connu, qui est celui qui permet son activité.

Certaines des difficultés que nous avons évoqué en ouverture de cet article provenaient en effet de que le web n'est pas une archive, même s'il est en partie archivable et donne accès à des archives, quoique la permanence de celles-ci ne soit pas toujours certaine. À ce titre il est utilisable, en conjonction avec d'autres sources, par les historiens du très contemporain qui n'ont d'ailleurs guère le choix. Ce matériau impose cependant des contraintes spécifiques, qui obligent à définir des stratégies d'enquête et des tactiques manipulatoires adaptés, tant aux questionnements propres de l'historien qu'aux caractéristiques de son matériau, dont nous savons qu'elles sont parfois inédites et souvent fluctuantes. Désorientation et surcharge cognitive guettent souvent l'utilisateur de données complexes et hypertextuelles et l'historien ne fait pas exception à la règle, qui s'y aventure sans le secours d'une sémiotique stabilisée<sup>42</sup> ni celui d'une diplomatique exprimant une tradition érudite. Si un usage raisonné de ces ressources suppose l'appropriation, coûteuse, d'une instrumentation nouvelle, il peut en espérer des bénéfices cognitifs, dont les moindres ne sont sans doute pas l'invitation à une réflexion sur ses propres pratiques et la possibilité de demeurer fidèle à sa vocation, qui est, par la manipulation inventive et réglée d'inscriptions, la production d'un savoir à la fois critique et réflexif.

Philippe RYGIEL

Centre d'histoire sociale du XX<sup>e</sup> siècle (Paris I/CNRS)<sup>43</sup>

<sup>42</sup> Les spécialistes de l'architecture des sites insistent ainsi sur la rapidité avec laquelle se transforment les modes d'organisation de présentation et de visualisation des sites web, qui sont en eux mêmes, l'expression à la fois d'une visée et d'une organisation productive, voir par exemple Paul Kahn et Krzysztof Lenk, *Mapping Web Sites*, Amsterdam, 2001.

<sup>43</sup> Je remercie, les premiers lecteurs de ce texte, Julien Alerini, Alain Dallo et Éric Guichard, pour leur relecture avisée et leurs remarques sagaces.