École des hautes études en sciences sociales

L'internet : mesures des appropriations d'une technique intellectuelle

Éric Guichard

Thèse de doctorat en sciences de l'information et de la communication
sous la direction de Monsieur Hervé Le Bras,
Directeur d'études à l'EHESS

Autres membres du jury :
Jack Goody, Christian Licoppe, Jacques Perriault, Serge Proulx

17 octobre 2002

Remerciements

Je voudrais tout d'abord remercier M. Hervé Le Bras, qui a accepté de diriger ce travail : dans ce cadre comme dans d'autres, relatifs à des passions communes, il m'a encouragé à combiner les instruments de diverses disciplines, et m'a offert de précieux conseils, notamment en matière de mesures des pratiques sociales.

Je voudrais aussi remercier Hervé Théry, à la fois mentor et complice, qui a le grand défaut d'être géographe — défaut si bien partagé par les lecteurs attentifs de cette thèse — et qui, ce jour, est au Brésil. J'évoque très brièvement la géographie dans ce travail, et pourtant mes proches savent à quel point elle m'est essentielle. Représentations (mathématiques, territoriales, collectives), cartes, voyages, boussoles... Je n'en dirai pas plus. D'autres amis comme Henri Desbois, Christian Jacob, Catherine Lefebvre et Jean-Claude Moissinac, qui m'ont tant aidé dans ce domaine, me comprendront.

Je mentionne avec plaisir l'apport intellectuel de Jean Bazin, Rupert Hasterok, Clarisse Herrenschmidt, Jacques Lajoie, Philippe Matherat, Paul Mathias, Luc Tournon et Russon Wooldridge, dont la grande culture et l'originalité intellectuelle m'ont beaucoup aidé.

Je n'oublie pas les amis qui ont relu attentivement ce travail : Vincent George, Michel Nguyen Thé, Philippe Rygiel (et d'autres, déjà cités), ni ceux qui m'ont encouragé à le réaliser : tout d'abord Jean-François Chanet, mais aussi Marie Bauwens et Paul Jouanneaux.

Ni encore les personnes qui, au travers de leurs institutions, m'ont permis de le réaliser dans les meilleurs conditions : Étienne Guyon, ancien directeur de l'ENS, et Brigitte Trousse, de l'INRIA. Ce qui me donne aussi l'occasion de remercier d'autres institutions, en premier lieu le CNRS et le RNRT, qui ont été de réels catalyseurs du type de travail que je présente ici.

Merci encore à Jacques Beigbeder, Roberto Di Cosmo et Catherine Le Bihan pour leurs conseils informatiques et leurs intuitions sociologiques en ce domaine. À Jean-Claude Chamboredon, à Gérard Noiriel et aux amis de *Pugwash*, dont la probité et l'exigence intellectuelles sont toujours réconfortantes dans l'exercice difficile de la recherche. Enfin, à Gabriel Akol Ayok et Halima Khaldi, qui furent les premiers à me donner concrètement le goût des sciences sociales, et à tous mes amis, même s'ils ne sont pas cités ici.

Et, bien sûr, mille et un mercis à Émilie.

Paris, le 26 septembre 2002

VRINTI·REGNANTE

Introduction

Peut-on définir un cadre conceptuel qui permette d'appréhender l'internet ? Tantôt ensemble de câbles, tantôt bibliothèque de protocoles, réseau dont les nœuds sont des ordinateurs, l'internet pourrait se comprendre comme un système technique de plus dans notre environnement. On pourrait en faire l'histoire, mesurer ses effets sur l'industrie, comme cela a pu être fait pour l'électricité et l'aviation. Mais les historiens des techniques nous rappellent qu'une telle approche est prématurée, car l'internet n'est ni stabilisé, ni massivement socialisé depuis plusieurs générations. La meilleure preuve en est la foison des approches et des discours à son sujet, qui embrassent tous les champs du social, de la politique et de l'économie. Comme les fonctionnalités de l'internet sont mal connues, et comme ses détournements — preuve de son appropriation — sont encore trop récents, celles-ci sont plus imaginées que décrites. Leur évocation s'obtient en abusant d'un vocabulaire approximatif, le flou des mots employés permettant une grande liberté dans la construction des raisonnements. Les médias, mais aussi de nombreux ouvrages, donnent à penser que l'internet est avant tout une machinerie à faire rêver, ou à détester. On y projette trop de social, trop de politique. De tels discours nous semblent d'autant plus pervers qu'ils masquent des projets politiques : « j'ai inventé un logiciel qui fabrique des communautés », nous affirmait un heureux informaticien, dont le produit venait d'être acheté à prix d'or par une banque désireuse d'investir dans la Net-économie en 2000. Bien sûr, notre interlocuteur ne se posait pas la question du type d'échange, construit, imposé — formaté, pourrait-on dire — par ses programmes qui prennent en charge le social. Mais, dans les faits, Paris et Bruxelles manifestent-ils plus de recul quand ils invitent Français et Européens à « entrer dans la société de l'information » avec les « nouvelles technologies », usant ainsi de deux expressions qui nous semblent dénuées de sens ?

Dans un contexte aussi bruyant que réducteur il semblait difficile de faire de l'internet un objet de science. Nous espérons malgré tout expliquer rationnellement ces appels si peu rationnels au fait social total. Nous nous sommes longuement demandé si la référence technique était pertinente pour analyser l'internet : les chantres du déterminisme technique, qui veulent nous persuader que la technique nous garantit un avenir (social) radieux, et les descriptions d'un internet comme un amas de fils et de machines, fonctionnant de façon relativement autonome, ne nous y incitaient guère. Bien plus, les représentations associées au mot *technique* combinent deux préjugés, liés l'un à l'autre : tout d'abord l'idée qu'une technique est associée à un ensemble de machines qui travaillent la matière, souvent en économisant l'énergie musculaire de l'homme ; ensuite, son corollaire, qui sous-entend qu'aux artisans et ingénieurs, s'opposent les intellectuels, qui ne se servent que de leur tête, voire de papier et de crayons qui, de toute façon, n'influeraient jamais sur leurs productions « cérébrales ». Aux uns le travail sur la matière, aux autres l'agilité conceptuelle. On verra que ces représentations simplistes, lointaines héritières de celles de Platon, sont très partagées et ont des effets redoutables.

Une technique est en fait un ensemble de procédés destinés à une fin particulière, requérant souvent des instruments, des savoir-faire et des connaissances. Si l'on considère les buts premiers de l'internet — transmettre des paquets d'information —, notre objet aurait droit au statut de technique. Mais on perçoit alors mal en quoi il se distinguerait du télégraphe ou du réseau SWIFT des banques. Si l'on s'attache aux ressources qu'il héberge, le cadre technique semble trop contraignant ; pour autant, les dérapages que l'abandon d'un tel cadre permet, sur les registres de la virtualité et de l'immatérialité, nous rappellent qu'un internet métaphorique, bouquet de protocoles ou réseau d'échange, n'a pas de sens sans les machines, lourdes, coûteuses et présentes, qui le rendent tangible : les ordinateurs. Ceux-ci sont des instruments relativement bien connus, car ils ont entre 60 et 80 ans d'existence. Les premières expériences de mises en réseaux de ces machines (ordinateur à imprimante, ordinateur à ordinateur, etc.) sont, elles aussi, suffisamment anciennes pour qu'on puisse les analyser avec un minimum de recul.

Que fait un ordinateur ? Il effectue avant tout des comptages, des calculs. Aux algorithmes élémentaires se sont ajoutées au fur et à mesure diverses fonctions mathématiques, de plus en plus élaborées à mesure qu'elles pouvaient être *écrites*. Ce point est essentiel : aujourd'hui, l'ordinateur est un instrument relativement souple, qui peut réaliser nombre de choses non prévues par ses constructeurs, pourvu qu'elles soient le fruit de lignes de commandes. Même si d'aucuns imaginent que dans un futur proche, nous pourrons transmettre des commandes orales à de telles machines (ce qui est aujourd'hui réalisable, mais encore rare), la condition essentielle d'une telle transmission sera, et à notre avis pour longtemps, l'existence de lignes de codes écrites qui permettront la reconnaissance, puis la traduction et la mise en application de ces flux oraux. Dès le début, l'ordinateur a été (et reste) une machine à traiter des symboles graphiques (nombres ou lettres) avec des symboles du même type, organisés en programmes.

On comprend alors que, même si l'ordinateur est aussi une machine à traiter la matière[1], ses fonctionnalités diffèrent grandement — et depuis longtemps — de l'ensemble des autres machines qui, depuis le XIX^e siècle, ont nourri notre imaginaire technique.

En ce sens, l'informatique, dans toute sa diversité actuelle, apparaît comme une technique centrée sur l'écriture; et l'internet n'est qu'un prolongement (un parmi des dizaines d'autres, concrètement expérimentés dans les années 1980) de cette technique. Pour construire un tel réseau, il faut assurément des câbles, mais surtout des machines qui en constituent les points nodaux, et qui, par l'intermédiaire de programmes pouvant contenir jusqu'à des millions de lignes de texte, trient, réadressent ou traduisent les informations qui leur arrivent, d'une manière ou d'une autre, sous une forme textuelle ; et aussi des êtres humains pour produire toutes ces formes de textes, de ceux qui servent à faire fonctionner ces machines à ceux qu'elles transmettent.

À l'évidence, l'internet devait être replacé dans un contexte à la fois plus large et plus traditionnel, celui de l'informatique, que l'on devait appréhender à son tour comme un élément d'une technique ancestrale : l'écriture.

1. Par exemple l'inscription d'un fichier sur une disquette correspond à une magnétisation du support, la frappe d'une touche illumine certains pixels d'un écran, et l'on sait relier des ordinateurs à des locomotives, des avions, des distributeurs de banque, etc.

Tel est l'objet de notre première partie : nous rappelons que l'écriture, à la définition si ambiguë[2], se structure autour de son système de signes, intimement lié à la matière (le support, mais aussi le calame, voire l'encre) comme à la pensée : l'organisation d'un raisonnement complexe, sa critique, sa diffusion se font bien plus aisément dans un contexte d'écriture que dans un contexte d'oralité.

Dans les faits, ce triplet matière / signes / psyché est inséparable sous peine de conduire à des erreurs de raisonnement, qui par exemple, opposent de façon binaire la matière et la technique à l'esprit, à la pensée pure. Plus, si l'on détaille l'ensemble des procédures, savoir-faire, actes physiques comme mentaux associés à l'écriture, son statut de technique apparaît clairement, même si ce fait est trop souvent euphémisé dans le domaine des sciences de l'homme. À la suite d'un Lucien Febvre, d'un Ignace Meyerson, nous appelons outillage mental cet ensemble de connaissances, d'instruments, et de potentialités préalables à un usage efficace de l'écriture.

Pour autant, l'usage optimal de l'écriture ne va pas de soi. La maîtrise — et le développement — de l'outillage mental nécessite un long apprentissage, dont on sait qu'il peut prendre toute une vie. Cet apprentissage ne se fait pas en solitaire : il se construit dans un environnement collectif, propre à l'individu, que nous appelons laboratoire ; mais il crée aussi des effets de sédimentation plus larges. En effet la transcription des savoirs, leur matérialité, comme leurs effets politiques — grâce aux avantages obtenus suite à une meilleure compréhension du monde — puis économiques, se réalisent dans une temporalité longue, qui est celle de la constitution des structures sociales. Ces effets peuvent aussi avoir des pendants pervers : valorisation d'une écriture archaïsante qui perd tout lien avec la langue parlée, constitution de castes attachées à leurs privilèges au point qu'elles en oublient leur mission de transmission de la connaissance, mise en place de monopoles dans la chaîne de production de l'écrit, etc.

2. « Représentation de la parole ou de la pensée par des signes » ; [Rob85].

La complexité des mécanismes intellectuels, puis sociaux, économiques et politiques construits autour de l'écriture a été détaillée par des auteurs comme Henri-Jean Martin, Christian Jacob, Pierre Bourdieu ou Jack Goody. Ce dernier définit comme *technology of the intellect*[3] l'ensemble de ces relations entre activités mentales (solitaires ou collectives), objets qui les permettent, et même structures (y compris industrielles) qui concourent à l'élaboration des uns et des autres[4].

Nous pouvons alors expliquer pourquoi l'internet ne peut être séparé de son contexte informatique. Depuis environ un siècle, l'homme apprend à différencier l'espace de visualisation du texte de son espace de stockage. Une telle rupture, manifeste avec l'internet, fait que des centaines de protocoles de recodage s'intercalent entre le support et la graphie. Nous donnons quelques exemples de telles traductions, qui, souvent, explicitent ce statut graphique du texte alors même que l'image obtenue reste codée avec notre système de signe habituel, et donc devient manipulable comme ce texte qu'elle porte. Il nous semblait essentiel de mettre en évidence, en nous appuyant sur le fonctionnement des ordinateurs, une première propriété réflexive de l'écriture, qui n'avait été jusque là attestée que dans des cadres très théoriques.

Nous mettons en évidence les effets intellectuels — fastes ou néfastes — rendus possibles par ces procédures de travail sur notre système de signes, ainsi que les appétits économiques qu'elles suscitent : c'est parce que l'informatique et l'internet s'intègrent dans un processus de transformation — ni le premier ni le dernier — de la technique d'écriture que leurs enjeux apparaissent si manifestes, si multiformes, mais aussi si peu compris. Si l'on s'accorde donc avec le fait que l'écriture est l'élément central de notre outillage mental, et que toute transformation de cet outillage génère des enjeux et des conflits considérables[5], la preuve que nous donnons de cette réorgani-

3. Ou *technologies*. Traduit en français par « technologie de l'intellect », quand « technique de l'intellect » aurait été plus approprié.

4. Pour un exemple, voir la note 66 de la page 43.

5. Ces deux propositions pourront être considérées comme démontrées ou admises à titre d'axiome ; c'est toute la difficulté du raisonnement en sciences sociales qui apparaît ici : un

sation de la technique d'écriture par l'internet explique la puissance virtuelle de ce dernier objet — et par conséquent les violences qui ne manqueront pas d'apparaître dans les courses au pouvoir qui s'ensuivront.

Cependant, même si nous proposons une grille de lecture de la compétition économique actuelle, notre but reste l'explicitation du fonctionnement de l'écriture. En effet, parer l'écriture de propriétés intrinsèques reviendrait à commettre une erreur trop répandue, qui consiste à se représenter une technique comme un objet autonome, en oubliant l'homme, qui en est à la fois le créateur et l'utilisateur.

Certes, notre outillage intellectuel se nourrit des interrogations que l'on porte à son sujet : l'analyse, explicite ou non[6], de la façon dont l'écriture agit sur elle-même accroît de façon surprenante nos capacités à comparer, imaginer, synthétiser. Mais, étudier cette réflexivité, profiter de sa dynamique, c'est évidemment aussi s'y engager soi-même : adopter une posture de recul face à l'instrument essentiel à la construction de notre pensée transforme notre représentation du monde, l'organisation de nos raisonnements. Autrement dit, l'étude d'un phénomène qui semble spontané nous implique dans une pratique volontaire, lucide et assumée, qui n'est autre que l'unique constituant de la pratique scientifique : Jack Goody et Pierre Bourdieu rappellent que celle-ci se réalise quand l'on prend acte de la dimension réflexive de l'outillage mental, quand on en tire le meilleur parti.

Or, c'est ce fait trop souvent perçu comme une hypothèse ou une intuition que notre thèse vise essentiellement à prouver : la réflexivité de l'outillage mental est le catalyseur de la production scientifique mais aussi sa condition, dans la mesure où la recherche est intimement liée à la possibilité de repousser les limites de nos connaissances, d'interroger leurs constituants et les relations

ensemble de faits historiques attestés peut toujours être refusé au titre d'une sensibilité qui fait plus la part belle au préjugé qu'à la rigueur scientifique.

6. Par exemple, la production d'une liste ne correspond pas toujours à un désir explicite de tester les capacités de l'outillage mental. Cependant, la liste procure un sens supplémentaire et différent de la somme des sens de ses constituants, ceci pour son auteur comme pour ses lecteurs.

qu'ils entretiennent entre eux. Il est alors tentant de tester cette proposition en prenant comme objet l'internet, qui transforme l'écriture, mais qui n'est pas pour autant vu comme un catalyseur de la production scientifique. Dans cette optique, le terrain le plus propice pour notre recherche est constitué des personnes qui ont la charge de transmettre et d'améliorer l'outillage mental, c'est-à-dire les universitaires et les chercheurs. Nous nous sommes principalement penchés sur les pratiques des spécialistes des sciences humaines, parce qu'ils ont cette particularité de valoriser et de commenter abondamment les effets secondaires de l'écriture (culture, histoire, patrimoine, identités, etc.) tout en affichant souvent un mépris pour la technique et un faible intérêt pour l'explicitation des relations entre écriture, outil et pensée.

Pour mener à bien notre démonstration, nous avons réalisé deux études.

La première porte sur la prestigieuse division littéraire de l'École normale supérieure. C'est l'objet de notre seconde partie. Nous y avons privilégié la méthode ethnographique : il nous semblait essentiel de montrer le plus précisément possible, non seulement comment s'organisaient les prises de position variées autour de l'écriture informatique, mais les conditions sociales et culturelles permettant l'appropriation d'un outillage contemporain qui n'a rien d'évident. Ce qui a permis aussi de montrer à quel point l'écriture est aujourd'hui — dans sa forme élémentaire comme sophistiquée — dépendante de l'informatique.

L'autre étude (notre troisième partie) repose sur une enquête par questionnaire auprès des universitaires les plus impliqués dans les divers processus de l'internet, avec comme activité centrale la publication électronique. L'organisation du monde scientifique autour des formes traditionnelles de l'écrit y apparaît de façon flagrante, en même temps que les pionniers enquêtés détaillent clairement leur perception de l'outillage mental et du laboratoire, en montrant comment s'y articulent, dans diverses disciplines des sciences humaines, les mécanismes de réflexivité.

Ainsi, dans les deux situations, l'internet sert de révélateur au fonctionnement de la recherche. L'engagement dans cette technique favorise l'analyse sociologique de l'univers scientifique, qui apparaît indispensable pour garantir des bénéfices intellectuels lorsqu'on est confronté à des résistances nourries par une référence constante à un outillage mental archaïsant.

Soucieux de montrer concrètement comment cette réflexivité de la technique d'écriture se réalise à partir de ses formes les plus classiques, et comment l'informatique et l'internet ne se concevaient pas l'un sans l'autre dans notre conception de l'outillage mental des chercheurs, nous avons décidé d'étudier une liste, issue de l'internet, que nous traitons en usant au mieux des instrument d'écriture actuels que sont les ordinateurs. C'est l'objet de notre quatrième partie[7], qui montre dans le détail, certes, les avantages variés qu'offre le traitement des requêtes que des centaines de milliers d'internautes ont adressées durant trois mois à un moteur de recherche; mais aussi la nécessité d'acquérir de nouvelles compétences, de développer de nouvelles méthodes, pour produire ces résultats scientifiques.

L'étude de cette liste permet tout d'abord de découvrir des faits sociaux et intellectuels plus précieux que ce qu'on aurait pu tirer de la lecture — ici impossible — de chaque élément de la liste. Ensuite, nous montrons comment un tel matériau met en perspective des thèmes de travail prometteurs pour les chercheurs en sciences sociales, à condition de tirer le meilleur parti des ressources des machines actuelles, et de défricher des méthodes aux confins des sciences humaines et exactes. Ce qui prouve d'une part la pertinence des études sociologiques fondées sur l'activité de machines, et d'autre part la nécessité pour les ingénieurs de s'approprier des pans conséquents de la culture des sciences humaines avant de tirer du sens de ces formes d'écriture — les archives électroniques — qu'ils ont inventées. Enfin, il apparaît que la production de mesures interroge le chercheur sur les implicites des disciplines : la découverte de phénomènes très rares, mais essentiels, qui souvent sont oubliés au profit de pratiques massives, mais dénués de valeur informative, invite à repenser les catégories sociales — ainsi que leur distribution,

7. Divers résultats relatifs à cette partie, mais aussi aux précédentes, sont aussi disponibles dans l'annexe.

ce qu'ont déjà mis en évidence des chercheurs comme Hervé Le Bras ou Maurizio Gribaudi — sur lesquelles nous fondons notre compréhension du monde : nous découvrons alors à quel point la sociologie reste dépendante d'un cadre gaussien implicite, qui confond norme et moyenne; nous proposons alors des méthodes pour reconstruire des catégories homogènes et pertinentes; ici, les approches quantitatives et qualitatives, censées s'opposer, se renouvellent en fait l'une l'autre. Certes, la nature de l'objet que nous étudions, trace écrite par excellence, nous a aidé à questionner les manières de compter et d'agréger les pratiques humaines. Mais au final, l'étude de cette liste nous a permis d'expliciter l'ensemble des procédures intellectuelles, matérielles, graphiques, impulsées par l'écriture, et la façon dont elles s'articulent avec des disciplines et des cultures variées, invitant à recomposer leurs frontières.

En définitive, les pages qui suivent voudraient proposer une grille de lecture efficace du phénomène internet, qui, avant toute chose, participe, avec l'informatique, d'une modification de notre outillage mental. Celle-ci ne se produit jamais sans heurts ni douleurs, mais elle reste le fait de l'homme. Unique artisan — en collectif — de la technique de l'intellect, il n'a qu'un moyen pour accroître sa compréhension du monde, et donc sa liberté : user et abuser de toutes les mises en correspondance que lui permet cette technique, ce qui revient à la remettre sans cesse en chantier, le cadre optimal d'une telle activité étant le champ de la science. Il est alors logique que l'internet, en tant qu'un des derniers produits de cette attitude réflexive de l'écriture, ait une incidence directe sur la production scientifique comme sur l'analyse de cette production — élément d'importance pour les chercheurs que nous sommes. Ce contexte intellectuel nous semble essentiel pour comprendre les enjeux actuels relatifs à l'informatique et à l'internet, y compris dans les situations (jeu, commerce, etc.) où la référence à l'outillage mental est soigneusement gommée, évacuée.

Première partie

Écritures

Dans cette partie, nous définissons le cadre conceptuel de l'outillage mental et du laboratoire, à partir de multiples exemples attestés par l'histoire de l'écriture. Ensuite, nous montrons à quel point l'informatique, passée et contemporaine, recompose la technique d'écriture. Les avantages que l'on peut attendre d'une telle transformation passent par une actualisation de notre outillage mental. Mais cette dernière est d'autant plus délicate à mettre en place que toute mention à la réflexivité de l'écriture est souvent gommée, ce qui permet de donner une seconde vie à des conceptions technicistes désuètes.

Chapitre 1

Technique et technologies

Quelle est la différence entre l'École polytechnique et les instituts universitaires de technologie ? La première serait-elle condamnée à enseigner les travaux manuels et l'artisanat, et les seconds voués à l'étude des sciences les plus modernes ? On sait bien que non, et cet exemple rappelle que le mot « technique » a subi des variations de sens au fil des siècles, pour acquérir une signification de plus en plus dépréciée. Au contraire, le mot « technologie » est de plus en plus utilisé ; on l'emploie maintenant au pluriel, et on l'adosse à la nouveauté. Ainsi, les « nouvelles technologies » s'imposent, et plus personne n'ose parler de technique.

Il semble donc utile de commencer un travail sur l'internet par une mise au point sémantique, qui passe par une analyse des origines et des implications des concepts aujourd'hui à la mode.

1.1 Définitions

L'adjectif « technique » renvoie à ce qui « appartient tout d'abord à un domaine particulier, spécialisé de l'activité ou de la connaissance » [1], dans un contexte distinct de la science et de l'abstraction, induisant l'idée d'une maîtrise pragmatique, dédiée à une finalité précise. On entre alors dans le domaine

1. [Rob85].

de l'application, du savoir-faire, et on s'éloigne de celui de l'« inspiration ». C'est ainsi que le substantif est défini par un « ensemble de procédés empiriques employés pour produire une œuvre ou obtenir un résultat déterminé ». Suivant les situations, ces processus peuvent être « plus ou moins régularisés et structurés », ou ordonnés suivant une méthode scientifique rigoureuse.

La technologie, est, dans la tradition grecque, « l'inventaire des procédures qui caractérisent les savoirs et savoir-faire qui échappent aux déductions spéculatives, mais reposent sur des connaissances expérimentales »[2], ou encore, un « traité ou exposé des règles d'un art »[3]. Par un retour des modernes à l'étymologie ($\tau\epsilon\chi\nu o - \lambda o\gamma\iota\alpha$), on en arrive à un savoir (ou à une réflexion) relatif aux techniques : « théorie d'une technique », ou « réflexion philosophique sur les techniques, leurs rapports avec les sciences théoriques ou fondamentales, les conséquences politiques, économiques, sociales et morales de leur développement [...] qui s'exerce ainsi du point de vue pratique sur les sciences expérimentales et humaines, notamment sur la cybernétique »[4]. Ainsi, les techniques contraceptives du XX^{e} siècle ou les techniques biologiques actuelles se prêtent-elles à une étude technologique. Mais aussi les systèmes automatisés introduisant la notion d'information, qui ont effectivement généré toute une réflexion éthique, sociologique, en même temps que de vivaces idéologies[5]. Gilbert Simondon propose aussi le terme de « mécanologie », et, dans cette conception philosophique, le métier de « technologue », ou de « mécanologue ». Il promeut une « attitude technologique, qui fait qu'un homme ne se préoccupe pas seulement de l'usage d'un être technique, mais de la corrélation des êtres techniques les uns par rapport aux autres »[6].

Entre ces sens précis de la technique et de la technologie, un sens commun s'intercale, considéré comme abusif par les défenseurs de la langue française

2. « Étymologie et histoire du mot technologie », François Charpin : http://www.info.unicaen.fr/bnum/jelec/Solaris/d04/4charpin.html
3. [Mor80].
4. [Mor80].
5. [eSP96].
6. [Sim89], p. 145.

que sont les Québécois : le Robert précise que le terme « technologie et ses dérivés » sont fréquemment utilisés pour signifier « technique de pointe, moderne et complexe [...] souvent pour des raisons d'emphase publicitaire ». Le Grand dictionnaire terminologique précise : « C'est un abus, et une erreur très répandue, non exempte d'un certain snobisme, que de remplacer systématiquement l'une par l'autre (la méthodologie et la technologie) : dire à tout propos *la technologie du collage*, la *technologie de la haute-fréquence* n'apporte rien, sinon une satisfaction vaine, et une complication supplémentaire pour les traducteurs »[7].

Dans les faits, les deux définitions rigoureuses de la technologie — ensemble de procédures non spéculatives ou philosophie des techniques — sont proches de celles que l'on rencontre aux États-Unis. Dans ce pays, où l'on emploie *handicraft* pour signifier « technique manuelle », le mot *technology* recouvre deux sens : le premier renvoie à la définition philosophique, épistémologique qui vient d'être précisée, l'autre à la notion de système technique : « *industrial arts collectively* »[8], que l'on pourrait traduite par « ensemble de règles et de méthodes appliquées à l'industrie dans son ensemble »[9], incluant aussi l'ensemble des connaissances et des savoir-faire associés. On pourra remarquer que ce second sens est très proche de celui de notre mot « technique », surtout si l'on prend comme référence la définition de *technology* proposée par la *National Academy of Sciences* : « moyens codifiés de manipuler délibérement l'environnement pour réaliser un objectif matériel »[10].

Hélas, de telles définitions semblent inadaptées à l'usage courant en France : peut-on imaginer définir une réflexion sur l'internet comme une technologie appliquée aux « nouvelles technologies » ? Il semble au demeurant que ces « nouvelles technologies de l'information et de la communication » et leur acronyme « (N)TIC » servent surtout à masquer une incapacité à proposer une réelle réflexion sur les techniques passées et contemporaines, un peu

7. http://www.granddictionnaire.com
8. [AM95].
9. Traduction personnelle.
10. Traduction personnelle. [Goo00], p. 132.

comme l'expression « société de l'information ». Mais, plutôt que se battre contre la francisation de telles expressions, on rappellera la façon dont elles se sont construites et deux effets de leur diffusion.

1.2 Usages

1.2.1 Prime à la nouveauté

La focalisation sur l'innovation ne date pas de l'invention de l'informatique ou de l'internet. David Edgerton explique qu'elle est spécifique à tout le XX^{e} siècle, et a deux fonctions : la construction sociale de l'inventeur, devenant héros national — au détriment de ses concurrents étrangers —, et une « canalisation des énergies sociales », afin d'éviter la critique d'un ordre social qui limite la redistribution [11]. Autrement dit, la valorisation du futur sert l'idéologie nationaliste et les tenants du pouvoir établi. Faire miroiter au citoyen des « lendemains qui chantent » l'incite à moins se plaindre. En fait, on entre, consciemment ou inconsciemment [12] dans le registre du déterminisme technique, qui prétend que la technique transforme la société.

Une telle valorisation de l'avenir date, pour l'internet, de 1992, quand le vice-président Al Gore a promu les « nouvelles technologies [de l'information] » et leur pendant, la *National Information Infrastructure* [13].

Sur le plan économique, on peut souligner que ce rêve, dont la rhétorique relève bien sûr du libéralisme, a été impulsé en faisant appel à des méthodes très keynésiennes de relance de l'industrie nationale, et donc à grand renfort de financements étatiques. Par ailleurs, son argumentaire est faible : il repose

11. [Edg98].

12. Si la démarche de l'homme politique peut être consciente, on s'étonne du fait qu'un grand nombre d'historiens des techniques se lancent sans recul dans une étude des innovations qui ont réussi, sans attendre que celles-ci fassent l'objet d'une réelle appropriation en profondeur — et donc des détournement qui prouvent celle-ci —, sans porter le moindre intérêt aux autres techniques qui « n'ont pas réussi ». Ici encore, voir [Edg98].

13. Cf. http://clinton4.nara.gov/WH/New/html/20000121_3.html et http://clinton1.nara.gov/White_House/EOP/OVP/html/nii1.html

sur l'illusion de la résolution du paradoxe de Solow, qui affirmait qu'« on voit des ordinateurs partout, sauf dans les statistiques de productivité ». Parce que l'informatique aurait été dans les années 1980 le principal secteur qui tirait la croissance américaine[14], il convenait de développer ce domaine, ce qui permettait par ailleurs de contrecarrer les initiatives de l'industrie japonaise.

Cependant, la fragilité de ces raisonnements — qui développaient l'illusion d'une croissance illimitée et exponentielle — n'a pas empêché la diffusion de l'expression « nouvelles technologies ».

1.2.2 Traversée de l'Atlantique

L'Europe allait aussitôt adopter le programme américain et son vocabulaire. En 1994, le rapport Bangemann vante les « technologies de l'information et de la communication » , « une révolution fondée sur l'information, elle-même l'expression de la connaissance humaine ». Il s'agit « pour l'Europe, d'exploiter les opportunités offertes par les nouvelles technologies »[15]. D'autres rapports suivent, en France notamment, aussi lyriques et prophétiques qu'ils manquent d'imagination ; la plupart reprennent un autre paradigme du déterminisme technique, la notion de « retard permanent », en insistant sur l'« urgence » qui s'impose si l'on veut combler en France ce fameux retard. Les journalistes, les universitaires adoptent les NTIC. Le mot « technique » disparaît de leur vocabulaire.

Bien sûr, cette appropriation s'est accompagnée de l'importation des représentations sociales qui l'accompagnent (foi en l'avenir, en l'industrie). Mais elle a aussi conduit à troubler le débat en affaiblissant la valeur des analyses développées sur notre continent sur le thème des sciences et des techniques : puisque l'objet change de nom, les anciennes analyses à son sujet perdent de leur légitimité ; les nouvelles technologies nous faisaient entrer dans une « nouvelle ère », qui n'avait plus de rapport avec notre ancien monde

14. Pour plus de détails, voir Éric Brousseau, http://panoramix.univ-paris1.fr/ATOM/pdf/eric/wp/new/EBAFRIE-eEconomie.pdf

15. Pour ces extraits, cf. http://europa.eu.int/ISPO/infosoc/backg/bangeman.html#chap1

des techniques. On se contentera de mentionner, sans les détailler ici, les premiers effets de cette terminologie : nouvel essor du déterminisme technique ; profusion de discours empruntant à la philosophie et la sociologie de bas niveau ; mais aussi — point positif — interrogation sur les objets techniques qui nous entourent.

Cet emprunt a aussi un autre effet, plus ambigu, au sein de nos professions. Le volontarisme politique a permis d'élargir les formes de financement de la recherche publique : de nombreux budgets de la recherche en sciences humaines sont maintenant prévus pour l'étude de l'internet. Parfois même, les laboratoires publics des sciences exactes ne peuvent obtenir de tels financements s'ils ne s'associent pas avec leurs équivalents des sciences humaines[16]. Étonnamment, l'obtention de tels budgets n'implique nullement de restreindre la recherche à un domaine appliqué. Mais les chercheurs en sciences humaines restent méfiants. Deux type de raisons expliquent cette forme de résistance :

— Tout d'abord, on comprend que des universitaires habitués à travailler leurs théories au corps refusent de cautionner le déterminisme technique ; de façon plus pratique, certains chercheurs se sont rendus compte que de tels travaux ne permettaient pas d'accroître leur crédit universitaire, ni celui de leurs étudiants[17].

— Par ailleurs, cette insistance sur les « nouvelles technologies » les obligeaient à repenser la relation de leurs disciplines à la technique. Celle-ci est souvent circonscrite à une application subordonnée[18] : dans le monde des sciences humaines, la distinction entre le chercheur, qui sollicite des processus mentaux complexes pour en déduire des conclusions écrites, et le technicien

16. Voir http://www.recherche.gouv.fr/appel/2002/societeinfo.htm pour un tel exemple.
17. [d'I99].
18. Les constructions mentales élaborées qui peuvent donner lieu à des thèses universitaires, comme en statistique, deviennent pour les sociologues des « techniques », comme c'est le cas des analyses factorielles ; la cartographie n'est qu'une « technique » de géographes, et eux-mêmes sont présentés comme de faibles théoriciens (cf. [Bou84], p. 45). La documentation est un « service » dédié aux chercheurs en sciences humaines.

ou l'ingénieur, qui doit appliquer et reproduire diverses méthodes et techniques, mais qui a un accès moindre à la publication, traduit une domination des premiers sur les seconds, et rares sont ceux qui, comme François Furet, ingénieur de travaux devenu président de l'EHESS, arrivent à s'en échapper. Accepter les injonctions relatives aux « NTIC », fussent-elles accompagnées de financements, risquait de redonner une autonomie à la technique, de lui permettre de s'émanciper de son inféodation à la production scientifique « noble ». Nous montrerons que ce n'est pas seulement l'excès de naïveté des discours sur les NTIC qui explique pourquoi les chercheurs en sciences humaines n'ont pas tiré parti de ces opportunités.

1.3 Sobriété

Par souci de précision, mais aussi parce que ce statut de la technique au sein des sciences humaines nous semble soulever une problématique passionnante, nous éviterons donc l'emploi du mot « technologie » et de ses variantes, sauf en cas de citation. Dans tous les cas, la définition grecque sera rejetée, pour cause de scission entre la pensée et le concret. Comme définition d'une technique, nous nous satisferons de celle du dictionnaire, à peine remaniée : c'est un « ensemble de procédés empiriques plus ou moins explicites destinés à une fin particulière ».

Il nous semblait important de faire ces détours historiques et géographiques pour rappeler les implicites idéologiques de la terminologie associée à l'internet, et pour préciser ce qui nous semble être le vocabulaire le mieux adapté à notre projet scientifique.

Chapitre 2

L'écriture comme technique

La pratique d'écriture semble aux antipodes d'une technique : elle est très valorisée socialement sans être pour autant précisée. Elle renvoie au génie solitaire de l'écrivain, au scientifique plongé dans ses équations, voire à l'éditeur, protecteur et garant de la culture. Il semble que ces représentations simplistes et affectives masquent l'interdépendance entre les objets matériels qui participent de l'écriture, l'activité intellectuelle et les constructions sociales.

L'écriture requiert un large ensemble d'instruments : tablette de cire, stylet, puis plume, papier, ruban de machine à écrire, etc. Elle s'appuie sur la connaissance d'un système graphique[1] qui peut dépasser un alphabet, par exemple en intégrant des pictogrammes, ou en devenant logographique comme dans le cas des mathématiques.

Ces instruments et ce système de signes facilitent la production de processus mentaux ainsi que le transfert des connaissances, avec ses corollaires en matière sociale (le langage est l'acte de l'échange). Le tout s'accompagne de la production d'autres objets matériels, qui peuvent être suffisamment sophistiqués pour n'être utilisés qu'au terme d'un long apprentissage[2], et

1. Définition certes un peu large, mais en accord avec les spécialistes. Par exemple, Nina Catach propose, comme définition de l'écriture, un « ensemble de signes discrets, articulés et arbitraires, permettant de communiquer n'importe quel message construit dans une langue donnée » ([Cat88], p. 253).

2. Par exemple, la lecture d'un texte au X^e^ siècle, avant la « révolution » liée à l'appari-

qui s'inscrivent naturellement dans des processus économiques tendant à la réorganisation des pouvoirs.

Ainsi, on remarque un premier effet de boucle entre les trois constituants de l'écriture que sont les instruments, le système graphique, et les processus mentaux : réunis, ils mènent à la conception de nouveaux objets matériels, comme les livres, qui stimulent l'activité intellectuelle. Nous allons préciser en quoi les interférences entre ces trois catégories sont bien plus importantes qu'on ne l'imagine souvent.

2.1 Premier type d'interactions

2.1.1 Effets du système graphique

L'écriture, parce qu'elle permet la transmission de l'information culturelle (le face à face ne s'impose plus), l'accumulation du savoir et surtout la multiplicité des comparaisons, est un préalable à la « défaite de la magie »[3] : les formes de réfutation et d'argumentation peuvent alors se transformer en scepticisme généralisé, et donner naissance à l'exercice d'une rationalité. D'où l'ouverture vers des opérations arithmétiques, l'expression du syllogisme, et l'élaboration d'une pensée scientifique.

Telle serait une première vertu du système de signes et de sa permanence.

Mais il peut aussi avoir une influence plus directe sur la manière de penser. Dans le cas limite du phénoménologue Husserl, qui, pour poser plus vite ses idées sur le papier, utilisait un système dérivé par ses soins de la sténographie Gabelsberger, au point qu'il est presque impossible à déchiffrer[4], on assiste à une « forme spécifique d'écriture pensante »[5] : la césure entre signe et processus intellectuel devient floue. Dans un contexte plus collectif, le choix

tion de l'index, lui-même considéré comme *illisible* au début de son introduction ([Mar96], pp. 152–153) .

3. [Goo97], p. 78.

4. Preuve que ce système de signes peut ne pas être partagé.

5. Conférence de Jean-François Courtine, Atelier Internet, 3 avril 1998, http://barthes.ens.fr/atelier/debats-et-CR/synt-3-4-98.html

des signes transforme les représentations : Jean-Pierre Vernant explique ce que la prouesse de Clystène, qui instaure la démocratie à Athènes, doit au rationalisme géométrique, et comment l'adoption d'un système décimal casse la tradition politique ionienne ; elle permet d'homogénéiser le temps civique — qui se distingue du temps religieux —, mais aussi l'espace. Émerge alors une rationalité qui va se traduire dans l'organisation d'une cité désormais gouvernée par le principe d'isonomie, qui fait de chaque citoyen le membre d'une communauté d'égaux[6].

L'existence et le type du code[7] ont donc des conséquences importantes sur l'organisation de la pensée, solitaire comme collective, et elle-même contribue à remodeler le système de signe sur lequel elle s'appuie.

2.1.2 Effets de l'instrument

Le support a une incidence sur le code graphique. L'écriture serrée des papyrus, mais aussi leur disposition physique, invitent à imaginer des pictogrammes : pour commenter, critiquer des passages répétitifs ou contradictoires dans un même rouleau, repérer des graphies peu courantes, signaler une altération par rapport à un texte considéré comme original, voire une citation ou une analyse criticable. Les « bibliothécaires » d'Alexandrie ont fini par multiplier en marge des textes une série de plus en plus complexe de signes au fur et à mesure des lectures.

Mais l'instrument oriente aussi la dynamique de la pensée : « [Avec le *volumen*,] la matérialité du livre et les contraintes de son maniement affectent les modalités d'appropriation du texte, le processus de construction du sens, et ceci est d'ailleurs vrai pour le livre manuscrit, imprimé ou affiché sur l'écran d'un ordinateur »[8]. La localisation des 490 000 rouleaux de papyrus de la bibliothèque d'Alexandrie, conçue par une autorité politique pour être

6. [Ver82], pp. 207–229.

7. Dans ce chapitre, « code » sera pris comme synonyme de « système graphique » — ou de « système de signes ». Dans le prochain chapitre, la distinction entre signes et codes sera détaillée.

8. [Jac96], p. 56.

un « dépôt de livres » sans public, dans diverses salles thématiques, elles-mêmes organisées suivant une représentation de l'œkoumène, les procédures de rangement physique des rouleaux sur des étagères, ont engendré autant de catégories conceptuelles qui sont à l'origine de nos disciplines actuelles.

Plus près de nous, la transcription écrite, puis la comparaison de mythes oraux comme celui du Bagré étaient quasiment irréalisables avant la diffusion du magnétophone, outil indispensable de l'anthropologue.

Ces exemples prouvent que les constituants les plus matériels de l'écriture contraignent notre système de signes comme notre pensée.

2.1.3 Construction du sens et de la culture

Jusque là, on a mis en évidence des relations binaires : le support influe sur le code, comme sur la pensée. Le système graphique et la pensée agissent explicitement l'un sur l'autre. Cette dernière conditionne le support, au moins dans ses formes dérivées (taille, structure).

Sans pour autant confondre les trois pôles de l'écriture, on peut maintenant prendre un peu de recul pour évaluer leurs effets sociaux et culturels. Déjà, les indices inscrits en marge du texte, qui parfois étaient très personnalisés, s'intégraient dans la culture des lecteurs de ces papyrus[9] annotés, qui pouvaient commenter ces décisions éditoriales. La complétion du code graphique, grâce à la mémoire et à l'évocation du signe, créait un réel tissu hypertextuel dont les fils reliaient les textes entre eux.

Christian Jacob montre comment cette combinatoire des supports et des codes génère toute une série d'opérations mentales, qui ont évidemment une grande influence sur la façon dont les savoirs sont organisés, présentés, voire oubliés : « cette accumulation va induire des effets intellectuels particuliers, fonder des pratiques érudites de lecture et d'écriture, et une manière savante de gérer la mémoire de l'humanité »[10].

9. Le caractère invariable de ce mot, qui définit avant tout une matière, est attesté par exemple chez Christian Jacob : « certains des papyrus littéraires... » ([Jac96], p. 59).

10. [Jac96], p. 67.

Tout d'abord se construisent des listes. Pour s'approprier l'espace de la bibliothèque, il incombe aux savants un travail de classement (repérer, archiver, ordonner) qui se complète vite par un travail d'édition, de commentaire. Les lexiques, organisés suivant divers critères, plus ou moins rationnels, « constituent une constante de l'érudition alexandrine »[11].

Jack Goody montre en quoi ces catégories qui peuvent sembler parfois vaines ou gratuites s'intègrent dans des techniques « cognitives »[12], et que le mot, la liste, le tableau, la matrice, ne sont pas des inventions récentes d'informaticiens peu motivés par le sens des mots. Pour lui, un « mot » est un « logogramme » enfermé dans une case ou sous une rubrique, un « terme linguistique sorti de son contexte »[13]. Le plus important est que cette dissociation entre la forme graphique, qui ne se distingue de sa voisine que par une absence (par exemple notre « espace » dans l'imprimé), et son sens, s'est toujours réalisée dès qu'il y a eu apparition de l'écriture, et de façon majoritaire, quand bien même ces séparateurs n'étaient pas visualisés[14]. Au final , « ce qui est significatif dans l'emploi de la langue dans les premiers systèmes d'écriture est qu'une grande partie présente une structure syntaxique très différente du discours parlé »[15]. En d'autres termes, l'écriture permet une combinatoire qui décontextualise les mots, qui les objective et les sépare de leur sens, et une telle opération produit à son tour du sens.

2.1.4 Outillage mental

Ce travail de recombinaison permanente d'informations, décontextualisé ou organisé, ne se limite par à la production textuelle : apparaissent les cartes, qui sont autant de mises en *table*[16] d'informations sériées. Ainsi, les formes graphiques produites par un tel travail de synthèse s'émancipent du texte

11. [Jac96], p. 67.
12. [Goo97], p. 281.
13. [Goo97], p. 282.
14. [Goo97], p. 262.
15. Cf. note 13.
16. La carte est appelée *pinax* en grec, ce qui signifie table ([Jac96], p. 71).

seul. L'arrangement du code appelle la création de nouveaux codes. On assiste ici à un premier processus réflexif de l'écriture sur elle-même qui nous apparaît fort important.

L'activité intellectuelle modifie donc le système graphique. Cela se rencontre fréquemment en mathématiques, mais aussi dans des situations plus classiques : les accents et les esprits grecs, chers aux antiquisants contemporains, n'apparaissent qu'au Ier siècle après JC, 600 ans après le siècle de Périclès, « sous l'influence des éditeurs de textes des bibliothèques alexandrines » désireux d'en faciliter la lecture à leur public[17]. Elle incite aussi à modifier la structure du livre, comme on l'a vu avec l'introduction de l'index[18].

Au final, les objets matériels qui interviennent dans le processus d'écriture, les savoir-faire liés à la manipulation de ces objets, l'organisation de raisonnements, la mise en place de concepts, leur classification, ne sont pas distinguables : de multiples allées et venues, trop souvent oubliées, font que l'écriture, instrument indispensable au développement de l'activité intellectuelle, apparaît comme une technique, et que la césure communément proposée entre l'une et l'autre ne peut être qu'artificielle. Ceci dit, nous reconnaissons que dans le choix de nos propres mots et catégorisations, nous subissons l'effet d'un *cadre de pensée* vieux de plusieurs siècles, dont nous avons du mal à nous déprendre, certainement parce que trop peu d'érudits ont défriché ces interactions. Or, la compréhension de ces dernières nous semble essentielle pour comprendre les formes et les incidences de l'écriture informatique.

Aussi, à la fois pour tenir compte des contraintes de notre vocabulaire, et pour ne pas trop rompre avec l'histoire de nos concepts, nous proposons une première définition : on appellera *outillage mental* (ou appareillage intellectuel) la somme et la variété des objets qui nous permettent d'organiser notre pensée. Jean-Pierre Vernant parle d'outillage conceptuel[19], Christian Jacob d'outillage intellectuel[20].

17. [Mar96], p. 68.
18. Cf. note 2 p. 26.
19. [Ver82], p. 215.
20. Lucien Febvre est sûrement le premier à proposer en 1942 ce concept d'outillage

2.2 Interactions sociales

L'écriture, puis la concentration d'objets écrits, ont des effets intellectuels considérables, et l'on voit comment code, support, et pratiques mentales sont étroitement liés : l'un ne détermine pas les autres. Mais l'on n'est pas pour autant dans un contexte de solitude. Les activités suscitées par la bibliothèque, comme par l'écrit en général, sont à la fois individuelles et collectives. Individuelles car elles renvoient à des mémorisations, des sélections, des organisations qui peuvent évidemment dépendre de la culture et de la personnalité de chacun. Collectives, car l'écriture est un moyen de communiquer, et, comme la langue, ne s'acquiert qu'au contact d'autrui. Mais aussi parce que l'activité éditoriale est parfois très physique : le commentaire de plusieurs rouleaux ne pouvait se faire sans la participation de cinq à dix personnes.

Ainsi, cet outillage mental ne se suffit pas de textes, de listes, de graphes, ni des opérations intellectuelles facilitées par ces objets. Il lui faut un terreau social pour se déployer : l'outillage mental d'Aristophane prend aussi appui sur l'ensemble des auteurs qu'il lit attentivement, sur ses collègues, ses visiteurs, et ses professeurs, comme Eratosthène à qui il succéda[21].

Dans sa relation à l'écrit, l'individu est plongé dans un collectif, qui est relativement restreint : son environnement proche (famille, amis) ; les classes primaires et leurs maîtres pour les enfants qui apprennent à lire et à écrire ; les « collègues », morts ou vivants, proches ou lointains, pour les érudits d'Alexandrie ou les chercheurs actuels. A cet environnement social restreint, en ce sens qu'il ne met pas en jeu l'ensemble d'une société, et que nous appelons *laboratoire*, correspond une temporalité locale, associée à l'espace d'un

mental, encore qu'il soit très orienté vers le vocabulaire dont disposaient les philosophes du XVI^e siècle. Prenant l'écriture pour acquise, il se demande quelles absences limitent ou contraignent la production intellectuelle de l'époque : « quelle efficacité pouvait avoir la pensée d'hommes, de Français qui, pour spéculer, ne disposaient encore dans leur langue d'aucun de ces mots usuels qui reviennent d'eux-mêmes sous nos plumes dès lors que nous commençons à philosopher — et dont l'absence n'implique pas seulement de gêne, mais vraiment déficience ou lacune de pensée » ([Feb68], p. 328.)

21. [Jac96], p. 60.

individu : temps d'acquisition des connaissances, propre au développement intellectuel d'une personne, plus que chronologie.

2.2.1 Laboratoire

Le laboratoire renvoie communément à un espace d'expérimentation typique des sciences exactes et appliquées, avec une organisation sociale partiellement régulée par les instruments. On imagine vite une paillasse carrelée, des pipettes, un microscope et un technicien en blouse blanche. Dans les faits, la situation est bien plus complexe, pour deux raisons : déjà, les physiciens et les chimistes ont plus fréquemment le nez dans leurs livres, sur leurs écrans d'ordinateur, ou sur des feuilles blanches, pour repérer des générateurs de groupes d'isométries, qu'au dessus d'une éprouvette. En second lieu, le mot laboratoire est devenu synonyme de centre de recherche, même quand il relève des sciences humaines. Pierre Bourdieu s'insurge contre un tel emprunt aux sciences exactes, imposé selon lui par le CNRS[22]. Mais il rappelle néanmoins l'importance qu'ont eu les laboratoires dans le développement d'une pensée originale au sein de l'École des hautes études[23], manifestant par là son désaccord avec ses collègues littéraires, qui valorisent la pensée originale et le travail en solitaire.

Ainsi, le laboratoire est l'ensemble des personnes et des « instruments » sollicités pour mener à bien une démonstration scientifique[24]. Si Bruno Latour va jusqu'à qualifier des chercheurs et des institutions comme l'INSEE d'instruments[25], nous n'utiliserons ce dernier mot que pour évoquer des objets inanimés. Ceux-ci intègrent évidemment les outils de la production écrite[26] : livre, article, code, écran d'ordinateur, image de synthèse, etc. On

22. [Bou84], p. 77, et la note 22 de la même page.
23. [Bou84], p. 145.
24. Ou une vérification, comme dans les nombreux « laboratoires de tests » industriels, qui renvoient alors à cette même image d'une technique appliquée.
25. [Lat89], p. 166.
26. Et se limitent certainement à ceux-ci : un appareil de mesure ne mesure rien s'il n'inscrit pas quelque chose qui se donne à lire.

a vu qu'on ne pouvait dissocier ces derniers de l'ensemble des procédures intellectuelles qu'ils permettaient.

Pour Bruno Latour, le centre de calcul et la bibliothèque sont des archétypes du laboratoire en tant qu'ils sont des lieux de concentration d'inscriptions commensurables les unes avec les autres[27], et il insiste sur leur fonction sociale : les rencontres imprévues, les dialogues débridés ont une part importante dans la production de la science. Christian Jacob rappelle qu'« on pourrait ainsi reconnaître dans Alexandrie l'un des prototypes de ces 'centres de calcul' »[28].

On peut alors donner un sens plus général au mot laboratoire : c'est le lieu du déploiement de l'outillage mental, c'est-à-dire de sa rencontre avec la collectivité nécessaire à son développement. Le laboratoire est l'ancrage social de l'outillage intellectuel. Sa temporalité est celle du groupe ou de l'individu.

Les modes d'organisation (politique et spatial) du laboratoire dépendent évidemment du contexte géographique et historique : aujourd'hui, en France, on conçoit avant tout le laboratoire comme un réseau de personnes (chercheurs, étudiants, techniciens, administratifs, etc.), d'objets (qui constituent souvent des nœuds du réseau, puisque l'on s'agrège autour d'eux, comme on le fait autour d'un écran d'ordinateur ou à l'entrée d'une bibliothèque) et de méthodes qui permettent l'élaboration d'un travail intellectuel. L'allusion de plus en plus fréquente au « laboratoire hors les murs » et de sa variante « collaboratoire »[29], à la fois corollaires d'une mise en réseau de plus en plus serrée des chercheurs (avec l'internet, mais tout simplement avec le rétrécissement du monde) et d'une volonté politique d'investir de moins en moins d'argent dans la construction de lieux scientifiques, pourrait faciliter la critique face aux métaphores géographiques précédemment évoquées : le laboratoire est-il un lieu ou un réseau ? L'un et l'autre : les érudits n'ont pas

27. [Lat96], p. 32. Voir aussi p. 45.
28. [Jac96], p. 70.
29. Contraction des mots américains *collaboration* et *laboratory*, introduite en France par William Turner. Cf. http://www.olf.gouv.qc.ca/ressources/internet/fiches/8388786.htm

attendu le XX^e siècle pour voyager et pour entretenir des correspondances. Les scribes égyptiens usaient déjà de ce système d'échange[30] en 2250 av. J.-C. Pour les lettrés de l'Écosse du XI^e siècle, la norme était le nomadisme entre les divers laboratoires qu'étaient les bibliothèques monastiques[31]. Même s'il est souvent concrétisé par un lieu, lui-même nœud d'un réseau, le laboratoire est en fait la condition de possibilité de la rencontre du social et du cognitif. A ce titre donc, tout l'outillage mental de l'historien solitaire, incluant la lampe de son bureau[32], son ordinateur, ses lectures, sa connaissance précise d'archives disséminées en Europe, son réseau social[33], et lui-même constituent son « laboratoire ». Suivant les formes de la production scientifique, des souris, des molécules d'ADN, des textes, ou même des êtres humains peuvent faire partie du laboratoire. Ainsi, l'anthropologue fait-il partie intégrante de son laboratoire, qui ne se limite pas à son « terrain » et à ses instruments[34].

Il faut signaler une acception récente du mot laboratoire, dont on pourrait penser qu'elle se rapproche de notre définition, mais qui nous semble trop large : des sociologues comme Paul Bernard décrivent la société québécoise comme un « laboratoire en grandeur nature », notamment grâce aux possibilités qu'offre aujourd'hui l'agrégation des bases de données informatisées des divers ministères[35] (santé, éducation, etc.). C'est aussi à cette conception du

30. [Goo97], p. 240.

31. [Goo97], p. 170.

32. Les instruments visuels, de la bougie à la lunette mal taillée, sont eux-aussi essentiels pour comprendre les conditions de possibilité des pratiques sociales et mentales associées à l'écriture, comme l'a démontré Armando Petrucci lors de sa conférence « écrire et ré-écrire au Moyen-Âge » (13 juin 1998, séminaire « Mondes lettrés », ENS).

33. Fût-il maigre, il comprend au moins son épouse, souvent remerciée dans ses dédicaces ou préfaces, son éditeur, quelques responsables de revues savantes et des étudiants ou des disciples.

34. On sait que l'anthropologue ne peut étudier une société donnée — exotique ou pas — sans tenir compte de sa propre présence au sein de cette dernière.

35. Communication personnelle. Ces méthodes sont plus fiables et moins coûteuses que les enquêtes longitudinales, avec lesquelles on perd vite la trace des personnes que l'on espère suivre. Or, même quand ils déménagent, la majorité des Québécois préviennent au moins leur centre de santé.

laboratoire que pense Jean-François Abramatic dans son rapport au secrétariat d'État à l'industrie sur le « développement technique de l'Internet »[36] : « en ce qui concerne le cas particulier de la recherche en sciences et technologies de l'information, en plus d'être un formidable outil de travail, l'internet est un remarquable sujet d'études et un laboratoire d'expérimentation en grandeur réelle absolument unique ». Dans les deux cas, il est clair que les auteurs sont persuadés que les processus d'écriture permis par l'informatique et l'internet permettent d'analyser — de lire, voudrait-on dire — une société tout entière, dans laquelle ils s'incluent naturellement. Pour eux, l'internet est un élément supplémentaire du laboratoire du chercheur, tant par ses effets sociaux de mise en réseau que par ses effets intellectuels, en permettant d'autres manipulations de l'écrit.

Cependant, le changement d'échelle induit par cette dernière conception du laboratoire nous fera préférer notre définition, à la fois étroite et large : étroite, parce qu'associée à un réseau social de taille réduite ; large, parce que son statut de condition de possibilité de déploiement de l'outillage intellectuel ne limite pas le champ d'application du laboratoire à un groupe d'érudits.

2.2.2 Temporalités longues

L'écriture est donc une technique qui se confond avec l'outillage mental en tant qu'elle associe codes, supports et procédures mentales. Parmi celles-ci, certaines sont apprises, d'autres inventées, toutes s'oublient, se réorganisent et se transmettent à la fois, dans des cadres sociaux et temporels restreints que nous appelons laboratoires, même si l'on met surtout en avant les institutions, comme l'école. Mais ces dernières appartiennent à une autre temporalité. On peut néanmoins décrire ce second régime d'interaction sociale.

Si, au début, « l'écrit ne fige pas la pensée, il la dynamise »[37], au fil du temps, l'exploitation des sources finit par créer une langue qui se dissocie complètement de l'oralité, « qui cultive souvent l'obscurité, l'archaïsme ou la

36. http://mission-dti.inria.fr
37. [Jac96], p. 73.

forme rare, et qui resémantise avec délectation des mots épinglés dans les collections lexicographiques »[38]. Plus généralement, l'écriture limite aussi la créativité et la liberté d'expression[39]; les procédures d'apprentissage et d'érudition portent en elles les germes de pratiques sociales figées, routinisées, où le lettré, qu'il soit scribe, prêtre ou juriste, tire son pouvoir de ses connaissances et de sa maîtrise technique, où l'obligation de référence aux textes du passé réduit la dynamique intellectuelle; il s'ensuit une résistance à l'évolution qui porte en elle les germes de révolutions, au moins linguistiques[40].

On reconnaît là un phénomène typique de toute technique: suivant les modalités de son appropriation sociale, elle apparaît dynamisante ou source de régression. Ceci est d'autant plus clair dans ce contexte que la capacité de l'écriture à valoriser certains usages de la psyché la propulsent au centre des relations de pouvoir.

Et les statuts des experts s'imposent très tôt: celui de l'auteur, de l'écrivain — même quand ce dernier, comme Homère, est virtuel —, suivi de celui de l'éditeur, « intermédiaire entre l'auteur, le texte et les lecteurs »[41] — et inventeur des pictogrammes —, dont les choix vont engendrer autant de commentaires que la variété des textes qu'il a sélectionnés, éliminés, organisés. Statut du libraire, qui prend en charge à la fois la copie et la vente des textes des auteurs, ces derniers étant souvent ses amis[42].

Mais la construction de tels archétypes s'étend aussi au livre et au texte.

Le livre diffusé acquiert vite un statut de marchandise. Pour assurer sa vente, on en fait la publicité par voie de prospectus, et l'on va même jusqu'à en mettre des extraits en circulation[43]. Le texte, centre névralgique de ce triplet instrument/code/pensée, gagne ou perd sa légitimité au fil du temps: l'éditeur et l'auteur contribuent à l'obsolescence des livres qui leur servent à

38. [Jac96], p. 69.
39. [Goo97], p. 194.
40. [Goo97], p. 286.
41. [Jac96], p. 63.
42. Par exemple, Atticus était le libraire et le correspondant de Cicéron ([Mar96], p. 61).
43. Les éditeurs qui allèchent les lecteurs avec la publication sur le web d'un chapitre du livre qu'ils espèrent vendre n'ont rien inventé ([Mar96], p. 62).

en faire de nouveaux, ce qui reconstruit et réordonne le patrimoine culturel. Et ce travail critique, cette volonté de synthétiser produisent à leur tour d'autres effets intellectuels, qui — on l'a vu — vont s'enraciner dans toute une série de pratiques, plus ou moins codifiées, avec une invocation complexe des « autorités » choisies parmi les auteurs du passé pour asseoir la légitimité d'un nouveau traité[44].

Ainsi se construisent des statuts sociaux, des formes littéraires, des normes, qui se diffusent dans un espace social et géographique bien plus large que le laboratoire, et qui pourra s'étendre jusqu'à la totalité de l'Empire romain. Cette seconde temporalité, certes plus lente, mais surtout « délocalisée » par rapport à l'individu, est à distinguer de la première : appliquée à une société tout entière (ou à plusieurs), elle génère des effets proprement structurels.

Et donc, même si l'on néglige ici l'usage de l'écriture dans l'administration en se limitant aux pratiques intellectuelles, on mesure à quel point ses aspects proprement matériels, sociaux, intellectuels et politiques se combinent et s'entrelacent : Aristote est censuré au XII^e^ siècle[45] ; à cette époque, le code se simplifie, et la lecture peut redevenir silencieuse[46]. Les encyclopédies (ré)apparaissent, et les universités se multiplient dans le Nord de l'Europe tandis qu'à Florence, en 1334, un enfant sur deux apprend à lire[47]. Au renouveau de l'écrit est associé un renouveau des savoirs. Deux siècles plus tard, l'orthographe du français se stabilise, principalement sous l'influence des imprimeurs... hollandais : si Geoffroy Tory est à l'origine du *ç*, « Le *j* et le *v*, ainsi que les accents aigus et circonflexes remplaçant les *s* superflux devinrent plus communs dans leurs impressions, qui pénétrèrent à flots en France même. Ils contribuèrent donc largement à accoutumer les lecteurs à ces innovations »[48]. La Réforme protestante facilite le développement de l'instruction en même

44. [Jac96], p. 73.
45. [Mar96], p. 172.
46. Jesper Svendro date l'apparition en Grèce de la lecture silencieuse du V^e^ siècle avant Jésus-Christ ([Sve90], p. 263).
47. Soit 10 000 enfants; [Mar96], p. 185 et p. 310.
48. [Mar96], p. 288.

temps qu'elle permet d'échapper à la « tyrannie de Rome »[49]. Paralèllement, la Bible de Luther fixe la langue allemande.

2.2.3 Complexités

On redécouvre donc que la pensée se nourrit du livre, qu'elle recompose l'instrument et le code, comme eux-mêmes se réorganisent de façon réflexive, et dynamisent l'activité intellectuelle. Ces multiples processus sont d'autant plus complexes qu'ils interagissent aussi avec le social, suivant deux temporalités. Hélas, souvent, seuls certains de ces processus sont mis en avant, au détriment d'autres.

Pour ne citer que deux exemples, on rappellera que l'invention de la presse à imprimerie s'interprète plus comme le résultat d'une exigence intellectuelle que comme la cause d'un nouvel humanisme[50], même si, bien sûr, elle aura aussi des effets intellectuels, mais inattendus. Par exemple, cette invention n'eut aucune influence sur le code et la transcription même de l'écriture : « [le propos de Gutenberg] était qu'on ne puisse pas déceler son intervention et qu'on prenne chacun des exemplaires de sa Bible pour un manuscrit »[51]. On se souviendra aussi de l'importance des logiques économiques et politiques sur la psyché et la culture : les préoccupations industrielles et commerciales ont plus contribué à la stabilisation de l'orthographe du français que les serviteurs de l'Académie française appointés par Richelieu[52] ; le succès littéraire des écrivains français du XIXe siècle (Chateaubriand, Victor Hugo, Dumas, Lamartine, Balzac, etc.) doit beaucoup à l'introduction de la publicité dans les journaux par Émile de Girardin en 1836[53].

Ainsi, pouvons-nous définir l'outillage mental généralisé, d'un groupe ou d'un individu, comme la réunion de son outillage mental et de son laboratoire,

49. [Hal90], p. 270.
50. [Pet90], p. 267.
51. Intervention de Jérôme Peignot ([AL91] p. 289). Cf. aussi [Mar90], p. 158.
52. [Mar96], p. 258.
53. [Mar96], p. 388.

dans une temporalité à l'échelle de l'homme. Mais l'ensemble de référence des outillages mentaux généralisés ne peut être appréhendé alors que dans une autre temporalité, à l'échelle de l'histoire et de la société. Apparaissent alors des enjeux complexes — culturels, politiques, religieux, économiques. Pour analyser ces derniers, au regard de l'histoire de l'écriture et de nos activités intellectuelles et sociales, nous reprendrons le concept de « technique de l'intellect » défini par Jack Goody[54]. Celle-ci définit un régime d'interaction « extérieur à l'acteur », où les notions de matérialité et d'esprit intérieur perdent de leur sens au profit d'un *plongement* des hommes dans l'univers entier de l'écriture, de ses instruments, de ses artisans.

Cependant, une telle conception est très rarement partagée dans le monde des sciences humaines des XX^e^ et XXI^e^ siècles.

2.3 Technophobies

Les spécialistes des sciences humaines — disciplines faiblement cumulatives, et donc particulièrement dépendantes du texte — tendent, dans leur grande majorité, à valoriser la création, la culture, le génie aux dépens de l'environnement matériel des artistes, auteurs, et scientifiques. On compte dans ce paragraphe souligner les raccourcis et les incohérences qu'impliquent de tels discours en matière de raisonnement, puis mettre en valeur les débats et les enjeux sous-tendus par la place des techniques[55].

54. « Technologies of the intellect » dans sa version anglaise ([Goo97], p. 73; [Goo00], p. 132. Nous utiliserons parfois l'expression dérivée de la traduction française, « technologie de l'intellect », et c'est dans ce seul contexte que nous emploierons désormais ce mot « technologie ».

55. On notera cependant qu'au moment où ces pages sont rédigées, deux des plus importants bailleurs de fond des universitaires et des auteurs valorisent l'internet comme médium incontournable de la transmission de la culture française, vue comme patrimoine. Ce sont le ministère de l'Éducation Nationale — associé bien sûr à celui de la Recherche — et celui de la Culture. Aussi peut-on s'attendre prochainement à des changements opportunistes de discours de la part des détracteurs des techniques.

2.3.1 La technique tuerait la Culture

Certains des représentants actuels des mondes lettrés ne sont pas loin d'accepter cette notion d'outillage mental en ce sens qu'ils explicitent l'influence de la technique sur le code, et par là sur la langue prise comme identité culturelle. Se sentant à la fois dépositaires et responsables de la culture, ils expriment leur inquiétude, comme Jérôme Peignot à l'occasion d'une table-ronde sur l'écriture et l'informatique : « [en un siècle] nous avons perdu les accents sur les capitales, les petites capitales, les italiques sont menacées, les ligatures donc aussi... A quand... la langue elle-même ? »[56]. Il précise cette menace sur la langue : « ainsi en sommes-nous arrivés à ce paradoxe que non seulement nous ne savons pas à quel sort notre langue est vouée mais encore que nous contribuons nous-même à son saccage »[57]. Bertrand Poirot-Delpech rappelle comment, sous pression des fabricants de claviers soucieux d'économies, l'Académie française s'apprête à supprimer le *ù*, parce qu'il n'existe que dans notre langue[58]. La technique, prise dans son acception commune, apparaît clairement comme l'ennemie de la culture, si proche de l'écriture, elle aussi prise dans son sens le plus courant, c'est-à-dire le plus « noble » alors que seuls les dessins de son système de signes changent. En fait, la relation entre instruments, système graphique et pensée n'est pas explicitée. L'industrie, l'appât du gain, sont prétendus responsables d'une évolution de la graphie, qui aurait une incidence directe sur les « belles » lettres.

Certains universitaires sont sensibles au fait qu'un outillage intellectuel modernisé permet d'obtenir des résultats inconcevables avec celui d'une génération antérieure et qu'il infléchit les thématiques de recherche. Mais peu d'entre eux explicitent ce fait.

D'autres professeurs refusent simplement les ordinateurs, ou les confient à leur secrétariat, et affirment haut et fort qu'aucun outil, aucune machine ne peut avoir le moindre effet sur leurs pratiques intellectuelles[59]. Leur discipline

56. [AL91], pp. 289–290.
57. [AL91], p. 290.
58. [AL91], p. 293.
59. Ce point sera détaillé dans la seconde partie.

serait donc éternelle et exclusivement théorique. De tels discours sont assez étonnants : déjà, la catoptrique[60], la rhétorique et la grammaire se sont effondrées au cours des siècles, bien que ces deux dernières — techniques au sens le plus classique du terme — soient encore au fronton du grand amphithéâtre de la Sorbonne.

Cette relégation de la technique dans un univers matériel, méprisable car loin de la pensée pure, et censé la menacer, pose problème : ou se positionnent l'éditeur et l'imprimeur dans la chaîne de transmission de la littérature ? Par quel artifice peut-on distinguer un dictionnaire, un usuel d'une œuvre considérée comme majeure ? Le *Littré* est-il vraiment à mille lieues de *La Recherche du temps perdu* ? De façon générale, de telles réactions apparaissent étranges de la part de spécialistes de l'écriture et de son histoire.

2.3.2 Main de l'homme, bras de la machine

Il semble que le débat contemporain sous-jacent à de telles prises de position soit celui de la frontière entre l'intériorité de l'homme — spirituelle —, et l'extériorité de l'outil (dictionnaire, machine).

L'étroite relation entre l'objet technique et l'opération qu'il permet de réaliser (trier des mots, couper une haie, etc.) est plus importante que la dichotomie entre l'activité intellectuelle et l'exercice manuel. Jack Goody reprend à ce propos la métaphore de Norbert Wiener[61] : le bras articulé qu'utilise un mécanicien pour effectuer une réparation prolonge-t-il celui de l'homme ou n'est-il qu'une machine ? De même, pourrait-on compléter, la souris de nos ordinateurs est-elle un objet extérieur ou un complément de notre main[62] ? Celle-ci, comme les langages de programmation, nous permet

60. Science des rayon lumineux, disparue quand les lois de l'optique sont formalisées. Voir [Per89], p. 26 et p. 74.

61. [Goo97], p. 260.

62. Quand ils doivent recréer un fond de carte vectoriel à partir d'une image scannée, les géographes disent alors qu'il leur faut « faire la carte à la main ». L'objet souris est oublié au profit des très nombreuses manipulations nécessaires à la formation du contour des régions, villes, rivières et autres frontières qu'il leur faut redessiner, recopier, réécrire.

de « manipuler » des mots, des listes, en un sens bien concret. Ces ordonnancements, ces catégorisations produisent du sens, qui va pouvoir prendre corps à l'aide des mêmes outils (souris, ordinateur, etc.), sous forme d'articles universitaires, par exemple.

Ainsi, la carte, le *Littré*, et la bibliothèque, réserve de la mémoire, ne doivent pas être vus seulement comme des objet externes[63]. Il importe de mettre les objets qui permettent le déploiement de notre intelligence au cœur même des processus qui l'organisent, plutôt que d'opérer une brutale dichotomie entre la pensée et les objets matériels. Sinon, on est condamné à une réflexion stérile : si on définit « les compétences et les savoirs comme 'mentaux' au sens où ils sont entièrement contenus à l'intérieur de l'esprit, [...] on atteint rapidement un point où il est impossible de s'engager dans l'étude de l'interaction humaine »[64] : l'étude de la psyché ne se satisfait pas de la frontière définie par le corps humain et la technique ne s'oppose pas au mental.

La carte, apport heuristique qui surprend toujours son auteur car elle donne plus à lire et à interpréter que ce qu'il comptait y inscrire, met bien en évidence l'absurdité d'une telle frontière : « quand une carte ou un livre interviennent entre l'objet [explicité par la carte ou le livre] et le sujet, nous avons affaire à l''esprit' extérieur aussi bien qu'intérieur. [...] L'écriture nous offre un instrument capable de transformer nos opérations intellectuelles de l'intérieur ; ce n'est pas une question de *compétence* au sens limité, mais un changement dans les *capacités* »[65]. Il faut comprendre par là que l'on n'est pas dans un registre de simple potentialité, comme celle que pourrait offrir l'accès à un mode d'emploi, mais de net accroissement de la puissance intellectuelle et de ses effets réflexifs.

On en conclut que les produits de l'écriture, livres, bibliothèques, journaux, calculatrices, tout comme leur forme, leur présentation, et les industries

63. Jack Goody va jusqu'à douter que les techniques de l'écrit s'opposent aux « processus psychologiques de la mémoire » ([Goo97], p. 261).
64. [Goo97], p. 260.
65. [Goo97], p. 262.

qui y sont associées[66] ne sont pas distinguables de l'invention proprement humaine de cette écriture et de son renouvellement. Les plus matériels de ces objets participent d'une re-création de sens, d'une nouvelle appropriation des connaissances, et remodèlent la forme comme le contenu de l'écriture. On l'a vu avec les papyrus, avec les journaux du XIX[e] siècle et la littérature française, on le voit avec la ré-introduction des graphiques, dont la diffusion est facilitée par l'apparition des calculatrices et des logiciels.

L'opposition entre les pratiques proprement intellectuelles et le système technique qui environne l'individu est inopérante : il y aurait d'une part, la littérature, la création, et les sciences cognitives ? D'autre part, la sociologie, l'économie, qui se limiteraient aux phénomènes concernant de nombreuses personnes ? L'écriture, dans son abstraction intime, dématérialisée, face à la matière de l'ingénieur ? La lettre contre le chiffre ? Certes, cette dernière opposition est fortement ancrée dans le paysage intellectuel français. Mais, outre que la rupture entre les sciences de l'homme et les sciences de la nature est récente, datant du milieu du XIX[e] siècle[67], elle est de moins en moins tenable : l'organisation des disciplines dans les pays étrangers, mais aussi les appels répétés de chercheurs comme Jean-Gabriel Ganascia, Christian Jacob ou Jean-Louis Lebrave témoignent en faveur d'une plus grande souplesse intellectuelle, d'une dissolution de ces anciennes frontières universitaires. Il y aurait là le moyen de créer un dialogue fécond entre les sciences de la cognition et l'anthropologie, l'histoire et l'informatique ; de faire converser ces disciplines de l'esprit avec la linguistique et l'étude de la littérature. De penser les mondes actuels avec nos outils contemporains d'écriture, et d'enrichir la réflexion sur la communication qui souffre souvent d'un trop faible ancrage technique.

66. On ne peut abstraire une activité intellectuelle spécifique et *a priori* proprement intérieure et solitaire comme la résolution de mots croisés, fréquente dans les transports en commun, de l'ensemble des ouvriers qui concourrent à la fabrication de journaux. Cet « exemple permet d'illustrer la façon dont les développements dans la technologie de l'intellect (et plus spécifiquement dans les systèmes graphiques) agissent sur les procédés cognitifs » ([Goo97], p. 297.).

67. Philippe Descola, conférence inaugurale au Collège de France, 29 mars 2001.

En revanche, la valorisation de l'intériorité, du mental en son sens le plus restrictif, ne conduit pas qu'à nier l'outillage mental et le laboratoire, en reléguant la technique. Elle conduit aussi à nier le social, vite réduit à une somme d'individualités. Elle participe d'une idéologie, celle de la caste de l'écrit [68], et, paradoxe scientifique d'importance, elle mène à des représentations intellectuelles erronées.

On découvre alors que l'écriture, en tant que processus intimement lié à certaines formes de l'intelligence, produit évidemment des représentations mentales et sociales sur elle-même, qui à leur tour génèrent des effets sur la façon dont elle est perçue et utilisée, et donc sur la façon dont sont pensés, catégorisés, hiérarchisés les savoirs.

Malheureusement, l'idéologie des sciences humaines au sujet de l'écriture reste vivace. Et on ne peut que s'étonner, à la suite de Jack Goody, devant la posture intellectuelle des chercheurs en sciences humaines qui la véhiculent : « il est étrange qu'un groupe d'êtres humains qui passent probablement plus de temps à lire et à écrire qu'à parler et à écouter aient été si oublieux des implications psychologiques et sociales de leur profession » [69]. Du coup, on comprend que seuls les aspects matériels les plus significatifs [70] aient attiré l'attention des chercheurs, qu'ils soient linguistes, anthropologues ou historiens, alors même que « le penchant vers une science sociale 'mentaliste', souvent encouragé par un attachement à l''individualisme' » [71] les incitait à négliger les effets de cet outillage mental.

68. [Goo97], p. 289.

69. [Goo97], p. 268.

70. Jack Goody évoque le changement de support — argile, pierre, papier, etc. — et l'usage bureaucratique de l'écriture. Au sujet des effets structurels liés à l'organisation du pouvoir autour de l'écrit et de sa diffusion, on pourrait étudier l'écrit des lettrés romains comme l'écrit des télévisions américaines.

71. Cf. note 69.

2.3.3 Exemplarité

Ces représentations et leurs enjeux, qui seront détaillés dans la seconde partie, montrent vite leurs limites : le pouvoir intellectuel, et même la *doxa* universitaire, sont toujours très sensibles aux changements dans les capacités précédemment évoqués. La meilleure preuve en est donnée par l'ascension professionnelle du sociologue Pierre Bourdieu. Certes, on peut, à sa suite, invoquer le déterminisme, et affirmer que l'université française est le lieu par essence de « la complaisance résignée et de la complicité soumise »[72], où la « conformité ostentatoire » garantit la respectabilité scientifique[73], où le pouvoir est accordé à ceux qui sont les « plus inclinés et les plus aptes à reproduire sans altération » le système qui les élit[74]. Mais comment alors parvenir, comme lui, à une telle consécration universitaire et intellectuelle ? N'est-il pas, dans la forme comme dans le fond, dans son style[75] comme dans son emploi systématique de tableaux statistiques, de graphiques factoriels, un des meilleurs représentants de ceux qui créent de la pensée à force de jongler avec les techniques d'écriture, et en montrant tous les avantages intellectuels que procure cette nouvelle combinatoire ? N'est-ce pas parce qu'il a délibérément rejeté une technique d'écriture au profit d'une autre, qu'il a d'ailleurs imposée avec succès, qu'il a pu « faire école » ?

Cet exemple prouve que le monde scientifique contemporain, même vu comme lieu de prédilection du temps long, comme espace de reproduction d'une caste aveugle à ses propres outils, ne peut rejeter systématiquement ceux qui le régénèrent lorsqu'ils appliquent les principes et les méthodes qu'il promeut : il risquerait sinon de s'effondrer sous le poids de ses contradictions. Aussi, la démarche de Pierre Bourdieu apparaît-elle lucide et généreuse, mais peut-être n'explicite-t-elle pas assez les conditions de son propre succès, étroi-

72. [Bou84], p. 15.
73. [Bou84], p. 46 et p. 47.
74. [Bou84], p. 112.
75. A la fois en rupture avec le « beau style » qu'il critique si vertement, et en empruntant beaucoup à la terminologie des économistes et des marxistes : l'analyse textuelle de ses ouvrages permettrait de mesurer précisement les effets rhétoriques de l'auteur.

tement liées à une sollicitation systématique de l'outillage mental généralisé dont il pouvait disposer quand il travaillait à *Homo Academicus.*

En tant que technique intellectuelle, l'écriture est prise, sollicitée, convoitée, appropriée, et aussi transformée dans l'incessante bataille des rapports de force. Mais dans le cas de Pierre Bourdieu comme dans celui des grands intellectuels, le « génie » n'explique pas tout. Comme l'a signalé Jack Goody, l'écriture est l'ingrédient central d'une telle production, et peut donner une puissance inégalée.

En espérant avoir réconcilié la technique et la production intellectuelle, nous prenons cependant acte de nos limites, en ayant conscience qu'il faudrait solliciter trop de savoirs, notamment linguistiques et philosophiques, et ce sans garantie de succès, pour aborder dans le détail toutes les dimensions de l'écriture. Nous retiendrons que l'acceptation de l'écriture comme technique permet de ne pas couper l'individu de sa société, sur les plans de son autonomie intellectuelle comme politique : au contraire, elle le re-socialise, lui donne une singularité universalisable qui l'inscrit dans une triple dimension humaine, matérielle et conceptuelle.

Chapitre 3

Écriture informatique

Nous montrerons ici en quoi l'informatique et l'internet réorganisent la technique de l'écriture et quelles potentialités en découlent. Nous ne comptons pas raisonner en termes de transformations futures — qui ne pourront être analysées qu'à la lecture de lentes appropriations — mais repérer quelques changements quantitatifs et qualitatifs du fait de l'existence des ordinateurs, des réseaux, et des automates que nous utilisons aujourd'hui pour écrire.

Ces changements ne sont pas « révolutionnaires ». On sait d'une part que la technique de l'écriture est en perpétuelle reconstruction ; d'autre part, on verra que les objets et les méthodes informatiques se sont mis en place depuis plusieurs décennies.

Pour une meilleure compréhension des potentialités — fastes ou néfastes — et des enjeux politiques et économiques de cette technique, nous limiterons volontairement notre étude aux « professionnels de l'écriture », c'est-à-dire aux personnes confrontées au quotidien avec du texte, même si toutes ne s'interrogent pas sur les fonctions de l'écrit : universitaires, érudits et chercheurs actuels, personnes des métiers du livre (journalistes, graphistes, correcteurs, secrétaires de rédaction, etc.), ingénieurs et informaticiens. Bien sûr, ces transformations touchent aussi un public plus large, mais il semble raisonnable dans un premier temps de se cantonner à des pratiques et des outillages mentaux relativement homogènes, de façon à pouvoir réaliser des mesures et des analyses qui permettent la comparaison et la généralisation.

Nous mettrons tout d'abord en évidence des transformations élémentaires, concrètes, des outils d'écriture contemporains, puis détaillerons les transformations intellectuelles qu'ils induisent, sans négliger les formes spécifiques de culture nécessaires à l'actualisation de ces dernières. Enfin, nous exposerons les enjeux économiques et politiques tels qu'ils se dévoilent à la lueur de cette compétition intellectuelle.

3.1 L'ordinateur

Dans la suite, un ordinateur est pris dans son sens commun contemporain : nous oublierons donc l'ENIAC et ses trente tonnes (1943) pour ne considérer que des objets constitués d'un écran, d'un clavier, d'un système de traitement des signes et d'un système d'archivage. A cette configuration minimale, dont nous demandons qu'elle permette de lire et d'écrire [1], peuvent s'ajouter divers autres « périphériques », comme la souris ou le *joystick*, l'imprimante, etc., mais aussi d'autres éléments de base : deux ou plusieurs unités centrales reliées par un réseau, deux écrans, d'autres imprimantes, etc. Ces divers éléments ne sont pas nécessairement en un même lieu.

3.1.1 Rappel historique

L'ordinateur n'est pas une invention récente : Blaise Pascal construit sa machine à calculer en 1642 [2], et Charles Babbage imagine un calculateur en 1822. Si celui-ci ne fut jamais réalisé, divers programmes ont été réalisés à son intention par Augusta Ada Lovelace, la fille du poète Byron. Pour ses métiers à tisser, Falcon utilise des cartes perforées en 1728, déjà employées

1. Ce qui distingue *a priori* l'ordinateur des machines dédiées au jeu et des téléphones portables (mais pas du Minitel), même si on pourrait remettre en cause, dans une étude plus générale, cette classification, assez liée à la catégorie d'usages que l'on a choisie.

2. Pour cette référence et les suivantes, cf. http://www.maxmon.com

à piloter des automates et des boîtes à musique[3]. Joseph-Marie Jacquard perfectionne et industrialise le procédé en 1804[4]. Cette façon de transmettre de l'information à une machine, tant dans sa matière (le papier ou le carton) que dans son code (clairement binaire) se maintiendra jusqu'à la fin des années 1970. La machine à écrire, dont on verra qu'elle n'a aucun rapport avec l'ordinateur, même si les claviers de l'une et de l'autre semblent proches, fut, elle-aussi imaginée il y a longtemps : en 1714, soit 150 ans avant sa première réalisation (1867).

Plus en rapport avec l'internet, une imprimante commandée à distance — pour les besoins du télégraphe — est testée en 1902[5] ; Édouard Belin fabrique son bélinographe[6] en 1907. Le tube à faisceau cathodique est conçu en 1897 par l'allemand Karl Braun[7]. Vannevar Bush construit un réel ordinateur en 1927, mais on oublie que celui-ci est analogique. Il faut attendre 1940 pour que le codage numérique, décrit comme l'essence de l'informatique, apparaisse avec la machine de George Stibitz reliée par téléphone à un *teleprinter*.

On pourrait considérer que les réseaux informatiques actuels, au premier lieu desquels l'internet, ont donc plus de 60 ans, et qu'à l'exception de la souris, les composants de nos ordinateurs sont centenaires.

Certes, quelques améliorations techniques se sont produites, dues en grande partie aux progrès de l'électronique (semi-conducteurs, circuits intégrés, microprocesseurs, etc.). Mais le résultat actuel est plus le fruit d'associations inédites d'objets connus et de lentes améliorations que de pure « innovation ». Par exemple, les travaux sur l'interface graphique se développent dans les années 1970, parce que l'écran de télévision devient abordable et peut être utilisé comme terminal d'un micro-ordinateur[8].

3. [Mar96], p. 438
4. 1801 d'après [Mar96], p. 438.
5. Par Franck Pearne, et fonctionne en 1910.
6. Ancêtre du télécopieur.
7. Prix Nobel en 1909 avec Gugliermo Marconi. En 1931, René Barthélemy transmet à distance des images de télévision en France ; la définition n'est alors que de 30 lignes.
8. Terme dont l'apparition date de l'invention du Micral, en 1973. Mais cette machine française, créée par André Truong, n'a pas eu le succès escompté.

Pour éviter une surenchère dans l'hagiographie confinant au culte du progrès technique, nous nous contenterons de rappeler quelques faits supplémentaires en liaison avec notre recherche : en 1977, Steve Wozniak et Steve Jobs commercialisent l'Apple II [9], premier micro-ordinateur massivement vendu [10]. Fait essentiel pour notre propos, le tableur « Visicalc » (1979) est à lui seul responsable du quart des ventes [11]. Les premiers ordinateurs à système d'exploitation commandé par interface graphique sont en 1983, le Lisa, puis en 1984, le Macintosh, tous deux produits par la marque Apple. Le statut de l'ordinateur peut alors être comparé à celui de la voiture, puisque tous deux peuvent se piloter sans expertise « informatique » ou mécanique. Mais cette analogie fortement répandue étouffera tout questionnement sur les transformations des pratiques d'écriture, tandis que les objets se multiplient : avec le langage PostScript (1982) et les imprimantes laser (1984), les conditions sont réunies pour que l'ordinateur puisse pénétrer dans le cercle fermé des mondes lettrés. Parallèlement, d'autres ordinateurs sont développés, comme ceux qui utilisent le système d'exploitation Unix ; mais ils resteront longtemps réservés aux professionnels de l'informatique scientifique, notamment à cause de l'absence d'environnement graphique.

3.1.2 Calame

Pris séparément, chacun des objets matériels qui composent l'ordinateur a peu de chances de modifier notre outillage intellectuel. Nous rencontrerons à chaque fois des évolutions marginales, dont seule la combinaison finira par avoir des conséquences sur nos manières de faire et de penser. Le clavier est un bon exemple de telles modifications à la marge : ce n'est pas un nouvel objet, puisqu'il est quasiment identique aux machines à écrire ; les seules

9. http://www.blinkenlights.com/classiccmp/apple2history.html

10. Son prédécesseur ne s'était vendu qu'à 175 exemplaires.

11. Qui s'élèvent en 1980 à 50 000 machines, pour un prix moyen unitaire de 1 200 $ de l'époque. IBM, associé à l'entreprise Microsoft, comprend la menace et sort son premier IBM-PC en 1981. Mais la force commerciale du géant se fait vite sentir, puisque le même nombre de « PC » est vendu en 74 jours seulement (même référence).

transformations d'importance ne résident pas dans l'ajout de touches permettant l'inscription de nouveaux caractères (comme l'arobase, par exemple), mais dans l'apparition de touches qui, *a priori*, n'inscrivent pas de signes visibles : ces touches, dites de fonction [12] permettent de transmettre des messages spécifiques à la machine. On pourrait penser qu'elles sont un résidu du monde des programmeurs et que l'apparition de la souris, complément indispensable des interfaces graphiques, les condamnerait [13]. En fait, ces deux manières d'adresser des commandes se combineront [14]. Les fonctions de l'outil d'inscription sont donc étendues à des séquences d'instructions dédiées aux moteurs d'écriture que sont l'ordinateur et le logiciel.

La souris [15] donne l'impression formidable d'être toujours au cœur du texte, de le manipuler, de le réorganiser en le pensant comme objet graphique, au point que certains traducteurs, comme Jean-René Ladmiral [16], considèrent que l'ordinateur propose un retour au stylo, antérieur à la distanciation du texte induite par la machine à écrire. On en oublierait presque les commandes que rend accessibles la souris, par le biais de la sélection des « menus ».

Ces deux objets ne servent donc pas qu'à saisir ou reproduire du texte, mais à le manipuler, le réordonner, grâce à d'autres fonctionnalités permises par la machine. On est dans le registre du « bras articulé », et la proximité entre texte et pensée est accrue.

12. « Control », « Alt » ou « Option », « Commande », « Meta » principalement, mais aussi flèches, « Escape », « Tabulation », « Suppression », touches « F1 » à « F12 », etc.

13. Tout comme on s'étonne de ne pas voir encore diffusé le stylet électronique associé à une tablette graphique, qui fait son apparition avec les assistants numériques personnels (terme canadien), appelés par les Français assistants électroniques. Il ne serait pas difficile de concevoir une écriture informatique qui renoue avec l'écriture manuscrite.

14. Cf. paragraphe 3.4.2 page 77.

15. Et ses variantes, boule de commande (*trackball*) et pavé tactile (*touchpad*), cf. toujours le même dictionnaire québécois http://www.granddictionnaire.com pour la traduction.

16. [AL91], p. 272.

3.1.3 Traitement du texte

Ces possibilités de réorganisation du texte n'ont pas été imaginées pour les littéraires. Leur histoire est intimement liée à celle de l'informatique. Non pas qu'avant, elles n'existaient pas (la table de montage des journaux des années 1970, où les épreuves étaient hachées menu, recollées soigneusement avant le passage à la photogravure, en est la preuve), mais elles sont la condition du développement de cette science : dès que les ordinateurs ont pu être pilotés avec un clavier, et qu'il a été possible d'archiver le code nécessaire à l'exécution d'une tâche, se sont développés les « éditeurs » (de texte) [17].

En effet, la majorité des « programmes », ces successions de phrases interprétables par la machine, se ressemblent, et il fallait pouvoir réorganiser de façon rapide — et si possible systématique — les unes pour en réaliser d'autres. D'où le besoin de rechercher du texte, de le remplacer, de le déplacer, de le dupliquer. Cette production textuelle massive destinée à multiplier le nombre de tâches que peut réaliser un ordinateur a donné lieu à la construction de « bibliothèques de ressources » [18], indispensables aux programmeurs : ce sont en fait des milliers de routines, de programmes spécifiques qui peuvent être aisément sollicités par d'autres programmes.

Ce besoin impérieux d'écrire, de transmettre du code coûte que coûte, s'est encore amplifié avec le développement des réseaux : non seulement il faut « traverser » les diverses strates d'un ordinateur, mais, depuis 20 ans à 30 ans, les différentes « couches » de ces réseaux. Partant de leur hétérogénéité, ces couches ont été définies de façon à normaliser les opérations qui permettent l'interconnection des machines : « les concepts de base de cette architecture constituent le modèle de référence pour l'interconnexion de systèmes ouverts — dit modèle OSI — dont les modalités sont réparties en [...] couche physique,

17. Un éditeur, comme *vi*, *emacs*, *pico* ou *joe*, s'apparente à un traitement de texte, mais il est à la fois plus primaire (la notion d'enrichissement typographique n'existe pas ou est automatique) et plus performant (en matière de manipulation du texte, de programmation, etc.).

18. Pour une évocation détaillée des emprunts du vocabulaire informatique au registre du livre (ligne, page, bibliothèque, etc.), cf. [Ani98], p. 30.

couche liaison de données, couche réseau, couche transport, couche session, couche présentation, couche application »[19].

Ainsi, les textes de l'informaticien et de l'utilisateur sont-ils appelés à subir de multiples transformations. Nous ne doutons pas que ces nouveaux statuts du texte, comme le développement des « outils » conçus pour l'appréhender, finiront par être perçus par un grand nombre d'auteurs, au-delà du monde des ingénieurs. Nous retrouvons là un aspect du processus réflexif de l'écriture, d'autant plus difficile à cerner que le respect des usages traditionnels incite à le masquer au mieux. Par exemple, la programmation, nouveau genre d'écriture, permet de fabriquer des outils qui donnent à l'utilisateur l'illusion que le traitement de texte modifie son mode de travail de façon minimale.

Ce type de paradoxe nous apparaît constant dans les discours et analyses relatifs à l'innovation : celle-ci ne se produit pas là où on l'évoque. Avec l'informatique, la technique reste dans le registre de la machine-outil qui allège des tâches et qui sert avant tout à reproduire — sans imagination — des pratiques enracinées[20]. Cette logique d'effort physique, de travail « manuel », incite à négliger le caractère intellectuel des pratiques quand on est confronté à un ordinateur. Il semble pourtant que ce soit là que l'on puisse parler d'innovation, et non pas en expliquant aux érudits que le traitement de texte leur permet d'accéder au statut bien peu convoité de secrétaire. Pourtant, les discours sur l'innovation adhèrent doublement à la métaphore de la presse de Gutenberg : d'une part, en euphémisant toute forme de changement, comme l'inventeur l'a fait lui-même[21] ; d'autre part, en colportant une histoire mythique de cette invention, sous sa forme la plus réductrice[22] : la technique génère des outils,

19. Source : le grand dictionnaire terminologique.

20. Un très bel exemple de cette fixation sur la matière classique est le format PDF, qui donne l'illusion que la page sur le web est en tous points conformes à celle d'un imprimé, et qu'elle est difficilement copiable.

21. Cf. page 38.

22. Il conviendrait de faire une étude des mythes occidentaux relatifs à la technique comme on fait celle des mythes d'autres sociétés.

qui, par magie, vont amplifier les potentialités de la pensée intérieure. Ici, le système de signes est oublié au profit d'une conception binaire de la matière et de l'esprit, incompatible avec une étude précise de l'invention et du fonctionnement des ordinateurs : certes, la combinaison d'instruments — d'usines, en fait — et d'un travail conceptuel permet de concevoir d'autres machines ; mais celles-ci ne transforment pas la matière, elles sont avant tout destinées à la production de signes, peu ou prou associés à notre système graphique. Ces symboles ont deux fonctions entrelacées : la plus visible, destinée à l'auteur ou au lecteur, est aussi la plus traditionnelle. La seconde, assez masquée, est destinée à la gestion de ce système graphique. Cette double fonctionnalité permet le travail de production et de réorganisation des signes sans lesquels la pensée ne peut se déployer. Mais on remarque aussi un saut conceptuel, fruit d'une audace intellectuelle qu'on aurait tort de négliger, et générateur d'une nouvelle relation à l'écriture, qu'il convient de détailler.

3.1.4 Support et codes

Ces machines à produire du signe génèrent en effet une rupture conséquente dans nos habitudes : à la grande différence des textes qui étaient réalisés sur papyrus ou à l'aide d'une machine à écrire, celui-ci n'est plus directement accessible : l'ordinateur entérine la séparation du lieu de l'inscription et du lieu de la visualisation. Là encore, il serait dangereux de parler d'innovation. Les premières machines à projeter des images[23], puis la photographie et surtout le cinéma ont déjà introduit l'écran comme dispositif essentiel de la lecture : on ne regarde pas un film dans la pièce du projectionniste. Mais avant la diffusion des ordinateurs individuels, les professionnels qui avaient un accès distancié à l'écrit étaient engagés dans des processus industriels suffisament complexes pour qu'on puisse considérer qu'ils fussent rares. Aujourd'hui, il est quasiment impossible à un non-aveugle de concevoir

23. Van Musschenbroek est l'auteur de la première projection d'images animées en 1725 ([Per89], p. 32).

un texte électronique sans écran[24], alors même que ce texte est enregistré sur un disque plus ou moins proche ; l'internet est le meilleur exemple de cette délocalisation radicale, même si elle fut préfigurée par le télégraphe. La meilleure preuve en est donnée par les *webmail*, qui permettent de rédiger son courrier alors que la machine qui le stocke et le distribue peut être située à des centaines, voire des milliers de kilomètres du scripteur.

Ce dédoublement du support, entre disque dur composé d'éléments magnétisés positivement ou négativement (d'où la notion de codage binaire), et écran, somme de points rouges, verts ou bleus, auparavant ou verts ou jaunes sur fond noir, induit un effet essentiel dans le code de l'écrit : au système graphique, mais aussi à la structure du texte, se surimposent des centaines de codages, au point que ces derniers priment sur les premiers.

Dans une acception simple, on pourrait supposer que le codage entre l'homme et la machine soit unique, et parfaitement réversible. L'ordinateur assurerait les deux traductions nécessaires, même si le système de signes serait étendu à un alphabet qui dépasse un cadre strictement national[25]. On pourrait admettre que cet alphabet subisse des altérations, consécutives à des rapports de force culturels, comme cela se produit déjà : l'absence d'accents dans le premier code ASCII[26], la disparition du *œ* dans la graphie sur le Web en sont autant d'exemples[27].

Dans les faits, la situation est bien plus inquiétante ; cette fonction de transcodage n'est plus maîtrisée par le scripteur, et finit par devenir opaque,

24. Si on reste impressionné par l'exploit de nombreux érudits qui recopient au moyen d'un ordinateur — et parfois au moyen de machines encore plus rudimentaires, et sans écran — des textes en bibliothèque, les yeux toujours fixés sur l'ouvrage, il n'en reste pas moins qu'à un moment, ce texte sera visualisé sur un écran.

25. Comme les accents ou la ponctuation en espagnol, par exemple, ou la possibilité d'inscrire des caractères tels que le *Å*, le *Ø* ou le ¶.

26. American Standard Code for Information Interchange.

27. Qui peuvent aussi correspondre à des choix « culturellement suicidaires ». Par exemple, les fondateurs du Minitel ont conçu un alphabet réduit aux lettres non accentuées, dans leur désir de fabriquer un objet populaire. Les Américains ne sont pas toujours les premiers à malmener la graphie nationale ([Ani98], p. 96).

même pour les informaticiens les plus érudits : déjà, des dizaines d'instructions s'intercalent entre le texte sur le disque et son affichage (déchiffrement des secteurs du disque, transmission à l'unité centrale, interprétation par un logiciel, commande d'une carte graphique pour afficher des dessins correspondant à une police de caractères, etc.). Si celles-ci sont encore vaguement « normalisées », les codages induits par les logiciels, souvent appelés « formats », sont tous différents, même s'il existe parfois des passerelles entre eux. Ce sont donc des centaines de codes qui s'interposent entre ce que le scripteur croit écrire et ce qui est véritablement inscrit[28] : l'industrie ne propose pas toujours que des solutions rationnelles ou universelles...

On découvre alors, comme corollaire de cette distance entre espace de visualisation et espace de stockage, que le support n'existe pas sans codage, sans texte. Ce fait est essentiel, et il induit une association inattendue entre l'instrument et le système de signes : si le même texte, ici pris comme succession de caractères alphabétiques, pouvait être indifféremment écrit sur la pierre, la cire, le bois, le papier, aujourd'hui, son inscription sur un écran passe par un support qui est définitivement illisible si l'on ne dispose pas de toute une panoplie d'objets, physiques comme moins matériels : disque, écran, et leurs pilotes, logiciel, etc. Un même disque dur ne sera pas lisible par un autre système d'exploitation que celui qui l'a « formaté » ; un document QuarkXPress écrit dans sa version 3.32 n'est pas lisible par la version 3.31 du même logiciel ; une cartouche amovible n'entre pas dans un lecteur qui accepte un support de taille différente.

28. « Les codages sont relativement normalisés en ce qui concerne l'écriture sur disques, et le code machine des processeurs, même si le nombre des plus répandus dépasse la vingtaine : Intel a réalisé une dizaine de processeurs (DX, DX-2, Pentium II, III, IV...), Sun quelques-uns, HP aussi (processeurs HPPA). Puis il y a les machines parallèles de Convex, Fujitsu, Cray, IBM. Dans chaque cas, il y a un 'code machine' différent. En revanche, pour ce qui est des cartes graphiques, audio, et des modems, les codages se comptent à chaque fois par centaines. Restent enfin les logiciels, qui souvent ont un codage propre » (communication de Jacques Beigbeder, responsable du service de prestations informatiques de l'ENS).

Cette dissociation entre lieux *géographiques* de la visualisation et du stockage induit toute une série de traductions qui deviennent de plus en plus difficiles à maîtriser, et qui finissent par unifier les notions, historiquement distinctes, de support et de système de signes, de matériel et d'écriture.

3.2 Exemples de codages

Afin de montrer concrètement comment ce type d'agrégation s'organise, nous proposons quelques exemples bien connus des utilisateurs de l'informatique — à l'exception peut-être du premier —, qui mettent en évidence le nouveau statut du texte et la façon dont ce dernier est pris dans des logiques monopolistiques.

3.2.1 Instructions pour processeurs

Le premier exemple met en évidence la façon dont des instructions sont transmises au processeur de la machine. Le langage utilisé est l'« assembleur » et on demande à deux processeurs différents de réaliser l'opération « $2 + 5$ ».

Le tableau 3.1 de la page 58 donne la liste de ces commandes, destinées à un processeur Intel et à un processeur Sparc (machines Sun), tout en négligeant les lignes du début et de la fin des programmes, similaires, et souvent simples commentaires.

Pour un processeur, on distingue la mémoire (espace de stockage) des registres (capables de réaliser des opérations comme l'addition). La logique du programme d'addition proposé consiste donc à mettre 2 et 5 en mémoire, puis à glisser ces deux nombres en des registres, à mettre la somme dans un troisième registre, puis à renvoyer ce résultat en mémoire.

Sans rentrer dans les détails, on constate que le programme destiné au processeur Sparc est plus lisible que l'autre : quand le premier dispose de registres dénommés `%o0`, `%o1`, etc., le second a des registres du type `%eax`, `%edx`, `%ecx`, sans que l'on comprenne leur ordre d'apparition. Ensuite, les opérations pour le Sparc sont explicitées : `mov2, %o0` envoie le chiffre 2 dans

Intel	Sparc	Commentaire Intel	Commentaire Sparc
.file"intel.c"	.file"sparc.c"		
...	...		
.globl main	.type main,#function		
.type main,function	.proc04		
main :	main :		
pushl %ebp	!#PROLOGUE# 0		
movl %esp,%ebp	save%sp, -128, %sp		réserver 128 octets de mémoire
subl $12,%esp	!#PROLOGUE# 1	12 octets réservés	
movl $2,-4(%ebp)	mov2, %o0	mettre 2 en mémoire	mettre 2 dans le registre o0
movl $5,-8(%ebp)	st%o0, [%fp-20]	idem pour 5	mettre le contenu de o0 en mémoire
movl -4(%ebp),%eax	mov5, %o0	mettre 2 dans le registre eax	mettre 5 dans le registre o0
movl -8(%ebp),%edx	st%o0, [%fp-24]	5 dans edx	mettre le contenu de o0 en mémoire
leal (%edx,%eax),%ecx	ld[%fp-20], %o0	addition	ramener 2 dans le registre o0
movl %ecx,-12(%ebp)	ld[%fp-24], %o1	7 en mémoire	ramener 5 dans o1
pushl $0	add%o0, %o1, %o0		addition
call exit	st%o0, [%fp-28]	fin	renvoyer 7 en mémoire
...	mov0, %o0		vider le registre o0
...	callexit, 0		fin

TAB. 3.1 *Comparaison des assembleurs pour processeurs Intel et Sparc*

le premier registre, `st%o0, [%fp-20]` sauvegarde cette valeur dans la pile de la mémoire. Ainsi ce premier registre est-il libre pour accueillir le nombre 5 : `mov5, %o0`, qui lui-même sera stocké ensuite un peu plus haut en mémoire : `st%o0, [%fp-24]`. Pour le processeur Intel, les deux opérations sont réduites à une : `movl $2,-4(%ebp)`.

L'addition se produit en glissant les valeurs en mémoire dans des registres : `ld[%fp-20] %o0` et `ld[%fp-24], %o1`, pour le Sparc, qui réalise l'addition dans le même registre qui contient le nombre 2 : `add%o0, %o1, %o0`. En revanche, l'Intel utilise un troisième registre : `leal (%edx,%eax),%ecx`.

Ensuite, les deux processeurs renvoient le résultat de l'addition en mémoire : `st%o0, [%fp-28]` pour le Sparc, `movl %ecx,-12(%ebp)` pour l'Intel[29].

Nous réalisons que, même pour une opération très simple, la série de commandes adaptée à un processeur ne fonctionnera pas pour l'autre.

3.2.2 Format RTF

Ce format[30] a été inventé par Microsoft, et a été maintenu tant que cette firme voulait accroître son monopole sur l'écrit avec la diffusion de ses logiciels de type Word. C'est un format « d'échange », conçu pour qu'un traitement de texte distinct de celui qui a été utilisé pour la rédaction d'un document puisse interpréter ce dernier. Il n'est pas attaché à un système d'exploitation donné. Si l'on ne dispose pas du traducteur qui permet de visualiser à l'écran le texte mis en page, on peut néanmoins, après avoir abandonné les commandes absconses du début du fichier, en retrouver le sens. À partir de l'extrait qui suit, tiré du rapport Silberman[31], on arrive à relire sans trop de difficultés cette phrase :« Accompagnant la croissance rapide de la demande d'expertise aux sciences sociales, la production de données s'est considérablement développée ».

29. Ce qui fait que pour l'un comme pour l'autre, le nombre 2 est en bas de la pile de la mémoire, le nombre 5 juste au dessus, et le 7 encore au dessus.

30. RTF est l'acronyme de *Rich Text Format*.

31. http://www.education.gouv.fr/rapport/silberman/table.htm

```
[...] \par \}\pard \sb80\widctlpar\intbl\adjustright
\{\i\fs22 Accompagnant la croissance rapide de la demande
d\rquote expertise aux sciences sociales, la production de
donn\'e9es s\rquote est consid\'e9rablement d\'e9velopp\'e9e.
```

On reconnaît là un langage par balises[32] et, malgré sa complexité, le lecteur peut accéder au texte, décider de le remanier, de tenter de le réorganiser, de le copier (comme on le fait ici), etc.

3.2.3 Formats Microsoft

En général, les logiciels qui prétendent libérer l'utilisateur des contraintes techniques favorisent des modes de surveillance particulièrement dangereux grâce à l'usage d'un codage binaire, *a priori* illisible. Toute une série d'informations est inscrite sur les fichiers écrits par l'utilisateur sans qu'il en ait connaissance : nom du propriétaire du logiciel — situation toujours délicate pour celui qui l'a piraté —, divers outils de l'utilisateur (ordinateur, imprimante[33], etc.) et très fréquemment, les divers épisodes de la rédaction du document[34]. L'application sociale de ce dernier fait peut-être très cruelle : c'est ainsi que certains chercheurs ont pu découvrir que certains de leurs collègues, qui leur envoyaient des fichiers attachés produits avec ce logiciel, pensaient le plus grand mal d'eux.

Voici par exemple ce que donne à lire un fichier qui a été créé en deux étapes : Tout d'abord, nous avons écrit — avec le logiciel Word 5 pour Macintosh[35] — « Monsieur X est un universitaire borné et dangereux », et demandé à ce que ce document soit enregistré sur disque. Ensuite, nous avons rem-

32. Cf. le paragraphe 3.2.6 page 64.

33. Exemple tiré d'un document reçu en 2000 :
`Confidential^@rrentControlSet\Control\Print\Environments\Windows 4.0\Drivers\EPSON EPL-5700`.

34. Dans le cas de l'« enregistrement rapide », toujours sélectionné sauf si l'utilisateur le refuse explicitement.

35. Cette version de Word fut la dernière sans « macros », donc à garantir une non-transmission de virus.

FIG. 3.1 *Le Petit Robert : dessin obtenu (le bord rouge devient ici gris).*

placé les termes insultants de façon que le texte devienne « Monsieur X est un universitaire d'une grande finesse », et avons enregistré les modifications du texte. Le résultat est éloquent :

```
[...] ^@#^@^@^ [...]  Monsieur X est un universitaire born<8E>
et dangereux^M^M^Qu^@x^BÅ{^Bá|^F∀<83>^D5<84> ^D5d'une grande
finesseD [...]
```

Bien sûr, pour tout utilisateur du même logiciel Word, seule la version finale et polie apparaîtra. Cet exemple donne la mesure de la confiance que l'on peut accorder à un tel produit comme aux formats incompréhensibles.

3.2.4 Format PostScript

Ce format — propriété d'Adobe — est devenu un « standard » pour imprimer du texte[36], et plus généralement pour l'afficher sur un écran. Il est clairement graphique, mais a l'avantage d'être *vectoriel* : une figure sera décrite par les lignes et les courbes qui la composent, et non pas comme une simple somme de points.

Contrairement au format précédent, il est intégralement lisible, en ce sens qu'il utilise notre alphabet et que l'on perçoit l'existence d'une syntaxe, même s'il peut apparaître difficile à déchiffrer : c'est un format documenté associé à un langage que l'on peut s'approprier.

36. Même par une imprimante qui ne comprend pas ce langage, par exemple avec des logiciels qui savent l'interpréter et transmettre le code *ad hoc* à l'imprimante.

Le premier exemple propose l'inscription du texte *Le Petit Robert, Tome 2* sur deux lignes, inséré dans un carré aux bords rouges[37]. Pour une lecture plus aisée, sont ajoutés quelques commentaires, qui commencent, comme dans le cas de ce langage, par des symboles de pourcentage :

```
%!PS-Adobe-3.0 % en-tête du fichier
%%BoundingBox : 199 199 301 301 % Encadrement du dessin (non indis-
pensable)
/m{moveto} def % m est défini comme une abréviation de moveto
newpath % ici commence un chemin (un carré)
200 200 m % on se positionne au point de coordonnées 200, 200
0 100 rlineto % déplacement relatif de 100 points vers le haut
100 0 rlineto % idem vers la droite
0 -100 rlineto % vers le bas
closepath % fin du carré
1 0 0 setrgbcolor % choix de la couleur rouge ; rgb = red, green, blue
3 setlinewidth % épaisseur du trait en points
stroke % réalisation du dessin
% début du texte
/Helvetica findfont 14 scalefont setfont % choix de la police
203 260 m
(Le Petit Robert) % premier texte entre parenthèses
0 setgray show % écrit en noir (0 = noir, 1 = blanc)
227 235 m
(Tome 2) % second texte, toujours entre parenthèses
0 setgray
show % afficher le texte
showpage % visualiser le résultat
%EOF. % fin du fichier
```

Même si l'on n'est pas expert, on comprend aisément comment changer *Petit* en *Grand*, remplacer le bord rouge par un bord vert, etc.

37. Voir sa représentation figure 3.1.

Le second exemple est le résultat de la mise en page d'un texte pour la flasheuse[38] d'un imprimeur : à cause de la position minutieuse des lettres, et donc de l'optimisation de l'interlettrage, un mot peut être fragmenté, et les opérations de recherche/remplacement peuvent apparaître là plus délicates, même si on sait maintenant que le texte est toujours entre parenthèses :

```
71.095 100.379 M 0.817(Une telle comple)W
-0.167 N 0.817(xit\216 f)W
-0.11 N 0.817(acilite les analyses partiales. )W
```

Ici, on retrouve sans trop de difficultés l'expression *une telle complexité facilite les analyses partiales.* Cependant, la recherche du mot *complexité* ne donnera aucun résultat : il est tronqué, de façon que le x soit rapproché du e qui le précède. En sus, l'accent aigu est codé.

Un tel « langage » transforme le texte en image, sans toutefois empêcher sa manipulation : la page imprimée acquiert un statut graphique, au même titre que les « figures » que contient cette thèse ; mais le travail sur le texte reste possible : l'index intégral de l'ouvrage « Comprendre les usages de l'internet »[39] a été réalisé à partir du fichier PostScript de l'imprimeur, dont on vient de présenter un extrait.

Ce détour par le dessin nous ramène à la calligraphie : on sait qu'une fonte donnée n'offre pas le même dessin des lettres qu'une autre, et que les typographies des automates diffèrent grandement de l'écriture manuscrite. Jacques Anis traite cette dernière de continue[40], et infinie dans sa variété, à l'opposé de la frappe discrète (non-continue) et mécanique des machines. Mais on peut aussi redessiner chaque caractère avec un logiciel dédié[41], ou plus communément avec un logiciel de dessin[42]. Ces procédures de définition

38. Imprimante haute définition.

39. [Gui01] ; cet index est disponible en ligne à l'URL http://barthes.ens.fr/internet01/indexation.html

40. [Ani98], p. 28.

41. Et assigner ce caractère à une touche spécifique, comme on le fait pour les langues dites rares.

42. C'est ainsi que les panneaux publicitaires se couvrent parfois d'une typographie rappelant les machines à écrire en mauvais état.

du dessin d'un caractère peuvent être combinées à l'infini, comme le montre Jacques André[43] : on peut aisément réaliser un programme qui modifie la forme d'une lettre comme le *a* au fil de son usage dans le texte, ou qui modifie, de façon aléatoire ou organisée (comme dans l'écriture arabe), des approches de paires, des ligatures, etc. Il s'ensuit une relation étroite entre le message (la lettre telle qu'elle est lue) et son dessin, c'est-à-dire le pictogramme que l'on confond avec un élément de l'alphabet.

Avec le PostScript — si ce langage n'existait pas, on en aurait inventé un autre, aux fonctionnalités analogues —, la relation entre le signe et son dessin est rendue explicite, et devient aisément paramétrable.

3.2.5 Formats d'images

Il existe d'autres façon de gérer les images, qu'elles soient de type vectoriel[44] ou *raster*, c'est-à-dire définies par des séries de points (comme pour une photographie). Mais dans tous les cas, ces images sont codées sous une forme textuelle[45].

Comme le programme, l'image est en fait du texte alphabétique. Ce qui étend encore la notion d'écriture, et intensifie la relation entre système de signes et code.

3.2.6 Formats par balises

Pour faire retour au texte dans son acception courante, nous rappellerons l'existence de formats tels que LaTeX et html. Ceux-ci ont *a priori*

43. [And91].

44. Exemple: le format SVG (Scalable Vector Graphics), qui permet la réalisation d'images animées sur le web, tout en acceptant des formats de fichier ASCII.

45. Un ruban vert, enregistré au format Gif, commence par les caractères suivants : `GIF87a+^@E^@`, suivi de « signes » bien peu évocateurs (et bien sûr qui dépendent de l'éditeur et du système d'exploitation utilisé) ; mais on repère quand même quelques caractères plus courants, comme des *Ã*, des *ô*, des *f* ou des chiffres. Le remplacement, dans un éditeur, des caractères *3* par des *U*, fait que si on visualise cette fois-ci l'image avec un logiciel *ad hoc*, certains points de cette image passent du vert au jaune.

un inconvénient : le scripteur doit marquer lui-même, avec des balises explicites, les fragments de texte auxquels il désire donner un statut particulier. Par exemple, avec LaTeX, une mise en italique s'opère de la façon suivante : `\emph{l'expression à mettre en italiques}` ; pour signaler un appel de note à la suite d'un mot, on y accolle une expression de ce type : `\footnote{Le texte qui sera en note}`.

Mais ces formats ont un très grand avantage : ils facilitent la structuration du texte, sa mise en forme étant considérée comme secondaire (elle est d'ailleurs paramétrable). Il s'ensuit que ces « outils d'écriture » offrent une bien plus grande liberté intellectuelle que les formats de type Word. De plus, le choix d'un pur format ASCII offre deux garanties au scripteur comme au lecteur :

— possibilité de rédiger, voire de lire le texte[46] sur tout type de machine, avec tout type de traitement de texte ou d'éditeur ; on est donc protégé contre les pannes logicielles qui détériorent trop souvent le document de travail ;

— assurance que ce texte pourra être accessible et réutilisable sur de très longues périodes, puisque le format de base ne change pas.

On reste donc au plus proche du sens du texte, alors même que son caractère graphique est aussi entièrement assumé (par sa traduction automatique en PostScript par exemple).

Pour l'html, les balises sont plus simples, mais tiennent compte de la variété des codages des divers systèmes d'exploitations : un caractère accentué comme le *é* se transcrit sous la forme *é* . Un passage à la ligne se traduit par un *
*ou un *<p>*. L'italique se code avec la balise *<i>*, suivie du texte en italique, suivi de la balise fermante *</i>*.

En fait, ces langages par balises sont plus répandus qu'on ne le pense ; SGML, XML, XHTML sont peut-être des exemple peu connus ; mais le RTF est aussi un format par balises, et même le logiciel QuarkXPress permet un

46. Un texte « source » en LaTeX reste lisible, puisque c'est ainsi qu'on le saisit. Il en est de même pour un texte en html ; ce dernier est souvent consulté par l'intermédiaire d'un programme, qui est le navigateur.

tel usage des balises (les Xtags). Les balises correspondent à la façon la plus simple de transmettre un ordre à une machine, et on verra que la solution de facilité consistant à cliquer sur une icône donnée pour éviter leur saisie peut s'avérer très coûteuse[47].

On retiendra que de tels formats, qui sont d'autant plus limpides qu'ils n'insérent rien à l'insu du scripteur, sont malgré tout assez peu utilisés en sciences humaines, certainement parce qu'ils entrent en rupture avec la métaphore de la machine à écrire.

3.3 Capacités de l'écriture informatique

Nous commençons à percevoir les limites des analyses qui veulent réduire l'informatique à une technique séparée des opérations qu'elle permet. Contrairement à une métaphore trop relayée, l'ordinateur n'est pas une automobile, qu'on peut piloter sans être obligé d'en connaître le fonctionnement, qu'on amène chez le mécanicien dès le premier problème survenu[48].

Il n'est plus besoin de prouver en quoi notre relation au texte est transformée par les possibilités de recherche, de remplacement, de suppression, de modification, de copie. Il s'ensuit un travail de ré-écriture qui assouplit les conditions de travail du siècle dernier, et qui peut-être transforme les éléments classiques de son organisation : élaboration d'un plan, recherche de la phrase parfaite, vision d'ensemble de l'œuvre à produire, que celle-ci soit un article, ou une encyclopédie. Là encore, il ne s'agit pas de juger[49], ni de négli-

47. Cf. le paragraphe 3.4.3 page 79.

48. On pourrait aussi retourner cette métaphore, en abordant l'automobile comme on le fait pour l'écriture : cet objet technique pollue, réorganise toute une économie, et tue. Mais le fait qu'en une journée, on puisse parcourir 1000 kilomètres et non plus 300 transforme aussi notre relation au monde, au moins sous l'angle de notre perception géographique. Ainsi, il pourrait s'intègrer dans notre outillage mental ([Goo97], p. 261). Ceci dit, le point aveugle de cette métaphore réside dans l'absence de réflexion sur la socialisation de la technique, ce qu'on évoquera dans la seconde partie.

49. Jérôme Lindon se lamentait de l'incidence du traitement de texte sur le style littéraire ([AL91], p. 292).

ger les connaissances nécessaires à un bon usage de l'informatique. Mais les capacités de l'informatique ne se lisent pas uniquement « par défaut » : considérer que les ordinateurs permettent de gagner du temps, de se décharger de contraintes répétitives afin de permettre un meilleur épanouissement de la pensée pure revient à avaliser la frontière entre spiritualité et machinisme, dont nous avons montré l'inconsistance.

Pour mettre en évidence ces effets mentaux de l'écriture informatique, nous ne pourrons nous satisfaire de la notion d'usage dominant. Non seulement parce que celle-ci n'a pas de sens, comme on le montrera dans la quatrième partie, mais parce qu'une potentialité cognitive n'a pas à se réaliser à chaque instant pour être évaluée à sa juste mesure : par exemple, le théorème de Gödel constitue un acquis de taille, et sa connaissance distingue les mathématiciens d'aujourd'hui de ceux du XIX^e^ siècle. Pour autant, on ne s'en sert pas tous les jours.

Réciproquement, notre démarche n'est pas pour autant élitiste. Les instruments qui suivent sont utilisés par des informaticiens chevronnés comme par des jeunes chercheurs en sciences humaines. Et dans la prochaine partie, nous montrerons que l'appropriation de l'informatique n'est pas corrélée au statut social ou intellectuel de ses utilisateurs.

Trois objets, ou concepts, sont essentiels pour cerner ce saut quantitatif dans nos capacités. Deux sont usuels : ce sont la liste et le graphique. S'y ajoute l'hypertexte, qui lui non plus, n'a rien de novateur. Nous allons maintenant montrer en quoi ces figures de l'écrit, associées aux machines et aux réseaux dont nous disposons, transforment de façon conséquente notre outillage mental, notre production de sens. On se souviendra que l'écriture permet une décontextualisation de la forme graphique, et que cette distanciation du sens permet d'en produire du nouveau[50].

50. Comme on l'a rappelé page 29.

3.3.1 Expressions régulières

L'expression régulière élargit cette notion de forme graphique : un « mot » est alors un ensemble de signes — une chaîne de caractères, disent les informaticiens — séparé d'un autre par un groupe d'autres symboles : ce peut donc être un mot au sens commun, un nombre, une phrase (les séparateurs sont alors les signes de ponctuation), une instruction graphique, une balise, etc. Mais ce peut être aussi une série de formes graphiques non connexes — ou contiguës —, grâce aux variables de repérage de motif[51].

Les opérations textuelles étendues aux expressions régulières permettent de repérer des parties variables de texte, et de les modifier rapidement. À côté, les opérations permises par les traitements de texte ordinaires font figure de parent pauvre. Si la notion d'expression régulière est intégrée aux commandes Unix, divers éditeurs, comme par exemple BBedit sur Macintosh, la proposent aussi, et de nombreux linguistes et chercheurs en littérature l'utilisent aux États-Unis[52]. Bien sûr, une des premières applications des expressions régulières est l'internet, avec la recherche de mots — indépendamment de leur codage et de leur casse, voire de leur position. Une autre, toute simple mais très utile au scripteur, consiste à mettre en couleur le texte compris entre des accolades ou parenthèses, des mots-clés prédéfinis, etc. Ce que font la plupart des éditeurs de texte.

L'expression régulière s'applique à toute forme de fichier ASCII : html, PostScript, etc. On remarque qu'elle s'appuie sur une traduction du code en notre système graphique, et donc que là encore, les deux se confondent. Si certains peuvent être choqués d'apprendre qu'une œuvre de Zola se manipule comme un tableau statistique ou comme le logo d'une entreprise, on remarquera que ce type de traitement — bien sûr combiné à d'autres — finit par

51. $1 à $9 en Perl, \1, \2, etc. sous BBedit ou Nedit...

52. Il suffit par exemple d'interroger, à l'aide d'un moteur de recherche, les pages qui contiennent l'expression « expression régulière BBedit » (14 réponses en août 2001 sur Google) ou « regular expression BBedit » (715 réponses, la cinquième étant l'URL http://www.ling.udel.edu/colin/tools/BBEdit/grep_search.html du département de linguistique de l'université de Delaware).

s'imposer : le développement de laboratoires de littérature comme l'ARTFL[53] ou Frantext[54], mais plus généralement, d'une nouvelle industrie de la langue, en est la preuve.

Le traitement par expressions régulières permet une manipulation puissante de toute forme de texte ; il contribue grandement à la distanciation entre le sens de ce texte et les formes qui le constituent.

3.3.2 Tableaux

Avec ou sans expression régulière, on peut réaliser des listes variées : listes de mots, de nombres, d'images, de fichiers... En regard des éléments de ces listes, il est naturel d'inscrire des informations de types variés. On obtient alors des tableaux ; par exemple, en associant aux mots d'un texte leur fréquence. Sur ces tableaux, les opérations de tri, de sélection, d'agrégation, de catégorisation — et bien sûr toutes les formes de calcul — sont réalisées très rapidement ; dans toutes ces procédures, l'ordinateur est incontournable : aujourd'hui, ces tableaux et matrices sont de taille virtuellement infinie[55].

Les effets intellectuels inattendus produits par les listes sont amplifiés par les possibilités de traitement de ces listes et par l'accroissement de leur taille. Comme on le verra dans notre quatrième partie, la finesse des mesures obtenues avec le traitement informatique incite à interroger les catégories conceptuelles qui servaient auparavant de repères méthodologiques.

3.3.3 Graphiques

L'« outil » le plus communément utilisé pour traiter des tableaux est le... tableur[56]. Dans une vision naïve, qui entérine la distinction entre le chiffre

53. A Chicago. http://humanities.uchicago.edu/orgs/ARTFL

54. De l'ex INaLF. Si ce centre est scindé en deux depuis le premier janvier 2001 (cf. http://www.inalf.fr), en septembre de la même année, le site frantext est toujours répertorié à l'adresse de l'ancienne institution (cf. http://zeus.inalf.cnrs.fr).

55. Bien utilisé, un ordinateur courant peut aujourd'hui traiter des listes de millions de lignes.

56. A condition que la taille de ces listes ne soit pas trop importante.

et la lettre, entre la machine et la psyché, le tableur facilite les calculs statistiques. Or, depuis bien longtemps — la fin des années 1980 —, des tableurs comme Ragtime proposent une indistinction entre chiffres et textes, entre statistiques et mise en page ; mais ils n'ont de succès qu'auprès de rares experts (concepteurs de catalogues, etc.).

Même dans le registre étroit de la gestion des nombres, le tableur est un outil d'écriture fortement sollicité : on se souvient que l'essor de l'ordinateur Apple II provient en partie du succès rencontré par son tableur [57]. La société Microsoft l'a bien compris, qui a développé le logiciel Excel pour mieux capter le public des Macintosh.

Paradoxalement, le tableur n'est pas un outil statistique, mais graphique, et à deux titres : tout d'abord, en matière de présentation. À partir d'un séparateur de texte spécifique (communément tabulation, point-virgule ou espace), on visualise les fragments de lignes obtenus dans une série de petites boîtes, appelées « cellules », en faisant en sorte que les cellules d'une colonne donnée aient toutes la même largeur ; on obtient alors un confort de lisibilité difficile à obtenir avec des moyens traditionnels.

Ensuite, le tableur sert plus à réaliser des graphiques [58] que des calculs. La sélection d'une plage de cellules permet de produire des camemberts, histogrammes, nuages de points, etc. qui font aujourd'hui partie intégrante de notre écriture, comme les cartes et autres schémas. Ces graphiques amplifient de manière décisive les possibilités d'association mentale, toujours difficiles à réaliser face à des tableaux de nombres.

Le tableur facilite donc toute une série de procédures associées à des formes complexes du raisonnement, tant sur le plan heuristique que sur le plan démonstratif, et il socialise une écriture pictographique. Cependant, il existe aujourd'hui des assistants intellectuels encore plus performants.

57. Voir page 50.

58. Sc, disponible sur Unix et Linux, est l'un des rares tableurs qui ne permettent pas de réaliser des graphiques. Il est en cela conforme à la logique d'Unix, qui privilégie les petites briques logicielles plus que les « usines à gaz » — les *obésitiels* — censées tout faire.

Ceux-ci sont les logiciels permettant de produire des graphiques sans passer par l'organisation — manuelle, voudrait-on dire — des tableaux de nombres qui les fondent : autant, avec un tableur courant, on multiplie les calculs, les tris, et on demande ensuite la mise en forme graphique des résultats produits, autant c'est l'inverse qui se produit avec un outil comme Xmgr : à partir de listes, cet outil les trie, par exemple de façon à afficher le graphe de leurs distributions statistiques (éventuellement cumulées). Il propose ensuite l'application de diverses fonctions (logarithme, exponentielle, etc.) applicables à ces listes (donc aux axes du graphe) pour rendre plus lisibles les courbes produites. La meilleure approximation d'une courbe correspondant à des données empiriques est réalisée sans intervention de l'opérateur sur les listes : la fonction optimale est appliquée, à la demande — après que l'on a choisi une famille de fonctions. Cette étrange inféodation des calculs mathématiques complexes à la représentation graphique, qui les devance, est chère aux physiciens. Ces derniers avouent sans ambage que leurs méthodes de travail en sont transformées : ils se sentent libres d'élaborer des raisonnements sophistiqués, qui s'appuient sur des mécanismes de conviction — ou de contradiction — proprement visuels[59].

Ces formes de construction du raisonnement à partir de listes et de graphiques se répandent aussi en sciences humaines. Auparavant, bien peu d'institutions se dotaient — comme l'ancêtre de l'EHESS en 1954 — d'un laboratoire de cartographie[60]. Aujourd'hui, de nombreux étudiants ou chercheurs historiens, sociologues, etc. réalisent les cartes dont ils ont besoin sur leurs ordinateurs personnels. Et ces assistances graphiques au raisonnement, déjà développées par des chercheurs comme Pierre Bourdieu ou Hervé Le Bras, s'étendent aux disciplines de l'érudition.

59. Notre quatrième partie témoigne de telles méthodes, et ses graphiques sont produits par le logiciel Xmgr.

60. Fondé par Jacques Bertin. Préfigurant les transformations induites par l'informatique, ce centre devient, à la fondation de l'EHESS en 1974, un « laboratoire de graphique ». Cependant, ce laboratoire n'aura jamais le statut de centre de recherche, au grand dam de son fondateur.

Le graphe et la carte s'intègrent donc dans l'outillage mental du chercheur contemporain sous deux formes : la première, assez répandue, est liée à l'usage massif du tableur (et dans une moindre mesure, du logiciel de cartographie) comme producteur d'images. Elle entérine le remplacement des laboratoires dédiés par des logiciels. Bien sûr, les abus de cette rhétorique graphique peuvent agacer, quand elle n'est pas maîtrisée. Mais ce fait entérine la diffusion de l'écriture pictographique.

La seconde, plus rare, place le graphe au centre du raisonnement, de l'activité d'écriture. La rupture, en termes de pratiques intellectuelles (et donc de connaissances requises), est ici encore plus manifeste.

C'est ce type de production, plus que la série de petits dessins inscrits sur les touches des machines à laver ou sur les panneaux du code de la route, qui explique à notre avis en quoi notre écriture redevient pictographique. Ce ne sont pas non plus les pictogrammes que l'on rencontre de plus en plus dans les textes — comme la puce ou le triangle qui se substituent au tiret dans une énumération, le petit carré qui signale la fin d'un article de magazine, ou les flèches (voire les nuages blancs sur fond de ciel bleu) qui envahissent les présentations « Powerpoint » des industriels devant un parterre d'investisseurs — qui témoignent du statut de plus en plus pictographique de notre écriture, mais notre capacité accrue à produire des dessins à partir de listes pour organiser, synthétiser notre raisonnement.

Henri-Jean Martin rappelle que cette évolution de notre écriture date du XVII^e siècle, quand l'essor de l'« imagerie schématisante et documentaire [...] de la cartographie et [de] certains livres scientifiques »[61] se combine avec un accroissement du nombre des logographes en mathématiques, puis dans d'autres disciplines, comme la chimie au XVIII^e siècle[62]. La fin du XX^e siècle, avec la diffusion des ordinateurs, amplifie encore plus la production de ces pictogrammes qui interviennent directement dans l'acceptation ou le refus d'une hypothèse, et dont la variété est sans limites : la facilité de produire du

61. [Mar96], p. XV.
62. [Mar96], p. 450.

dessin qui fait sens est essentielle pour comprendre les formes contemporaines de notre outillage mental.

3.3.4 Hypertexte

Les objets d'usage courant qui offrent des listes organisées de mots sont l'index et le dictionnaire, la bibliographie, et plus généralement le fichier (matières, auteurs, etc.) de la bibliothèque. Ces listes sont maintenant aisées à manipuler, et appellent à la création de listes de listes, et de renvois entre les premières comme les secondes.

L'hypertexte, longtemps pensé de façon métaphorique — avec les pictogrammes sur les papyrus symbolisant des renvois entre œuvres — ou conceptuelle — suite aux travaux de Vanevar Bush et de Theodore Nelson[63] — trouve ses premières réalisations dans les années 1980, avec les CD-ROM, et deux exemples simples peuvent témoigner de sa vulgarisation.

Considérons tout d'abord un CD-ROM tel que le dictionnaire *Le Robert*. Au lieu de rechercher la définition d'un mot dans un lourd imprimé, il suffit d'en saisir les premières lettres dans une fenêtre électronique ; diverses possibilités s'offrent alors au lecteur : définition abrégée ou détaillée, étymologie, citations exemplaires, etc. La sélection de n'importe quel mot entrant dans une définition le transforme en vedette ; plus précisément, tous les mots communs[64] sont reliés entre eux par un tissu hypertextuel ; des *jokers* aident à construire des expressions régulières simples ; on retrouve aisément tous les mots qui contiennent une graphie donnée[65]. Un tel outil, propre à l'écriture informatique, permet une utilisation rapide de la liste ordonnée de mots que constitue un dictionnaire, et autorise les détournements d'usages (mots croisés, *Scrabble*, etc.).

63. [Ani98], p. 167.

64. La sélection du mot *Fustel*, qui apparaît dans une citation de Fustel de Coulanges, renvoie à la liste des entrées proches, comme *fustet*, *fustilable*, etc. Les liens ne fonctionnent donc pas avec les noms propres.

65. Par exemple, *déf*n*tion* renverra à *déféminisation*, *défenestration*, *défibrination*, et à *définition*.

L'autre paradigme de l'hypertexte est évidemment le web, avec un langage simple comme le html (et un logiciel bien plus sophistiqué, qui gère le protocole http d'un serveur) : contrairement au CD-ROM, dont la production exigeait une haute technicité, n'importe quel utilisateur peut construire un système hypertextuel au sein d'un même texte, ou entre divers textes de sa production[66]. La différence fondamentale avec l'exemple précédent est que les renvois peuvent aussi dépasser le cadre d'un seul support : le lien renvoie alors à un texte situé *ailleurs*. Il s'ensuit une réticulation textuelle qui n'est plus maîtrisée par ses auteurs : à partir d'un seul fichier, on peut rapidement dévider un très grand sous-ensemble des pages web du monde entier[67].

Ces successions de définitions, de renvois, ne pouvaient que donner le désir de construire des listes ou des dictionnaires. Ce qui peut se produire de façon alphabétique ou thématique — et bien sûr non exhaustive — avec les listes de pointeurs, si simples, mais si fondamentales aux chercheurs, jusqu'aux listes... de listes de listes[68].

La réticulation de ces pages par l'hypertexte et le caractère électronique de leur contenu textuel ont par ailleurs permis de développer un nouvel outil de travail, aujourd'hui incontournable : le moteur de recherche. Deux cheminements coûteux ont ainsi pu être évités : celui qui mène à la bibliothèque, jamais assez vaste pour contenir toutes les pages web actuelles, celui de la lecture en diagonale, pour retrouver une expression dans le texte. Bien sûr, les moteurs ont aussi leurs défauts : aspiration non exhaustive du web (ce qui serait difficile), modes de repérage purement lexicométriques, et mauvais

66. Voici la syntaxe paradigmatique du renvoi : `<a href="adresse de l'objet vers lequel pointe l'expression qui suit">expression dont la sélection organise le renvoi</a>`. L'objet visé peut être indifféremment du texte, un programme, une image, etc. L'objet source est souvent un fragment de texte, mais peut aussi être une image.

67. On a démontré récemment qu'au moins le quart des pages web était relié entre elles de façon étroite et réciproque : cf. l'article *Graph structure in the web*, http://www.almaden.ibm.com/cs/k53/www9

68. Par exemple, listes de listes de dictionnaires en ligne : http://www.admin.ch/ch/f/bk/sp/dicos/dicos.html

repérage de l'imprimé (malgré les tendances contemporaines à la numérisation, de très nombreux textes, anciens ou récents, ne sont pas accessibles sur le web). Mais là encore, on voit comment une pure production de listes, qui s'associe aux moyens de les traiter, amplifie considérablement nos capacités mentales. Et les outils créés pour travailler ces listes transforment le statut du texte, qui devient lui-même... liste[69].

3.4 Cultures informatiques

Jusqu'ici, nous avons décrit l'évolution du système de l'écrit et le type de capacités qu'il offrait en termes d'outillage mental[70]. Mais une technique requiert aussi des méthodes et des savoir-faire. Nous allons en détailler ici quelques-uns, qui renvoient autant à des pratiques intellectuelles qu'artisanales, et qui poseront à nouveau la question des représentations dominantes relatives à la technique.

3.4.1 Formats et logiciels

La multiplication des formats de fichier (formats de texte, d'image, de son, etc.), leur évolution permanente et leur adaptation plus ou moins grande aux divers systèmes d'exploitation (formats de compression, par exemple) impose au « grand-lecteur informatique » une virtuosité de tous les instants. Certes, depuis quelques années, l'internet est devenu le mode principal de transfert de tels formats, et les navigateurs sont en général dotés de suffisamment de *plug-in* pour décoder la majorité des fichiers ou sinon conseiller l'utilisateur. Cependant, un chercheur en sciences humaines qui refuse de se laisser désarçonner par le premier fichier attaché venu doit connaître l'existence et les

69. Par exemple, la comparaison lexicométrique de plusieurs ouvrages s'obtient en réduisant chacun à une (longue) ligne.

70. On restait donc dans le registre du premier type d'interactions décrit au chapitre précédent.

fonctionnalités d'une quinzaine de logiciels[71]. À plus forte raison s'il s'engage dans une politique d'édition sur l'internet.

La maîtrise de ces assistants d'écriture est d'autant moins aisée que souvent, le scripteur est confronté à des systèmes d'exploitation qui se partagent entre deux types profondément opposés : le premier, intuitif, laisse une grande liberté à l'utilisateur, mais nécessite une maintenance régulière (réorganisation du système, contournement des incompatibilités, etc.) et suppose que ses correspondants aient le même type de machine que lui[72]. Pour que sa machine soit fiable, il doit développer au quotidien une lourde activité technique, au sens classique du terme : l'instabilité de ces systèmes d'exploitation et leur prétendue simplicité d'utilisation les transforment en réelle passoires en matière de sécurité. C'est encore à l'utilisateur final qu'incombe cette surveillance : il doit aussi se prémunir contre diverses formes d'agressions, des virus aux destructions de serveurs web.

Le second type de système d'exploitation est plus contraignant, en matière de droits — on ne peut détruire aisément le fichier qu'un collègue a déposé sur la même machine —, mais, parce que les usages collectifs d'une même machine y sont pensés d'avance, infiniment plus robuste. Il est basé sur un grand nombre de logiciels spécifiques, particulièrement bien « écrits », et donc très efficaces. Mais encore faut-il en connaître l'existence... Sinon, il offre deux avantages appréciables : cette connaissance des outils et protocoles permet à son utilisateur de maîtriser sans trop de peine les ordinateurs du premier type — alors que l'inverse n'est pas possible — et surtout, explicite la culture de l'écrit informatique.

71. Exemples : un traitement de texte banal, un éditeur, un tableur, plusieurs logiciels de (dé)compression, de conversion de textes, d'images, deux ou trois logiciels graphiques, pour ne parler que du traitement hors ligne. Un navigateur, un gestionnaire de courrier électronique, un gestionnaire de transfert de fichier (protocole ftp), etc. Et ne sont pas évoqués ici les logiciels permettant de travailler le son ou la vidéo, ceux qui permettent l'optimisation des systèmes d'exploitation grand public (antivirus, défragmenteurs...), les émulateurs, ni la myriade d'outils de type Unix.

72. On aura reconnu les systèmes d'exploitation de la firme Microsoft, et dans une moindre mesure, de la firme Apple.

Dans les deux cas, les potentialités de l'écriture électronique invitent à l'acquisition d'une culture informatique qui dépasse largement le mode d'emploi d'une machine à transformer la matière.

3.4.2 Culture digitale

À cette forme de connaissance fondée sur la compréhension, et que l'on pourrait encore traiter de « mentale », s'en ajoute une autre : la manipulation de l'ordinateur génère une nouvelle gestuelle, une culture proprement digitale ; la mémorisation parfaite des touches du clavier, caractéristique des dactylographes comme des écrivains, apparaît secondaire face à celle de leur combinaison ; pour les machines Unix, la tabulation permet la complétion des « mots » (qui sont ici, des commandes, des noms de fichiers ou de dossiers), les flèches du clavier rappellent les précédentes commandes, et souvent, un ordre se transmet en tapant plusieurs touches ensemble ou successivement[73].

Ce retour au passé d'avant la souris s'est particulièrement développé avec l'internet, dont la fonction première est de permettre la communication entre machines hétérogènes : du coup, l'environnement graphique est souvent inutilisable, puisqu'il est associé à un système d'exploitation donné[74]. Mais si les informaticiens ont eux-aussi vite adopté la souris (par exemple, elle peut intégrer les fonctions du copier/coller, sans recours à une combinaison de touches ni à un menu), leurs pratiques digitales se sont répandues : tout le monde découvre assez vite qu'un « raccourci clavier » permet d'enregistrer un document, ou qu'une grande sélection de texte s'opère plus facilement si l'on utilise de façon combinée la souris et la touche majuscule.

73. Pour donner un exemple simple, la suppression d'un texte, d'un endroit donné jusqu'à sa fin se réalise ainsi avec l'éditeur « joe » : « ^KB, ^KV, ^KK, ^KY » est la succession de gestes à effectuer pour définir le début d'un bloc, aller à la fin du texte, définir la fin dudit bloc à cet endroit, et demander sa suppression (le signe « ^ » symbolise la touche « Control » du clavier).

74. L'évolution des systèmes d'exploitation comme des débits changera bien des choses. Mais ici, toute prospective est vaine, du fait de la violente concurrence industrielle qui s'exercera pour imposer des « standards ».

Cette culture digitale des programmeurs — qui, au demeurant, ne permet plus de lire le texte qu'ils saisissent en ayant les yeux fixés sur leur clavier — se retrouve aussi chez les professionnels de l'édition, graphistes ou spécialistes de la mise en page : pour modifier le contour d'une image, ou pour réduire les approches de paires ou de groupe [75], il faut alors appuyer jusqu'à quatre touches en même temps [76]. Chez les cartographes, la culture digitale s'affirme devant la culture iconique : par exemple, ils connaissent le raccourci clavier pour fusionner des points appartenant à des polygones distincts, mais ne savent plus retrouver le sous-menu correspondant à ce raccourci.

Ce savoir-faire gestuel renvoie à une conception plus traditionnelle de la machine : avec le clavier, on travaille physiquement le texte ou l'image, pris comme objets graphiques, comme matière, pourrait-on dire, mais on le fait plus rapidement qu'avec la souris ou les menus, qui sont les instruments valorisés par les systèmes d'exploitation grand public.

Ces combinaisons de touches ont une dernière fonction : elles permettent, parfois au terme de réelles « acrobaties digitales », de saisir des caractères plus sophistiqués que ceux disponibles sur une simple machine à écrire, comme les guillemets français en forme de chevrons, ou allemands, les ligatures rares, les diacritiques sur des capitales, les symboles de redirection [77], etc.

Ainsi, pour les professionnels de l'écriture, le savoir-faire de la production et de la transformation du texte s'accompagne-t-il d'une culture gestuelle nécessaire à son appropriation : le texte reste inaccessible à qui ne sait pas le *pétrir*. Ce type de connaissances requises est rarement évoqué, car on ne sait dans quelle catégorie culturelle le placer : il n'est pas universel, il varie au fil du temps et suivant les machines ; il s'inscrit dans la série des « astuces », forme de savoir-faire secondaire *a priori* non indispensable au profane, ce qui

75. Que l'on nomme aussi « crénage ».

76. Des logiciels professionnels comme Illustrator ou QuarkXpress complètent leur documentation d'un réel dictionnaire de ces combinaisons de touches, qui peuvent être de l'ordre de 300.

77. Par exemple, il n'est pas toujours aisé d'inscrire le signe « | » sur une machine non Unix. Cf. aussi la note 25 de la page 55.

conforte ce dernier dans l'idée que l'ordinateur n'est qu'une « machine », sans relation avec la pensée pure[78].

3.4.3 Environnement graphique

Cette incapacité à positionner la culture informatique (logiciels, protocoles, méthodes, gestuelle) est grandement favorisée par la rhétorique de la convivialité, qui joue sur la confusion d'objets et de protocoles aux statuts bien différents, et qui, sous couvert de simplicité, présente l'ordinateur comme une machine utilisable par le premier venu, comme le serait l'aspirateur : les ordinateurs « conviviaux » abusent de la métaphore du « bureau virtuel », composé de diverses images (dossiers, fichiers, logiciels, disques, imprimantes, etc.) sur lesquelles on peut « cliquer » pour les « ouvrir » à loisir. Là encore, on retombe dans le registre de l'action sur la matière. La fonction de ces icônes est prévue pour être « évidente », mais n'a rien d'explicite[79]. À ces icônes s'ajoutent des formes plus textuelles, comme les instructions qui l'on transmet avec les menus d'un logiciel. Bien sûr, ces codes graphiques changent suivant les systèmes d'exploitation, les logiciels, etc. Tous créent les conditions d'une culture pictographique dont on ne peut plus faire l'économie si l'on veut se servir d'un ordinateur.

Ces pictogrammes imposés[80] contribuent à donner l'image d'une informatique aisée d'emploi, évidente, indolore, et donc à nier la complexité des processus intellectuels et des connaissances sous-jacentes. Ils ont un avantage

78. De façon générale, la matérialité de l'activité intellectuelle, avec ou sans ordinateur, a toujours été sous-évaluée : la bonne connaissance d'une bibliothèque passe aussi par une série de gestes pour effectuer une recherche dans un fichier matières, parcourir des rayonnages, repérer l'ouvrage désiré. Mais rares sont les chercheurs qui insistent sur les interactions entre nos faits et gestes, entre nos efforts musculaires et notre production intellectuelle, même si, dans le cas de l'informatique, on remarque un intérêt croissant pour l'ergonomie appliquée à l'interaction homme-machine.

79. Il a fallu une vingtaine d'années pour socialiser les icônes de photocopieurs.

80. A ne pas confondre avec les graphiques que l'on crée soi-même, comme on l'a remarqué plus haut.

certain : leur caractère intuitif est satisfaisant pour une première familiarisation et nous ne nions pas qu'ils ont beaucoup aidé à la diffusion des ordinateurs. Mais ils induisent des gestes répétitifs qui finissent par encombrer l'esprit du scripteur et le fatiguer physiquement : pour produire une carte animée avec des logiciels « clickodromes », il faut répéter environ 1000 fois une dizaine de gestes. Taylor n'est pas loin.

Or, si nous nous souvenons que les outils permettant un tel travail graphique ont été écrits, et que tout résultat informatique n'est jamais que du code, c'est-à-dire une succession de signes crées par l'homme, nous pouvons nous douter qu'un travail sur ce code permet de faire l'économie de ces procédures manuelles.

C'est la fonction des « langages de programmation » qui, non seulement décuplent effectivement nos possibilités de tris, de calcul, de réorganisation textuelle, graphique, etc., mais donnent aussi à comprendre en quoi l'informatique est associée à l'écriture.

Certes, on est alors condamné à travailler un peu : lire des documentations, imprimées ou sur le web, expérimenter des logiciels pas toujours explicites, apprendre des méthodes, etc. Mais à chaque métier ses instruments et sa culture, son temps d'apprentissage. Or, la façon dont les professionnels de l'écrit abandonnent leur production à des outils qui nient la spécificité de leur activité et prétendent tout faire est stupéfiante : tout comme on ne traduit pas Kant avec un guide Berlitz, on n'est pas obligé de confier sa production intellectuelle à un logiciel ou un système d'exploitation destiné « aux nuls »[81].

Paradoxalement, les éditeurs d'informatique grand public incitent aussi à l'insertion de commandes écrites au cœur de leurs outils prétendument universels. On le voit avec les « macros », les lignes de code des tableurs et des gestionnaires de base de données, et même avec les langages grand public comme Visual Basic ou Applescript. Ces éditeurs ont donc conscience des limites de leur approche fondée sur l'icône et le geste.

81. Titre d'une collection d'ouvrages de vulgarisation informatique qui encombrent les rayonnages des libraires.

La métaphore du bureau virtuel et de la convivialité finit par effectuer des ravages : elle réduit l'ordinateur à l'état d'instrument, aux dépens de ses fonctionnalités d'écriture. Bien sûr, elle conduit à nier l'effort préalable à l'acquisition de la culture qui permettrait de tirer intelligemment parti de telles potentialités. Enfin, et c'est peut-être le plus grave, elle incite les professionnels de l'écrit à confier leur travail à des produits de bien piètre qualité, sans qu'ils s'interrogent sur la façon dont leur activité est transformée par la technique, sans qu'ils recherchent des logiciels performants et clairement adaptés à leurs besoins.

3.4.4 Illettrisme

Nous venons de repérer trois facteurs de désappropriation du texte par les personnes qui obéissent à l'injonction d'utiliser l'ordinateur comme une machine intuitive. Parce que les connaissances indispensables à une bonne maîtrise de l'informatique sont sous-estimées[82], son usage est réduit à une série d'activités physiques, mécaniques au sens traditionnel du terme[83]. Bien sûr, le discours commun assure que de tels malaises sont un effet de génération. Mais l'argument prétendant que les « jeunes » auraient moins de difficultés que leurs parents avec les ordinateurs ne tient pas[84]. Ceux-ci ne sont pas des machines à laver, mais à penser.

Toute technique a besoin d'un cadre social pour se diffuser. En l'occurrence, apprendre les fonctionnalités des logiciels, optimiser ses gestes, se fa-

82. Et aussi en partie parce qu'il est difficile de les expliciter et de les ordonner.

83. Une majorité de personnes sont désemparées devant un fichier attaché qui « résiste » à leur « double-clic ». Très souvent, le document est pourtant rédigé dans un format standardisé, comme le html, ou le RTF de Microsoft, mais est déclaré illisible par le système d'exploitation (Windows en l'occurrence). S'ensuit toute une série de communications téléphoniques pour expliquer aux destinataires les démarches physico-logiques à effectuer : « Cliquez sur 'ok'. Notez dans quel dossier du disque vous sauvegardez le fichier attaché. Lancez l'application Word. Dans le premier menu de gauche sélectionnez le sous-menu 'ouvrir'. Parcourez l'arborescence de votre disque de façon à retrouver l'emplacement du fichier sauvegardé. 'Double-cliquez' sur l'icône du fichier reçu... ».

84. [Gan99], pp. 49–50.

miliariser avec les codages les plus répandus, s'y retrouver dans la variété des contenus du web, tout cela renvoie à une culture spécifique qui, comme toujours, ne peut s'acquérir que dans le cadre d'un laboratoire[85]. Mais ces savoirs s'appuient aussi sur la production d'autrui : l'acquisition de telles connaissances est ralentie si l'on ne dispose pas des textes électroniques, des fonds de carte ou des données chiffrées déjà numérisés par d'autres. Comme toujours, le chercheur est « pris » dans l'univers de la production de ses collègues.

On arrive donc à une situation inconfortable pour certains érudits, qui, magnifiant leur culture classique, finissent par devenir proprement illettrés : ils ne savent pas effectuer une recherche sur le web, ni rédiger un document hypertextuel ; bien sûr, ils ne peuvent lire un texte si l'icône de ce dernier résiste à leur double-click.

Il est peut-être alors temps de redéfinir la culture: cette multiplicité de connaissances prétendument élémentaires, et notamment la connaissance des multiples codages, risque fort d'être au chercheur d'aujourd'hui ce qu'était à l'intellectuel la connaissance de diverses langues et de divers systèmes d'écriture durant le Moyen-Âge. Cependant, il n'est pas dit qu'en France, ces premières connaissances témoignent d'« intelligence » ou de « culture », au sens où on l'entend pour les secondes.

3.5 L'écriture, enjeu économique

Ces transformations de la technique d'écriture ont évidemment des effets qui ne peuvent laisser indifférent : sa maîtrise, au stade le plus élaboré, donne du pouvoir, d'abord intellectuel, ensuite économique ou politique. Il est logique que la modification de notre outillage mental suscite des appétits féroces et fasse l'objet d'une violente compétition marchande.

85. Cf. le paragraphe 2.2.1 page 32.

3.5.1 Etouffer l'outillage mental

Le phénomène le plus visible en est l'apparition d'une multiplicité d'intermédiaires bien décidés à tirer une rente des possibilités d'écriture actuelles. Celle-ci s'obtient en général en imposant un format standard[86], qu'il s'agit de modifier régulièrement, sous couvert d'amélioration « technique » ; ainsi, le producteur doit acheter les nouvelles versions du logiciel d'écriture, le lecteur doit télécharger à ses propres frais celles de l'outil de lecture associé, et les éditeurs concurrents subissent toujours un retard qui les empêche de devenir compétitifs[87]. De telles stratégies ne se militent pas à la production des outils d'écriture « nobles » : grâce à un jeu subtil d'alliances, les composants des ordinateurs et les supports de stockage sont aussi « améliorés », de façon que l'obsolescence soit aussi très rapide pour ces objets matériels[88].

Les conséquences d'une telle stratégie sont cruelles[89] : ce changement régulier des normes du format, du support, et aussi du code[90], doublé d'un mépris pour la compatibilité ascendante[91] fait qu'au bout de cinq années, les

86. WordXX pour le texte hors-ligne, PDF pour le web, Flash pour l'animation, etc.

87. Pour ce dernier point, les effets d'annonce sont essentiels : peu importe que le nouveau produit soit à la hauteur de ce que son éditeur prétende ; souvent, la nouvelle version d'un logiciel est inutilisable tellement elle est « boguée », et il faut attendre six mois avant l'apparition d'une mise à jour opérationnelle. Mais un tel discours publicitaire valorise toujours le tenant du monopole : pourquoi acheter un produit efficace chez X si Y assure que dans un futur proche, le sien fera mille fois mieux ?

88. La méthode la plus simple consiste à développer des logiciels ou des systèmes d'exploitation très gourmands en mémoires — vive ou de stockage —, ce qui fait qu'un ordinateur de plus de trois ans ne peut faire fonctionner les nouveaux logiciels. Cette pratique s'étend aux écrans et imprimantes, principalement par le biais de nouveaux protocoles de communication avec ces périphériques (exemple récent : l'USB).

89. Pour une description détaillée des faits explicités dans ce paragraphe, l'ouvrage [DCN98] constitue une excellente référence.

90. Au-delà des changements de format réguliers, de WordXX en Word XY, on a pu remarquer, il y a trois ans, l'abandon du code ASCII appelé IBM-PC, qui était associé au DOS, et qui a fait que le format d'échange RTF, longtemps perçu comme un standard, n'est plus totalement compatible avec les systèmes d'exploitation autres que Windows.

91. Par exemple, un document écrit avec Word 4 pour Macintosh n'est pas lisible par la version 6 du même logiciel.

anciens textes électroniques sont illisibles ou inaccessibles, donc perdus, s'il n'ont pas été imprimés. Parce que le support magnétique n'est plus reconnu par les ordinateurs actuels, parce qu'il est abîmé, parce que le logiciel utilisé pour saisir le texte n'est plus compatible avec les systèmes d'exploitation du moment, etc. Les services spécialisés d'informatique dans le monde universitaire témoignent de centaines d'expériences de ce type. Il s'ensuit un effet de contraction de la mémoire, particulièrement préjudiciable, alors que l'on aurait pu penser que celle-ci se serait épanouie avec la diffusion des ordinateurs ; tout simplement à cause de l'appropriation par des firmes informatiques privées du système graphique et de ses codes, que l'on pourrait, à juste titre, considérer comme du bien public[92].

L'absence de fiabilité est le second effet de cette course à la nouveauté : d'une part, un logiciel performant sera oublié car vite incompatible avec les nouveaux disques, systèmes d'exploitation, etc. D'autre part, on l'a déjà dit, les logiciels sans cesse « révolutionnaires » sont d'une mauvaise qualité affligeante. Comme le faisait remarquer Eric Brousseau[93], acheterait-on une voiture si, au moment de freiner, s'affichait une annonce du type « cette opération est incompatible avec votre automobile. Veuillez redémarrer » ? Or, de tels messages sont monnaie courante avec les logiciels grand public. On ne dispose hélas pas de statistiques sur le temps que perdent leurs usagers à redémarrer leurs machines, reformatter leur disque, réinstaller ces logiciels, voire réécrire intégralement les documents perdus à l'occasion de ces opérations. Mais il est sûr que ces machines, censées nous faire gagner du temps, en font perdre à la majorité de leurs propriétaires.

En revanche, on commence à mesurer les effets de la confiance aveugle en de tels outils. Les virus agressifs se multiplient, obligeant les scripteurs à une attention de tous les instants. Les serveurs web utilisant le logiciel *Inter-*

92. On a vu que, de tout temps, les objets matériels associés à l'écrit (papyrus, parchemin, plume, etc.) étaient une marchandise. Mais l'extension de ce statut à notre système graphique est préoccupante.

93. Conférence « Économie du numérique », DEA de sciences sociales ENS-EHESS, 28 mars 2002.

net Information Server[94] plutôt qu'Apache ont été dans leur grande majorité rendus inaccessibles, voire détruits, au dernier semestre 2001. Et aujourd'hui, le mythe même de l'accès universel à l'information est mal en point : du fait de ces virus, les parcs informatiques de plusieurs universités françaises sont désorganisés, et la plupart du temps inutilisables. Parce que de nombreux responsables informatiques sont mal formés, mal payés, ils choisissent eux-aussi des systèmes d'exploitation grand public. Ne sachant pas les protéger des trop faciles intrusions, ils privilégient des solutions draconiennes, en interdisant l'usage de protocoles qui permettraient d'utiliser efficacement l'internet. C'est ainsi que certains campus sont coupés du monde. On se souvient de leur existence quand... leurs membres nous envoient des flots de virus.

3.5.2 Démarches incantatoires

Pourtant, la sécurité devient le maître mot de l'internet, surtout depuis que les réseaux peuvent transmettre des biens culturels, lesquels ont parfois des propriétaires. Dans ce cas là, lorsque la technique n'est pas infalsifiable, le droit vient à sa rescousse, et d'une façon particulièrement préjudiciables à la science : la récente loi américaine du *Digital Millennium Copyright Act* interdit la diffusion de techniques qui cassent la protection du *copyright*. Ainsi, la publication d'un article scientifique démontrant l'inefficacité du cryptage vidéo proposé par Intel pour protéger les « droits » des distributeurs de films rend-elle son auteur passible de poursuites[95]. Rappeler que le « Roi est nu » mène donc à la prison. En revanche, cette rhétorique propriétaire réduit les droits du scripteur : d'une part en favorisant l'encryptage sous toutes ses formes, d'autre part, en laissant les *lobbies* faire : un format comme le SVG est menacé par les vendeurs de systèmes d'information géographiques, qui s'effraient de voir leurs fonds de cartes apparaître en « clair ».

94. IIS. Livré en standard avec les systèmes d'exploitation Microsoft.

95. Le système est appelé « High-bandwidth Digital Content Protection » (HDCP), cf. http://www.macfergus.com/niels/dmca/cia.html

On en arrive donc à ce tour de force qui fait que, sous couvert de protéger des droits toujours mal défendus (propriété intellectuelle, droit d'auteur), on protège ceux des éditeurs[96] (de textes, de cinéma, de — mauvais — logiciels) et de quelques ayants-droits[97].

Nous réalisons à quel point cette appropriation de *notre* écriture par des producteurs de formats, et de codes, pour lesquels la sécurité n'est qu'affaire incantatoire, étouffe nos capacités intellectuelles. Notre mémoire, nos pratiques d'écriture et de lecture sont fortement malmenées par les méthodes et les outils même qui devraient leur servir de béquille. L'appel à la notion de propriété[98] permet de masquer cette réalité douloureuse, en légitimant l'usage de formats inaccessibles au lecteur-scripteur, et en lui interdisant de vérifier par lui-même leur prétendue fiabilité. Aussi ne peut-on que passer du registre de la confiance étayée par la preuve à celui de la croyance, de la foi, ce qui est déjà un inquiétant paradoxe pour ceux qui pensent que la rationalité ne peut se développer sans une écriture robuste : capacité de se relire, de se corriger, de vérifier ses calculs, etc. Enfin, cette commercialisation d'un outillage mental de mauvaise qualité s'étend à notre propre production, individuelle comme collective. À force de faire monnayer tout droit d'écrire, mais aussi de consulter — notre pensée ne se développe pas sans les textes d'au-

96. On rappellera que le « patrimoine » des éditeurs ne leur appartient pas : ils en sont dépositaires, locataires plus que propriétaires ; s'ils paniquent devant les facilités de transmission permises par la numérisation, il faut garder à l'esprit le fait qu'ils n'ont pas peur du lecteur ; ils n'ont peur que de l'arrivée d'un nouvel éditeur, encore plus mercantile qu'ils ne le sont eux-mêmes : toute l'histoire de l'édition est faite de contrefactions, organisées par les éditeurs, les imprimeurs ou les journaux, comme elle est faite d'auteurs morts dans la pauvreté.

97. Suite à divers conflits avec ces derniers, Olivier Bogros, conservateur de la médiathèque de Lisieux, ne s'autorise à publier sur le web une d'œuvre que si son auteur est décédé depuis au moins 86 ans. Cf. http://www.bmlisieux.com/maj.htm

98. Le logiciel serait là pour nous *protéger*, il se substituerait donc à la police ; notre production intellectuelle serait donc réduite à l'équivalent d'une maisonnette de campagne, menacée par des intrus donc on ne connaît pas le visage... Une étude approfondie du caractère proprement totalitaire, voire misanthropique des rhétoriques publicitaire et juridique relatives aux documents électroniques mériterait d'être menée.

trui —, à force de multiplier des barrières de douane tous azimuts, on risque fort de tuer dans l'œuf une rationalité collective indispensable, mais aussi très fragile en ces moments délicats de socialisation d'un nouvel outillage mental.

3.5.3 Écritures libres

Fort heureusement, il existe des possibilités de profiter au mieux de la technique contemporaine de l'intellect. Mais celles-ci sont peu connues, et sollicitent des ressources complexes : mentales, bien sûr, mais aussi matérielles, et sociales.

Leurs caractéristiques sont en tout point opposées à celles qui sont proposées au grand public.

Tout d'abord, le format de fichier (du texte, mais aussi du résultat) reste toujours lisible : le plus souvent, en ASCII. Ainsi, la relecture et l'archivage sont aisés. Ce choix d'un format simple permet de traiter des fichiers de taille conséquente[99].

Ensuite le système d'exploitation est particulièrement robuste, et peut-être aisément modifié si l'on y découvre une fragilité : la panne de courant est quasiment le seul facteur d'arrêt d'une machine Linux ou Unix. Les logiciels sont dédiés à une tâche précise ; du coup, leur efficacité est 10 à 1000 fois plus grande que celle de leurs homologues grand public. Le fait que, la plupart du temps, leur « code source » soit d'accès libre et gratuit facilite la maintenance (collective) et la réactivité des personnes qui ont décidé de s'en charger.

C'est ainsi que des produits dits « libres » comme Apache (serveur web), Perl (programmation) et Linux s'imposent chez les professionnels de l'informatique du fait de leur très grande qualité, comme l'internet l'avait fait en son temps face aux protocoles réseaux développés par des entreprises comme IBM ou Compuserve.

Certes, il faut apprendre à se servir de tels « instruments ». Mais on n'apprend pas à lire ou à compter sans efforts, ni seul : le réseau social construit

99. La taille d'un document est communément multipliée par quatre quand il passe d'un format ASCII à un format de logiciel propriétaire.

autour de tels modèles de travail informatique est essentiel pour en tirer le meilleur parti. Il permet aussi, en termes de production, de favoriser un mode d'échange qui ne soit pas fondé sur l'achat naïf de solutions logicielles ou de données. On en arrive à un étonnant paradoxe : les objets intellectuels les meilleurs sont les moins chers [100].

100. On connaît déjà cette situation dans l'économie du numérique : le prix d'un produit dépend, non pas de son coût marginal — souvent quasi-nul —, mais de ce que l'utilisateur accepte de débourser ([Ser01]).

Chapitre 4

Conclusion

L'internet n'est pas une nouvelle technologie mais un dispositif supplémentaire qui s'inscrit dans la technique multimillénaire qu'est l'écriture. Celle-ci, préalable à l'édification et au développement de la rationalité, s'appuie sur trois ingrédients *a priori* distincts, mais en perpétuelle interaction : des objets matériels, un système de signes, et la pensée.

Ces trois catégories, et l'ensemble des procédures associées, constituent ce que nous appelons l'« outillage mental ». Celui-ci, combinaison d'instruments, de méthodes, de savoirs et de raisonnements — du passé ou du présent — est donc l'ensemble des ressources qui sont à la disposition d'un individu donné à une époque donnée. Il se renouvelle constamment, du fait de la transmission des connaissances, de la production des commentaires et des synthèses ; ces pratiques ont des répercussions sur l'ensemble de l'écriture : les annotations (note, index, etc.) mettent les textes en réseau, les synthèses invitent à imaginer de nouveaux objets graphiques (carte, symbole, etc.), et la décontextualisation résultant de la mise en listes de formes graphiques (mots, expressions, etc.) recompose les catégories et les concepts.

Ces dynamiques sont attestées depuis les débuts de l'écriture, tout comme le caractère réflexif de cette technique — comment l'explicitation des procédures qui la constituent et qu'elle engendre réorganise la pensée, le système de signes, la matière et la structure du support.

I.4 Conclusion

L'outillage mental, essentiel au développement de nos facultés de compréhension, n'est pas pour autant éthéré : il s'inscrit dans les pratiques de l'échange entre les hommes, et à ce titre, se déploie suivant deux temporalités, celle de l'individu et celle de la société.

Nous avons appelé laboratoire le premier « espace-temps » : il n'y a pas de transmission de la connaissance, de développement d'outils et de méthodes sans un voisinage social, celui des maîtres, collègues, élèves, parents, amis. Le laboratoire, peuplé de personnes et d'instruments, est l'ancrage social de l'outillage mental. L'analyse des transformations de ce dernier ne va pas sans l'étude attentive de l'organisation du laboratoire.

L'autre temporalité génère des sédimentations, des archétypes, des statuts, des normes, des écoles, des cultures. Elle est bien plus complexe à décrire, et traduit la façon dont le pouvoir de l'intellect se distribue au sein des autres formes de pouvoir (économique, religieux, politique) ou s'accommode de leur existence. Mais ce ne sont pas que des données sociologiques qui font que l'écriture est prise, enchassée dans ces autres contextes. On a pu voir à quel point des systèmes graphiques, mais aussi la diffusion d'idées, peuvent dépendre de logiques industrielles ou commerciales. Et si les lettrés et les savants arrivent si facilement à s'intégrer dans les structures du pouvoir, c'est bien que la transmission et l'amélioration de l'outillage mental apparaît indispensable à toute société structurée autour de l'écriture ; quitte à accepter qu'apparaissent des effets contraire aux buts annoncés, quand se constituent des castes développant des pratiques intellectuelles routinières.

Dans ces deux temporalités, l'outillage mental est pris en des enjeux dont la meilleure preuve de l'importance est signalée par la complexité et l'irrationalité des discours qui en témoignent. Le débat contemporain est, sur ce point, éloquent.

D'un côté, on aurait la Culture, légitime et nationale, héritage étonnant de génies solitaires, désocialisés, dont la production est exempte de tout contact avec la matière. Toute transformation mineure du système de signes et de sa graphie mettrait cette culture en péril, traduirait la volonté de la transformer en marchandise. Cette situation, typique des défenseurs de notre patrimoine, vaut aussi pour la plupart des enseignants-chercheurs en sciences humaines, qui ont souvent pour fonction de perpétuer et d'analyser cette culture. Si les sciences sociales se distinguent partiellement de telles attitudes — en grande partie parce qu'elles sont sensibles aux structures collectives et aux effets d'institutionnalisation —, elles privilégient néanmoins une attitude mentaliste qui oblitère le caractère technique de l'écriture. En effet, dans ces deux situations, la technique est ravalée au rang d'opération mécanique, donc prévisible et répétitive, sur la matière, et ne peut en aucun cas prétendre à une quelconque interaction avec les nobles activités de l'esprit.

De l'autre, on aurait des techniques, modernes, et donc par nécessité toujours changeantes, qui nous rendraient plus intelligents, plus fraternels, plus cultivés. Il est ici déconseillé de solliciter l'outillage mental pour expliciter les raccourcis d'un tel raisonnement : il est en effet inutile d'apprendre, car tout savoir devient intuitif ; la machine sait tout, l'utilisateur n'a qu'à « cliquer ». En cas d'échec, voire de panne, celui-ci est censé se reporter au mode d'emploi qui lui est dédié : « l'informatique pour les nuls ». Le plus surprenant dans ces discours typiques de la publicité informatique — hélas trop repris par les journalistes et les hommes politiques — est que, malgré l'évocation confuse d'un outillage mental réactualisé (références à la modernité, aux capacités que nous offre la machine), ils nous renvoient à la vision techniciste du XIX[e] siècle : par l'intermédiaire de la métaphore du bureau, table de travail constituée de dossiers, de chemises, de feuilles, voire de *post-it*, mais aussi par la gestuelle imposée par les systèmes d'exploitation les plus courants, l'utilisateur de l'ordinateur est contraint à une série d'activités manuelles, répétitives, aliénantes. La taylorisation des pratiques — et peut-être de la pensée ? —, s'universalise, avec la pénétration des ordinateurs dans nos espaces domestiques.

Or, l'ordinateur n'est ni une machine à écrire, ni un objet pensant qui nous réduit à l'état de dactylographe. C'est une machine, certes, mais qui travaille sur notre système de signes, par l'intermédiaire de ce même système de signes. Peu importe, à la limite, qu'on use pour cela d'un carton perforé ou d'une plaquette de silicium. On écrit des instructions pour que la machine réalise des opérations, des plus simples comme l'addition, jusqu'aux plus complexes (calcul de la trajectoire d'une fusée ou production d'un roman). Le résultat n'est qu'une succession de caractères, que leur tout constitue un programme, une image ou un son. Tous ces objets se travaillent comme du texte, avec des outils qui amplifient plus qu'auparavant les procédures de distanciation entre la forme graphique et le sens du texte. Ces caractères, ces textes, qui ne sont rien sans ces outils, voient leur statut tranformé par la présence des seconds. Ainsi, le système graphique et le codage, la pensée et ses instruments de lecture sont indissociables.

Il s'ensuit une approche conceptuelle de la connaissance et de son ordonnancement d'un type — osons le mot — nouveau. Non pas révolutionnaire, car il est clair que toute l'histoire de l'écriture témoigne de telles transformations, qui touchent autant à nos catégories mentales (produit des mesures et classifications à notre disposition) qu'aux possibilités d'analyse des processus de l'écrit induites par la forme de notre outillage mental. À ce titre, l'informatique explicite toute une série d'opérations, puisqu'il a fallu les énoncer avant que la machine ne puisse les reproduire.

Mais avant de pouvoir amplifier nos possibilités de calcul, de synthèse, et de questionnement sur nos activités intellectuelles, il faut suer et souffrir. Là encore, les mots ne sont pas trop faibles. Abandonner les anciens réflexes, en assimiler de nouveaux, apprendre des méthodes, des savoir-faire, des tours de mains ou des procédures qui ne sont pas toujours d'une rationalité évidente — ou clairement documentés —, assumer son inculture, constituer un réseau social d'entraide, tout cela prend du temps. Ce n'est qu'au terme de tant d'efforts que l'on profite des capacités d'écriture des ordinateurs. On a vu que ceux-ci ne nous aident pas qu'à gagner du temps. Ils nous permettent de faire un usage judicieux de ces trois constituants de l'écriture que sont la liste, le graphique, et l'hypertexte. Il s'ensuit une production de sens dont on aurait tort de se priver : après tout, cet appareillage composé par les logiciels de qualité évoqués précédemment a été conçu par des hommes ayant des besoins explicites en matière de manipulation du code, d'organisation du raisonnement, de production de synthèses.

Entre les marchands de solutions toutes faites (et mal faites) qui étouffent nos capacités intellectuelles à force de multiplier les formats, les cryptages et les supports, et les défenseurs de l'outillage mental classique qui acceptent trop souvent les produits des premiers au prétexte qu'une simple machine à écrire leur suffit, et que la technique ne peut participer à la reconstruction de la culture, le chemin, le cheminement intellectuel pourrait-on dire, est parsemé d'embûches.

D'autant que les enjeux sont de taille : on a brièvement évoqué leurs aspects économiques, liés aux constructions de rentes, de monopoles, qu'élaborent quelques industriels en s'appropriant notre système de signes, et plus largement notre outillage mental. Mais ces enjeux sont avant tout intellectuels, et structurent donc le monde des professionnels de la connaissance, enseignants et universitaires. C'est pourquoi nous devons maintenant décrire la façon dont s'expriment et se réorganisent les rapports de force, les dynamiques de recherche dans les sciences humaines, suite à la diffusion d'un outillage mental qui s'appuie sur les ordinateurs et les réseaux.

Deuxième partie

Socialisation de l'internet à l'ENS

Afin d'expliciter les enjeux de l'outillage mental au sein du monde universitaire, nous commencerons dans cette partie par décrire précisément les modes de diffusion de l'informatique et de l'internet au sein de la division littéraire de l'École normale supérieure[1]. L'influence majeure des informaticiens, l'amusement ou le désarroi des premiers expérimentateurs, puis la constitution de laboratoires et la mise en place d'une réflexion sur les processus de l'écrit — à l'occasion de la découverte du web — seront détaillés.

Enfin, viendra le temps des conflits entre les partisans de la machine à écrire, d'une informatique à usage administratif, d'un outillage mental immuable, et entre ceux qui auront décidé d'aller aux limites des processus d'écriture permis par le renouvellement de cet appareillage intellectuel.

Cette histoire brève mais tumultueuse ne se termine pas comme un conte de fée. Elle est cependant riche d'enseignements : un des effets réflexifs de l'écriture consiste en l'explicitation de nos pratiques, intellectuelles, gestuelles et, pour les professionnels de l'écrit, sociales. Il s'ensuit que l'organisation du monde universitaire et les tensions qui le traversent se donnent « à lire » au travers de l'étude de la socialisation de l'internet.

1. Cette division regroupe dans ses grandes lignes les disciplines les plus traditionnelles des sciences humaines. À l'ENS, la division entre « littéraires » et « scientifiques » correspond aux deux types de concours dans lesquels les départements espèrent puiser leurs futurs élèves — c'est ainsi que l'on appelle les étudiants de l'ENS, alors qu'ils ont entre 19 et 25 ans —, même si des jeunes mathématiciens deviennent parfois économistes ou philosophes. Les principaux départements (ou sections) de l'École littéraire recouvrent les disciplines suivantes : histoire, géographie, littérature et langages, philosophie, sciences de l'Antiquité, sciences sociales (économie incluse). Quelques laboratoires du CNRS s'ajoutent à ces départements, et parfois, les relations des uns et des autres se limitent à des relations de pur voisinage, quand même leurs thèmes de recherche peuvent être proches.

Pour décrypter de tels usages, un instrument s'impose. Il va de soi que son emploi et la maîtrise des savoir-faire associés peut — sans boutade — se résumer en l'application d'une technique sociologique. C'est celle que Pierre Bourdieu a développée dans *Homo Academicus* que nous privilégierons[2], tant pour sa mise en valeur de l'économie symbolique du monde universitaire que pour ses mises en garde méthodologiques.

Aussi, commencerons-nous par proposer une grille de lecture des normes et des codes qui régissent les chercheurs en sciences humaines.

2. Même si nous nous en écarterons parfois.

Chapitre 1

Sociologie des chercheurs

Deux écueils se présentent quand nous désirons analyser notre propre univers professionnel et apporter une contribution scientifique qui dépasse le simple témoignage — même si celui-ci peut être fort utile : le premier est celui de la prise de distance d'avec son propre monde, de « la rupture avec la connaissance indigène » ; il n'est pas propre à un terrain donné : que l'on soit universitaire, médecin, ou ouvrier, l'analyse des tensions, des conflits, des enjeux, au sein de son univers professionnel, ne peut se réaliser si la description proposée reste dépendante des codes implicites de son fonctionnement. Mais ensuite (et là apparaît le second écueil, propre à la sociologie des chercheurs), il faut « revenir », c'est-à-dire réussir à exprimer en des termes scientifiques — et donc conformes aux critères de rigueur et de rationalité appris et développés au sein de l'Université — les connaissances obtenues à l'occasion de cet éloignement [1]. Autrement dit, après s'être écarté de son univers pour mieux en saisir le fonctionnement, l'explicitation du savoir acquis par cette prise de distance ne peut se réaliser qu'en effectuant un retour, en reprenant les normes scientifiques de la démonstration et du raisonnement typiques du monde que l'on étudie [2].

1. Ce que Pierre Bourdieu nomme la « restitution de la connaissance scientifique obtenue au prix de cette rupture ». [Bou84], p. 11, comme la citation précédente.

2. Il s'agit notamment d'éviter les défaillances du langage ordinaire, qui permet trop facilement de reconstruire une histoire journalistique, en usant de stéréotypes semi-savants,

1.1 Proximités

Si l'on désire mesurer les interactions entre pratiques sociales et intellectuelles des chercheurs, cet aller et retour est d'autant plus délicat que le chercheur est pris dans un collectif, celui de son laboratoire, où ses collègues, étudiants, amis et aussi ennemis contribuent au développement de sa pensée, de ses analyses. Et, avant de montrer comment l'expression d'une pensée solitaire et originale est valorisée, ce qui rend difficile la synthèse puisque chaque chercheur est alors un cas particulier, puis, de rappeler comment les contraintes propres à la reproduction du monde universitaire favorisent à l'opposé une forme de conformisme, surtout quand on s'attache à décrire son fonctionnement, il nous semble important d'insister sur une forme spécifique d'autocensure : celle-ci résulte d'un respect des règles de scientificité. Or, la production d'énoncés en matière d'outillage mental impose une position en surplomb qui peut donner le vertige.

1.1.1 Estimes

Quels que soient le degré de spécialisation que l'on atteigne, la qualité des experts dont on s'entoure, il y a toujours quelque difficulté à interpréter et à généraliser leurs témoignages lorsque ces travailleurs de l'intellect explicitent leurs méthodes. Peut-on comprendre précisément les techniques d'écriture, de comparaison, d'analyse qu'un physicien ou un linguiste met en œuvre sans en avoir soi-même une pratique régulière ? On en doute, d'autant que l'enquêteur qui affirmerait accéder à une telle compréhension pourrait alors prétendre au statut scientifique des érudits et des savants qu'il étudie.

Par ailleurs, on sait que cette analyse de l'outillage mental doit se doubler de celle du laboratoire : pour élucider les modes de socialisation au sein de ce dernier, il faudrait être suffisamment au fait de l'institutionnalisation des disciplines, ce qui est loin d'être acquis.

---

comme on en rencontre souvent (le développeur libertaire, le professeur de grec réactionnaire, le jeune scientifique, le vieux lettré, etc.).

Face à de telles questions, l'humilité est la règle. De plus, le fait que les meilleurs débats sur l'outillage des uns et des autres se produise souvent dans un contexte d'amitié ou d'admiration (parce que l'on touche au cœur de la productivité intellectuelle) génère lui aussi un important effet de pudeur chez celui qui prétend en rendre compte.

Autrement dit, la possibilité qu'offre le monde universitaire de créer un réseau d'estime est spontanément génératrice d'autocensure pour qui veut produire des énoncés scientifiques sur les comportements, les méthodes, les productions de ses collègues.

1.1.2 Contradictions

Cependant, le passage du champ de la prospective intellectuelle à celui de l'analyse sociale peut faciliter l'énonciation de tels raisonnements : la description minutieuse du fonctionnement de son laboratoire, et la découverte de pratiques qui contredisent les postulats censés régir le monde scientifique, offrent au chercheur un nouveau souffle, une nouvelle vigueur pour aborder ce concept d'outillage mental et pour comprendre la façon dont il peut être sollicité, approprié, détourné.

Pierre Bourdieu explique bien ce fait : assumant le caractère fortement auto-analytique de son analyse des ressorts du monde universitaire, il conclut son ouvrage en détaillant comment le serviteur de la science qu'il était s'est senti obligé d'obéir aux axiomes qu'on lui avait inculqués : « Et la place qu'occupe dans mon travail une sociologie assez particulière de l'institution universitaire s'explique sans doute par la force particulière avec laquelle s'imposait à moi le besoin de maîtriser rationnellement, au lieu de le fuir dans un ressentiment auto-destructeur, le désenchantement de l'oblat devant la futilité ou le cynisme de tant de prélats de curie et devant le traitement réservé, dans la réalité des pratiques, aux vérités et aux valeurs que professe l'institution et auxquelles, étant voué à l'institution, il était voué et dévoué »[3].

3. [Bou84], p. 307.

On a déjà vu comment la technique de l'écriture induisait des effets proprement réactionnaires[4], du fait de la valorisation de ceux qui la maîtrisaient : l'obtention d'un fort pouvoir intellectuel invite à trouver les moyens de le conserver par des arguments de type intellectuel. Mais les recettes d'une telle valorisation ne peuvent évacuer les formes d'intelligence permises par le déploiement de l'écriture : les plus grands privilèges obtenus grâce à l'usage de l'outillage mental disponible sont toujours menacés si un nouveau venu arrive à mieux l'exploiter. Certes, l'histoire regorge de mises à l'échafaud de savants, mais, dans les sociétés qui valorisent la connaissance au point qu'elles en font leur priorité[5], les institutions qui sont chargées de la diffusion de cette connaissance doivent, peu ou prou, accepter d'être remises en question par toute forme de production conforme à ce qu'elles promeuvent.

Pour ce faire, elles ont acquis une indépendance, de façon à être le moins sensibles possible aux discours d'astrologues populaires[6], ou aux préjugés d'un dictateur. À la fortune, à la considération publique sont préférées l'évaluation, la critique et la considération par les pairs, plus discrètes, plus lentes, mais aussi plus conformes aux exigences de rationalité. Aux menaces que fait peser le pouvoir politique répond le droit de régir les conditions de sa propre reproduction. Là encore, cette autonomie a ses limites : le professeur de lycée Claude Lévi-Strauss fut révoqué suite à la promulgation des lois de Vichy et dut partir aux États-Unis. Marc Bloch sera lui torturé et abattu. Mais dans les périodes de paix, le monde universitaire définit ses propres règles de fonctionnement, étant entendu qu'il est aussi censé s'y plier.

1.2 Economie symbolique

Ces formes d'auto-régulation génèrent aussi des effets pervers. Dans ce monde *a priori* abstrait du matérialisme, que Pierre Bourdieu définit comme

4. Cf. paragraphe 2.2.2 page 35.
5. L'enseignement représente en France le quart des dépenses de l'État, comme le rappelent nos ministres des Finances dans les textes qui accompagnent nos fiches d'imposition.
6. Même si parfois, certains d'entre eux deviennent docteurs en sociologie...

« pré-capitaliste »[7], s'instituent des formes symboliques de l'échange. La multiplication des dons et contre-dons, la variété de leurs champs d'application[8], rendent particulièrement complexe l'analyse du groupe social ayant légitimité à transmettre l'outillage intellectuel, particulièrement dans le domaine des sciences humaines.

Ainsi, la valorisation d'un individu se mesure autant par le respect que l'on porte à son œuvre, à sa culture ou à ses découvertes que par les titres qu'on lui accorde pour lui donner le droit ou les moyens de reproduire un système fondé sur cette valorisation du travail intellectuel. Certaines légitimités ont des effets directs sur le salaire : par exemple, lorsqu'on est appelé à participer au jury d'un concours prestigieux (agrégration, etc.), ou lorsqu'une position de directeur donne accès à une voiture et à un appartement de fonction.

Pierre Bourdieu rappelle que, pour accéder à de pareilles consécrations, il faut composer au mieux avec son réseau professionnel, de façon à obtenir des responsabilités qui permettront de gérer un budget de recherche — moyen de faire pression sur ses subordonnés —, de diriger une revue savante ou une collection éditoriale — façon de promouvoir, de remercier des collègues en les invitant à y participer, et d'en écarter d'autres —, et bien sûr d'obtenir un réseau d'étudiants, si possible juste avant le passage initiatique de la thèse. Les participations aux jurys permettant la sélection des futurs professeurs et chercheurs, aux commissions ministérielles ou plus largement scientifiques chargées d'attribuer des budgets de recherche, complètent cette panoplie de l'échange qui articule les relations entre universitaires. Il s'ensuit une logique de clientélisme, où la lenteur devient un mode de fonctionnement proprement politique : le responsable perd de son temps à participer à de multiples réunions où se constituent alliances, dons et contre-dons, temps qu'il fait payer aux impétrants en reportant le moment où ils auront le droit de publier dans la revue du maître[9].

7. [Bou84], p. 128.

8. En France, les activités de recherche, d'enseignement et d'administration sont — sauf au CNRS — étroitement liées.

9. [Bou84], pp. 118–119.

On le voit, cette logique non marchande est l'occasion de multiplier les conflits, les rancœurs, de façon d'autant plus importante que, souvent, la personne évincée d'un poste, d'un budget, etc., prend la mesure de son indignité en même temps qu'elle subit un « des procédés classiques de la polémique universitaire qui ne désigne les adversaires que par des allusions, des insinuations ou des sous-entendus compris seulement des détenteurs du code, c'est-à-dire, en plus d'un cas, des seuls adversaires visés »[10] : la mise à l'écart, souvent sournoise, ne peut être publiquement interprétable qu'en termes de production intellectuelle de mauvaise qualité[11].

1.3 Ethique

Comment, dans un tel cas de figure, le chercheur peut-il garder le recul nécessaire à l'énonciation scientifique et éviter de se plier à la logique dominante, d'entrer dans ce jeu de services rendus, de solidarités automatiques, afin d'obtenir ce qu'on lui fait espérer ?

Dans un cadre étroit où les conflits de personnes se confondent inconsciemment avec des conflits d'intérêt, la volonté d'exemplification peut dériver vers une dénonciation qui peut ne plus avoir de rapport avec le travail scientifique. Encore faut-il que ce dernier ne dérive pas vers un discours où les « raisons des adversaires » soient réduites à des intérêts sociaux[12]. Ainsi, il s'agit d'éviter le double piège de l'attaque *ad hominem*, même feutrée, sans pour autant réduire la personne visée à un pur produit de sa caste, c'est-à-dire à nier son propre attachement à une production scientifique.

Les meilleurs révélateurs des rapports de force universitaires restant les conflits, se pose alors le problème de leur restitution, dans un cadre qui implique souvent l'analyste lui-même. À l'époque où l'internet était diabolisé

10. [Bou84], p. 39.
11. C'est-à-dire en termes d'agression très personnelle, au vu de l'axiomatique propre au monde de l'enseignement et de la recherche.
12. [Bou84], p. 28.

en France[13], notre engagement dans cette technique nous a fait subir diverses agressions. La rigueur méthodologique nous obligera à les expliciter, afin que le lecteur ait les moyens de comprendre de la façon la plus précise les enjeux, les tensions que mettait en évidence cette technique d'écriture.

Reste la question de l'attitude à choisir face à la nomination des personnes qui ont participé, parfois inconsciemment, à l'ensemble de ces conflits qui ont secoué le monde universitaire. Pierre Bourdieu choisit de « donner en toutes lettres les noms des universitaires étudiés »[14], partant du principe que la réduction aux initiales ne pouvait que favoriser la dénonciation implicite qu'il condamne. Il compte par là aussi faire œuvre d'historien, puisqu'il se dit conscient qu'un des intérêts de l'analyse du champ universitaire qu'il propose réside dans le fait qu'elle éclaire le lecteur sur « la production culturelle française des vingt dernières années »[15].

N'ayant pas de telles prétentions, nous avons fait le choix opposé. De plus, il nous semblait inutile de nous apesantir sur celles et ceux qui abandonnent le travail scientifique parce qu'ils sont trop pris dans les logiques de contrôle de l'institution et trop attachés à l'accroissement de leur capital symbolique. Sans pour autant négliger le poids des institutions, nous croyons encore que les fossoyeurs de la recherche, s'ils peuvent ralentir la productivité scientifique, ne peuvent la paralyser. Ces individus ne sont donc pas cités, et tout a été fait pour qu'ils ne soient pas reconnaissables. En corollaire, les personnes qui ont fait preuve d'audace intellectuelle ne sont pas non plus nommées. Ce qui a deux avantages : nous ne risquons pas d'induire l'idée d'une évaluation scientifique implicite, en mettant les uns en avant et en renvoyant les autres aux poubelles de l'histoire ; ensuite, nous évitons de faire encourir des risques aux militants de l'internet, qui subissaient — et souvent subissent encore — le pouvoir de ces fameux « prélats de curie » que critique Pierre Bourdieu. Aussi, par souci de protéger celles et ceux que nous appellerons des « pionniers »[16],

13. 1996–1999. Cf. [Gui97] et le paragraphe 2.1 de l'annexe (p. 397).
14. [Bou84], p. 297.
15. [Bou84], p. 297.
16. Ce mot est à prendre en un sens dénué de toute option morale : « défricheur »,

dans cette partie comme dans la suivante, leurs actes et leurs témoignages seront entièrement anonymisés.

Après ce premier panorama centré sur les personnes, reste maintenant à expliciter comment s'articulent les relations entre singularités et normes sociales, et comment l'outillage mental est sollicité dans le cadre de la compétition collective au sein du monde universitaire.

1.4 Compétition entre disciplines

Si on se limite encore une fois à la France et aux sciences humaines, il importe de comprendre en quoi les positions relatives des disciplines expliquent leurs relations à l'informatique et plus largement à l'internet. Aux disciplines « théoriques » et « pures », comme la philosophie, le français, les lettres classiques, s'opposent dans les années 1970 les disciplines pratiques, « appliquées », « empiriques », « impures » comme l'économie, la sociologie, l'ethnologie, la linguistique et la géographie[17]. Aux premières la conceptualisation, aux secondes l'appel aux techniques. Reste que les « techniciens » vont s'employer à valoriser leur statut d'infériorité, pour arriver à inverser la tendance. Comment s'y prendront-ils ?

Tout d'abord, il suffit de laisser faire : les seigneurs de l'Université vont se focaliser sur leur reproduction, qui s'étend jusqu'à l'enseignement secondaire[18]. L'outillage intellectuel ainsi dispensé finira bien par être critiqué pour son archaïsme, et pour la façon tout aussi anachronique de le diffuser,

« personne qui est parmi les premières à se lancer dans une entreprise, qui fraye un chemin » ([Rob85]). Il est synonyme du mot « innovateur », et les dictionnaires précisent qu'ils s'opposent aux adjectifs « archaïsant » et « routinier ». Un pionnier ou un innovateur n'ont pas nécessairement de meilleures raisons que d'autres à agir comme ils le font. Ce qui importe ici, c'est le caractère, statistiquement rare, de leurs engagements.

17. [Bou84]. Les adjectifs entre guillemets sont tirés de la légende du graphique de la page 160.

18. Les « disciplines canoniques, histoire de la littérature française, lettres classiques ou philosophie [contrôlent la reproduction de l'enseignement secondaire] à travers les programmes, les cours et les sujets de concours ». [Bou84], p. 134.

notamment en termes pédagogiques[19]. Évidemment, les critiques les plus virulentes viendront des nouvelles disciplines, dans un état de domination, qui tendent néanmoins à s'emparer des thèmes intellectuels — en perte de vitesse vu le temps passé à gérer l'institution plutôt qu'à produire une œuvre scientifique — que se sont appropriées les disciplines légitimes. La linguistique, la sémiologie, la sociologie utilisent au mieux les conceptualisations et les formalisations qu'elles développent, souvent grâce à leur plus grande proximité avec les sciences exactes : l'outillage mental est ainsi réactualisé. Il suffit alors de quelques coups de boutoirs contre les formes de l'écriture des anciennes disciplines, réduites à l'état de « beau style littéraire », en en proposant de nouvelles, qui assument leur inélégance du moment en « empruntent aux rhétoriques les plus puissantes [...], celle de la mathématique [...] ou celle de la philosophie »[20] pour entériner l'inversion des rapports de force.

Bien sûr, de telles transformations du paysage universitaire ne peuvent se réaliser qu'en abandonnant la conception de l'artisanat classique de l'intellectuel au profit d'une valorisation du travail collectif. Le laboratoire est d'autant plus proche qu'une telle dynamique intellectuelle est facilitée par la faible institutionnalisation des jeunes disciplines, qui peuvent se consacrer à la recherche, et acquérir par là une notoriété internationale[21]. La dépossession des anciens se finalise quand un individu incarne l'inversion des rapports de force : la philosophie est reléguée derrière les sciences sociales quand Claude Lévi-Strauss annihile la figure exemplaire de Jean-Paul Sartre, « réhabilite ces disciplines traditionnellement méprisées par les normaliens philosophes et les institue en modèle de l'accomplissement intellectuel »[22].

19. [Bou84], p. 305.
20. [Bou84], pp. 45–46.
21. [Bou84], p. 140 et p. 155.
22. [Bou84], p. 300. On remarquera que les intitulés ont plus de force que l'on croit dans de telles luttes entre disciplines : Pierre Bourdieu rappelle qu'au mot « ethnologie », un peu désuet, Claude Lévi-Strauss a préféré l'expression « anthropologie structurale », qui témoignait d'une plus grande universalité, en combinant des emprunts à l'école anglaise et à la philosophie allemande de Kant.

Ainsi, alors même qu'elle se refuse souvent à l'admettre, l'Université est particulièrement dépendante de l'outillage mental qu'elle façonne et distribue, au point qu'elle vit de façon très conflictuelle les réordonnancements de ce dernier[23], tant sur un plan manifeste (les « mandarins » contre les « hérétiques ») que plus discret et d'autant plus cruel : les « bons élèves » comme les normaliens fidèles à la philologie, par exemple, « restés étrangers [...] du haut de leurs certitudes statutaires, à l'évolution des sciences du langage [...] se sont trouvés soudain dévalués, puis relégués [..] devant l'irruption de la linguistique, importée et défendue par des marginaux, souvent non normaliens, provinciaux ou issus de disciplines 'inférieures' »[24].

Les bons élèves sont d'autant plus menacés que les « anciens dominants [...] qui se trouvent conduits, à leur insu et malgré eux, à une position dominée, contribuent en quelque sorte à leur propre déclin en obéissant au sens de la hauteur statutaire qui leur interdit de déroger et d'opérer à temps les reconversions nécessaires »[25].

Ce rappel de la complexité de la sociologie du monde universitaire précise en quelque sorte le statut fragile des tenants du pouvoir, et montre en quoi l'arrogance est leur seule défense face à des méthodes de travail en perpétuel renouvellement.

*

L'université n'est pas aussi figée qu'elle peut sembler l'être : elle est aussi dépendante de sa propre logique, qui la conduit à accueillir ses meilleurs élèves, et à accepter un renouvellement de l'outillage mental.

Mais les formes d'optimisme que peut exprimer une personne sensible depuis une dizaine d'années à l'incidence de l'informatique sur les pratiques de recherche en sciences humaines ne doivent pas faire oublier que l'usage

23. Pierre Bourdieu en donne plusieurs exemples : conflit entre ancienne et nouvelle Sorbonne à la fin du siècle dernier, essor des historiens avec les *Annales*, consécration de Claude Lévi-Strauss en 1955, mouvement de mai 1968.

24. [Bou84], p. 166.

25. [Bou84], p. 167.

d'un outillage intellectuel renouvelé sert autant les appétits de pouvoir que les projets plus proprement scientifiques. Certains philosophes du langage n'hésitent pas à user de l'aura procurée par l'ingénierie de la langue pour améliorer leur « image de marque ». Depuis peu, les mêmes qui méprisaient l'internet font tout pour obtenir la responsabilité de la publication web de leur université. Par ailleurs, les relations hiérarchiques entre disciplines ne sont jamais fixées. Si certaines, comme la sociologie, à la suite de Pierre Bourdieu, ont considérablement modifié leurs méthodes dans les années 1970, les héritiers de ce dernier n'ont pas toujours cherché à prolonger un tel effort ; d'autres, ayant une image plus poussiéreuse, peuvent être confrontées, par choix ou par nécessité, aux techniques d'écriture contemporaines : avec leurs corpus de CD-ROM et l'implémentation sur des ordinateurs de graphies rares (grec ancien, araméen, etc.), les lettres classiques semblent jouer [26] une carte moderne. En revanche, les spécialistes de littérature qui dédaignent l'internet n'ont donc que le style et la graphie classique pour défendre leur science. Ce qui peut les mettre en fâcheuse posture d'ici quelques années.

On se rappellera que l'écriture contemporaine offre des potentialités plus que des solutions : cette technique peut accroître nos capacités intellectuelles ; mais encore faut-il qu'elle fasse l'objet d'une appropriation, et que celle-ci soit accompagnée d'une explicitation des processus mentaux ainsi générés.

26. Au sens premier du terme. Il est surprenant de voir à quel point cette discipline fait un constant appel à l'écriture, au système de signes, pour légitimer une attitude de ségrégation intellectuelle (qui ne sait pas écrire en grec ancien est exclu) sans pour autant — à de rares exceptions près — tirer un quelconque parti de son rapport privilégié à divers systèmes graphiques.

Chapitre 2

Arrivée de l'internet

Le détail de la façon dont l'internet s'est progressivement diffusé au sein de la division des lettres de l'École normale supérieure dans les années 1990 répond tout d'abord au besoin de décrire le plus précisément possible une lente et coûteuse appropriation, voire des choix difficiles en termes d'investissement intellectuel pour les étudiants et les chercheurs qui l'ont choisie. À ce titre, ce travail participe d'une démarche que l'on peut qualifier d'ethnographique. Mais celle-ci est aussi historique : en effet, dans le domaine de l'internet, la tendance est à la « mémoire courte », et bien des pratiques, des réactions, voire des stratégies, partagées ou décidées il y a peu d'années sont déjà oubliées. Ce fait est amplifié par le dogme du déterminisme technique, qui veut toujours faire du récent passé table rase : le culte de l'éternelle nouveauté nous engage dans un présent perpétuel qui fait abstraction de son terreau social. En donnant un témoignage détaillé de la façon dont un ensemble de chercheurs a pu s'approprier un outil d'écriture qui semble accepté en 2001, nous espérons mettre en évidence les principales tensions politiques et intellectuelles à l'œuvre dans les sciences humaines.

Reste alors à savoir si une telle description permet ou non la généralisation : l'École normale est une institution prestigieuse en France, et ses membres n'ont pas besoin de solliciter les « nouvelles techniques » pour en améliorer l'image. C'est par ailleurs un établissement interdisciplinaire, où le poids de la division scientifique est très fort, en moyens financiers, en nombre de chercheurs, ou en matière de renommée sientifique : plusieurs prix Nobel et médailles Fields sont passés par l'ENS, et les informaticiens de l'ENS ont vite assis leur réputation, malgré la jeunesse de leur discipline. Ces théoriciens, et bien d'autres chercheurs en sciences exactes, utilisent évidemment des ordinateurs, et leurs besoins comme leurs pratiques profilent des normes et des usages qui s'étendent bien au-delà de leurs départements. Une telle situation se rencontre rarement dans une université qui se consacre exclusivement aux sciences humaines.

Pourtant, malgré cette « marginalité » de l'ENS, la confrontation de sa partie littéraire avec l'informatique et l'internet est révélatrice d'un fonctionnement aisément généralisable.

Tout d'abord parce que l'ENS littéraire reste un espace de production d'auteurs universitaires : ses anciens élèves monopolisent de nombreux postes dans les universités, et leurs nombreuses publications (ouvrages scientifiques, manuels, etc.) leur garantissent des relations privilégiées avec les éditeurs. Malgré ses faibles effectifs, l'ENS elle-même, en tant qu'organe de reproduction, valorise beaucoup l'imprimé : ses professeurs sont souvent responsables de collections ; sa bibliothèque, réservée aux élèves et anciens élèves, est de fait un lieu de travail pour « grands lecteurs », qui lisent pour écrire. Aussi peut-on s'attendre à ce que les « pouvoirs de l'imprimé » réagissent de façon manifeste à la diffusion de l'informatique et de l'internet, même si ces réactions ne sont pas toujours explicites. En ce sens, la situation à l'ENS peut être considérée comme exemplaire de celle d'un groupe social spécifique, que l'on désigne parfois comme les intellectuels ou les érudits, mais qui regroupe assurément les héritiers du système éditorial et universitaire qui s'est mis en place au XIX^{e} siècle.

Ensuite, parce que la situation privilégiée de l'École joue bien sûr en sa faveur : le contact avec les « scientifiques », avec les étrangers, les dotations budgétaires conséquentes, l'autonomie des chercheurs font que des choix techniques, pédagogiques ou politiques s'y imposent moins fortement et moins uniformément que dans une petite université sans grands moyens ; par suite, les membres de cette institution peuvent expérimenter de nouvelles pratiques inimaginables dans beaucoup d'espaces d'enseignement et de recherche. Aussi peut-on penser que si la diffusion des techniques informatiques a pu générer des tensions, des incompréhensions, des conflits au sein de l'École littéraire, ceux-ci ont dû être encore plus importants au sein de la majorité des laboratoires et universités de sciences humaines.

L'intention n'est pas de sous-entendre que l'ENS littéraire a abordé les questions de la relation de l'informatique et de l'écrit avant les autres, ou qu'elle s'est livrée à des expérimentations qui servent de référence : ces tentatives s'inscrivent dans un contexte plus large, mais le fait d'en décrire une de façon détaillée donne les moyens de comprendre un mouvement diffus et complexe. Et c'est surtout par ses formes de résistance à l'informatique et à l'internet que la partie littéraire est exemplaire : celles-ci ne témoignent pas seulement d'un conflit entre « anciens » et « modernes », mais servent de référence pour les autres institutions tentées par de tels refus de repenser l'outillage mental, au vu de la légitimité de cette institution à définir et à reproduire les normes universitaires[1].

2.1 Un projet d'informaticiens

En 1990, peu de personnes utilisaient l'internet en France : environ 3000 machines, réparties sur 50 sites[2], étaient interconnectées. Les protocoles les plus utilisés étaient le courrier électronique[3], la connexion sur une machine

1. Les lecteurs étrangers seront peut-être surpris de voir réunis sous une même casquette de « littéraires » un ensemble d'étudiants et de professeurs spécialisés en des disciplines aussi variées que l'archéologie, l'économie, l'histoire, etc. Mais le choix des mots traduit aussi des enjeux essentiels : l'affirmation du caractère « littéraire » des activités d'enseignement et de recherche en sciences humaines à l'ENS signifie aussi que les représentations mentales du XIX^{e} siècle perdurent encore aujourd'hui en ce lieu.

2. Domaines, comme celui de l'ENS (*ens.fr*) ou de l'INRIA (*inria.fr*)

3. SMTP, ou *single mail transfer protocol.*

distante (telnet) et le transfert de fichiers (ftp[4]). Le confort d'utilisation était alors rudimentaire. Tout d'abord, les débits étaient très réduits ; l'ENS, qui, après avoir utilisé le réseau Bitnet d'IBM, s'est raccordée à l'internet en 1988, disposait alors d'une connexion vers l'INRIA[5] de 9600 bauds[6], soit un débit d'environ 1000 caractères par seconde, au lieu de 2 millions de bauds à partir de février 1993 (date du raccordement au réseau RENATER[7]) et 155 millions depuis 2002. Ensuite, il n'existait aucune forme d'annuaire pour repérer l'adresse d'une personne ou d'une machine ; ces informations s'obtenaient par le bouche à oreille ou par le courrier électronique, lequel évidemment ne contenait pas de caractères accentués et dépassait rarement la dizaine de lignes. Enfin, les commandes informatiques, tout comme leur syntaxe, étaient absconses : on ne retenait que les plus utiles, comme celles permettant l'usage du courrier électronique.

En ces temps où un disque dur d'un giga-octets tenait dans une armoire et se louait 5000 F par mois[8], seuls quelques groupes de chercheurs en sciences exactes qui travaillaient dans des institutions privilégiées en matière d'équipement, pouvaient envoyer un bref message à un collègue ou lancer une procédure sur une machine distante (du CNRS, par exemple).

À cette période, l'internet était concrètement associé à des machines Unix[9], reliées entre elles par un cablâge de type Ethernet, et réputées non-

4. *File transfer protocol.*

5. Institut National de Recherche en Informatique et en Automatique.

6. Bits par seconde, le « bit » étant la plus unité informatique, composé d'un « 0 » ou d'un « 1 ». Huit bits composent un octet (« byte » en anglais), qui code un caractère.

7. Réseau national de télécommunications pour la technologie, l'enseignement et la recherche, créé officiellement en janvier 1993. Cf. http://www.renater.fr/GIPRenater/historique.htm

8. Ce que coûtait le disque dur du second ordinateur IBM du Centre de calcul de l'ENS, qui fut utilisé jusqu'à la fin 1992.

9. Unix est un système d'exploitation multi-tâches et multi-utilisateurs. Le succès d'Unix est lié à sa puissance, à sa sécurité, et surtout à l'intégration entière, à partir du début des années quatre-vingts, des protocoles TCP/IP nécessaires au fonctionnement de l'internet : IP (Internet Protocol) pour le chemin permettant d'aller d'une machine à une autre, TCP (Transfer Control Protocol), pour la gestion des services entre les machines.

conviviales : déjà l'obligation de se signaler avec un *login* et un mot de passe empêche une utilisation spontanée de la machine ; ensuite, la complexité des commandes requiert une bonne connaissance du manuel, en ligne, et rédigé en anglais. Enfin, l'interface graphique, souvent rudimentaire — les instructions sont principalement transmises à la machine par l'intermédiaire du clavier — rebute les profanes.

Cependant, l'utilisation de l'internet faisait doucement son chemin, et, « au sein du département de mathématiques et d'informatique à la fin des années quatre-vingts, il était tacite que tout le monde lisait son courrier [électronique] »[10].

Si d'autres départements de l'ENS, comme le département de physique, avaient eux aussi une utilisation précoce de l'internet, et si des chercheurs en sciences humaines utilisaient dès les années quatre-vingts des connexions à distance pour effectuer des calculs jugés à l'époque lourds (comme des géographes le faisaient avec les machines du Circé), l'apparition de l'internet au sein de l'École littéraire date de janvier 1992 : le laboratoire de sciences sociales venait d'acquérir une généreuse dotation ministérielle pour faciliter sa structuration en département. Pour témoigner de son dynamisme, et pour prouver son engagement dans une modernisation de l'outillage mental, il se devait d'acquérir un équipement informatique prestigieux : une telle démonstration devait permettre au futur département de d'asseoir sa domination sur les autres, conformément aux nouvelles donnes initiées par Claude Lévi-Strauss et Pierre Bourdieu.

Or, l'usage de l'informatique en sciences humaines faisait déjà l'objet d'un débat à l'ENS littéraire : certains chercheurs avaient connu les gros systèmes, tout en restant à distance des machines, gérées par des ingénieurs. Cette forme de dépendance s'atténua dans les années quatre-vingt, quand les chercheurs découvrirent la micro-informatique, et notamment les Macintosh qui,

Linux est un système d'exploitation dérivé d'Unix.

10. Communication personnelle.

grâce à leur système d'exploitation iconique et leurs logiciels *wysiwyg*[11], leur offraient un fort sentiment d'autonomisation. vers 1990, quelques sociologues et économistes de l'ENS faisaient usage d'un tableur et de quelques logiciels de statistique[12], quand certains de leurs collègues antiquisants utilisaient des logiciels d'extraction de formes graphiques grecques ou latines. Les autres, dans leur grande majorité, découvraient l'ordinateur, en se familiarisant avec le traitement de texte. Cependant, les déboires que les uns et les autres rencontraient avec ces machines les incitèrent à demander à la direction de l'ENS de mettre sur pied une cellule informatique littéraire qui satisferait les besoins spécifiques des littéraires — étudiants et enseignants — en informatique, notamment en matière de formation et de maintenance de machines. Cette démarche témoignait d'un désir d'autonomie face au service de prestations informatiques (SPI), qui était considéré comme trop inféodé aux mathématiciens : préconisant des stations Sun, des outils d'écriture comme $\TeX$, et peu tenté par des logiciels clickodromes, ce service avait la réputation d'être sourd aux préoccupations des littéraires.

Deux amateurs de l'informatique furent invités à fonder cette cellule informatique littéraire, dont l'auteur de ces lignes, ce qui nous invite localement à nous exprimer à la première personne. J'avais découvert l'informatique depuis peu, avec les Macintosh, dont j'appréciais, comme tout profane, à la fois le caractère convivial, et les potentialités en matière de traitement graphique du texte et de l'image. Mais j'en imaginais peu d'autres que ludiques ou esthétiques, étant assez peu au fait de la recherche en informatique pour les sciences humaines. J'étais néanmoins considéré, du fait de mes

11. *What you see is what you get.*

12. L'un d'eux, Spad, donne une idée de la complexité de la relation texte-image de cette période : ce logiciel permettait de réaliser des analyses statistique à l'aide d'une interface graphique. Mais les graphiques factoriels obtenus étaient en fait du texte, et des espaces. Par exemple, les axes étaient représentés par des points d'exclamation. Pour visualiser ou imprimer les graphiques, il fallait, dans un éditeur ou traitement de texte, sélectionner l'ensemble du document, et demander la sélection d'une police de caractères à espacement fixe, seul moyen de rendre ces axes alignés et la position du texte en accord avec les valeurs numériques...

cultures mathématique et anthropologique, comme un « informaticien » ouvert au dialogue avec les sciences sociales. Mon collègue, parce qu'il avait une compétence en termes de production de graphies rares (aussi sur Macintosh), apparaissait comme le représentant des antiquisants.

J'eus à prendre en charge l'équipement du futur département de sciences sociales. Au fil des dialogues avec les membres du SPI, j'étais de plus en plus sensible à leurs arguments, même si je conservais une passion pour les machines de la firme Apple. Ces collègues — comme d'autres chercheurs en informatique — penchaient, on l'a vu, pour des machines Unix, qui semblaient aussi puissantes que coûteuses, mais qui requéraient des connaissances susceptibles de rebuter les chercheurs en sciences humaines. Ces experts n'avaient qu'un mot à la bouche : « Internet ». Au fil des discussions, ce néologisme soulevait d'étranges questions: quel lien ce réseau américain pouvait-il avoir avec la dynamique de la recherche en sciences sociales en France ? Était-il si utile d'investir dans des systèmes techniques complexes quand l'enjeu essentiel consistait à former des esprits brillants à la conceptualisation et à la théorie ? Fort de tous ces préjugés, je résistais tant bien que mal, assez peu impressionné par l'intérêt du ftp, du courrier électronique, ou encore du *talk*[13].

Au final, mes conseillers, dont l'érudition m'impressionnait, m'aidèrent à trouver un compromis entre leurs préférences et les miennes, en m'incitant à choisir des machines NeXT, qui étaient seules à combiner l'Unix, l'internet et un système d'exploitation à interface graphique. Le fait qu'elles avaient été conçues par une équipe réunie par Steve Jobs, co-concepteur des ordinateurs Apple[14], semblait garantir une adaptation rapide des chercheurs et des élèves à ces nouvelles machines: les Macintosh étaient très répandus à l'ENS.

13. Permet un dialogue (écrit) en direct entre deux personnes si elles sont connectées sur le réseau au même moment, ce qui donne l'illusion d'un échange synchrone.

14. Le système « X » du Macintosh est en filiation directe avec les systèmes d'exploitation NeXT, puis NeXTStep, de ces machines, ce qui prouve qu'un système d'exploitation peut avoir dix ans d'avance sur les autres, contrairement aux discours sur l'innovation qui nous assurent qu'un retard de six mois n'est jamais rattrapable.

Cependant, la direction de l'École littéraire était réticente à l'idée d'un tel investissement : les NeXT étaient chers, et leur acquisition contredisait la volonté d'indépendance des littéraires face au SPI : il fallait acquérir des machines différentes de celles des scientifiques. Là encore, un compromis — plus financier que scientifique — fut trouvé : la dotation budgétaire servirait aussi à équiper la nouvelle cellule informatique, on achèterait des Macintosh en sus des NeXT, et les machines du département de sciences sociales seraient mises à la disposition de l'ensemble de la division littéraire de l'École.

Ainsi l'apparition officielle de l'internet au sein de l'École littéraire fut historiquement liée aux choix d'acquisition informatique d'un département. La console IBM du laboratoire de sciences sociales et le Macintosh II enfermé dans une ancienne chambre d'élève étaient remplacés en janvier 1992 par plusieurs Macintosh, et par quatre NeXT : 3 dans le département de sciences sociales, et un dans le bureau de la cellule informatique littéraire.

Ces quatre machines, très esthétiques, dotées de grands écrans — à la différence des hublots des Macintosh —, trônaient en majesté en deux lieux proches de l'ENS littéraire et témoignaient de son entrée dans la modernité : au-delà des conflits entre partisans des Macintosh et tenants des PC, elles symbolisaient aussi la spécificité des sciences sociales, qui savaient si bien s'approprier les outils de travail des sciences exactes. Du coup, les mots « Unix », « NeXT » et « Internet »[15] entraient dans le vocabulaire des chercheurs de l'ENS littéraire.

Aussi, du fait des informaticiens, les étudiants et les chercheurs en sciences humaines qui le désiraient ont pu se familiariser avec l'internet dès 1992. Point important, ces personnes étaient alors confrontées à deux systèmes d'exploitation distincts, ce qui les aidait à mieux comprendre le fonctionnement des ordinateurs.

15. En majuscule à cette période, et admirablement confondu avec « Ethernet ». L'ENS scientifique avait déjà acquis une vingtaine de NeXT, qui symbolisaient le web aux yeux de certains informaticiens (voir page 130).

2.2 Rejets

Bien sûr, pour expérimenter les nouvelles machines, il fallait se déplacer sur les lieux qui les hébergeaient ; c'est ainsi qu'au « réseau internet » se superposa un réseau social, constitué des membres du département de sciences sociales, de l'informatique littéraire, et de quelques enseignants-chercheurs et élèves en sciences humaines curieux des techniques informatiques d'alors : un mois après l'installation des NeXT, en février 1992, une vingtaine de personnes y disposaient d'un « compte ».

Mais celles-ci, comme l'auteur de ces pages, avaient en général assez peu de recul par rapport aux ordinateurs : le rejet pur, ou au contraire l'émerveillement et le prosélytisme naïf, l'emportaient sur une analyse raisonnée de leurs avantages et inconvénients. A de très rares exceptions près[16], la relation entre l'ordinateur et l'écriture n'était pas pensée, et le besoin de culture technique était euphémisé, juste réduit à une somme de tours de main.

Globalement, ces machines avaient trois inconvénients :

— elles s'imposaient aux personnes qui n'en voulaient pas ;

— elles ne faisaient pas ce qu'on attendait d'elles ;

— elles pouvaient réaliser des tâches impensées ou refusées.

2.2.1 Diacritiques et culture Unix

Les Next disposaient bien sûr d'éditeurs de texte, et l'on pouvait en acquérir d'autres en recherchant leurs sources... sur l'internet puis en les compilant ! Mais on a vu dans la première partie que de tels outils sont avant tout destinés aux informaticiens. Un seul traitement de texte clickable était disponible : WriteNow. Mais il était difficile d'utiliser un texte produit par ce logiciel sur un autre type de machine : les fondateurs des NeXT avaient choisi un code ASCII étendu bien à eux, ce qui fait que la communication avec les Macintosh, les Sun ou les PC obligeait à faire usage de logiciels

16. Comme le colloque de Nanterre, [AL91].

de recodage[17]. En revanche, deux tableurs étaient disponibles. Mais aucun n'acceptait pas les caractères accentués. Aussi, les personnes les plus rétives à l'informatique — majoritaires dans l'encadrement — allaient-elles bien vite rejeter ces machines.

Pour les autres, la curiosité était souvent au rendez-vous, tout comme une sorte de fascination devant la machine, joliment dessinée et capable de transmettre des courriers vocaux et des images. Par ailleurs, les concepteurs des NeXT avaient pensé aux littéraires, en introduisant les éléments clés de la culture... anglo-saxonne : le *Webster*, une Bible (en anglais) et les œuvres complètes de Shakespeare étaient disponibles, avec textes, images, procédures de recherche dans l'intégralité du corpus consulté et liens hypertextuels. Mais les potentialités de ces sources, comme celles des outils qui les accompagnaient n'étaient pas imaginées[18] en 1992.

Les ordinateurs restaient des machines à écrire et à compter, et leur organisation, comme leur système iconique apparaissaient déroutants : l'idée que la configuration des machines dépende des choix de l'utilisateur, et qu'un fichier personnel puisse être accessible à partir des quatre ordinateurs, troublait les esprits ; pour ce dernier point, l'architecture client / serveur d'Unix contredisait la conception classique de l'unicité du support papier.

Restait à maîtriser le nouveau système d'exploitation, qui souffrait de sa réputation de convivialité, avec ses menus déroulants et ses logiciels clickables : dans de telles conditions, il était difficile d'inciter les utilisateurs à suivre des longues formations[19].

17. Dont certains étaient explicitement prévus pour convertir le codage des Macintosh en codage NeXT et réciproquement. Nous nous étonnons rétrospectivement que de tels outils de transcription, si aisés à mettre en œuvre, aient été si peu développés et socialisés à cette période. De même pour la production de polices de caractères spécifiques, plus aisée qu'on ne le pense.

18. Alors que quelques années plus tard, ces formes d'hypertextes et les procédures de traitement qui les accompagnent apparaîtront symptomatiques des transformations de l'écriture.

19. De façon générale, un chercheur qui s'est construit une réputation d'excellence a toujours quelques réticences à subir un apprentissage dans un domaine où sa compétence

Enfin, la faible diffusion des NeXT faisait que peu de logiciels étaient écrits à leur intention. Par exemple, un seul logiciel statistique fut (plus ou moins bien) implémenté pour ces machines. Ce qui, encore une fois, incita les chercheurs ayant des besoins — ou des habitudes — spécifiques, à retourner vers des ordinateurs de type Macintosh ou PC.

2.2.2 Usages impensés

Cela peut sembler d'autant plus surprenant que de multiples logiciels mathématiques (et aussi graphiques ou dédiés au son) étaient livrés avec les NeXT. On aurait pu aisément en écrire d'autres, ou en télécharger : il aurait alors suffi d'entrer en contact avec la communauté internationale des développeurs. Mais les sociologues étaient des utilisateurs de « techniques », et ne se voyaient pas comme des concepteurs de méthodes.

Dans les faits, un protocole comme le ftp apparaissait aussi inutile que complexe[20]. L'utilisation de logiciels comme Archie ou Gopher, particulièrement bien conçus sur les NeXT, restera toujours fort rare.

Il en était de même pour les fenêtres « Terminal » permettant de transmettre des commandes ou de se connecter sur une autre machine : les informaticiens avaient évoqué la possibilité de travailler à distance, mais l'intérêt d'une telle pratique n'était pas perçu.

Même le courrier électronique ne générait qu'un intérêt amusé pour les plus motivés, réduits à une poignée de personnes malgré des formations régulières proposées aux enseignants comme aux élèves : à qui donc envoyer un *mail* en France en 1992, si ce n'est à la personne assise devant la machine voisine, quand on était chercheur ou étudiant en sciences humaines ? L'ENS littéraire était un des toutes premières institutions françaises de ce type à se doter d'un tel outil, environ cinq ans avant les autres.

est réduite. Ce fait, compréhensible, explique la difficulté à enseigner l'informatique à des universitaires patentés.

20. Même dans sa version simple, qui permettait d'effectuer des transferts de fichiers entre les Macs et les NeXT, en oubliant donc l'usage de la disquette.

De fait, quand les réseaux regorgeaient déjà de logiciels, les seuls documents intelligibles pour les chercheurs en sciences humaines étaient... des recettes de cuisine rédigées en anglais. Les groupes de *news*[21] semblaient dédiés à des professionnels de la programmation. L'apparition des bases Wais[22] en France enchanta les informaticiens, qui rendirent interrogeables, dès mai 1992, le catalogue de la bibliothèque et les *pre-prints* des chercheurs du département de mathématiques et d'informatique d'alors (DMI). Mais les chercheurs en sciences humaines ne disposaient ni des ressources informatiques, ni des pratiques éditoriales leur permettant de s'approprier de tels outils de lecture et d'écriture. Et ce n'est qu'en mai 1993 qu'ils découvrirent une utilité concrète de l'internet, avec la possibilité d'interroger des catalogues électroniques de bibliothèques[23]. Sinon, les rares audacieux qui, dès 1992, grâce à leurs compétences linguistiques, s'abonnèrent à des listes de discussion dédiées à la littérature étrangère, espérant s'affranchir du réseau trop fermé des littéraires de l'ENS (qui n'avaient quasiment aucun correspondant extérieur), s'avouèrent désespérés par la vacuité des messages qui s'emmagasinaient dans leurs boîtes à lettres : ces listes se multipliaient, mais, tributaires de leur jeunesse, restaient, de l'avis des expérimentateurs, fort naïves.

On peut penser que ces machines Unix sont venues trop tôt, et que leurs destinataires n'étaient pas préparés aux méthodes de travail spécifiques à un système d'exploitation qui privilégie les petites briques logicielles qui s'emboîtent les unes dans les autres, aux dépens des logiciels capables de tout faire. En fait, c'est la relation des chercheurs à l'informatique dans son ensemble qui était non pensée. Tout d'abord, la variété des besoins et des connaissances — même au sein d'un même département — était trop importante pour que puissent être définis des profils d'usage[24]. Aussi, le déterminant commun aux

21. Système d'échange public d'informations thématiques. En général, l'utilisateur consulte, quand il le désire, les documents publiés avec un outil logiciel approprié. Ainsi, les *news* se distinguent des listes de discussion.

22. *Wide Area Information Service*. Cf. page 129.

23. Évidemment américains. Par exemple, en faisant un « telnet melvyl.ucop.edu ».

24. Ce fait témoigne à lui seul du luxe de l'ENS : chaque enseignant, chaque élève est un

ordinateurs redevenait-il la machine à écrire. Même les personnes qui pouvaient tirer profit d'instruments banalisés comme les Macintosh se refusaient à imaginer leurs potentialités éditoriales, alors qu'elles avaient en charge la production d'ouvrages. Et nombreuses étaient les autres à s'avouer désemparées par ces ordinateurs, prétendus conviviaux et en fait trop intuitifs pour qu'une culture cohérente puisse se construire au fil de leur pratique[25].

Ainsi, c'est tout un accompagnement qu'il fallait prévoir, en tenant compte des besoins et des résistances de chacun, et plutôt que de jeter l'anathème sur tel ou tel système d'exploitation, il convient de prendre conscience que l'arrivée des machines Unix et de l'internet au sein de l'ENS littéraire eût pu être jugée suffisamment menaçante pour être niée.

2.3 Premières appropriations

Cependant, au fil du temps, les potentialités de ces machines — et leurs fonctions pédagogiques[26] — facilitèrent, au sein d'une poignée d'expérimentateurs, l'acquisition d'une culture informatique contemporaine.

2.3.1 Minitel

Le travail à distance se développa quand on apprit qu'il était possible de lire et de rédiger son courrier électronique à partir de chez soi. Si l'on disposait d'un ordinateur à domicile (situation rare), il fallait alors acquérir un modem (lent et coûteux), qu'il fallait alors configurer, ce qui pouvait prendre assez vite des dizaines d'heures d'essais infructueux avant de finir par transmettre

cas particulier, ce qui stimule beaucoup les activités de recherche — et l'achat d'ordinateurs —, mais complexifie l'approche pédagogique.

25. Au sens d'une « incorporation parfaite des contraintes de l'instrument », selon Pierre Bourdieu, qui rappelle à quel point la maîtrise d'un instrument par un scientifique nécessite « beaucoup de théorie et de routines pratiques » ([Bou01], p. 83.).

26. Liées par exemple à la possibilité d'apprendre des commandes Unix en lisant les effets d'une manipulation iconique sur une fenêtre Terminal, et bien sûr à l'accès direct à l'internet.

un SOS à un expert (par téléphone, ou en sollicitant un rendez-vous). À défaut, on pouvait utiliser le terminal vidéotexte (et son modem) disponible dans de nombreux foyers français : le Minitel. Encore devait-on en choisir un spécifique — de type 1b ou 2.

Cette pratique finit par se répandre. Et l'on constate que le Minitel n'a pas été un concurrent de l'internet. Au contraire, la possibilité de louer à coût modique des Minitels 1b (20 F par mois) a facilité à l'ENS la domestication du courrier électronique chez ceux qui n'avaient pas les moyens de s'acheter un ordinateur personnel et un modem[27].

Cependant, les exemples suivants donnent une idée de la difficulté à traduire le jargon des informaticiens, et à organiser les diverses connaissances permettant un tel usage. En effet, éprouvant moi-même de sérieuses difficultés à paramétrer ce terminal, je sollicitai une assistance technique qui, naturellement, me parvint par courrier électronique et envoyée d'un Minitel. En voici un extrait, en date du 22 janvier 1992 :

```
   Je r'ecris depuis mon Minitel, a la maison [...].
La connexion fonctionne donc pour peu qu'on ai taper
Fnct+T normal 1 !
   [...] [28]
   Voila pour ce soir. Excuse les fautes de frappes, j'ai du mal
avec certaines touches comme les fleches !...
   A bientot !...
   Olivier.
```

Ce message n'est pas très compréhensible, du fait de la difficulté à corriger les fautes de frappe sur ce type de terminal. D'où l'allusion aux flèches, qui,

27. De nombreux collègues d'outre-Atlantique prétendent que la socialisation de l'internet a été ralentie, non pas du fait de la culture technique requise, mais par les décideurs français qui voulaient favoriser le développement du Minitel (source : liste de discussion du CIRASI). Cet argument est peut-être légitime pour comprendre les usages du grand public, mais on voit ici qu'il n'est pas recevable si l'on s'intéresse au monde des chercheurs, y compris en sciences humaines.

28. Les points de suspension sont ajoutés par mes soins.

combinées avec d'autres touches du Minitel, permettaient d'effacer les caractères[29]. Par ailleurs, la combinaison de touches — essentielle pour obtenir un affichage correct[30] — n'était pas « Fnct+T 1 » mais « Fnct+T a ».

Après de multiples expériences à domicile, je rédigeai une notice à destination des possesseurs de comptes électroniques qui désiraient lire leur courrier à partir d'un Minitel ; les extraits qui suivent mettent en évidence la complexité des paramétrages à effectuer et la culture digitale spécifique, préalables à l'envoi de commandes absconses :

[...]

Allumez le minitel, puis tapez « Fnct T a » (i. e. maintenez appuyée la touche « Fnct » pendant que vous appuyez sur la touche « t », relâchez les deux touches, puis appuyez sur la touche « a ») pour obtenir un clavier ascii et un écran 80 col.

Composez le numéro de l'Ecole : 43 25 02 09.

Attendez le tîîît *de connexion, puis appuyez sur la touche « connexion/fin ». Un « C » apparaît en haut à droite de l'écran. Peu après, un « I » s'affiche à côté du « C ». La connexion est établie.*

Tapez alors sur la touche « retour » (retour chariot : et non pas ENVOI. Cette touche est sur le clavier en bas à droite).

[...]

Envoyer votre courrier

[...]

Corriger les fautes de frappe.

En cas de faute de frappe, la touche d'effacement n'est pas « Correction », ni flèche vers la gauche, mais « Ctrl <— » (touche « Ctrl » et touche « flèche vers la gauche »).

[...]

29. On réalise ici comme l'entreprise nationale France Télécom a tout fait pour, d'une part compliquer la transcription des caractères français, d'autre part, maintenir les Français dans un niveau d'usage particulièrement bas. Cf. note 27 page 55.

30. Avec un écran à 80 colonnes au lieu de 40.

Effet de l'anarchie des protocoles diffusés par les informaticiens et les industriels, on constate ici à quel point l'ensemble des savoirs à engranger avant de pouvoir s'approprier l'informatique est non structuré ; et l'on comprend alors la difficulté pour un profane à classifier de telles connaissances, ce qui ne facilite en rien leur acquisition.

2.3.2 Échanges de fichiers et programmation

Après la correspondance électronique à domicile et les catalogues de bibliothèques californiennes, un troisième avantage de la mise en réseau apparut avec les téléchargements et les interrogations de bases de données. Le protocole ftp permit notamment de réceptionner des fichiers de l'INSEE trop conséquents pour tenir sur une disquette : mais les transferts ne s'opéraient correctement qu'après de nombreuses communications téléphoniques avec un informaticien éloigné.

La structure particulière de ces données — les notions de ligne et de séparateur de champ des documents de l'INSEE ne correspondaient pas aux normes Unix — donna l'occasion de se familiariser avec les rudiments de la programmation Unix et des expressions régulières : 1993 fut l'année de la découverte de *sed* et *awk*, 1994 celle de Perl. Bien sûr, les personnes qui proposaient ces méthodes de travail sur le texte (et qui les enseignaient) étaient les mêmes informaticiens.

D'autres formes d'acquisition, comme celles résultant de l'interrogation à distance de bases de données telles que Questel et Frantext, allaient aussi inciter quelques personnes[31] à se familiariser — toujours avec l'aide des informaticiens — avec ces outils de téléchargement, de travail à distance et de nettoyage de fichiers sur des machines Unix. Il s'ensuivit une familiarisation avec la programmation qui allait générer la production de programmes informatiques dédiés à l'analyse textuelle. Du coup, certaines spécialistes de littérature pouvaient aussi tirer parti des machines Unix.

31. En fait, quatre, dont deux de l'informatique littéraire.

2.3.3 Laboratoire

Malgré les formes de rejet précédemment décrites, l'usage des NeXT se répandait : les formations — souvent personnelles — y contribuaient. Un nombre croissant de personnes profitait du courriel : quelques élèves, ayant des amis partis aux États-Unis, et des étudiants ou chercheurs invités à l'ENS.

Ainsi se construisit, dès 1993, un réseau social autour des machines reliées à l'internet. Des experts de disciplines variées venaient se retrouver autour d'elles, ce qui fut l'occasion de nombreux dialogues « médiatisés par l'ordinateur ». Non pas que les machines fussent des agents incontournables de l'échange : tout simplement parce qu'on se retrouvait devant les NeXT pour lire son mail, ce qui était l'occasion de discuter à bâtons rompus avec un voisin qui faisait de même, mais aussi de profiter de ses connaissances pour se familiariser avec les nouvelles machines et les protocoles de l'internet, ou à l'inverse, de jouer à son tour le rôle du tuteur face à une personne qui demandait de l'aide.

La présence des Macintosh voisins, avec leurs logiciels et codages distincts, aidait aussi à l'acquisition collective d'une culture informatique. On peut donc affirmer que, durant les années 1992 à 1994, les littéraires de l'ENS, par l'intermédiaire d'une vingtaine d'entre eux, ont construit un nouveau laboratoire.

Ces années semblent être celles de l'exploration et de l'expérimentation d'Unix et de l'internet. Les besoins des élèves et des chercheurs n'étaient pas explicités. L'idée essentielle était que les uns et les autres finiraient par tirer parti de leur familiarisation avec les nouvelles machines, même si elles n'intéressaient qu'un petit nombre de personnes. Par ailleurs, ces réseaux sociaux autour des machines valorisaient le département de sciences sociales qui affichait une proximité — toujours à définir : intellectuelle ? méthodologique ? fondée sur des emprunts de vocabulaire et d'ordinateurs ? — avec les sciences exactes, en même temps qu'une générosité complaisante face aux départements sous-informatisés.

2.4 Le web

Avant 1996, l'internet restait encore très lié à la culture et aux pratiques des informaticiens, et n'avait aucune raison d'intéresser, si ce n'est à la marge, les spécialistes des sciences humaines : à cause de la culture informatique requise, de la rareté des objets susceptibles d'intéresser ces chercheurs, de la difficulté à repérer de tels objets s'ils existaient.

2.4.1 Bref historique

Pour préciser ces faits trop vite oubliés, nous décrirons brièvement les divers « outils » qui ont permis la socialisation des techniques d'écriture et de lecture propres au web. Ce protocole de l'internet [32] est, comme beaucoup de solutions de mise en réseau, fondé sur une architecture client-serveur, et permet d'éditer, de lire et de manipuler des objets hypertextuels de façon étonnamment aisée [33]. Mais surtout, c'est, parmi une longue liste d'inventions, celle qui s'est imposée.

Ce point sera bref, car il existe une abondante littérature sur l'histoire du web et de ses ancêtres [34]. Il nous aidera à critiquer les analyses prétendument historiques, qui confondent souvent date d'invention et date de socialisation. Le but est aussi de rappeler comment, en quelques années, les conditions d'un dépassement de l'imprimé ont pu se réaliser : auparavant les productions multimédia étaient réalisées par un petit nombre de personnes (auteurs de textes, graphistes, informaticiens) et n'étaient pas transformables par le lecteur. L'adaptation d'un hypertexte aisé d'emploi à un réseau mondial de machines, et sa rapide appropriation par un grand nombre d'individus, ont été les deux raisons principales d'une transformation massive des techniques de l'écriture, et de la prise de conscience de cette transformation.

32. http, c'est-à-dire hypertext transfer protocol.

33. Cf. le paragraphe 3.3.4 de la partie I et notamment la page 74.

34. Tant sous forme imprimée que sur le web lui-même. Les lecteurs désireux de comprendre le fonctionnement des protocoles et logiciels évoqués ici pourront obtenir des informations dans l'ouvrage [And95].

On a vu que, jusqu'au début des années 1990, les protocoles les plus connus (telnet, ftp, mail, news) comme le contenu des échanges qu'ils permettaient restaient réservés au monde des informaticiens. Il s'ensuit que le développement des protocoles et des outils qui ont fait le succès de l'internet actuel est somme toute assez récent.

Le premier logiciel permettant de rechercher par mot-clé[35] des fichiers situés sur des machines distantes en interrogeant un ordinateur lui aussi lointain — et sur lequel on ne disposait pas de compte — fut Archie, en 1990. Développé à l'université McGill de Montréal et basé sur le protocole ftp, ce serveur de bases de données était naturellement complété par des logiciels clients[36].

En 1991 paraît Gopher, qui préfigure nos navigations actuelles sur les réseaux : sorte de web en ASCII, avec déjà des pointeurs qui relient des documents divers entre eux. Très vite, sur les NeXT puis sur d'autres machines, une version graphique du client Gopher sera implémentée, et les quelques personnes qui le découvrirent à l'ENS littéraire s'en servirent jusqu'en 1996. Gopher, « conçu par l'université du Minnesota pour bâtir un service d'information de campus, s'est rapidement imposé sur tout l'internet dans des environnements très divers, jusqu'à devenir le produit le plus populaire aux États-Unis ». Veronica, le moteur de recherche qui lui est dédié[37], apparaît en 1992.

En 1991 aussi, le protocole « Z 39.50 » introduit les sites et bases « Wais » (Wide Area Information Service), qui permettent alors d'effectuer des indexations, et donc des requêtes, sur le contenu propre des documents[38]. L'innovation est certaine, mais si les utilisateurs y trouvaient leur compte, ceux d'entre eux qui savaient construire et rendre publique une telle base étaient en nombre réduit.

35. En fait par leur nom.
36. Autrement dit des systèmes d'interrogation destinés à l'utilisateur.
37. Tout comme Archie indexe les sites ftp, Veronica indexe les sites Gopher par nom de fichier.
38. Passage à l'interrogation dite « full-text », ou « plein texte ».

L'outil qui allait populariser l'internet est le « www ». Il est décrit sommairement dans une note interne au CERN[39] par Robert Cailliau et Tim Berners-Lee en mars 1989. En octobre 1990, ce dernier développe son premier programme « WordWideWeb » (sur une machine NeXT) et le premier logiciel client (appelé souvent navigateur, ou encore *browser*, fureteur, butineur...) est disponible en mars 1991 pour certaines machines Unix. La version stabilisée (numéro 1.1) apparaît en janvier 1992, mais elle reste en « mode ligne » (ASCII, à la manière du logiciel Lynx), donc encore peu commode. Le 3 novembre 1992, 26 serveurs expérimentent ce nouveau protocole ; on en compte 50 en janvier 1993. Et si, dans l'historique du W3C[40], il est dit qu'« en juin 1994, ce nombre dépassait 1500 machines »[41], les responsables de tels serveurs apparaissent rétrospectivement comme une minuscule minorité de pionniers à cette date pourtant récente.

Il suffisait alors que le CERN abandonne ses droits sur l'invention de Tim Berners-Lee et laisse en 1994 l'INRIA et le W3C améliorer ce protocole d'écriture dans un esprit de « service public » pour que le web puisse se développer.

2.4.2 Lire, écrire, trouver

Le premier outil de lecture associé — le navigateur — un tant soit peu agréable, c'est-à-dire permettant une mise en forme du texte et des images, apparaît en février 1993 : Mosaic, développé au NCSA[42]. Cette date est aussi importante pour le développement du web, car il devenait possible de profi-

39. Organisation européenne pour la recherche nucléaire, voir l'URL http://cern.web.cern.ch/CERN/CERNName.html pour l'origine de l'acronyme.

40. World Wide Web Consortium, fondé en juillet 1994 sous la double tutelle de l'INRIA et du CERN. http://www.W3.org

41. Informations disponibles à l'URL http://www.W3.org/History.html
Voir aussi http://www.let.leidenuniv.nl/history/ivh/frame_theorie.html

42. *National Center for Supercomputing Applications* de l'Université d'Illinois à Urbana-Champaign ; Marc Andreessen, le fondateur de Mosaic quitte avec ses collègues le NCSA en mars 1994 pour fonder *Mosaic Communications Corporation* ; Mosaic deviendra plus tard le fameux Netscape.

ter des innovations apportées par le nouveau protocole : lire de façon confortable des pages entières, dotées d'un enrichissement typographique[43] et intégrant des images, passer simplement d'une page à l'autre (grâce au codage « hypertexte »), etc. Ce processus de lecture s'étendait à la consultation du code informatique du texte[44] — en clair —, dont on découvrait alors la simplicité, tant sur le plan de la structure des documents que de leur mise en forme : la juxtaposition des deux formes du texte (texte destiné au lecteur et texte destiné à l'éditeur), et la simplicité du balisage de la seconde permettait d'apprendre à réaliser une page par le simple fait d'en lire le code source. Ce choix, politique et pédagogique, allait inciter de multiples utilisateurs, souvent étudiants des écoles d'ingénieurs, à demander à leur ingénieur système d'installer (ou de les laisser installer) le logiciel serveur qui leur permettrait d'accéder au statut d'éditeur électronique.

Mais il manquait encore un système de repérage. Pour les uns, c'est l'apparition de l'annuaire Yahoo ![45] en 1994 qui allait faciliter la recherche d'information sur le web ; mais Yahoo ! ne faisait que recenser des pages web qui étaient ensuite regroupées par de nombreux employés suivant des rubriques prédéfinies et trop généralistes. Pour d'autres, c'est celle du moteur de recherche Alta Vista[46] en novembre 1995, qui est déterminante : on pouvait effectuer des requêtes complexes sur une machine[47] et obtenir en retour les adresses des pages web qu'un algorithme avait jugées les plus pertinentes.

En résumé, le protocole http offrait une facilité d'écriture déconcertante ; les profanes de l'informatique pouvaient apprendre à réaliser une page en hypertexte en 20 minutes, inclusion d'images comprise. Chaque « page » dis-

43. Même si les balises html, dérivées du sgml, étaient plus conçues pour structurer les documents que pour les mettre en forme.

44. En demandant à visualiser le « code source » des documents consultés.

45. http://www.yahoo.com

46. Actuellement http://www.altavista.com

47. Le terme « machine » est ici inadéquat, car des dizaines d'ordinateurs étaient sollicités pour explorer les pages web existantes, les indexer, et répondre aux utilisateurs qui, quelques mois après la diffusion d'Alta Vista, y déposaient deux millions de requêtes par jour.

posait d'une adresse simple et était rédigée dans un format de fichier universel (indépendant du système d'exploitation) et domaine public (donc gratuit). La lecture des documents (et de leurs sources) était rendue agréable et la « navigation » aisée. Enfin, les systèmes d'indexation permettaient d'effectuer des recherches documentaires sophistiquées.

Ainsi, la date de conception du web date-t-elle de 1989, celle de sa première implémentation de 1991. En 1994, l'invention a convaincu quelques milliers de professionnels, qui expérimentent cet outil éditorial et qui l'améliorent, par exemple en le dotant de logiciels de lecture. Mais il faut attendre 1996 pour disposer d'un outil de recherche qui lui soit adapté, et donc en imaginer un usage qui dépasse celui des professionnels de l'informatique.

On est donc dans la situation d'une « invention qui a réussi »[48], pensée en 1989, qui ne pouvait être socialisée avant 1996. Il n'y aurait pas d'intérêt à préciser ces évidences si la tendance générale des « historiens » de l'informatique n'incitait pas à confondre ces dates, obéissant par là aux exigences du poème épique, parant l'inventeur des vêtement du héros, toujours rattaché à une nation précise[49], et sollicitant la rhétorique la plus simpliste du déterminisme technique pour culpabiliser le lecteur, qui se sent à la fois « en retard » et à l'écart d'un monde dynamique.

2.4.3 Web, écriture, et revues savantes

Ensuite, ces nouvelles techniques de l'écrit avaient, dans le domaine des sciences humaines, à s'enraciner dans des pratiques sociales avant de pouvoir se déployer.

Tout d'abord en termes de contenus. Encore aujourd'hui, la discipline la mieux servie par l'internet reste l'informatique : une requête du type « ps2pcl »

48. Pour reprendre l'expression de D. Edgerton, qui rappelle que la majorité des inventions ne « réussissent » pas.

49. Par exemple, les États-Unis se sont appropriés le web, qui reste néanmoins une invention très européenne.

adressée à un moteur de recherche comme www.google.com renvoie immédiatement les « scripts » permettant de faire imprimer des fichiers PostScript par une imprimante qui ne l'est pas. De plus, on peut être assuré que les 39 réponses à cette requête proposent toutes le même programme, garanti par l'expérience de centaines d'informaticiens. Ainsi, l'existence de nombreuses pages dédiées à l'informatique sur le web et la spécificité des mots clés de cette discipline facilitent la tâche de l'utilisateur. Une requête propre aux sciences humaines doit souvent être beaucoup plus élaborée pour donner aussi vite un résultat aussi précis et réduit : la requête « Jack Goody » renvoie 21 700 réponses.

Par ailleurs, même si tous les renvois à cette requête étaient de la même qualité, son auteur ne leur accorderait pas nécessairement une grande valeur, car les normes d'usage du web sont plus récentes, moins stabilisées que dans le domaine de l'informatique : la confiance est aussi affaire de contruction sociale.

Enfin, les pratiques d'écriture propres aux sciences exactes facilitent l'usage de différents codages, y compris l'html : quand les chercheurs en sciences humaines avaient adopté les *wysiwyg*, la majorité des auteurs des sciences exactes utilisaient la combinaison LaTeX-PostScript, qui distingue clairement le code du texte de sa forme[50]. Il leur était donc aisé d'assimiler un nouveau codage comme l'html, fonctionnant lui-aussi sur le principe des balises. Certes, des chercheurs en sciences humaines utilis(ai)ent LaTeX et des machines Unix. Mais ils constituent une minorité en France[51].

En conclusion, les techniques mises en place par les informaticiens ne pouvaient être appropriées rapidement par l'ensemble des chercheurs et des professionnels : les publications du web couvraient surtout l'informatique et

50. Voir les paragraphes 3.2.4 et 3.2.6 de la partie I pages 61 et 64.

51. Il faut noter ici une constante : au sujet des formes de l'écrit, comme plus largement au sujet de l'appropriation technique, l'évolutionnisme n'est pas opérant ; si chez les informaticiens, le cheminement usuel fut le passage « ASCII, LaTeX, html », dans le domaine des sciences humaines, c'est souvent la découverte d'un html simple qui a incité les auteurs à expérimenter, après, LaTeX.

les domaines proches, comme la physique et les mathématiques, et ce n'est que dans ces disciplines que des mécanismes de confiance se mettaient en place[52]. Les outils de recherche d'information favorisaient naturellement les disciplines qui proposaient des contenus. Sur le plan de l'écriture, les personnes qui subissaient la dictature des formats de fichiers hétérogènes ne comprenaient pas l'intérêt d'un codage supplémentaire (fût-il aisé à manipuler) et, dans les sciences humaines, l'organisation sociale et politique de la publication imprimée n'avait pas encore prouvé ses limites comme dans les disciplines précédemment évoquées.

En effet, dans les sciences exactes, des chercheurs renommés prenaient conscience du fait que les éditeurs de revues savantes ne tenaient plus leur rôle : les articles étaient rédigés et mis en page par les chercheurs. Leur évaluation était effectuée par d'autres chercheurs. Ainsi, ils ne comprenaient pas pourquoi leurs laboratoires devaient payer très cher des revues savantes, alors que le travail des éditeurs se limitait à l'impression et la diffusion. Par exemple, en 2000, la bibliothèque des DMA et DI[53] de l'ENS payait plus de 900 000 F pour ses abonnements à 166 revues, dont 30 000 F pour la plus coûteuse, qui paraît moins de 12 fois par an[54]. L'arrivée du web mettait crûment à jour ces problèmes : les articles étant souvent prépubliés sur l'internet avant leur acceptation par une revue, l'activité de l'éditeur consistait en fait à faire payer très cher une validation sociale, qui survenait, comme le document imprimé, bien tard (souvent un an après la soumission de l'article).

52. L'exemple le plus probant reste le site http://xxx.lanl.gov, qui est un serveur d'articles dans le domaine de la physique, et financé par la National Science Fondation (NSF) depuis 1991.

53. Département de mathématiques et applications et département d'informatique, fruits de la scission du DMI.

54. Voir le site http://www.publiclibraryofscience.org pour une évaluation actualisée de la contestation universitaire internationale à ce sujet : plus de 25 000 personnes ont déjà signé une pétition demandant que les articles scientifiques soient gratuitement mis à la disposition des chercheurs six mois après leur parution et s'engageant à ne publier que dans les revues acceptant cet état de fait. Voir aussi l'article du Monde daté du 20 avril 2001 (Pierre Le Hir) à ce sujet.

On comprend que l'idée d'une « bibliothèque mondiale décentralisée », qui permettrait d'accéder de façon instantanée à l'ensemble des connaissances, fût déjà établie. Elle n'était pas une utopie, mais correspondait à une réalité vécue par les chercheurs. Les personnes les plus engagées évoquaient déjà un « monde meilleur », qui verrait la disparition des monopoles de l'écrit : chez les éditeurs de revues scientifiques comme chez les éditeurs de logiciels. Elles imaginaient l'avènement des « logiciels libres »[55], en prenant comme exemple l'internet lui-même : quand de nombreuses entreprises avaient essayé de développer des protocoles réseaux « propriétaires », le seul qui s'imposa fut celui construit par des scientifiques enthousiastes et capables d'engager des travaux collaboratifs.

Ces enjeux éditoriaux du web n'étaient pas perçus au sein des sciences humaines. La majorité des membres de l'ENS littéraire ne disposaient pas d'accès à l'internet, et le groupe de pionniers précédemment décrit subissait les effets de la complexité informatique plus qu'il ne pouvait en profiter. Néanmoins, la présence de plusieurs élèves et chercheurs informaticiens créa les conditions minimales d'un débat, souvent informel au commencement, sur ces enjeux : alors que la presse française était étonnamment silencieuse[56], divers groupes de l'ENS, qui finiront par tous se connaître, entamaient — parfois de façon malhabile — une réflexion sur le monopole de Microsoft sur l'écrit, sur l'organisation des bibliothèques, sur l'édition, sur la façon dont l'informatique et l'internet transformaient les méthodes de travail des chercheurs.

55. Un logiciel libre n'est pas nécessairement un logiciel gratuit. C'est un logiciel dont le code source est livré en même temps que l'outil. Deux avantages à cela : il est possible de comprendre le fonctionnement de l'application, ce qui est très important en matière d'éducation et de sécurité ; ensuite, on peut l'améliorer, l'adapter, le détourner. Même quand le logiciel est « libre » et gratuit, ses auteurs peuvent en tirer des ressources financières avec la vente du manuel d'utilisation et avec les conseils, installations etc. LaTeX, Perl, Pine, Apache et Linux sont les exemples les plus connus de logiciels libres (voir la première partie). Netscape, qui n'est pas un logiciel gratuit pour les entreprises, est un logiciel libre depuis 1998.

56. Sinon carrément imbécile, cf. le chapitre 2 de l'annexe (page 397).

Chapitre 3

Essor de la publication électronique

L'internet est donc apparu à l'ENS littéraire en 1992 sous l'impulsion des informaticiens. Son lien étroit avec les machines Unix, la rapide évolution de ses protocoles et de ses potentialités entre 1992 et 1995, ont fait que le nombre d'enseignants ou d'élèves littéraires qui se familiarisaient avec ce système technique ne pouvait être que très réduit, malgré la participation d'étudiants étrangers. Par ailleurs, les chercheurs en sciences exactes, très impliqués à toutes les étapes de la pratique éditoriale, mettaient en évidence les limites d'un système monopolistique de diffusion des revues savantes, et en préfiguraient le contournement, notamment par le biais de la publication électronique. La construction d'un réseau social autour des machines Unix permettait la diffusion de tels débats, notamment auprès des littéraires « téméraires » qui s'impliquaient dans l'informatique et qui donc, participaient à ce réseau. Pour ces derniers, l'ordinateur n'était plus une machine à écrire.

3.1 Codes de l'édition savante

Avant de décrire les pôles du développement du web au sein de l'ENS littéraire, il convient de souligner un amusant paradoxe : quelle que soit leur discipline, les personnes qui s'engageaient dans une réflexion sur les pratiques éditoriales universitaires disposaient des ressources pour impulser un travail de publication — ou savaient où les trouver. Mais, dans le domaine des sciences humaines, la situation semblait verrouillée car aucune n'avait la légitimité lui permettant de réaliser un tel travail dans un cadre scientifique, ou même d'en faire état [1]. En effet, les personnes étiquettés sciences humaines étaient élèves, documentalistes, sinon plus enseignants que chercheurs et n'avaient pas accès à la publication scientifique. Quant aux informaticiens, on attendait d'eux qu'ils publient dans les revues de leur discipline et non qu'ils s'approprient des débats qui *a priori* ne les concernaient pas.

À l'opposé, les professeurs patentés de l'École littéraire affichaient sans complexe leur pouvoir, en narrant pas le menu les affres de la création quand ils travaillaient à un livre, puis en se « faisant voir » dans les journaux, radios, voire télévisions une fois l'ouvrage publié. En revanche, l'idée d'une publication sur le web ne les effleurait pas, les bénéfices leur paraissant nuls. Bien sûr, ils mettaient en avant l'absence de validité scientifique de tels projets pour éviter d'y participer.

Cette notion de « légitimité », jamais explicitée, tout comme l'absence d'invitations à publier, pouvaient donner l'impression que l'édition, même savante, avait dans le domaine des sciences humaines des codes secrets qui ne pouvaient être dévoilés qu'au terme d'une initiation particulièrement complexe, voire douloureuse. Le pouvoir de l'imprimé semblait aussi inaccessible qu'occulte [2].

1. On peut nuancer cette assertion en évoquant les travaux de l'institut des textes et manuscrits modernes (ITEM). Mais ce laboratoire du CNRS était assez mal vu des enseignants de l'ENS littéraire, au point que certains voulaient le voir quitter la rue d'Ulm.
2. [Bou84], chap. 1.

3.2 Des informaticiens indispensables

On a vu que l'ENS avait ses partisans de l'internet avant 1988, et que dès 1994, l'internet apparaissait comme un espace éthique de production des savoirs : il devenait possible de publier de nombreux documents, de la fiche pédagogique à l'ouvrage érudit, en passant par les sources archivistes et bibliographiques nécessaires à la compréhension du dernier, sans pour autant se « commettre » à acheter un logiciel payant, au format propriétaire, sans cesse changeant et souvent mal conçu, et sans devoir faire appel à un imprimeur ou à un diffuseur. Cette prise de conscience apparut chez quelques élèves, d'une part, qui fondèrent le premier serveur de l'ENS en mai 1994[3], mais aussi chez des informaticiens[4]. Au moins deux d'entre eux se sont toujours montrés très disponibles vis-à-vis des personnes qui leur demandaient conseil en matière d'édition. Nous appellerons « passeurs » ces deux personnes, qui ont été grandement sollicitées.

En effet, la simplicité de l'écriture html masque deux pièges : tout d'abord, l'installation et la maintenance d'un serveur web requièrent l'assistance d'un expert. Si ce dernier n'est pas très compétent, il verrouillera des passerelles d'usages pour limiter les risques ; ce qui se produit hélas trop souvent au sein des universités françaises[5]. L'ENS n'a jamais souffert de telles contraintes.

Ensuite, la mise à disposition de documents de grande taille (bibliographies, bases de données, etc.) donne envie à leur auteur-éditeur de les rendre interrogeables pour accroître le confort de ses lecteurs, surtout s'il est sensible au fait que ces derniers disposent parfois d'une connexion à faible débit. La construction de telles pages dynamiques[6] renvoie l'auteur aux logiques de la programmation : il lui faut un minimum de connaissances en matière de « droits Unix » et de programmation des « scripts » d'interrogation[7].

3. http://www.eleves.ens.fr

4. Un « informaticien » est ici une personne qui a une thèse en informatique.

5. Cf. page 85.

6. Appelées ainsi parce qu'elles ne résident pas sur le disque dur du serveur, mais parce qu'elles sont créées à la volée par le logiciel d'interrogation.

7. Par exemple de type *cgi* (*common interface gateway*).

Une fois de plus, la culture informatique reprend ses droits. Or, quel que fût le type de problème qu'on leur soumettait, les passeurs ont toujours proposé des solutions techniques robustes, simples, élégantes, gratuites, en orientant leurs collègues et étudiants vers des logiciels du « domaine public » qui répondaient exactement à la question, en diffusant la culture « petites briques » d'Unix ; la proximité de ces experts était essentielle pour rendre autonomes les élèves et chercheurs non informaticiens dans le domaine de l'édition électronique.

Au-delà de l'internet, ces conseillers ont donc permis le développement d'un savoir-faire en matière de manipulation du texte : à partir des questions relatives à ces formes élaborées de la publication sur le web, des recherches typiquement littéraires ont pu se développer, et il apparut qu'en sciences humaines, la majorité des activités informatiques relèvent de l'analyse textuelle : il s'agit souvent de traiter des chaînes de caractères, de les remanier, bref de considérer des mots ou des expressions comme des éléments d'une base de donnée qu'il faut transformer, présenter, agréger. Le traitement d'une enquête statistique, la production d'une carte, le résultat d'une interrogation bibliographique, une recherche d'occurrences dans un corpus relèvent en fait de la même méthode. Ainsi l'apprentissage de la programmation sur le web développa de façon décisive les études lexicométriques à l'ENS littéraire [8].

Enfin, la compétence des passeurs en matière de construction de logiciels d'écriture — au sens large, incluant tous les protocoles de l'internet et leurs variantes — leur offrait une étonnante « intuition sociologique » : ils replaçaient dans un contexte simple et explicite tous les mythes et discours que la presse commençait à répandre à cette période et savaient en analyser les enjeux économiques comme politiques avec une grande finesse [9].

Le profil de ces personnes est suffisamment atypique pour qu'il soit rappelé ici : elles ont beaucoup contribué à la socialisation de l'informatique et

8. Tout comme le développement de l'internet allait, avec les moteurs de recherche, inciter à des développements analogues ; mais bien sûr à une échelle sans commune mesure.

9. Preuve supplémentaire que l'écriture est une technique et qu'il est toujours bon d'écouter les techniciens.

à l'émergence de débats de qualité à son sujet. Or, pendant longtemps, les informaticiens ont été vus, à tort ou à raison, comme des personnes avec qui le dialogue était difficile, voire impossible.

Cette conjonction favorable d'informaticiens compétents et rompus depuis longtemps aux pratiques de l'écriture informatique, ouverts aux problématiques sociales et attachés à un fonctionnement optimal de la recherche, allait pouvoir se matérialiser en un espace de pratiques et de débats grâce à une demande extérieure à l'ENS.

3.3 L'Atelier Internet

En 1994, une chargée de mission à la DISTB [10] pressentait que l'internet allait transformer notoirement les pratiques des chercheurs. Recherchant une équipe qui puisse réaliser une étude approfondie dans ce sens, elle s'adressa au département de mathématiques et d'informatique, qui, peu intéressé, transmit le projet au département de sciences sociales. Celui-ci ne manifesta pas plus de curiosité pour une telle recherche. Par un hasard fortuit, je vis circuler cette proposition, qui me semblait tout à fait au carrefour des activités de la cellule informatique littéraire, des préoccupations des informaticiens et des compétences des chercheurs en sciences humaines, et décidais d'y répondre.

3.3.1 Mise en place

La demande du ministère portait sur l'accessibilité de l'information scientifique, et donc sur les transformations qui risquaient d'apparaître dans l'organisation du travail chez les bibliothécaires et les documentalistes. Mais elle proposait aussi d'étudier si des nouvelles formes éditoriales allaient infléchir les logiques de la compétition académique, voire celles de l'enseignement, et en quoi les échanges quasi-synchrones comme le courrier électronique dynami-

10. Direction de l'Information Scientifique et Technique et des Bibliothèques, dépendant du ministère de l'Éducation Nationale, de l'Enseignement Supérieur et de la Recherche d'alors.

saient les liens entre chercheurs ou aidaient à la construction de « laboratoires hors les murs ». La formation de linguiste et les convictions européennes de la chargée de mission apparurent clairement quand elle insista pour que l'équipe qu'elle recherchait s'intéressât à la façon dont les informaticiens s'organisaient pour défendre les langues minoritaires.

La réponse positive du Ministère à la proposition élaborée par le groupe que j'avais réuni permit de fonder un séminaire interdisciplinaire [11], l'« Atelier Internet ». A part un anthropologue, les membres fondateurs de l'équipe appartenaient au DMI, au SPI ou à la cellule informatique littéraire [12]. La majorité des partisans de ce projet, dont le caractère sociologique était explicitement formulé par le demandeur, étaient donc... des informaticiens. Au fil du temps, d'autres « littéraires » que ceux de la cellule informatique représentés dès le départ, et des scientifiques non informaticiens, élèves, jeunes anciens élèves ou chercheurs, participèrent aux séances bi-mensuelles de l'« Atelier Internet » qui devint assez rapidement un espace de débats interdisciplinaire où de fortes personnalités de l'ENS et de l'extérieur se rencontraient.

Ce séminaire était à ses débuts assez éloigné des logiques traditionnelles de la recherche, notamment en matière de publications et de participation à des colloques. Il était expressément admis que le groupe infléchirait à son

11. Voici le résumé du projet initial, dont le titre était « Les chercheurs et Internet ; analyse des discours et des pratiques » :

« Avec les réseaux électroniques, les chercheurs disposent de nouveaux modes d'accès à l'information : échanges instantanés, accès à des banques de données organisées, présentation de leurs travaux, mais aussi fédération de nouvelles équipes autour de projets associatifs. Nous comptons évaluer les incidences de cette mutation sur les formes de la recherche, analyser les enjeux qu'elle représente pour les chercheurs et les stratégies qu'elle leur permet de développer, vérifier si la présentation d'un centre de recherche sur les réseaux a réellement pour effet d'accroître sa visibilité internationale. »

12. Rétrospectivement, ce programme de travail a représenté une opportunité exceptionnelle pour les membres de la cellule informatique : ils étaient incités à entamer des recherches correspondant à leurs compétences, alors que jusqu'alors, ils n'avaient qu'un statut de techniciens. Cette légitimité était appuyée par la direction d'alors, puisqu'elle permit l'acquisition d'un budget de recherche interne, en sus de celui du ministère.

gré les objectifs de recherche, en essayant de reproduire au mieux l'esprit de convivialité et la liberté de ton qui semblaient être typiques du monde des développeurs de l'internet à travers le monde. Ces échanges réguliers assez informels entre philosophes, mathématiciens, informaticiens, physiciens, linguistes, géographes, etc. ont aidé les participants à aborder en toute franchise des questions qu'ils n'auraient *a priori* jamais osé poser devant des personnes qui n'étaient pas de leur milieu : les non-informaticiens ayant en général l'impression que leurs questions étaient trop banales pour oser déranger les informaticiens, et ces derniers n'osant en général pas aborder devant leurs collègues des sciences humaines les « questions des société » qui les rendaient soucieux.

Déjà, en 1995, trop d'usurpateurs commençaient à publier dans la presse de longs articles évoquant l'importance des « nouvelles technologies de l'information et de la communication » pour l'avenir scientifique, culturel et économique de la France et du monde, alors qu'il était notoire que ces « visionnaires » ne savaient pas se servir du courrier électronique. Pour éviter donc que l'analyse fût dissociée de l'expérience technique, fut installé en novembre 1995 un serveur web[13], sur lequel étaient publiés les comptes rendus des séances de l'« Atelier Internet », qui étaient toutes enregistrées. L'équipe s'offrait ainsi un laboratoire pour expérimenter l'édition numérique et sa réception, et pour se donner les moyens d'étudier l'évolution des pratiques de ses membres[14] : les premiers comptes rendus datent de février 1996, et les premiers articles — validés par l'équipe elle-même —, de novembre 1996. En effet, la commanditaire avait insisté pour obtenir un rapport intermédiaire (imprimé), et il était tentant de tester les effets de la publication électronique. Ainsi, dès la fin 1996, outre les comptes rendus, cinq articles étaient disponibles sur le serveur. Ce qui offrit à l'équipe une certaine notoriété ; la dynamique instituée incita à rencontrer d'autres spécialistes[15], qui, par le

13. Initialement sur la machine elias.ens.fr, aujourd'hui sur la machine barthes.ens.fr

14. Évidemment, toutes les invitations à participer à la prochaine séance étaient envoyées par courrier électronique.

15. Principalement du CNRS et des ministères.

biais de colloques ou de réunions, commandèrent aux membres de l'Atelier Internet d'autres articles pour des revues savantes [16].

Ainsi, à partir de rencontres informelles et d'une expérimentation concrète des pratiques d'écriture, se construisit très vite une légitimité savante. Insensiblement, ces travaux, qui mettaient les participants à l'Atelier Internet en prise directe avec les pratiques éditoriales — responsabilité en matière de publication, mise en place d'un comité scientifique, production — les incitaient à s'interroger sur les codes implicites de la publication universitaire. Ce qui, petit à petit, leur fit aborder des sujets tabous, comme la fonction sociale de l'écrit dans l'université.

L'Atelier Internet ne motivait pas réellement la communauté des sciences humaines de l'ENS : si le nombre de participants a parfois dépassé les quarante, il se situait le plus fréquemment autour de la dizaine, experts en sciences exactes inclus. Mais il facilita le contact avec les élèves les plus impliqués dans le développement de l'internet, et permit aux informaticiens et aux littéraires de l'ENS d'engager un dialogue concret et approfondi, facilita les rencontres avec les autres chercheurs et praticiens qui avaient des préoccupations analogues.

3.3.2 Effets sur la recherche

Au regard de la multiplicité et de la richesse des sites web français en 2002, les productions de l'Atelier Internet — et du réseau qu'il a généré — à son origine peuvent faire sourire. Cependant, on peut en citer quelques-unes, dans la mesure où elles ont eu une incidence — à court ou à moyen terme — non négligeable sur l'organisation de la recherche en sciences humaines.

– *La documentation et l'avenir des bibliothèques.* La documentaliste de la cellule informatique littéraire a vite proposé plusieurs articles, précis et clairs, sur la relation entre la recherche et la documentation [17].

16. Pour les comptes rendus, cf. http://barthes.ens.fr/atelier/debats-et-CR
Pour le rapport intermédiaire, cf. http://barthes.ens.fr/atelier/articles
17. Cf. ici encore le site http://barthes.ens.fr

Sa spécialisation en histoire médiévale lui permettait par ailleurs de proposer une réflexion élaborée sur les archives, leur accessibilité, leurs modes de classement. Sa connaissances des bases de données électroniques d'alors, son intérêt pour les sites web proposant de réels contenus et pour les listes de discussion professionnelles, contribuèrent à enrichir les débats au sein du séminaire et, plus généralement, dans l'Université française[18].

– *Les sciences de l'antiquité.* Un participant, spécialiste d'études anciennes, avait réalisé en novembre 1996 un annuaire quasi-exhaustif des outils informatiques destinés aux antiquisants, incluant plus de 100 références de sites web, *newsgroups*, listes de discussion, revues électroniques, catalogues de bibliothèques. Il était animé d'un franc-parler, quand il s'agissait d'expliquer les « réticences et attirances des antiquisants face à l'informatique » ou la « marginalisation structurelle des littéraires »[19]. Son document électronique fut en quelque sorte le *best-seller* de cette série de publications puisqu'il avait été consulté plus de 250 fois un an après sa publication ; et plusieurs pages web le référençaient très élogieusement[20].

  Cet auteur décrivait la panoplie des outils indispensables au chercheur, montrait qu'une grande partie d'entre eux est disponible sur l'internet, et son annuaire raisonné projetait ces activités de recherche dans un cadre non plus parisien ou français, mais proprement mondial.

18. Les nombreuses publications qu'on lui a demandées depuis témoignent de la reconnaissance qu'elle a enfin obtenue, alors même qu'à l'ENS, elle était toujours considérée comme une assistante.

19. Toujours disponible : http://barthes.ens.fr/atelier/articles/ArticleInternetnov96.html

20. Et bien qu'il soit maintenant « daté », il a été consulté 875 fois entre octobre 1999 et juillet 2000. Pour se faire une idée du type d'éloges au sujet de cet article, on peut consulter, parmi de nombreux exemples, les URL
http://callimac.vjf.cnrs.fr/Ant_class.html,
http://perso.club-internet.fr/octavia0/classiques.htm ou
http://www.unine.ch/antic/ressources.html

- *La géographie.* Les collègues géographes [21] ont été très vite été sensibles aux potentiels de l'internet : les bases de données et les fonds de cartes disponibles, la possibilité de publier à moindres frais des cartes en couleur étaient pour eux des arguments déterminants [22]. Plusieurs d'entre eux participaient à l'Atelier Internet. En mars 1997 fut décidé de compléter l'atelier de cartographie [23] d'un site web [24], sur la même machine que celle qui hébergeait l'Atelier Internet. Un an plus tard, étaient expérimentées les nouvelles possibilités d'écriture du web (cartes clickables, publication d'atlas, puis de logiciels de cartographie en ligne), tandis qu'une réflexion sur les « territoires de l'internet » se mettait en place [25]. Cette expérience de publication a facilité l'explicitation de la relation entre texte, système de signes et graphiques.
- *L'histoire.* En septembre 1997, suite à la publication d'un ouvrage collectif relatif à l'histoire sociale de l'immigration [26], je fondais avec deux collègues un site dédié à ce thème [27]. Ce site, sur lequel fut publié tout l'éventail des outils nécessaires aux chercheurs [28], devint vite une référence pour la communauté des historiens. Son lectorat doubla chaque année pour attendre 400 consultations par mois en 1998, et plus de 500 par jour en 2001. Le succès de cette entreprise, comparable à celle de

21. Parmi lesquels un élève de l'ENS avait fondé le premier serveur de l'École littéraire à la fin de 1994.
22. Sans compter l'éparpillement du réseau de chercheurs, souvent en longues missions lointaines.
23. Fondé en 1992 par l'auteur de ces lignes et par Hervé Théry.
24. Actuellement : http://barthes.ens.fr/carto
25. Cf. http://barthes.ens.fr/atelier/articles/desbois-mai-98.html
26. [GN97].
27. Les objectifs d'alors étaient les suivants : « impulser une réflexion collective sur l'outil internet lui-même [et] faire en sorte que ce site puisse être un moyen de renforcer la solidarité entre les chercheurs en devenant un instrument de travail au service de tous ceux qui le souhaitent ». http://barthes.ens.fr/clio
28. Bibliographies thématiques, sommaires de revues, textes et archives, bases de données, cartes, comptes rendus de séminaires, articles, systèmes d'interrogation, et de production automatique de résultats à partir de sources primaires.

> l'Atelier Internet, offrit donc — avec le serveur de cartographie — un troisième laboratoire pour étudier *in vivo* l'évolution des pratiques professionnelles des chercheurs avec le développement de l'internet. Très vite, ses participants acquérirent une compréhension très précise de la différence entre une revue savante électronique et son équivalent imprimé, ainsi que de l'économie de la recherche universitaire.

Bien sûr, les animateurs de ces projets étaient souvent les mêmes. Mais ces travaux, sur l'internet lui-même, ou liés à des disciplines déjà institutionnalisées, eurent plusieurs effets :

— Ils incitèrent leurs participants à expliciter les processus en œuvre à l'occasion de leurs recherches. Et notamment, à mettre en évidence les relations entre recherche « pure », récolte, agrégation et présentation des sources primaires ou secondaires (avec leur pendant de mise en liste), et enfin instruments d'écriture (l'ordinateur et ses codages).

— Ils leur firent prendre conscience des effets de mise en réseau dans le monde de la recherche, même si à cette période, l'incrédulité des collègues prévalait.

— Et, en conséquence des deux points précédents, ils les aidèrent à imaginer puis à définir de nouvelles pistes de recherche.

De façon plus générale, les échanges et les témoignages au sein de l'Atelier Internet ont instauré un débat collectif qui incitait les participants, quels qu'ils soient, à se familiariser avec les protocoles de l'internet. La mise en commun des expériences de consultation, d'abonnement et de production a donné aux participants des idées pour amorcer d'autres réflexions collectives. Et l'exploitation concrète du potentiel éditorial du web par des spécialistes de toutes disciplines a permis d'aborder sans naïveté l'ensemble des questions propres à l'histoire de l'écriture et de ses supports, à l'organisation des connaissances, et aux modes de fonctionnement de la recherche.

3.4 Les élèves

Quelques centres de recherche du CNRS rattachés à l'ENS, comme le laboratoire d'archéologie, l'ITEM[29], le Delta[30] et le Centre d'études des religions du livre[31] expérimentaient eux-aussi de façon concrète l'incidence de l'internet sur les pratiques de recherche et sur les formes éditoriales. Mais ces laboratoires n'avaient que peu de contacts avec les élèves ou les enseignants. En revanche, les sections ou départements qui fédéraient les grandes masses des agrégatifs de l'ENS (histoire, philosophie, littérature et langages[32], centre d'études anciennes) n'ont publié leurs premières pages web qu'en avril 1998, en se cantonnant au début à reproduire partiellement la plaquette papier de l'École littéraire[33].

Restaient les élèves. Comme la plupart des départements n'avaient pas d'équipement informatique à usage collectif, les élèves rédigeaient majoritairement leurs travaux dans les salles informatiques qui leur étaient dédiées, l'« infirmatique » et la salle « S ». Cette dernière avait divers inconvénients : non climatisée, elle était surchauffée, et ne proposait que des machines « Sun », accessibles donc par un *login* et un mot de passe ; pire, elle était dans les locaux des mathématiciens et des informaticiens. Avec de tels « handicaps », bien rares étaient les élèves littéraires qui osaient s'y aventurer : l'infirmatique, dédiée aux littéraires au début avec ses Macintosh, proche de la cafétéria, semblait nettement plus conviviale[34], malgré la même absence de climatisation.

29. Cf. page 138 et le site http://www.item.ens.fr

30. Département et Laboratoire d'Economie Théorique et Appliquée. Ce laboratoire d'économie est sous la triple tutelle ENS-EHESS-CNRS, et installé dans le 14^{e} arrondissement ; bien que formellement rattaché au département de sciences sociales, il fonctionne de façon très autonome ; depuis 1999, les liens entre économistes et sociologues de ce département sont devenus plus concrets. http://delta.ens.fr

31. « Nouvelle gallia judaica », http://www.ngj.ens.fr/ngj/accueil.html

32. Aussi dénommé « Lila ».

33. Le site http://www.ens.fr/lila, avec ses deux pages, était jusqu'à la fin 2001 symptomatique de ce type de publication minimaliste.

34. En 1994, pour remédier à ces formes de ségrégation spatiale comme culturelle, et

Mais les rares élèves littéraires qui se laissèrent convaincre par leurs amis informaticiens de pénétrer dans la salle S finirent par acquérir une solide maîtrise de l'informatique Unix, de LaTeX et de l'internet grâce aux conseils délivrés par leurs camarades[35]. Ainsi, le noyau construit initialement par les « administrateurs du réseau », vite devenu le groupe « gourous »[36], s'élargit progressivement en accueillant des étudiants en sciences humaines et l'on assista en 1997 à la naissance du groupe des « tuteurs » ; composé d'une douzaine de personnes, il commença à proposer régulièrement à l'ensemble des élèves de l'ENS des stages d'informatique[37].

Ce groupe de tuteurs comprenait évidemment des jeunes informaticiens, mathématiciens et physiciens, mais accueillit rapidement quatre élèves issus des départements d'histoire, des sciences de l'antiquité, du Lila.

Parmi eux, une historienne, convaincue des avantages apportés par LaTeX devint rapidement la personne-clé de ce réseau, en facilitant les échanges entre les tuteurs et l'ensemble des élèves littéraires, et en proposant diverses formations autour du système d'exploitation des Sun, de LaTeX, et de l'internet. La proximité des gourous, d'une très grande compétence informatique, réduisait toute question complexe en un exercice amusant pour ces experts qui savaient donner des frissons aux administrateurs-système[38] et qui pas-

pensant que la rencontre entre littéraires et scientifiques faciliterait les échanges en matière de culture informatique, je proposais d'installer des machines Unix (NeXT et Sun) dans ce local. Malgré les critiques de certains élèves littéraires, qui affirmaient que l'infirmatique était transformée en salle de jeux pour scientifiques, il est possible que cette décision ait eut quelques effets fastes.

35. Les relations extra-professionnelles, nouées notamment lors de participations à des clubs de loisirs de l'ENS (jonglage, cirque), furent essentielles pour diffuser une telle culture informatique.

36. L'un et l'autre ont des droits spécifiques qui leur permettent d'installer des logiciels sur le réseau des élèves, d'améliorer les configurations de base, etc.

37. Cf. http://www.eleves.ens.fr :8080/tuteurs

38. Les jeunes informaticiens de l'ENS n'ont pas toujours eu bonne presse : comme certains d'entre eux n'hésitaient pas à démontrer que des systèmes prétendus fiables ne l'étaient pas, ils firent l'objet de plusieurs plaintes de la part des institutions qu'ils « pénétraient ». La communication entre les machines des élèves et l'extérieur de l'ENS fut

sèrent beaucoup de temps à développer le système d'exploitation Linux. Avec ce groupe s'est diffusée une culture de l'internet, qui a néanmoins tardé à pénétrer les instances institutionnelles de l'ENS, en grande partie parce que les échanges entre élèves et professeurs étaient très réduits, dès que l'on s'écartait des enseignements traditionnels[39].

Même si le nombre d'élèves intéressés par les stages que proposaient les tuteurs était au début assez réduit, l'importance de leur action dans un lieu aussi atomisé que l'ENS n'est pas négligeable. Ces tuteurs — au nombre de 18 en 2001 — ont construit un réseau social qui a permis l'appropriation d'une informatique adaptée aux besoins des érudits, en favorisant l'usage de logiciels et de codages fiables, malgré le désintérêt des responsables de l'ENS. La constitution du premier serveur web de l'ENS, la construction dès 1995 d'un réseau Ethernet entre chambres d'élèves[40], l'implication des tuteurs dans les formations à l'informatique, ou plus simplement la fréquence d'apparition des pages web d'élèves quand on interrogeait l'indexeur du CNRS[41] sur un domaine donné, témoignent d'une dynamique qui a souvent peu de liens avec les rapports d'activité des départements.

3.5 Situation en 1996

Ainsi, au sein des étudiants et des chercheurs en sciences humaines de l'ENS, les personnes qui avaient peu ou prou la possibilité d'utiliser l'internet étaient en nombre très réduit. On peut les décomposer en trois groupes — chacun de 10 à 15 personnes — assez éloignés les uns des autres : d'une part

alors filtrée entre 1990 et 1993. Mais les gourous sont aussi ceux qui ont adapté LaTeX à l'écriture du grec ancien.

39. Il serait illusoire de ne comprendre les transformations intellectuelles au sein de l'ENS — qu'elles soient liées à l'internet ou à d'autres problématiques — qu'au travers des cours et séminaires des enseignants. Le rôle des élèves est essentiel.

40. Aussi appelé « rézo », cf. [Tou99].

41. http://www.urec.cnrs.fr/AV

un noyau d'élèves littéraires et scientifiques, les tuteurs[42], qui avaient absorbé la majorité des gourous, très expérimentés en informatique. Mais à quelques exceptions près, ce réseau ne communiquait pas avec les enseignants.

Ensuite quelques laboratoires du CNRS, qui avaient des relations privilégiées avec les sciences exactes, comme le Delta et ses chercheurs souvent issus des mathématiques, ou l'ITEM, qui pratiquait déjà le dialogue avec des informaticiens[43], ou encore le laboratoire d'archéologie[44]. Mais, à l'exception du Delta, les membres de ces laboratoires avaient peu de contact avec les enseignants et avec les élèves de l'ENS.

Enfin, au sein de l'École littéraire proprement dite, l'Atelier Internet offrait les conditions d'un dialogue interdisciplinaire entre chercheurs et élèves. Mais souvent, ses séminaires n'atteignaient une masse critique qu'avec l'apport de participants extérieurs à l'ENS.

Rétrospectivement, ces trois types d'expériences apparaissent fort timides : en 1996, les seuls serveurs de l'ENS littéraire se limitaient à celui des géographes et à celui des sciences sociales — qui ne publiait en fait que les activités connexes à celles de l'Atelier Internet.

En revanche, la grande majorité des personnes de l'École littéraire qui avaient une réelle culture en matière d'histoire et de sociologie de l'écriture ou de la lecture, et qui étaient impliquées dans des activités éditoriales, n'intervenait ni dans ces débats, ni dans les processus de production[45].

Le seul protocole de l'internet à avoir un minimum de succès était le courrier électronique. Mais quand l'ENS littéraire comprenait plus d'une centaine d'enseignants[46] en 1996, 23 d'entre eux possédaient une adresse électronique,

42. Les tuteurs ne s'adressaient pas qu'aux littéraires, même si des cours *ad hoc* étaient réalisés à l'intention de ces derniers.

43. Par exemple, à l'occasion de la conception de postes de lecture qui permettaient de visualiser un manuscrit numérisé en mode image en regard de sa transcription numérisée en cours afin de produire une édition cohérente.

44. Qui s'était lancé dans la conception logicielle, en développant des outils de reconstitution de fouilles comme Archéoplan.

45. A l'exception d'un professeur de philosophie.

46. Chercheurs CNRS exclus, ici.

et seulement 9 envoyaient en moyenne au moins un message par jour, ce qui est fort peu[47] ; certes, ces enseignants ne disposaient en moyenne que d'un ordinateur pour trois personnes, et rares étaient, hormis les NeXT, les machines connectées à l'internet. Sur ces dernières, les plus grands utilisateurs du courrier électronique étaient de fait des étrangers qui avaient demandé à profiter d'un confort électronique qui intéressait finalement assez peu ceux pour qui il avait été conçu.

De façon analogue, le nombre d'élèves littéraires qui rédigeaient en moyenne plus d'un courrier par jour, si l'on regroupe tous ceux qui avaient des comptes sur les divers serveurs[48] de l'ENS, ne dépassait pas la quarantaine.

Autrement dit, à la fin de 1996, plus de 90 % des élèves et des enseignants de l'ENS littéraire ne voyaient pas l'intérêt du courrier électronique, quand bien même ils en connaissaient l'existence.

Aussi, les personnes qui s'engageaient dans des explorations qui dépassaient le simple usage du courriel, en publiant des pages web et en développant une réflexion sur leurs pratiques, étaient-elles non seulement rares, mais surtout, elles ne disposaient même pas de la masse critique de collègues qui leur aurait permis d'organiser un débat dans un contexte plus large que leur propre cercle.

3.5.1 Technique et réseaux sociaux

Le premier constat qui s'impose est l'étroite corrélation entre la présence de machines Unix et la possibilité de découvrir l'internet : les départements d'histoire, de littérature et langages, des sciences de l'antiquité, la bibliothèque littéraire ne disposaient d'aucune machine Unix, alors déterminantes pour garantir un contact rapproché avec les derniers développements de l'internet entre 1990 et 1996[49]. En revanche, on trouvait de telles machines,

47. Le détail de cette enquête et de la suivante, relative aux élèves, est donné au chapitre 1 de l'annexe (page 385 et suivantes).

48. Machines « clipper », « elias », etc.

49. Ce n'est que tardivement qu'on a proposé des serveurs internet sur les systèmes d'exploitation des ordinateurs personnels.

installées dans des salles d'accès libre, pour les élèves, au Delta, chez les sociologues, les géographes, et les philosophes. Il est clair que ces configurations ont d'autant facilité la socialisation de l'internet et d'une informatique qui explicite la relation au code, au texte, et aux outils qui font le lien entre l'un et l'autre, qu'elles facilitaient des rencontres, organisées ou dues au hasard : autant de façons de diffuser la culture de l'écrit contemporain et d'entretenir une réflexion sur l'outillage mental. Dans ce dispositif, les informaticiens avaient un rôle essentiel.

À l'opposé, la très grande majorité des personnes qui découvraient les ordinateurs et en acceptaient l'image commune de machine à écrire améliorée ne pouvaient concevoir que ces machines à produire du signe avec du signe puissent transformer un tant soit peu leurs pratiques. Elles s'inféodaient à des codages abscons, ne connaissaient pas l'internet, disposaient d'une très faible culture informatique, restaient — souvent isolées — dans leurs laboratoires traditionnels, et prenaient le risque de devenir rapidement illettrées[50].

3.5.2 Interdisciplinarité

Rétrospectivement, les années 1994 à 1996 furent, à l'ENS, particulièrement fastes, reconstituant les prémisses d'un échange de qualité entre littéraires et scientifiques. Un groupe d'élèves et de jeunes chercheurs expérimentait des activités variées, de la confection d'ouvrages imprimés à la publication électronique, de la lexicométrie à l'histoire, de la cartographie à l'économie, dans une ferveur créative, et en toute insouciance politique.

Au sein de l'ENS littéraire, le laboratoire le plus institutionnalisé de telles pratiques était l'Atelier Internet, où se mettaient en place des enseignements et des programmes de recherche collectifs. C'est aussi là que s'élaborait une réflexion sur l'édition et la revue savante et sur l'incidence des techniques sur les processus intellectuels, les liens entre ces deux thématiques étant alors vaguement perçus, mais pas encore explicités.

50. Cf. le chapitre 3.4.4 de la partie I page 81.

C'est en ce sens que l'on peut comprendre le caractère interdisciplinaire de ce séminaire, comme il se prétendait être : non pas que ses participants désiraient aborder plusieurs disciplines[51] ; mais au contraire parce qu'ils comprenaient que les dispositifs d'écriture, tels qu'ils étaient transformés ou malmenés à l'occasion du développement de l'informatique et de l'internet, influaient sur les modes de fonctionnement implicites de nombreuses disciplines ; et aussi parce que — osons le mot — la stupidité des arguments qu'ils rencontraient dans la presse écrite ou électronique les incitait, en bons serviteurs de la raison qu'ils étaient, à faire appel à des registres précis de l'histoire des techniques, de la sociologie, de la philosophie, ou simplement de l'informatique, pour contredire ces allégations qui allaient servir de socle aux religieux de toute sorte, détracteurs ou passionnés de l'internet.

En effet, l'expérimentation de l'internet invitait tout d'abord, à une difficile appropriation de la culture informatique, ce qui incitait, on l'a vu, au débat. Celui-ci, au départ orienté vers des questions techniques, s'élargissait vers les possibilités de travail scientifique que permet l'informatique (tris, comptages, graphes, etc.) et vers les conditions d'utilisation de ce travail (lecture, écriture, récolte d'informations). À partir de là, et avec des exemples concrets, se posait la question de « publiciser » les résultats scientifiques obtenus, qu'ils soient d'ordre méthodologique ou statistique.

Mais cette interdisciplinarité peut aussi être lue comme une rencontre entre des chercheurs de disciplines différentes qui avaient la double particularité de ne pas se satisfaire, ni des hiérarchies ni des démarches intellectuelles qui prévalaient au sein de leurs propres disciplines. En ce sens, leur désir d'interdisciplinarité signifie aussi le refus des vieilles frontières, des chapelles poussiéreuses, et des trop fortes dominations. Et il n'est pas étonnant que ce soient les enseignants, les chercheurs et les étudiants les plus marginaux qui se sont intéressés aux outils ici évoqués : les personnes trop à l'étroit intellec-

51. On sait que l'interdisciplinarité est d'un emploi difficile auprès des chercheurs, qui l'associent souvent à un abandon de la spécialisation qui leur est chère au profit de connaissances éparses et non approfondies.

tuellement dans une discipline, et aussi trop peu reconnues par leurs pairs, ne pouvaient qu'expérimenter ensemble — au moins sous forme de pari — les avantages et inconvénients de ce système d'écriture, dont ils savaient qu'on ne pouvait plus faire l'économie de son apprentissage.

Ils avaient découvert l'internet quand l'écrasante majorité de leurs collègues connaissait à peine l'informatique. Forts du soutien des informaticiens, leur appropriation efficace des ordinateurs les amenait à produire des résultats scientifiques qu'ils pouvaient publier sans attendre les contraintes de temps imposées aux impétrants. Leur apparition sur le « marché scientifique », au mépris des codes universitaires, se traduisait, sur un plan collectif, par une remise en cause des équilibres disciplinaires au sein de l'ENS. Par exemple, l'histoire et la géographie étaient valorisées aux dépens de la sociologie. Enfin, au travers de leurs analyses, ils concevaient l'internet comme un objet scientifique, et, pour s'affranchir de tous les faux débats et discours de comptoir à son sujet, développaient — certes inconsciemment — l'idée d'un coût d'entrée comme préalable à une compréhension rationnelle des enjeux de cet objet.

Ces oblats, qui jouaient si sérieusement avec la science et ses symboles, allaient vite être traités d'incapables et de fainéants.

Chapitre 4

Deux conceptions de l'informatique

Trois pôles restreints étaient donc à l'origine de la créativité collective consécutive à l'appropriation de l'internet. Mais les activités scientifiques ou éditoriales de deux d'entre eux n'avaient pas d'incidence sur la politique locale : les élèves, d'une part, maintenus dans une situation de domination objective (au moins par le fait que l'obtention d'un poste en université à l'issue de leur « scolarité » dépendait principalement de la direction de l'ENS et de celles des départements), les chercheurs du CNRS, d'autre part, qui n'avaient de comptes à rendre qu'à cette institution. Restait donc le réseau de l'Atelier Internet, très lié à la cellule informatique littéraire.

On pouvait imaginer que l'implication dans la recherche de cette dernière correspondait à une légitime autonomisation de cette structure, après cinq années d'existence. Tout comme le désir de ses membres de développer des activités de formation et d'enseignement, dans un domaine qui leur apparaissait en plein essor. Mais ces formes d'émancipation entraient en contradiction avec les conceptions des professeurs qui avaient demandé la fondation de cette cellule, ou qui héritaient de sa tutelle.

4.1 L'informatique littéraire avant 1997

En effet, la cellule informatique littéraire a été fondée dans un esprit de service aux élèves et chercheurs de l'ENS littéraire sans que son rôle n'ait jamais été défini par écrit. Au fil de la demande, ses membres comprirent qu'ils avaient pour fonction d'enseigner l'informatique — vaste programme —, d'aider les demandeurs à résoudre des problèmes simples (conversion de disquettes, nettoyages de fichiers, remise en état d'ordinateurs personnels, etc.) ou un peu plus abouties (formation à un système d'exploitation ou à un logiciel, travaux statistiques, recherches sur des bases de données, etc.). L'expérience avait prouvé qu'il était difficile d'inciter les professeurs comme les étudiants à prendre une quelconque autonomie vis-à-vis de l'informatique si on se limitait à un enseignement général : les premiers préféraient que l'on résolve à leur place les problèmes qu'ils rencontraient, considérant que les « techniciens » de la cellule étaient à leur service ; les seconds ne voyaient pas l'intérêt de cours d'informatique qui ne servissent pas immédiatement leurs travaux du moment, quand ils n'avaient pas déjà pris les habitudes des premiers. Aussi, le désir de relier l'informatique à des recherches précises était-il à l'origine de la (co-)fondation de divers « ateliers-séminaires » comme l'Atelier Internet, l'atelier cartographie, et le séminaire d'histoire sociale de l'immigration [1]. Cette solution, combinée à des enseignements d'initiation à l'interrogation de bases de données textuelles (grecques, latines, ou françaises) s'avéra efficace : les élèves littéraires qui y participaient prenaient conscience des avantages de l'automatisation dans leurs propres disciplines et acquéraient une culture informatique satisfaisante. Et l'internet entrait naturellement dans leur panoplie des outils, sources et critiques.

Avec l'expérience acquise dans le domaine de la programmation, ces formes d'assistance aux élèves et chercheurs purent alors se compléter par des collaborations nettement plus complexes : mise en forme des corpus, anciens ou récents, numérisés par un élève ou obtenus grâce aux réseaux, dénombrement

1. Cf. paragraphe 3.3.2 page 144.

et classification d'expressions singulières, et enfin mise en forme graphique des résultats obtenus. Bien sûr, les intéressés[2] étaient en nombre restreint, de l'ordre de la poignée chaque année pour chacune des trois personnes de la cellule. Ce faible nombre était compensé par la longueur des séances de travail, qui se prolongeaient souvent durant plusieurs permanences, engageant les « informaticiens littéraires » et ceux qui faisaient appel à eux dans des problématiques intellectuelles et techniques souvent passionnantes.

L'informatique littéraire s'émancipait donc de sa fonction initiale : la mise en place d'ateliers autonomes, le dialogue avec leurs participants, mais aussi avec quelques élèves, incitaient à la prise en charge de recherches sur le thème de l'écriture informatique.

Ces activités, notamment l'Atelier Internet et cette curiosité lexicométrique, furent tout d'abord accueillies avec bienveillance, par la direction de l'École, incluant son pendant littéraire[3] : celle-ci tenait un discours très séduisant, centré sur la « liberté » intellectuelle (liberté d'engager les recherches que l'on veut, de monter une équipe, d'attirer des élèves à son projet), et se montrait très sensible au prestige de l'École. Du coup, les membres de la cellule se sentaient légitimés dans leur statut d'oblats dynamiques par le sommet de la hiérarchie de l'ENS, ce qui est toujours réconfortant dans une institution qui fonctionne encore de façon princière.

4.2 Homo academicus

Mais cette cellule n'était peut-être qu'un pion dans la partie d'échecs que se jouaient les tenants du pouvoir. 1995 est l'année de la fin du terme de la fonction directoriale, d'une durée de cinq ans. Bien sûr, les deux soutiens de la

2. Élèves, enseignants, anciens élèves, voire étudiants extérieurs à l'ENS : par souci de prosélytisme, la porte de la cellule n'était fermée à personne.

3. A cette époque encore, si le directeur de l'ENS était issu d'une discipline « scientifique » (ce qui était alors le cas), le sous-directeur provenait d'une discipline « littéraire », et vice-versa. Ainsi s'effectuaient les conditions d'un équilibre auxquelles la communauté normalienne est très attachée.

cellule candidatent au même poste, et c'est l'ancien directeur qui l'emporte. Le *challenger* décide alors de quitter l'ENS, et, en 1996, la division littéraire se retrouve décapitée — par choix du réélu[4] —, ce qui la rend plus sensible aux formes de compétition entre ses différents professeurs. Ces derniers allaient déployer un acharnement contre les membres de la cellule informatique littéraire tel que la violence propre au monde universitaire allait en perdre jusqu'à son caractère symbolique.

Parce que ces tensions dévoilent en fait des attitudes opposées en matière de conception de l'écriture, je suis dans l'obligation d'évoquer quelques péripéties personnelles. Tout d'abord parce que tout effet de style tendant à les masquer ne conduit qu'à un discours général très fragile sur le plan méthodologique : « l'effort du sujet connaissant pour s'abolir en tant que sujet empirique, pour disparaître derrière le protocole anonyme de ses opérations et de ses résultats, est voué à l'échec »[5]. Ensuite, parce que le statut de « bouc émissaire » que j'ai vécu me semble symptomatique du malaise du monde universitaire face aux conséquence de la socialisation aussi massive que peu pensée de l'informatique et de l'internet en son sein : ma situation est exemplifiable, tout comme l'est celle de mes adversaires du moment. Leurs réactions s'inscrivent dans une logique parfaitement compréhensible de défense d'un corps aux abois, logique que je détaillerai en termes d'enjeux de pouvoir liés au statut de l'écrit. Aussi les quelques lignes qui suivent mettront en évidence, non pas la singularité de l'ENS littéraire au sein des sciences humaines françaises, mais au contraire, sa parfaite adéquation avec des pratiques de résistance, voire de panique largement répandues.

4. Après le départ du sous-directeur régissant la division littéraire de l'ENS, le directeur (scientifique) gouverna seul, avec simplement deux « directeurs de la recherche et des études » à ses côtés (une pour chaque division), tout en préparant une modification de l'organigramme de l'ENS. Depuis 2000, le directeur est secondé par deux sous-directeurs (un de chaque division), chacun d'entre-eux étant toujours assisté par un ou une directeur/rice de la recherche et des études.

5. [Bou84], p. 39.

4.2.1 Une période de conflits

Dès la rentrée 1996, les membres de l'informatique littéraire subirent des critiques indirectes, des attaques *ad hominem* colportées par des commérages de couloir, et des brutales mises à l'écart. Point focal de ces agressions, je fus relégué dans une ancienne chambre d'étudiant ; régulièrement, j'était informé oralement de mon départ prochain. Bien sûr, aucune évaluation scientifique des trois personnes visées ne fut jamais réalisée[6]. L'ingénieur documentaliste, désabusée par l'absence de choix clairs en matière informatique, par l'impossibilité de recevoir une définition précise des tâches, et par la multiplicité des rumeurs (projets de démantèlement de la cellule informatique littéraire, attaques *ad hominem*, etc.), finit par obtenir sa mutation à la Sorbonne.

Menacé, statutairement et scientifiquement[7], je tentais de maintenir et de développer les activités que j'avais contribuées à mettre en place, en m'engageant, avec des collaborateurs de plus en plus nombreux, dans une logique de production scientifique suffisamment intense pour que nos détracteurs, s'ils la niaient encore, prenaient alors le risque de prouver clairement que le travail scientifique qu'ils vantaient tant n'était qu'un slogan creux destiné à satisfaire leur soif de pouvoir. Pour légitimer ces activités autant que pour réduire mes doutes personnels quant à leur qualité, je recherchais, puis obtenais des évaluations positives — par exemple de la part du CNRS ou d'instances scientifiques étrangères —, même si elles n'avaient aucune incidence sur les critiques qui m'étaient adressées au sein de l'ENS.

L'incongruité de cette situation m'incita par ailleurs à affiner mon analyse de la relation du monde universitaire à l'écriture et à la technique. En effet,

6. Une des formes les plus sournoises de la domination symbolique à l'ENS est son abus de la référence à l'excellence. Elle peut alors refuser toute forme d'évaluation, puisque les « juges » ne sauraient en aucun cas être au niveau de ceux qu'ils doivent estimer. Un savant distillage d'allusions colportées par ses pairs fait alors comprendre au chercheur attaqué qu'il n'est plus *digne* de travailler dans cette fantastique institution scientifique au dessus de toute évaluation.

7. Cette situation de crise se prolongea jusqu'à la fin 1999, date à laquelle j'organisais avec mes collaborateurs un colloque international.

il m'apparut que ces pratiques sournoises, insidieuses, et épuisantes pour qui les subissait ne pouvaient se comprendre en termes de pur clientélisme : j'étais certes la seule des trois personnes attaquées à ne pas avoir de poste fixe, mais reconductible chaque année, et la volonté de me remplacer par une autre, à qui on aurait promis un poste à l'ENS, pouvait expliquait le désir de m'en expulser. Mais l'ENS souffrait bien moins que les universités des pénuries de poste, et la multiplication des embauches d'informaticiens invalidait ce type d'argument. En fait, il s'agissait de rappeler aux membres de la cellule informatique littéraire que leur mission consistait surtout à servir les professeurs, sans prétendre à une autre fonction, et sans proposer à ces derniers de renouveler leurs pratiques scientifiques.

4.2.2 Technique et rhétorique

Les activités de recherche documentaire et d'analyse lexicale incitaient les membres de la cellule informatique littéraire à interférer dans l'orientation des recherches qui les suscitaient, alors que la majorité des demandeurs aurait préféré une assistance silencieuse, qui ne remette pas en question des méthodologies parfois douteuses. Dans une contribution au rapport final transmis par l'Atelier Internet au Ministère de la recherche, on peut lire que de tels échanges relèvent non pas d'un service, mais d'une étroite collaboration : « le documentaliste doit éclairer les principes qui ont régi les réalisations techniques et le chercheur pour sa part doit préciser ses besoins, sa méthodologie afin qu'il y ait adéquation entre la demande et l'offre »[8]. Ce même auteur explicite aussi les tensions propres aux représentations sur l'informatique : « Trop souvent, la requête documentaire est faite dans l'urgence : une maîtrise à illustrer de graphiques et statistiques, une citation à vérifier car on doit rendre sa copie à un éditeur, une bibliographie à recomposer pour la rédaction de la thèse, etc. La demande s'apparente à celle du client de supermarché irascible dès que le vendeur ne lui propose pas de solutions toutes faites. Le documentaliste peut se sentir 'maltraité' par une attitude de ce

8. [Duc98].

type qui sous-entend que tout savoir technique n'est est pas un »[9]. Dans les faits, il semblait inconvenant de prétendre que la qualité d'un résultat informatique pouvait être liée à l'explicitation du projet scientifique qui amenait un visiteur à demander assistance aux membres de l'informatique littéraire.

Par ailleurs, l'énergie apportée par les participants à l'Atelier Internet et aux deux autres séminaires avait déjà des incidences sur les dynamiques de recherche à l'ENS, même si les chercheurs impliqués n'en avaient pas une conscience claire. Mais d'autres enseignants-chercheurs auraient aimé que les heures passées à construire ces séminaires servissent plus leurs propres recherches ou les fonctions de reproduction de l'institution[10].

Enfin, l'autonomie obtenue grâce à des financements extérieurs pouvait rendre jaloux des collègues moins bien dotés ; elle limitait aussi l'exercice du pouvoir répressif, puisque le chantage à la réduction budgétaire ne pouvait fonctionner.

Cette indépendance était aussi rendue visible par l'intérêt que des institutions extérieures aux travaux des membres de la cellule : ceux-ci étaient fiers de contribuer à la bonne image extérieure de l'ENS, en répondant aux demandes qui leur étaient faites d'organiser des enseignements à la Sorbonne et à l'EHESS, et de produire des communications à des colloques. Mais ils n'avaient pas compris que de telles activités n'avaient pas de valeur interne puisqu'elles ne rapportaient aucun bénéfice à un département précis ; au contraire, la légitimité croissante de la cellule à l'extérieur de l'ENS ne pouvait que renforcer son autonomie, et par là brouiller les cartes du jeu politique de l'ENS littéraire. Or, les trois trublions ne disposaient pas des titres

9. *id.*

10. Désireux de m'insérer dans le moule parfait des enseignants de l'ENS, j'avais pris en charge bénévolement un enseignement de l'agrégation de sciences sociales. Mais, au bout de deux années, pour manifester mon désaccord avec les pratiques du jury, qui proposait des énoncés de niveau trop variable, mal rédigés, et parfois faux — quand les candidats étaient tenus d'inverser des matrices non carrées, par exemple —, je décidais d'abandonner cette expérience, malgré les formidables succès de certains de mes étudiants. Ce qui, à coup sûr, a dû me faire gagner quelques inimitiés...

(doctorat, certes, mais surtout titres hiérarchiques, comme « directeur », etc.) de leurs adversaires déclarés.

Leurs interrogations sur la relation entre la machine et la pensée, leur engagement de plus en plus manifeste dans la production de revues savantes électroniques, finissait par dépasser les limites imposées par la bienséance universitaire. Il fallait rappeler aux audacieux que leur rôle était de tenir un tournevis (*sic*), non pas de s'immiscer dans les logiques éditoriales [11]. Il était exclu qu'à partir d'une expérience technique sur les modes les plus simples de l'écriture, à savoir compter des mots, les ordonner, les mettre en page, on puisse alors prétendre les combiner soi-même pour en faire du « beau langage », proposer ensuite qu'ils aient valeur scientifique, et finir par inverser l'harmonie de la hiérarchie universitaire en fondant des revues savantes, fussent-elles électroniques.

Pour mieux comprendre pourquoi des personnes dotées d'un solide capital universitaire, dont la fonction consiste à garantir le niveau d'excellence de l'institution dont elles ont la charge, ont adopté une stratégie de pure agressivité à l'encontre de ceux qui expérimentaient et socialisaient les techniques d'écriture contemporaines — alors qu'elles auraient pu tirer parti, pour elles, pour l'institution, de ces méthodes —, il nous faut détailler d'où cette direction tirait sa légitimité.

4.2.3 La force des traditions

Il convient en effet de rappeler le pouvoir de l'écrit, et de sa délégation dans une institution comme l'ENS, pour mieux comprendre le choc proprement culturel, et au niveau le plus simple, qu'a pu produire la diffusion de l'informatique auprès des élites françaises.

11. [Bou84], p. 134 à 136.

4.2.3.1 L'écrit et la plume

Jusqu'à la fin des années 1990, l'ENS — et surtout sa composante littéraire — vivait encore au rythme des « petits mots » manuscrits et des échanges téléphoniques. La hiérarchie ne répondait pas au courrier électronique : on le doublait alors de sa copie imprimée.

En revanche, les secrétaires ont toujours été très nombreux à l'ENS. Il ne s'agit pas de déplorer ce fait, mais de montrer comment une institution, qui a pu vivre à l'écart des suppressions de personnels, alors remplacés par des machines, reproduisait une organisation professionnelle fort classique. On est là loin d'une situation générale : par exemple, un étudiant du DEA de sciences sociales ENS-EHESS s'émerveillait du fait que cette structure d'enseignement, formant environ 30 étudiants par an, disposât d'une secrétaire propre, alors qu'on en comptait un ou une par millier d'étudiants dans son université d'origine. Mais ce confort propre aux institutions élitistes a certainement retardé l'obligation pour certains responsables d'utiliser un ordinateur pour écrire ou compter, *a fortiori* pour maîtriser les formes de l'échange électronique. D'où ce recours au secrétaire. Et si plusieurs responsables de l'ENS n'ont pas hésité, dès les années 1980, à se familiariser avec l'informatique, certains de leurs collègues ont pu continuer leurs activités en se servant d'un ordinateur de façon minimale. Ce qui a induit un ralentissement dans la socialisation du courrier électronique : tant qu'un collègue ou qu'un supérieur ne consultait pas son *mail*, celui-ci n'avait pas de valeur réelle aux côtés des autres modes d'échange. Il a par exemple fallu attendre 1999 pour être assuré que plus de la moitié des enseignants de la division littéraire de l'ENS lisaient leur *mail* et y répondaient [12]. Bien sûr, ce poids de la plume traduisait aussi des conceptions hiérarchiques. Par exemple, en 2001, un élève [13] qui se proposait d'envoyer un courrier électronique à la direction entendit

12. En 2001, il restait encore quelques professeurs qui ne lisaient pas leur courrier électronique, mais ceux-ci se comptaient sur les doigts d'une main.

13. Les élèves sont de fait les meilleurs informateurs pour connaître les pratiques électroniques des enseignants.

un de ses enseignants lui répondre : « vous n'y pensez pas ? Envoyez donc un courrier manuscrit, et proposez que la communication se prolonge par courrier électronique. Vous verrez alors si votre interlocuteur l'acceptera ou pas ». L'enseignant suggérait par là que l'usage du courrier électronique induisait une familiarité implicite qui pouvait froisser son destinataire.

4.2.3.2 Les Anciens

Ce malaise face aux ordinateurs était encore plus sensible chez les « anciens élèves ». Le laboratoire principal de l'ENS littéraire est encore sa bibliothèque : comportant d'innombrables ouvrages[14], parfois rares, elle est ouverte six jours sur sept aux élèves, enseignants et chercheurs, mais aussi à tous les anciens élèves de l'ENS (et de ceux de l'ancienne ENS de jeunes filles, dénommée « Sèvres »), et ce quel que soit leur âge[15]. En revanche, la bibliothèque est fermée à ceux qui « ne sont pas de l'École », c'est-à-dire à tous ceux qui n'ont pas réussi le concours d'entrée, sauf s'ils sont expressément « autorisés »[16]. Bien sûr, l'accès à la bibliothèque est gratuit pour tous les heureux élus, à qui est réservé un accueil de qualité doublé d'une liberté totale : ils peuvent aller et venir dans les rayons, choisir les ouvrages de leur choix, pour les consulter sur place ou les emprunter, parfois de nombreuses années durant.

Le projet d'informatisation de cette bibliothèque[17], initié il y a plus de 10 ans, a subi tellement de retard qu'il ne s'est amorcé qu'en 1998. Il n'est bien sûr pas terminé. Dans ce sanctuaire de l'écrit, fait d'une connaissance précise des rayonnages de la bibliothèque, de prises de notes sur des feuilles volantes,

14. Parmi ses effets pervers, l'informatisation renvoie à des pratiques de comptages : on prétendait que la bibliothèque recelait 500 000 volumes en 1995. Maintenant que son catalogue électronique est bien avancé, elle n'en contiendrait, approximativement, que 350 000.

15. C'est ainsi que des jeunes gens d'une vingtaine d'années côtoient des vieillards de plus de 80 ans.

16. Cette situation évolue depuis 2001 : les conditions d'accès à la bibliothèque sont plus souples ; et surtout, elles sont plus explicites qu'auparavant.

17. Qui, accessoirement, n'a jamais disposé d'un fichier matières.

l'introduction de l'informatique ne pouvait que s'avérer particulièrement difficile, malgré les efforts de quelques documentalistes et ingénieurs. D'une part, on menaçait une organisation sociale, centrée sur l'espace de convivialité que constituait cette bibliothèque, lieu de rencontre physique de plusieurs générations d'« archicubes »[18], en acceptant les conditions du ministère, disposé à financer cette informatisation à condition que la bibliothèque s'ouvre à un plus large public. Mais on s'exposait aussi à dévoiler l'incapacité d'une grande majorité d'anciens, souvent professeurs d'université, à manipuler des ordinateurs[19]. L'introduction de l'informatique ne pouvait donc voir le jour tant que la « communauté normalienne », plus précisément sa composante littéraire, risquait de manifester son mécontentement face à ces nouveaux outils de travail.

En effet, au-delà des conflits de personnes ou de disciplines tels qu'on les a évoqués précédemment, ces anciens normaliens constituent un pouvoir homogène[20] qu'aucune direction de l'ENS ne peut contrarier impunément.

Cette dépendance des responsables de l'ENS face à cette communauté d'anciens explique d'une part la « légitimité » dont ils peuvent se prévaloir, d'autre part leur difficulté à proposer — ou même à penser — un trop rapide renouvellement de l'outillage mental.

18. C'est ainsi que l'on appelle un ancien élève de l'ENS.

19. Bien sûr, à grand renfort d'anecdotes, on pourra prouver que cette situation n'est pas générale. Il y aura toujours un normalien des sciences humaines, enseignant ou pas à l'ENS, dont la compétence informatique, la réflexion sur les usages nouveaux de l'écrit sont exceptionnelles. Et on connaît la propension du monde universitaire à refuser les analyses globales, à jouer sur la singularité de ses chercheurs pour rejeter d'avance toute tentative de ce type. Il n'empêche que se produisent à l'ENS comme ailleurs des effets d'institution, indépendants des spécificités individuelles de ses membres, voire en pleine contradiction avec ces dernières. Le fonctionnement de la bibliothèque littéraire de l'ENS, qui repose sur une conviction tout à fait légitime — le fait qu'elle constitue un outil de travail extraordinaire pour les chercheurs —, tend paradoxalement à nier l'importance des méthodes de travail et de leur évolution.

20. Majoritaires aux jurys du concours d'entrée à l'ENS elle-même, des concours de l'agrégation, ils détiennent aussi nombre de postes de pouvoir au sein de l'université comme dans les commissions de programme des lycées.

En ce sens, l'ENS n'est plus une institution marginale, mais clairement représentative du fonctionnement du monde de l'enseignement et de la recherche en sciences humaines en France.

4.3 Apparition de l'informatique

Dans un tel contexte, l'acquisition d'un ordinateur, souvent pensé comme une machine à écrire, signifie une perte de standing, puisque qu'il témoigne de l'absence d'un ou d'une secrétaire. Et même une fois dépassée cette première attitude de rejet, il est difficile d'admettre qu'une machine puisse s'intégrer dans une quelconque panoplie de techniques intellectuelles : « que l'ordinateur soit une machine à écrire performante, cela est parfait. Qu'il permette l'accès à d'autres sources de savoir, cela ne peut être toléré. Il ne s'agit pas de 'blocage' mais bien plus d'une volonté de résistance »[21]. On conçoit alors qu'après avoir retardé l'introduction de l'informatique, l'ENS littéraire ait décidé de doubler ses investissements tout en les appliquant à des conceptions mécanographiques.

4.3.1 Administration

À partir de 1997 l'ensemble des services de l'ENS (gestion, direction, logistique, etc.) vécut une informatisation massive. Pour assister l'équipement des services, la direction mit sur pied une « cellule informatique de gestion », qui comprenait déjà huit personnes en 2001, alors qu'elles avaient la charge d'un parc d'environ 100 machines, le plus souvent individuelles (d'abord Macintosh, puis Windows). L'investissement logiciel suivit, parfois coûteux[22].

En revanche, l'informatique non administrative ne se développait pas : la cellule informatique littéraire était — on l'a vu — menacée régulièrement de

21. [Duc98].

22. Par exemple, l'ENS payait 300 000 F par an une entreprise pour la seule maintenance d'une application dédiée à la gestion du fichier informatique des élèves (communication personnelle, 1998) et comptait investir 240 000 F pour se protéger du bogue de l'an 2000.

dissolution ; les effectifs du SPI restèrent constants : en 2000 comme en 1995, ce service n'était composé que de trois personnes[23].

Ainsi se développait une informatique de service, comme si les ordinateurs, les logiciels prolongeaient la panoplie des fax, photocopieurs et autres objets techiques indispensables au bon fonctionnement d'une administration.

4.3.2 Enseignants et étudiants

À ce moment, les derniers départements à être faiblement informatisés (histoire, littérature et langages, études anciennes) accrurent aussi leur équipement de façon déterminante : en 2000, la majorité des enseignants disposait d'un ordinateur sur leur bureau. Mais ils ont été livrés à eux-mêmes, n'étant ni assistés par la cellule informatique de l'administration, ni incités à contacter la cellule informatique littéraire ou le SPI, qui avaient tous deux mauvaise presse. Ainsi, toute réflexion sur les avantages comparés des divers types de machines, sur le type de maintenance à prévoir ou sur l'organisation d'un minimum de formation était évacuée.

Évidemment, contrairement aux discours publicitaires, les chercheurs étaient confrontés à de nouveaux problèmes liés à leur découverte de l'internet : ils devaient installer ou faire installer des outils abscons, tels que des pilotes de modems ou de cartes Ethernet, configurer les protocoles, se familiariser avec des logiciels auparavant inconnus, comme Netscape ou Eudora, apprendre à convertir les fichiers illisibles qui leur étaient transmis par mail, et ne pas oublier les mises à jour régulières des anti-virus nécessaires pour compenser les fragilités des « obésitiels » comme les suites Office de Microsoft. On comprend que cette informatisation tardive, arrivant à un moment où la culture nécessaire au bon fonctionnement des ordinateurs ne se compare plus au maniement d'un traitement de texte et à la sauvegarde d'un fichier

23. On pourrait s'étonner de cette situation puisque le SPI gère aussi indirectement les 4200 comptes électroniques de l'ENS, ses 2400 adresses IP, et les multiples services attenants, dont les plus connus sont : mail, anti-virus, gestion de listes de discussion, news, web, ftp, serveur de noms de domaines, Pop, impression, routage et filtrage TCP/IP.

sur disquette, ait pu dérouter, voire gêner dans leur travail, la majorité des enseignants qui ne s'étaient pas préparés à une telle invasion d'automates.

Le manque de moyens dédiés aux chercheurs a d'autant freiné le développement de leur culture informatique qu'ils ne disposaient pas de parcs collectifs leur permettant de construire un réseau d'échange[24] ; une maintenance désorganisée les faisait travailler sur des machines en mauvais état, et ils interprétaient les bogues comme autant de preuves de leur propre incompétence. Leurs machines et logiciels obsolètes ne les incitaient pas à découvrir l'internet : en l'an 2000, quelques chercheurs travaillaient encore avec un écran à quatre niveaux de gris, et ne disposaient que de la version 2 de Netscape, ce qui amplifiait leurs problèmes de visualisation de pages web. On ne s'étonnera alors pas que de telles personnes arrivaient à lire leur courrier électronique, mais... ne savaient pas y répondre.

Quand la technique était au rendez-vous, c'est la connaissance des outils de travail qui était défaillante : un collègue assurait en 1999 qu'« il n'y avait rien sur l'internet » dans sa discipline. Mais il ne connaissait pas l'existence de moteurs de recherche autres que Voilà.

Les élèves, tant sur le plan des machines mises à disposition, que sur celui de la formation, étaient eux aussi laissés pour compte. C'est au vu de cette situation qu'il faut interpréter la fondation du groupe des tuteurs[25], ces bénévoles qui proposaient à leurs camarades de tirer au mieux parti des faibles ressources techniques mises à leur disposition. Une autre catégorie d'utilisateurs a aussi souffert de cette absence de réflexion sur les besoins informatiques : ce sont les anciens élèves. La majorité de ces derniers, plus encore que les élèves, étaient complètement oubliés de la direction littéraire, alors que leurs besoins étaient conséquents, pour au moins deux raisons : souvent AMN[26] dans des universités, ils avaient rarement accès à des ordinateurs

24. Le département d'histoire en obtiendra un en 2001. À cette date, le département littérature et langages disposait toujours de deux ordinateurs — dont un perpétuellement hors-service — reliées à un scanner dans une pièce fermée à clé.

25. Cf. partie 3.4 page 148.

26. Assistants moniteurs normaliens, aujourd'hui « allocataires ».

et essayaient donc d'utiliser le matériel de leur ancienne école ; débarrassés de l'agrégation, ils entamaient leur thèse et se trouvaient confrontés à la nécessité de réaliser des travaux originaux et conséquents, ce qui les incitait à recourir à des méthodes informatiques[27].

La « direction littéraire » limitait donc sa conception de l'informatique, entre 1997 et 2000, à celle d'un service : des ingénieurs étaient embauchés pour satisfaire les besoins de l'administration. En revanche, les élèves, les anciens élèves ne profitèrent ni d'un accroissement du nombre de machines, ni de l'assistance du nouveau service. Les chercheurs, mieux équipés, devaient se débrouiller tout seuls.

4.4 Bilan

Le premier bilan de cette politique est lourd. Sans même tenir compte des difficultés de la majorité des chercheurs à s'approprier diverses cultures informatiques, on peut se contenter de l'analyse d'un indicateur: la production sur le web. On obtient alors, en juillet 2001, les résultats décrits dans les tableaux 4.1 et 4.2.

Le département littérature et langages ne propose que deux pages web, dont une dédiée aux enseignants et chercheurs. Parmi ceux-ci, au nombre de 23, deux seulement donnent une adresse électronique permettant de les contacter. Dans ces deux pages, il n'est nullement fait allusion au laboratoire du CNRS, l'ITEM, qui pourrait pourtant proposer des sujets de recherche

27. Le propos n'est pas là de défendre une caste qui pourrait profiter à vie des services offerts par une institution dont ils sont les heureux élus parce qu'ils ont passé un concours difficile : le fait que les anciens élèves puissent profiter de la bibliothèque des lettres jusqu'à leur mort, nous semble plus de l'ordre du privilège que du droit. Mais on peut s'inquiéter de l'efficacité d'un discours valorisant la recherche et l'érudition quand rien n'est fait pour que des jeunes chercheurs, issus ou non de l'ENS, puissent profiter des instruments actuels pour mener à bien leurs travaux au moment où ils sont grands consommateurs de ressources techniques et intellectuelles. Heureusement, les quelques départements ayant construit des salles d'informatique communes ont une autre politique d'accueil des jeunes chercheurs.

intitulé	Nb pages web
Littérature et langages	2
Sciences de l'antiquité	6
Géographie	15
Philosophie	20
Sciences sociales	37
Histoire	76

TAB. 4.1 *Nombre de pages web des départements et sections de l'ENS littéraire en juillet 2001. Calculs obtenus après mirorisation des sites, et élimination des fichiers images.*

intitulé	Nb pages web
Nouvelle Gallia Judaica	24
Archéologie	148
Item	200
Delta	237
Équipe RST	581

TAB. 4.2 *Nombre de pages web de quelques laboratoires CNRS et équipes de recherche associés à l'ENS littéraire en juillet 2001.*

aux élèves de ce département. Le centre d'études anciennes partage sa page de garde avec le laboratoire d'archéologie du CNRS. Mais les deux serveurs sont différents, et celui du département ne comporte que 6 pages, dont aucune ne propose de contacter les enseignants. Il est à noter que ces deux départements sont les plus « littéraires » de l'ENS et attirent 40 % des effectifs des élèves de l'ENS littéraire.

Les autres départements ont déjà des sites plus fournis et proposent de contacter leurs enseignants par courrier électronique. Cependant, la géographie et les sciences sociales n'ont pas profité de leur avance initiale : ils étaient les seuls à disposer d'un site web en 1996 quand les autres départements n'en ont eu qu'à partir de 1998. Or, le département d'histoire, un des derniers à construire un site, est le plus fourni et le plus cohérent. On constate encore qu'il n'y a pas de prime à l'innovation. Ceci dit, tous les départements ont une caractéristique commune : aucun ne s'est lancé dans une politique de publication de sources primaires ou secondaires.

La politique des laboratoires CNRS est tout autre. Le site Nouvelle Gallia Judaica, maintenu par une seule personne, a déjà un nombre de pages tout à fait honorable, dont une quinzaine d'outils et d'articles téléchargeables. On voit là que la réflexion sur l'édition électronique est aboutie. Le Delta a une politique éditoriale bien différente de celle du département de sciences sociales, auquel il est rattaché : déjà, il a six fois plus de pages ; ensuite, la publication d'articles n'est pas négligée : quand les sociologues ne s'engagent pas au-delà du résumé, le directeur du Delta, Roger Guesnerie, professeur au Collège de France, publie l'intégralité de sa conférence inaugurale [28]. Malgré sa position exceptionnelle, il n'hésite pas à proposer un contact direct par courrier électronique. Le laboratoire d'archéologie compte 25 fois plus de pages que son homologue départemental (et l'on n'a pas compté ici les images), et l'Item en compte exactement 100 fois plus que le département littérature et langages.

À titre indicatif est donné le nombre de pages du site de l'équipe *Réseaux, Savoirs & Territoires* (RST), qui prolonge les travaux entamés par les premiers participants de l'Atelier Internet.

Ces statistiques pourraient être critiquées : on a compté le nombre de pages, mais on n'a pas évalué la valeur scientifique de leur contenu. Mais on peut imaginer que cette dernière doit être bien faible pour les sites des départements qui se sont contentés de reproduire sur le web l'annuaire de leurs enseignements, et on s'expose, en acceptant de tels arguments, à nier la qualité des travaux d'un Roger Guesnerie.

À la lecture de tels résultats, on doit reconnaître que les publications électroniques des départements témoignent d'un manque de réflexion sur les avantages et les inconvénients de la publication électronique, et de façon plus

28. http://www.delta.ens.fr/guesnerie/inaugural.html

générale, sur la façon dont les outils de recherche influent cette dernière. Cependant, il ne faut pas oublier que cette « avancée à reculons dans le cyberespace » risque fort de servir de modèle à la plupart des universitaires des sciences humaines, au vu de la forte légitimité de l'ENS dans ces domaines.

En matière d'usages, la situation n'apparaît pas meilleure.

En 2000, les enseignants ne sachant pas ou ne voulant pas utiliser le courrier électronique commençaient enfin à se faire rares. Mais les application de cet outils n'étaient pas pour autant entrées dans les mœurs. Les listes de discussion au sein de l'ENS littéraire étaient réduites à deux : une en géographie, l'autre en histoire. Le département de sciences sociales n'a pas su[29], comme on l'a vu, profiter de la dynamique qu'il a instaurée en 1992 : par exemple, le conseil pédagogique du DEA de sciences sociales ENS-EHESS décida en 1999 de se munir d'un *alias* permettant de joindre de façon électronique l'ensemble de ses membres (une quinzaine de personnes). Seul le coordinateur en fait usage, deux fois par an, pour convoquer les membres du conseil. Aucun abonné n'a jugé utile d'utiliser ce mode d'échange pour des débats. Malgré leur usage pionnier du courrier électronique, les sociologues restent plus attachés aux échanges inter-individuels proposés par le téléphone ou le fax qu'aux modes de communication collectifs comme les *alias* et listes de discussion.

L'utilisation des ressources du web est aussi très faible. À titre indicatif, l'article publié en novembre 1996, contenant une centaine de références précieuses pour les sciences de l'antiquité[30], a été consulté 2876 fois entre sa date de publication et janvier 1999. Mais il n'a été consulté que 92 fois par des machines de l'ENS[31]. En bref, ce document longtemps unique en son

29. Ou pas pu : l'EHESS, qui constitue pour ce département une référence, semble avoir eu dans son ensemble encore plus de mal que l'ENS littéraire à s'approprier l'informatique et l'internet.

30. Cf. page 145.

31. En excluant les machines de l'informatique littéraire, où l'auteur pouvait tester la mise en page de son article ou montrer à des visiteurs la qualité de son travail. Comme souvent dans les statistiques de cette thèse, les erreurs d'accès (≥ 400) et l'incidence des

genre n'était pas consulté plus de 4 fois par mois en interne, alors que son public potentiel représente plus du cinquième de l'École littéraire.

Il y a encore deux ans, la critique la plus fréquente face à l'énonciation de telles statistiques consistait à étiquetter leurs auteurs de prosélytisme irréfléchi. Aujourd'hui, de nombreux responsables s'alarment de cette situation d'indigence électronique au sein d'une institution qui a toujours proclamé son excellence.

On doit donc conclure qu'au moment où, en sciences humaines, l'informatique devenait un outil de recherche à part entière, en partie du fait que les réseaux proposaient des services efficaces, en partie grâce à la démocratisation de la programmation, les instances de décision de l'ENS multipliaient les résistances et freinaient l'appropriation d'un tel instrument.

Bien sûr, de tels retards constituent un préjudice pour la recherche, et en premier lieu pour les élèves, qui approfondissaient leur maîtrise d'une rhétorique traditionnelle sans pour autant pouvoir se familiariser avec d'autres outils de travail. On sait qu'un retard ne constitue pas nécessairement un handicap définitif, comme l'exemple des historiens le prouve ; mais il peut être problématique pour des jeunes chercheurs qui vont entrer en compétition avec beaucoup d'autres quand ils chercheront un poste. Enfin, la conception d'ordinateurs et d'informaticiens au service de l'administration et plus généralement du pouvoir étouffe aussi la possibilité d'une attitude réflexive par rapport à l'écriture, ce qui peut être un handicap pour les personnes qui ont choisi de mettre cette technique au cœur de leurs pratiques professionnelles.

caches n'ont pas été prises en compte ici. Ce qui sous-estime la disproportion entre les accès extérieurs et intérieurs à l'ENS. En revanche, les accès « PPP » de chercheurs à partir de leur domicile sont comptabilisés.

Chapitre 5

Conclusion

L'ENS littéraire a découvert l'internet quatre à cinq ans avant la majorité des institutions françaises en sciences humaines, en grande partie grâce à son voisinage avec des chercheurs en sciences exactes. De 1992 à 1996, quelques groupes de chercheurs, d'enseignants, d'étudiants (élèves et anciens élèves) se familiarisent avec les réseaux à une période où ceux-ci n'ont pas encore basculé dans l'espace marchand. Quand les « pionniers » découvrent la complexité mais aussi, vers 1996, les avantages de la culture informatique nécessaire à une maîtrise raisonnée de l'internet, 90 % de leurs collègues n'utilisent pas le mail, et sont très désemparés face aux ordinateurs. À cette période, les pressions sont telles que l'administration et les enseignants de l'École littéraire prennent conscience du besoin de s'informatiser en masse. Mais les enjeux intellectuels de l'écriture informatique sont niés au profit d'une vision bureaucratique, qui obéit autant aux injonctions publicitaires (les ordinateurs sont d'usage aisé, toute forme d'apprentissage est inutile) qu'aux conceptions dominantes en matière de technique (les informaticiens n'ont pas à réfléchir, mais à obéir aux professeurs d'université). Il s'ensuit des conflits de représentation au sujet de l'informatique. L'internet n'est pas pensé, ce qui peut à la limite être admissible quand on sait que son appropriation massive ne peut dater d'avant 1996. Mais le refus d'imaginer les potentialités de la publication électronique — traduisant une volonté de garder la main-mise sur l'édition savante —, de s'interroger sur le coût d'apprentissage de l'informatique à la période où elle se complexifie — avec les protocoles de l'internet — et de profiter des acquis des laboratoires qui avaient expérimenté l'informatique et l'internet, ont finalement été très préjudiciables à l'institution tout entière : les sites web des départements ne sont pas crédibles ; le fax reste préféré au courrier électronique. Le catalogue de la bibliothèque, lieu chéri des « anciens », n'est toujours pas informatisé. Mais surtout, les élèves, dans leur grande majorité, restent formés à d'anciennes méthodes, qui privilégient les beaux discours aux méthodes de travail contemporaines. Il ne s'agit pas de prétendre que l'ENS littéraire accuse un retard ou un préjudice dont elle ne se relèvera pas : on connaît la fausseté de tels arguments alarmistes. Mais il semble clair que quelques promotions d'élèves ont été sacrifiées sur l'autel du culte de l'omniscience des mandarins, qui peuvent, sans vergogne, diriger des collections imprimées dédiées à « l'étude des apports théoriques, méthodologiques et pratiques des nouvelles technologies dans le domaine des sciences humaines et de la littérature en particulier » alors même qu'ils ne savent pas se servir du courrier électronique.

Ceci n'est pas un procès, mais un constat. Le propos n'est pas de corréler l'internet à une certaine forme d'intelligence : certaines personnes, très attachées à l'outillage mental dont elles ont hérité, sont des chercheurs hors pairs, et par ailleurs des enseignants aussi efficaces que dévoués. D'autres favorisent le développement de centres informatiques, qu'ils destinent, suivant leur position du moment, à l'administration ou à la recherche, dans le seul but d'accroître leur pouvoir.

Il ne s'agit pas non plus d'imaginer que si les « réactionnaires » ont refusé l'internet à l'ENS, c'est parce que les nouvelles technologies[1] organiseraient

1. On comprend bien ici l'économie intellectuelle qu'offrent des expressions comme « nouvelles technologies » ou des acronymes comme « NTIC ». Comment dire, sans prêter à rire, que « les réseaux transforment les relations sociales » ? L'internet n'étant pas que des tuyaux, il faudrait alors dire que « la technique transforme le social ». Mais il y a tant d'hommes derrière cette technique (auteurs de protocoles, de pages web, etc.) qu'on en arriverait à la tautologie suivante : « les hommes transforment le social ». L'usage d'une

naturellement la promotion d'une organisation sociale plus souple que celle érigée en notre État par l'histoire, comme l'écrivait le sénateur Tregouet[2]. On sait que l'idée d'un tel assouplissement des relations hiérarchiques est un des mensonges induits par le déterminisme technique, et on doute que la direction de l'ENS y ait souscrit, en ayant peur de tels effets prétendus de la technique.

S'il y a une leçon passionnante à tirer de l'expérience de l'ENS, c'est bien la façon dont des personnes ayant délibérément choisi le pouvoir au détriment de la recherche sont restées prisonnières des registres de l'échange qui les ont fait accéder à ce pouvoir. Propriétaires d'un outillage mental d'autant plus désuet qu'elles n'ont pas cherché à l'actualiser par une recherche continue, préférant les honneurs administratifs à la compétition scientifique, elles ont confondu ordinateur et papier carbone, traitement de texte et travail sur le texte. En montrant leur attachement à une forme d'écriture bureaucratique (secrétaires, photocopieurs, circulaires, etc.), et leur refus d'imaginer un quelconque lien entre évolution d'un système de signes et organisation de la pensée, elles témoignent déjà *a contrario* de la relation entre outillage mental et production scientifique : pour garder le pouvoir, il convient de ne rien changer à ses habitudes ni à celles de ses confrères.

Mais, outre l'explicitation des processus intellectuels[3] que permet une technique d'écriture en transformation — déjà abordée, mais qui sera approfondie au chapitre suivant —, l'obligation d'opérer un retour sur la sociologie des tenants de l'outillage mental constitue déjà en soi un des aspects positifs de cette forme de réflexivité propre à l'écriture.

---

expression comme « nouvelles technologies » permet alors de masquer l'incohérence des raccourcis intellectuels.

2. « Structure hiérarchique, centralisme, dirigisme, cloisonnement [sont] aux antipodes, nous le verrons, des valeurs de la Société de l'Information ». Résumé de l'intervention de Réné Tregouet lors de la conférence de presse du 8 avril 1998 au Sénat. Le rapport complet est disponible à partir de l'URL
http://www.senat.fr/rap/r97-331-t1/r97-331-t10.html

3. Et donc, des pratiques professionnelles des artisans de l'écrit que sont les universitaires.

Troisième partie

Enquête au sein de l'Université

Mépris de l'informatique — et de ceux qui en faisaient usage —, refus de la notion de laboratoire, crispation sur les revues savantes imprimées, tels sont les déterminants de la réaction des responsables de l'ENS littéraire, prise à son propre jeu institutionnel d'organe de reproduction de l'élite universitaire. En même temps, émergeaient quelques petits groupes de personnes qui découvraient les potentialités d'un outillage mental associé à un usage réfléchi des ordinateurs, dont l'apprentissage requérait la construction d'un réseau social autour des informaticiens, qui expérimentaient les formes éditoriales permises par le web, et qui redécouvraient ainsi — effet réflexif de leur engagement — les conditions sociales de la légitimation scientifique.

Les résistances de l'ENS littéraire à l'internet apparaissent au final logiques, si l'on garde en mémoire son statut de « gardienne du temple » au sein des sciences humaines universitaires[4].

Que se passait-il en dehors de cet espace clos ? Les universités, moins prisonnières de leur image, étaient-elles plus dynamiques ? Leurs enseignants y étaient-ils plus libres, ou au contraire, plus conformistes, et donc encore plus portés sur les enjeux symboliques de la publication imprimée ? Au-delà des réponses à ces questions, nous voulions vérifier si notre argumentation centrée sur la difficile acquisition d'une culture informatique restait valide dans un cadre général[5]. En revanche, si nous ne nous trompons pas, la faiblesse

4. Statut qui peut aussi, sans risque de contradiction, s'accorder avec une fonction de pilote en matière d'adoption d'un outillage mental actualisé. En fait, au-delà de leurs clivages, modernistes et traditionalistes de l'ENS tiennent formellement le même discours : c'est toujours au nom de la position phare de l'ENS — réelle ou fantasmée — au sein de l'enseignement et de la recherche qu'ils défendent leurs positions.

5. Deux discours tendent à nier la complexité de la culture nécessaire à une telle appropriation, et donc la notion d'outillage mental : un discours commun, qui insiste souvent sur

du dialogue interdisciplinaire dans l'enseignement supérieur français aurait comme corollaire un engagement dans la publication électronique savante, et une réflexion à son sujet, fort timides.

Pour répondre à ces questions, mais aussi pour nous aider à affiner notre compréhension des transformations intellectuelles induites par l'informatique et l'internet, nous avons adressé un questionnaire à des fondateurs de revues savantes ou de sites savants électroniques[6] en sciences humaines.

A notre connaissance, notre enquête est la seconde du genre : au printemps 1999, Jean-Michel Salaün[7] en avait coordonné une première sur le thème « les usages et les besoins des documents numériques dans l'enseignement supérieur et la recherche »[8]. Son enquête déborde largement le propos de notre travail ; tout d'abord, elle intègre dans le numérique l'internet, les cédéroms, et aussi l'audiovisuel ; ensuite, elle examine les nouvelles formes de l'enseignement (incluant la formation à distance) permises ou imaginées, la réorganisation de la documentation, et étudie les méthodes de travail qu'adoptent les chercheurs — dénommés par la suite les « innovateurs »[9] — qui utilisent les documents numériques et les réseaux ; enfin, son champ d'étude concerne l'ensemble des disciplines.

Si, dans sa conclusion, Jean-Michel Salaün rappelle que « les enquêtes [...] menées sont trop parcellaires pour donner, à coup sûr, une photographie fidèle de la situation du document numérique dans l'université française », ce premier travail nous a malgré tout servi de repère. Par exemple, il nous a conforté dans l'idée que les réactions de l'ENS littéraire face à l'internet étaient en

une compétence spontanée des jeunes, et un discours d'ingénieurs, qui espèrent proposer tôt ou tard des outils informatiques aussi simples d'emploi qu'une machine à laver.

6. Pour une définition précise de ces notions et le détail de leur écart à l'imprimé, cf. la partie 3 de l'annexe (p. 401 et suivantes).

7. Professeur à l'ENSSIB (École nationale supérieure des sciences de l'information et des bibliothèques), installée à Villeurbanne.

8. Ce document, remis à son commanditaire, la Maison des sciences de l'homme (MSH) en septembre de la même année, est disponible dans son intégralité à partir de l'URL http://www.pner.org/html/activduprog/doc2b1.asp. Voir aussi [Sal01].

9. Sans qu'aucune mention morale soit attachée à ce mot. Cf. note 16 page 106.

fait assez communes, et que les premiers à s'engager dans l'internet soulèvent rarement l'enthousiasme de leurs collègues : « ces innovateurs [...] sont isolés dans leur université, même quand ils ont l'appui des instances dirigeantes. [...] Ils ne semblent pas beaucoup rencontrer de soutien de la part de leurs collègues. [...] Sans reconnaissance, ni validation scientifique, les réalisations peuvent être, à tort ou à raison, considérées comme non pertinentes. Certaines nous ont paru impressionnantes, d'autres discutables ». La façon dont se reconstituent alors les réseaux d'échange scientifique et d'évaluation nous semblait aussi intéressante : « par contre, les innovateurs ont trouvé ailleurs, parfois très loin, des contacts, des collègues, des laboratoires et même des infrastructures susceptibles de les aider dans leurs travaux. Ils participent à un réseau de spécialistes pointus dans leur domaine, qui se connaissent, discutent et s'entraident ».

Notre enquête prolonge donc en partie celle de Jean-Michel Salaün tout en ayant trois spécificités : tout d'abord, elle est restreinte aux personnes qui se sont lancées dans un projet éditorial d'envergure dans le domaine des sciences humaines. Nous pouvons donc espérer avoir touché des acteurs dont l'activité scientifique ne peut être remise en cause. Ensuite, elle explicite les potentialités de l'outillage mental associé à l'informatique et l'internet : de nombreuses questions portaient sur ce thème. Enfin, elle explore la relation entre enjeux intellectuels et enjeux sociaux au sein de l'organisation universitaire, de façon à vérifier concrètement la cohérence du cadre théorique que nous avons élaboré dans la première partie de cette thèse.

Chapitre 1

Présentation de l'enquête

Le questionnaire[1] proposait à ses destinataires de détailler leurs motivations quand ils se sont impliqués de façon concrète dans l'édition électronique, les contraintes et les opportunités qu'ils ont rencontrées, leur analyse des transformations intellectuelles produites dans leur champ disciplinaire suite à leur travail. Plus précisément, les interrogations portaient tout d'abord sur le moment et les raisons de l'implication initiale, puis sur la culture technique de la personne enquêtée, et sur les ressources dont elle disposait alors. Ensuite, celle-ci évaluait la qualité des contenus sur l'internet à cette période, et la réception de son initiative dans son univers professionnel. Enfin, ces questions étaient réactualisées de façon à permettre une comparaison entre période initiale et période d'envoi (automne 2000) et la personne interrogée détaillait ses bilans et projets.

La précision des réponses, et le statut professionnel des personnes interrogées donnent un panorama détaillé des contextes techniques, économiques et sociaux (voire politiques) des initiatives prises par ces spécialistes des sciences humaines qui, pour la plupart, ont construit des sites web aujourd'hui réputés dans leur discipline ; mais cet échantillon comprend aussi une poignée de personnes qui n'ont pas encore réussi à finaliser leur projet éditorial.

1. Celui-ci est présenté en page 409 de l'annexe.

1.1 Note méthodologique

Le questionnaire a été adressé en septembre 2000 à 51 personnes, dont 31 ont répondu. Trois réponses étaient trop succintes pour que l'on puisse en tirer des informations intéressantes. Il reste donc 28 personnes. On pourrait s'inquiéter d'un échantillon aussi restreint. Mais sa petite taille témoigne déjà des difficultés qu'il y avait à prendre une grande partie de son temps pour réaliser des activités qui n'étaient pas valorisées au milieu des années 1990 : explorer les contenus du web, construire un serveur ou un forum, inciter les collègues à publier, se constituer un réseau d'experts, dans sa discipline comme en informatique, n'allait pas de soi. Par ailleurs, le but de cette enquête n'est pas de recenser toutes les initiatives, mais de détailler de façon précise celles qui répondaient à une réelle exigence scientifique : celle-ci n'est pas l'apanage des professeurs d'universités, et c'est pourquoi des enseignants du secondaire ou des classes préparatoires ont été interrogés. Enfin, les 34 questions envoyées étaient ouvertes (à l'exception de cinq questions de base), et leur envoi sous forme électronique a permis aux destinataires d'y répondre sans limitation de taille et en développant leur expérience personnelle. Ainsi, on peut espérer que ces témoignages sont représentatifs, non seulement des préoccupations des pionniers de l'internet en sciences humaines, mais aussi de l'état d'esprit de leurs collègues qui ont tenté, avec un moindre succès, de participer à la diffusion de contenus structurés sur le web.

Certaines réponses sont d'une franchise étonnante : certains enquêtés n'ont pas hésité à critiquer vertement leurs collègues. Par souci de protection des informateurs, il importait donc d'anonymiser les réponses. Mais cela ne suffit pas : dans les « micro-mondes » de l'Universé, il est aisé d'identifier un chercheur, si l'on connaît son âge, son institution, et ses activités, surtout quand celles-ci sont marginales. Aussi le lecteur comprendra-t-il que certains profils aient été volontairement mélangés de façon qu'aucun enquêté ne puisse être retrouvé : un propos tenu par *une* spécialiste de la discipline A sera parfois transcrit comme étant celui d'*un* de la discipline B. De même, le terme « institution » sera uniformément employé pour évoquer une université, un lycée ou une grande école. Les structures plus petites, axées sur des recherches ou des enseignements thématiques, seront quant à elle présentées comme des « laboratoires », indépendamment de leur éventuelle appartenance au CNRS.

Ces choix ne résultent pas seulement d'une inquiétude exagérée face aux déboires que pourraient subir les informateurs (l'internet n'est pas une activité « criminelle », même dans le monde universitaire : au contraire, les pionniers sont aujourd'hui parfois valorisés). Il s'agit simplement de maintenir une relation de confiance : ayant reçu généreusement des informations confidentielles, nous nous attachons à garantir cette confidentialité.

1.2 Profil général des répondants

Les 8 femmes et les 20 hommes dont les réponses ont été retenues ont découvert l'internet relativement tôt : entre 1989 et 1998, dont 22 entre 1989 et 1995. De même, leur implication personnelle date d'avant 1997 pour 20 d'entre eux[2]. L'éventail des statuts est large, tout en restant typique du monde de la recherche[3] : quatre chercheurs CNRS (dont 2 directeurs de recherche), cinq professeurs d'université, sept maîtres de conférences, deux doctorants, trois professeurs agrégés détachés dans le supérieur et deux en classes préparatoires, auxquels s'ajoutent un médecin attaché hospitalier, trois ingénieurs d'études ou de recherche et un haut-fonctionnaire de l'Éducation nationale. Il est à noter que deux personnes de cet échantillon travaillent au Canada et ont des relations professionnelles avec quelques chercheurs français interrogés dans le cadre de l'enquête. Le risque de solliciter un réseau spécifique a été soigneusement évité : les 28 personnes retenues au final se décomposent en 16 sous-groupes d'interconnaissance disjoints[4], le plus important étant composé de cinq personnes (qui ne travaillent pas dans les mêmes secteurs de recherche et n'ont pas la même tutelle). Cet éclatement du réseau

2. Soit entre 1991 et 1996. Quatre autres personnes se sont investies en 1997.

3. La description qui suit tient compte des statuts actuels des enquêtés, et non pas de leur statut initial.

4. Du moins, à notre connaissance.

résulte de la volonté de voir représentées un grand nombre de disciplines des sciences humaines : l'archéologie (une personne) et les lettres classiques (deux personnes), la philosophie (deux personnes), les sciences politiques (deux personnes), la géographie (quatre personnes), l'histoire (autant), la littérature (deux personnes), la sociologie, l'économie, la communication, la médecine et la linguistique (une personne à chaque fois), auxquelles s'ajoutent trois « divers » (ingénieurs, etc.) et autant de spécialistes des bibliothèques et de l'édition. Bien sûr, ces profils ne sauraient témoigner de spécificités disciplinaires : le but de l'enquête consiste au contraire à mettre en évidence des situations homogènes au-delà des disciplines, tout en donnant la parole à des représentants de cultures différentes. Représentants, et non-représentatifs, puisque ces personnes, chacune à sa manière, se distinguent grandement de la moyenne de leurs collègues dans leur pratique de l'internet.

La majorité des Français travaillent à Paris ou en région parisienne[5] (21 personnes) : 5 personnes sont fixées dans diverses villes de province.

La plupart des enquêtés (18 personnes) se sont impliqués dans la réalisation de sites web hautement spécialisés. Quatre d'entre eux ont même fondé une revue savante électronique. Ce travail de production éditoriale se double souvent de participations intensives à des forums (voire à leur mise en place) ; parmi les personnes qui ne sont pas précisément impliquées dans un site web opérationnel, cinq ont publié (dans des revues ou rapports, électroniques ou imprimés) des analyses sur le contenu du web dans leur discipline, sur les recherches qu'il permet, sur ce qu'on peut en attendre : leur expertise est reconnue. Les cinq autres sont des « consommateurs éclairés » : deux d'entre elles n'ont pas finalisé leur projet, une se sert surtout de l'internet comme support documentaire pour son enseignement et ses recherches, et enfin, deux n'ont pas trouvé le temps de dépasser le stade de la consultation (sur le web ou sur des forums), malgré leur désir. Leur témoignage n'en est pas moins intéressant, mettant lui-aussi en évidence l'état de l'art dans leur discipline relativement au web, et l'état d'esprit de leurs collègues.

5. Ou ont une double affectation, à Paris et en province.

On peut imaginer qu'au vu du statut professionnel de la majorité des répondants, ceux-ci ne soient pas si jeunes : si deux personnes ont entre 25 et 30 ans, la majorité se déploie entre 35 et 55 ans [6], avec une exception pour une chercheuse qui a plus de 60 ans. Cette remarque contredit les discours prétendant que seuls les « jeunes » s'investissent dans les techniques contemporaines : l'organisation de revues savantes, la publication de textes érudits, de listes exhaustives de pointeurs, ou la participation à des forums professionnels impliquent une culture approfondie du champ sur lequel on s'exprime et une exigence scientifique qui nécessitent un minimum de maturité, intellectuelle comme biologique.

6. 8 entre 35 et 40 ans, 7 entre 40 et 45, 3 entre 45 et 50, 6 entre 50 et 55, une entre 55 et 60.

Chapitre 2

Débuts de l'engagement

2.1 Motivations

Tout d'abord, quelques personnes avouent leur engouement spontané pour l'internet : « émerveillement devant la capacité à diffuser facilement de l'information de qualité » ; « désir de sortir de l'hexagone, une sorte de voyage intellectuel » ; « pure et simple fascination » ; « remède à l'ennui ». Ces personnes assument leurs passions, voire leurs besoins de passions ; mais elles ne sont pas seules à témoigner de leur curiosité ou du caractère ludique de leur activité sur l'internet : trois autres chercheurs mentionnent aussi de tels aspects, parmi d'autres. Citons un exemple : « diffusion de travaux. Jeu. Puis appui pédagogique. ».

Au-delà de ces facteurs, à ne pas négliger, le déterminant commun est bien sûr l'exigence professionnelle, doublée d'un volontarisme scientifique : « à cette forte motivation relevant d'une sorte d'éthique professionnelle s'ajoutait la curiosité personnelle et le désir d'apprendre et d'acquérir un nouveau savoir ». Et si quelques personnes mettent en avant des arguments propres à leur statut (un anthropologue des milieux scientifiques, un chercheur chargé de promouvoir l'informatique, un bibliothécaire incité par ses collègues), la majorité des enquêtés a surtout ressenti le désir de dynamiser une communauté scientifique ; cette expression est parfois complétée par le détail de projets altruistes : désir de publier des textes difficilement accessibles, ou de dynamiser « la communauté en partageant le savoir, qui est la seule manière de le faire fructifier ».

Le goût de l'expérimentation est aussi prononcé (il va de pair avec le jeu et la curiosité) dans un domaine où les possibilités du travail collaboratif ont vite été pressenties. On lit par exemple : « volonté de découvrir les potentialités offertes, l'envie d'un mode de travail collectif », ou « caractère exceptionnel de l'outil, mêlant efficacité et souplesse. J'ai surtout apprécié son caratère décentralisé, ouvert, dynamique ».

En même temps, l'accent est mis sur la faiblesse et les limites des instruments classiques : « ayant vécu à l'étranger, je me suis rendu compte de la difficulté qu'il y avait de se tenir au courant de ce qui se passait [dans mon institution] quand on n'était pas sur place », ou « [je voulais] créer un espace de publication pour de jeunes chercheurs [...] qui n'ont pas souvent d'opportunité de publication dans les revues traditionnelles ». Et bien sûr, les possibilités de contact et d'échange sont mises en avant par plusieurs personnes, que ce soit par l'intermédiaire de listes de discussion ou de pages web de chercheurs : « nécessité de disposer d'informations à jour sur les activités (publications, conférences, etc.) de la communauté scientifique dont je fais partie et qui est principalement composée d'anglo-saxons ». Enfin, une personne cite expressément un argument économique : « il n'y avait plus d'argent pour financer les [...] diffusés sous forme de volumes imprimés » ; une autre, parmi les premiers précurseurs, évoque le désir de transférer sur l'internet des bases de données acessibles par Minitel.

Il apparaît donc que ces initiateurs cultivaient le moindre des défauts du chercheur : la curiosité, souvent associée avec une volonté de développer des réseaux d'experts ; sur ce point, déjà deux points apparaissent en filigrane : le réseau est international, et les étudiants et les jeunes chercheurs n'en sont pas exclus.

Pour la majorité de ces initiateurs, cette démarche était individuelle : 13 personnes le soulignent, auxquelles il faut en ajouter deux pour lesquelles, dans la pratique, ou sous pression de la hiérarchie, un projet collectif s'est retrouvé porté par une seule personne. Mais l'inverse s'est aussi produit : quatre personnes ont commencé leur travail en solitaire pour le finir collectivement, et neuf autres ont dès le début travaillé en groupe (et ont réussi à prolonger ce fait). Mais, par « démarche collective », la plupart des personnes évoquent plus un partage des tâches et des responsabilités décidé de façon informelle qu'une dynamique impulsée par leur institution.

2.2 Culture technique

Aux questions « votre culture informatique vous semblait-elle un gage de succès de votre entreprise » et « sinon, avez-vous ressenti le besoin de compléter cette culture technique, voire de l'acquérir entièrement pour mener à bien votre projet », six personnes ont répondu « non » ; mais quatre d'entre elles ont assuré disposer du réseau de compétences leur permettant d'éviter cet apprentissage : « non, étant donné le savoir et le savoir-faire d'autres membres de l'équipe » ; « [non,] par contre, je travaille en liaison étroite avec des informaticiens et je suis persuadé de l'importance réciproque de ces collaborations » ; « non, mais cela n'avait aucune importance compte tenu de la facilité de fonctionnement. Par ailleurs, mon mari — ingénieur généraliste — a assuré le côté vraiment technique » ; enfin, une experte en bases de données affirme ne pas vouloir s'« occuper de la 'cuisine' », car elle dispose pour cela d'un ingénieur.

Pour ces répondants, le faible besoin de culture informatique s'explique par la densité et la qualité de leur réseau social d'entraide.

Restent 22 personnes qui affirment avoir eu besoin de compléter leur culture technique, que celle-ci fût faible ou déjà conséquente. Certes, divers témoignages rappellent la facilité du code html ; d'autres manient sans vergogne l'euphémisme : « Elle [ma culture] était minimale, et l'est restée d'ailleurs », affirme une manipulatrice régulière des protocoles telnet et ftp, quand, enfin, certaines ont une consience aiguë des problèmes rencontrés :

« besoin de maîtriser protocoles rigoureux (syntaxe) ». Pour les uns, l'acquisition de ces connaissances s'est faite aisément : « oui, c'est même l'un des plaisirs et des enjeux de la création d'un site que de se retrouver contraint d'apprendre — ou de mettre à jour — des compétences nouvelles » ; « certes, il m'a bien fallu compléter cette culture, mais je l'ai fait avec plaisir et intérêt grâce aux publications papier actuelles et au soutien compréhensif des techniciens informatiques » ; « on apprend tous les jours en pratiquant, et c'est appréciable pour juger autant que faire se peut des différentes offres techniques / commerciales qui nous sont faites ».

D'autres ont mis en évidence les contraintes douloureuses de cet apprentissage, ou les conséquences prévisibles de leur trop faible investissement : « j'ai dû tout apprendre au fur et à mesure ; la volonté 'politique' et les fins visées pour la qualité du service ont porté le besoin et le souci d'apprentissage pour des outils et techniques qui par ailleurs ne m'attirent pas vraiment en tant que tels » ; « ce sont mes insuffisances techniques, et ma dépendance vis-à-vis des informaticiens qui empêchent la plupart de mes projets d'aboutir » ; « les lacunes de ma culture informatique risquent effectivement de me poser des problèmes ».

Évidemment, les personnes ayant le plus conscience de la complexité de la mise en place et de la maintenance d'un site web qui dépasse quelques pages affirment clairement avoir eu besoin de compléter leurs connaissances déjà solides : « j'ai toujours été un autodidacte en matière informatique... Je continue à m'autoformer constamment. En matière informatique où tout bouge constamment, qui n'avance pas, recule » ; « je me suis autoformé à la programmation en C, en bases de données... Je ressens le besoin de compléter ma culture, en autodidacte, car je n'ai que des lacunes » ; Cette culture, « obligatoire », plusieurs ressentent le besoin de la compléter « tous les jours ». Et le partage des tâches ne se décrète pas d'avance, au vu de leur complexité : « elle [la culture informatique] est indispensable, pour évaluer correctement ce qui est possible. C'est la culture d'une équipe car un tel travail met en jeu des compétences assez diverses : conception de logiciels (avec recherche d'information et sélection de pré-produits existants, veille technologique), achat et programmation du matériel, conception du produit scientifique, coordination entre les qualités scientifiques et les possibilités techniques ».

Ainsi, on peut admettre que l'implication dans un projet scientifique dans le cadre de l'internet nécessite de passer beaucoup de temps à assimiler des connaissances autant liées à l'informatique qu'à la discipline initiale du chercheur (ce besoin de double culture expliquant les réticences de nombreux chercheurs à utiliser l'internet dans un cadre professionnel). On retrouve aussi le besoin de constituer un réseau social dense pour arriver à de telles fins. La norme semble être la constitution d'une petite équipe motivée par un projet, équipe qui inclut des informaticiens. Cette référence à l'assistance des informaticiens revient très souvent, et on verra que plusieurs projets ont obligé les équipes à en embaucher. Ce clair besoin d'assistance nécessite parfois des efforts de socialisation : « beaucoup de sollicitations des collègues — à faire avec doigté ! ! ». Une seule fois, il n'a pas été entendu : « blocage total par l'informaticien, qui 'ne croyait pas à Internet[1]' ».

La maîtrise de l'informatique apparaît donc incontournable en même temps qu'elle conditionne une activité prenante et quotidienne.

2.3 Moyens disponibles ou obtenus

Les conditions dans lesquelles les enquêtés ont pu mener à bien leurs activités sont très variables suivant leurs institutions, que l'on peut regrouper en trois catégories. Tout d'abord celles qui étaient proprement « sinistrées », ou « rétive[s] à toute innovation », et les laboratoires dont les personnes qui désiraient consulter leur mail étaient condamnées à « s'enfermer dans la pièce sans fenêtre où résidait la seule machine connectée à Internet ». Ensuite, on constate une assistance passive des institutions : le réseau était présent,

1. Les expressions des enquêtés ont été laissées telles quelles. Aussi, on retrouvera parfois dans leurs citations des majuscules — et une absence d'article — pour les mots internet et web, alors que nous avons choisi l'option inverse dans le cadre de la rédaction de notre travail.

mais rien n'était fait pour inciter à la consultation ou à la production ; au contraire, l'inertie était de règle. Enfin, des laboratoires autonomes financièrement — souvent du CNRS —, dont les responsables ont délibérément choisi de s'impliquer dans la publication électronique.

En 1995, certains doivent tout construire : « [les ressources] n'étaient même pas inexistantes, mais négatives. La situation était sinistrée, au niveau micro-informatique. Il a même fallu créer un service. On n'avait pas d'ordinateurs, ou alors, sous-dimensionnés. Pas de logiciels. Pas d'imprimante, et pas de connexions réseau... L'institution était plutôt un obstacle. Tout ce que nous faisions au départ était informel (aide des informaticiens, conseils des copains, discussions pour les projets) ». Cette situation pouvait inciter les pionniers à quémander un accès au réseau auprès d'une autre institution : « en 1995, j'ai 'forcé' l'introduction de l'X dans le serveur de l'Y, et ai pu bénéficier d'un accès institutionnel. Cela dit, pendant près de 8 mois, j'étais l'unique détenteur d'un compte X, le responsable, le directeur, le rédacteur... ». Cette personne remarque de façon sarcastique : « la plus grande aide qui m'ait donc été apportée tenait à la totale indifférence de l'institution à son sort numérique ». On retrouve là le profond désintérêt des sciences humaines pour l'informatique.

L'intérêt pour les réseaux pouvait générer des réactions agressives : « l'aide de l'institution a surtout été matérielle, côté moral nous avons plutôt connu les attaques franches ou plus souvent encore lâches ». Un autre enquêté, qui disposait au départ d'un solide réseau social et technique, a vu lui aussi ses projets vite contrariés : « ensuite j'ai essayé de créer une petite équipe web [de quatre personnes, en 1995] qui, comme par hasard s'est trouvée rapidement dissoute avec l'aide efficace de la direction du X aidée par la direction du Y qui voyait d'un mauvais œil se créer une autre entité informatique ».

Sinon, l'inertie reste l'arme privilégiée des universitaires, même en 1999 : « j'ai été aidé de manière passive : j'avais accès à des ordinateurs, et un réseau, comme tous. J'ai été aidé de manière plus active par le responsable de l'informatique, qui m'a donné tous les accès nécessaires sans attendre que la lourde bureaucratie interne ait donné sur papier avec tampon tous les visas.

Cela m'a permis de travailller et de présenter un projet fini. Bien entendu, j'ai attendu que les autorisations soient réunies pour mettre en ligne le nouveau site web... ». Un autre enquêté témoigne d'une situation analogue : « je ne gérais directement aucune enveloppe de crédit permettant des investissements en informatique ; je devais donc négocier au coup par coup avec mon supérieur hiérarchique... Il a fallu susciter une véritable volonté politique au sein de la direction. Cela a pris environ 2 ans, avec des hauts et des bas ». Et pourtant, son projet était soutenu par l'État et la région. Un troisième décrit une situation aussi peu satisfaisante : « très peu d'aide de la part de l'institution. Budget pas prévu pour cet objectif au départ. Informaticiens très peu disponibles en interne pour la conduite du projet (soutien de principe, mais pas de possibilité d'aide réelle) ». Il est à noter que ces deux dernières personnes espèrent voir leur projet finalisé en 2001, alors qu'elles y travaillent depuis trois ans.

Cette inertie est telle que parfois, les enquêtés réagissent vivement quand on leur demande s'ils ont profité d'une assistance institutionnelle : « bien sûr que non ! », répond un membre actuel de l'Institut Universitaire de France, quand un autre rappelle l'absence d'aide à la formation en informatique : « il n'y avait pas d'encadrement, ni même de stage rapide. Il fallait épier discrètement les autres et essayer de comprendre ».

Au total, ce sont 21 personnes sur 28 qui ont dû compter majoritairement sur des ressources techniques et sociales extérieures pour mener à bien leurs projets, quand ceux-ci n'étaient pas ouvertement critiqués. Leurs statuts et âges étaient variables, tout à fait représentatifs de l'échantillon total, à une exception près : aucune d'entre elles n'appartient au CNRS.

Restent sept autres personnes, toutes en France, qui ont eu, semble-t-il, plus de chances ou plus de moyens. Mais cinq d'entre elles étaient responsables de leurs budgets de laboratoire : « [je disposais du] matériel du labo, et [d']un ingénieur » ; « [je disposais du] budget confortable d'un laboratoire que je dirigeais », et donc « [d'une] grande latitude sur le plan budgétaire », rappelle de son côté la personne qui a subi la résistance de l'informaticien qui refusait l'internet ; « je disposais d'un crédit suffisant au niveau direction générale du labo... plus les ordinateurs successifs nécessaires », « [nous disposions des ressources] d'un laboratoire CNRS... le financement d'un serveur a été aidé par le Ministère de la Recherche, puis le CNRS a donné un financement complémentaire... et nous a accordé un poste de technicien. [Avant,] nous avons pris sur le budget du laboratoire les vacations destinées à assurer l'élaboration du contenu de la revue » ; « la bibliothèque et le département informatique ont très vite compris l'intérêt de l'investissement dans ce domaine. Les difficultés ne sont ni financières, ni techniques. Elles concernent nos compétences, le temps qui file et nos capacités d'analyse, d'invention et d'organisation, compte tenu des autres obligations (enseignement notamment) qui pèsent sur nous... [nous avons aussi profité] de partenariats, d'appels d'offres gagnés etc. » ;

Restent enfin deux personnes, qui avaient de moindres responsablités, mais qui ont aussi trouvé aisément les moyens désirés : « les ressources ont été celles en provenance des sources publiques traditionnelles (ministère) mais aussi une très forte implication de l'institution régionale » ; « au début j'ai reçu une aide fondamentale : celle de ceux qui ont cru au projet et qui m'ont poussé à le continuer (directeur d'un côté, collègues, étudiants et collaborateurs de l'autre). Puis sont arrivés les moyens ». Cette dernière a réussi à impliquer huit institutions, françaises et étrangères.

Il apparaît que si l'on n'appartenait pas au CNRS, mieux valait vivre en province qu'en région parisienne quand on désirait s'engager dans l'internet universitaire : les institutions d'Ile de France (Grandes Écoles, universités, etc.) brillent par leur inertie. Eu égard au grand nombre de professeurs d'université à Paris, on est conduit à penser que beaucoup d'individus disposant d'un grand pouvoir scientifique se sont, entre 1995 et 2000, absolument désintéressés de l'internet et de ses potentialités pédagogiques, documentaires, ou éditoriales, quand ils n'ont pas tout fait pour en retarder la socialisation.

2.4 Un engagement coûteux

On imagine alors que la majorité des initiateurs aient dû dépenser beaucoup, en temps et en argent, pour mener à bien leurs projets.

2.4.1 L'argent

Neuf personnes considèrent que leurs dépenses étaient faibles ou nulles : certes, trois d'entre elles n'ont pas finalisé leur projet ; les autres n'évoquent que des frais téléphoniques, liés à leur connexion à domicile, qu'elles considèrent comme négligeables ou déjà intégrés dans leurs dépenses professionnelles habituelles (comme l'achat de journaux) ; on peut aussi supposer qu'elles disposent de bureaux correctement équipés et de budgets de recherche leur permettant d'éviter l'achat de machines sur leurs fonds propres, comme le rappellent deux personnes : « l'achat d'une machine personnelle remonte à 1986 à une époque où les implications financières n'étaient pas négligeables » ; « en fait, machine perso et modem livrés par le labo ; je ne paie que le téléphone ».

Ces frais, ajoutés à l'achat de documentations et parfois de logiciels, peuvent être estimés à environ 3000 F par an, si on respecte la sous-évaluation induite par les intéressés : « finalement, le montant est peu élevé si on le rapporte sur 5 années, mais j'avoue ne pas être en mesure de le chiffrer », rapporte une enquêtée qui avoue cependant avoir utilisé ses « fonds propres pour acquérir un modem, puis un ordinateur plus puissant, puis un modem plus rapide pour connexion ISDN. Idem pour l'abonnement au serveur ».

D'autres exemples témoignent de telles sous-estimations des dépenses : « difficile à dire. Nous n'avons jamais fait d'évaluation financière englobant tous les frais réels ou cachés (qui sont nombreux) » ; « non en théorie [je n'ai pas eu à faire appel à mon budget personnel pour mener à bien mes activités]. Néanmoins, par principe, j'ai aussi un abonnement et matériel payé sur mon budget personnel. Cela doit représenter une dépense annuelle d'environ 5000 F tout compris, mais très variable suivant les années (renouvelt de matériel) ».

Dans la majorité des cas, la moyenne se situe entre 5000 et 10000 F par an. la somme globale peut être explicitée par l'enquêté : « 5000FF/an. abonnements divers, notamment au câble » ; « dépenses considérables : 10000 FF en moyenne » ; « budget difficile à estimer, sans doute entre 5 et 10 000F par an ». Sinon, le coût du matériel acheté est détaillé[2] : « achat de machine pour avoir internet à la maison » ; « 1 machine, 1 modem et une imprimante (lors de mon arrivée à X en 1999) soit 9 000 francs » ; « achat de 2 machines personnelles depuis 1993 : bureau et portable, imprimante laser ; environ 10000 F/ an » ; « ma mise de fond personnelle pour mon équipement a été de l'ordre de 30.000 F TTC » ; « environ 7 à 8000 francs [par an] » ;

Parfois, le matériel personnel doit être transféré sur le lieu de travail : « je me suis longtemps (2 ans) servi de mon modem personnel au bureau (2500ff à l'époque, cela fait pleurer...). Plus l'ordinateur (mais cela j'en avais de toute manière besoin) » ; « on paie toujours de sa poche quand on veut foncer. Comme il n'y a plus de subventions, je paie moi-même actuellement un ordinateur personnel, un lecteur de saisie optique, une imprimante, une connexion Internet par câble. Disons $ 5000 CAN pour les douze derniers mois (1999-2000) ».

Deux personnes ont réussi à réduire les coûts en les intégrant dans leur comptabilité professionnelle : le médecin (« les dépenses peuvent entrer dans le cadre fiscal des frais professionnels — disons inférieures à 20.000f ») et un ingénieur, interrogé à propos de son bilan (« possibilité de bosser en freelance pour le privé donc 'beurre dans les épinards' largement réinvesti dans le matériel informatique personnel »). Mais ces situations atypiques ne doivent pas faire oublier que la passivité des institutions est souvent la cause de lourdes dépenses personnelles : « les collègues vraiment intéressés ne font pas confiance aux équipements de X ou de la fac : ils s'équipent à leurs propres frais et font de l'internet chez eux. En fait, l'institution universitaire est en retard par rapport aux comportements individuels » ; « sans ordinateur à la maison, avec un ordinateur pour tout le [laboratoire], rien n'était possible ».

2. Le délai d'amortissement des machines personnelles est alors estimé à quatre ans.

Ces témoignages prouvent que l'acquisition de l'appareillage et l'accès au réseau ont un coût conséquent, surtout pour des personnes qui utilisent l'internet depuis bien longtemps : « en 1994, une connexion internet coûtait 900FF/mois chez Oléane (avec une mise de départ de l'ordre de 5000FF si mes souvenirs sont bons) ». Ainsi peut-on affirmer que l'investissement des enquêtés sur cinq ans est plus proche de 30 000 F que des 15 000 F évoqués par les euphémistes. De telles dépenses ne sont pas négligeables pour des chercheurs qui ont souvent de modestes salaires et des promotions tardives.

2.4.2 Le temps

Ces chercheurs qui ne comptaient pas leur argent comptaient-ils leur temps ? D'entrée, certains manifestent une fois de plus leur engagement en critiquant la question « ce temps était-il pris sur le travail ou les loisirs ? », qui n'apparaissait pas très pertinente pour des personnes habituées à confondre les deux : « le travail d'un prof qui s'intéresse à son boulot tend à remplir tout l'espace disponible » ; « ce temps est pris sur mes loisirs bien que je pense que dans le cas du métier de chercheur, il devient difficile de séparer nettement travail et loisirs » ; « où se situe la frontière ? ».

Mais ce type de réponse traduit déjà une forme de motivation peu courante. Si les rares personnes qui se sont légèrement impliquées évoquent un « temps variable », ou « limité jusqu'à présent, et sur les heures dites de travail », une nette majorité déclare des propos tout autres.

Dix personnes consacrent une journée par semaine aux réseaux, autant prise sur l'activité professionnelle que sur les loisirs : « une journée par semaine, prise sur les loisirs » ; « 5 à 10 heures hebdomadaires [prises sur les loisirs], le plus important restant la production de contenu » ; « loisirs. 2 à 5 heures par jour », etc. « Je consacre à Internet (en dehors des autres activités informatiques) environ 10 heures par semaine, en confondant heures de travail et de loisirs : la frontière est impossible à définir, puisque je me connecte indifféremment depuis le bureau et le domicile » ; « environ 1/10e de mon temps de travail + loisirs » ; « sur les heures de travail et sur les loisirs soit à peu près une journée de loisirs par semaine », « très peu initialement. Beaucoup pour publication des cours sur pages web (5 h par semaine de plus qu'avant ; pris sur autres travaux) », etc.

Pour dix autres, la machine du temps s'emballe. Passe encore que les personnes effectivement payées pour travailler sur l'internet déclarent y passer la majorité de leur temps : « depuis 1996 c'est devenu mon travail à 100 % ». Un répondant, chargé d'imaginer l'informatisation d'une bibliothèque et de publier son catalogue sur le web, commence, lui à s'alarmer : « de plus en plus à partir de 1998 : c-à-d avec des périodes de pointe à plein temps (par exemple entre septembre et décembre 1999). En 2000 : en moyenne 3/4 jours par semaine (en équipe) et de plus en plus en 'heures sup.' (non rémunérées bien sûr) » ; mais les boutades du type « 25 heures sur 24 ! Travail _et_ loisirs. Une véritable catastrophe ! » vont de pair avec des propos plus mesurés d'un autre enquêté, qui évoque malgré tout « 16 heures sur 24, quotidiennement, aussi bien sur le travail que les loisirs, car le net procure un surcroît de plaisir à travailler et à partager ». L'acheteur du modem à 2500 F évoque, lui, un temps « considérable. [À mesurer] en milliers d'heures », quand le benjamin de l'échantillon répond : « beaucoup de temps ! Malheureusement, je ne peux pas donner d'évaluation. Ce temps était probablement pris à la fois sur mon temps de travail et mon temps de loisirs (je n'ai pas d'heures de bureau, cela m'est difficile de le dire) ». La personne qui rappelait qu'« on paie toujours de sa poche quand on veut foncer » précise à quel point la passion peut solliciter : « disons que je passe plus de temps à mon poste de travail chez moi que dans des cours ou comités sur le campus. [...] Un horaire 'heures de travail' n'existe pas. Créer un site web c'est en même temps boulot et loisir si on aime ce qu'on fait ». Enfin, un dernier témoignage est encore plus éloquent : « ça dépend des moments. Actuellement c'est en train de me bouffer la vie, il faudra que j'arrête ».

Ces pionniers se sont engagés dans un internet scientifique quelques années avant l'engouement des institutions publiques et des entreprises pour l'internet. D'ores et déjà, on a la preuve qu'ils ne pouvaient arriver à leurs fins sans posséder une solide culture informatique. Pour l'acquérir, mais aussi pour construire leurs sites, ils n'ont pas compté leur temps. Leur solitude, liée à leur intuition précoce, les a incité à acheter sur leur budget personnel du matériel, des abonnements, etc., autant de frais que leur institution refusait souvent de prendre en charge.

On pressent que les résistances à leurs projets n'étaient pas que le fait d'une hiérarchie peureuse ou traditionaliste, mais que de nombreux collègues considéraient cet engagement sur le web comme une chimère, voire une preuve du dilettantisme des pionniers.

Il nous reste donc à détailler et à expliquer le scepticisme collégial d'alors, vis-à-vis de l'initiative des enquêtés, comme vis-à-vis de l'internet en général.

Chapitre 3

L'environnement collégial

3.1 Un web désertique

Quand les personnes interrogées ont fondé leurs sites ou forums, la production électronique dans leurs disciplines leur apparaissait « inexistante ». Et souvent, l'insistance est mise sur l'opposition entre le « désert technique et conceptuel » français et les dynamiques d'outre-Atlantique.

Un historien se souvient qu'en 1997, « [l'état des contenus était] quasi inexistant en France et dans l'espace francophone ; les institutions anglo-saxonnes proposaient cependant déjà des sites très riches » ; un de ses collègues précise qu'alors, « la France n'en était qu'à ses balbutiements » ; un autre confirme : « plutôt décevant en histoire ; l'essentiel des outils indispensables étant toujours sur support papier ; toutefois, la situation évolue rapidement et bien : on commence à trouver du contenu ».

La même opposition se retrouve dans les autres disciplines : « en 1996, il n'y a avait pas vraiment de sites universitaires de géographie (hormis aux États-Unis) » ; dans le cas des lettres classiques « les ressources accessibles sur Internet sont devenues très rapidement excellentes, d'abord aux États-Unis, puis en Italie et en Allemagne, et récemment en Belgique. La France est à la traîne » ; un archéologue témoigne à sa façon de cette situation. Alors qu'il a fondé un site en 2000, il considère que son travail est encore « expérimental ».

En littérature, « [un] démarrage [se profilait en 1995] à Chicago et à Nice », propos confirmés par une autre enquêtée : « En donnant le sens '1995' au mot 'alors', je peux dire qu'il y avait à ce moment-là assez peu de choses dans le domaine des études françaises. Il faut pourtant mentionner le projet ARTFL (Université de Chicago) ».

Dans le domaine de l'économie ou de l'édition, la situation n'était pas meilleure en 1997 : « peu de pages personnelles riches en France parmi les universitaires de mon domaine (économie) ; beaucoup aux USA. Presque aucun cours ni en France ni ailleurs » ; « au départ quasi-inconsistant en France, sauf dans quelques créneaux très précis. Aux US, et très vite en GB, le réseau était déjà un outil naturel pour le monde scientifique et documentaire ».

Les philosophes, quant à eux, témoignent de « contenus assez épars, pas très consistants ». Mais l'un remarque « des listes de discussion plutôt vivantes » (évaluations entre 1994 et 1996). Cette importance des listes de discussion est aussi remarquée par le médecin : « les contenus officiels étaient débutants en France — les contacts se sont rapidement étoffés entre 'passionnés'. La masse d'information disponible actuellement (francophone et anglophone) est très importante et intéressante ». De même, « en documentation, le mail et les groupes de discussion constituent une véritable révolution : l'information circule vite et bien ».

Ainsi, ces précurseurs témoignent de contenus de qualité, mais localisés sur des sites étrangers, et de l'importance des listes de discussion pour débattre et s'informer. Autrement dit, l'internet était déjà un espace précieux pour les universitaires des sciences humaines dès 1995, à condition d'en exploiter toutes les ressources, d'y passer de longues heures, et bien sûr, de se familiariser avec les protocoles.

En 2001, il reste possible de réaliser des recherches sans utiliser l'internet, mais cela devient de plus en plus difficile : sans mail, on est « coupé du monde », et la non consultation du web prive le chercheur d'un type spécifique d'information. Ce qui apparaît aujourd'hui comme une évidence ne l'était pas auparavant : de tels propos étaient considérés comme absurdes et irritants entre 1995 et 1998. C'est bien pourquoi les enquêtés peuvent être présentés comme de véritables pionniers : ils ne se sont pas satisfaits de « surfer », mais se sont lancés dans une véritable politique éditoriale, alors que de telles pratiques n'étaient pas balisées. Aussi ne faut-il pas être choqué par des propos qui manifestent une claire conscience du caractère novateur de telles activités : « en géographie, en France, [contenu] inconsistant aux tout débuts, meilleurs aujourd'hui (notamment grâce à notre propre action) ».

Le caractère marginal, exemplaire de ces pratiques est indéniable, surtout à l'époque où, en sciences humaines, le web francophone n'était qu'une coquille vide.

3.2 Des collègues frileux ?

Afin de mettre en contexte leurs expériences, il était proposé aux enquêtés de préciser l'état d'esprit de leurs collègues, vis-à-vis de l'internet en général, et de leur initiative en particulier. Il va de soi que les réponses obtenues sont susceptibles d'induire deux types de biais contraires : d'une part, la valorisation des activités solitaires et la violence symbolique au sein du monde universitaire peuvent inviter les répondants à donner une image dépréciative de leurs confrères ; mais d'autre part, la reconnaissance obtenue par les pionniers, même si elle s'est produite dans des milieux qui n'ont pas les moyens de promouvoir la carrière des répondants, peut les amener à euphémiser les résistances — voire les agressions — qu'ils ont rencontrées dans leur environnement proche.

La perception de l'internet en cette tardive aube de sa diffusion en France était fort négative. Les maîtres mots sont « scepticisme » et « indifférence ». On se limitera ici au rappel des témoignages les plus explicites, sachant qu'ils rendent compte de la majorité des autres : « scepticisme largement partagé. Aucune attente (rien ne valant, comme chacun sait, une bibliothèque, un livre et un crayon) ». De telles positions n'évoluent pas avec le temps : « Mes collègues [...] de X étaient en majorité inertes. Si on les interroge, ils se montrent [encore aujourd'hui] dubitatifs ou ironiques sur l'utilité du web ». Un philosophe décline de la façon suivante les positions de ses collègues : « 1. détestaient (et détestent) Internet ; 2. désintéressés ; 3. intéressés mais incapables de maîtriser ; 4. n'avaient pas compris les possibilités du médium ; 5. quelques enthousiastes quand même ». La situation était à peine meilleure au Canada : « en rentrant en 1995 [de congé sabbatique], j'ai découvert avec plaisir que plusieurs jeunes collègues avaient déjà créé des pages et sites web, dont un qui était chargé de créer un site pour le département. Mais la très grande majorité de mes collègues restaient indifférents, voire hostiles ». Et une telle situation semble prévaloir en 2000, d'après un spécialiste de littérature : « en 1993 : scepticisme à peu près total de la part de la hiérarchie directe (le bâtiment a tardé à être câblé). En 2000 : scepticisme encore partiel d'une partie de la communauté littéraire (la moitié ?) ».

Certes, on peut excuser une telle inertie, ou s'en inquiéter : « En 1999, les professeurs du [laboratoire] se mettaient tout juste à utiliser le courrier électronique. Je dois dire qu'ils s'y sont mis avec une relative célérité la première année de la mise en place des listes électroniques. Il reste encore, cependant, un irréductible qui ne veut pas de compte email », rappelle un historien. Un autre rappelle l'ambivalence des sentiments vis-à-vis de l'internet : « le mépris (surtout chez les anciens) ou une fascination idiote et peu critique (chez les plus jeunes) Pas vraiment satisfaisant ! ». Un troisième complète ce panorama : « difficile à déterminer ou à reconstituer. Il semble que le terme le plus approprié serait une indifférence polie de la plupart, parfois une ignorance totale, voire une certaine condescendance chez certains, souvent les plus installés et les plus âgés, en partie, mais c'est une hypothèse un peu gratuite parce que dans mon univers d'origine tout ce qui peut se rapprocher de la technique est dévalué et peut difficilement être pensé comme création ou réflexion ». On retrouve là les « classiques » des représentations universitaires : mépris de la technique et du laboratoire, refus d'imaginer que l'outillage mental assimilé puisse se moderniser.

3.3 Une grande solitude

Si, comme le dit un professeur de philosophie, ses collègues étaient composés à « 99 % d'indifférents narquois, et 1 % de curieux 'prosélytes' », on peut concevoir que cette combinaison majoritaire de mépris et de condescendance ne facilitait pas la tâche des pionniers. Une telle attitude grégaire face à l'innovation — prise en un sens large et non pas dépendant de techniques nouvelles — s'intègre pleinement dans le registre de violence symbolique propre au monde universitaire déjà évoqué dans notre seconde partie. Mais on y a aussi vu que le monde universitaire ne pouvait étouffer toute forme d'engagement ou de marginalité scientifique.

On reste malgré tout étonné de la ténacité des chercheurs face aux formes collectives de dénigrement et de ségrégation. Une première personne les rappelle, même si elle s'affirme chanceuse : « [Au sujet de mon environnement professionnel,] j'ai été particulièrement encouragé par un de mes [anciens] collègues qui m'a incité à écrire, à tirer les leçons théoriques de mes expériences professionnelles, une chance et un vrai soutien ! » ; sinon, « chez les bibliothécaires [les réactions étaient] : 'c'est bien... mais est-ce bien sérieux ?' Vaguement condescendant mais cela change. [En 2000, notre site], notamment, a su convaincre les professionnels. »

Ces pratiques collectives de mise à l'écart ne sont pas sans effets sur la carrière. À la question « vos activités vous ont-elles nui professionnellement », quelques réponses sont sans appel : « oui, cela a servi de prétexte pour me placer première collée au concours de recrutement au CNRS. Par miracle j'ai réussi à passer l'année d'après » ; « Oui. Pour un [chercheur de la discipline] X, il n'y a pas de poste universitaire à profil informatico-X. Le temps passé à faire de l'informatique et de l'internet n'a pas été employé à la thèse d'habilitation. Cette compétence acquise et le temps passé ne peuvent se monnayer en diplômes ou titres universitaires » ; « au sein du Y, elles se sont traduites par une mise à l'écart ».

D'autres personnes, moins menacées professionnellement, car professeurs d'université ou équivalent, sont plus nuancées, mais ont assurément vécu des situations difficiles : « plutôt une sorte de relégation, qui conduit à accroître le sentiment de solitude. Il se peut qu'actuellement les choses changent un peu » ; « le travail sur Internet paraissait superflu à la plupart des littéraires, mais j'ai été soutenu par les collègues scientifiques » ; « [cela m'a nui] peut-être un peu auprès du CNU ou de la commission de spécialistes. Mais ce n'est pas sûr » ; « je ne crois pas, mais ce n'est pas impossible ».

Et quand l'universitaire n'est pas menacé par ses confrères, c'est d'ailleurs que viennent les tracas : « soutien moral au sein de l'administration de mon département. Opposition radicale de la part du responsable financier des X (imprimés), qui voyaient dans [leur version électronique] un concurrent dangereux (il a fait bloquer l'adresse Internet des X au niveau du serveur !) » ; « il y a eu au départ (1996) une tentative pour m'interdire de publier mes cours sur le web. Mais très vite, 'on' a été obligé de reconnaître que c'était contraire à mes droits fondamentaux, sans compter que c'était contraire à la liberté académique, et que j'avais suffisamment de contacts dans la presse spécialisée internet pour épancher mes états d'âme, ce qui aurait fait vilain. En fait, mon directeur était très favorable, sa direction, disons... moins ».

Ces réponses laissent imaginer toutes les difficultés qu'ont pu rencontrer les enseignants du second degré et les universitaires moins bien armés (en termes de culture informatique, de réseau social, ou simplement de force de conviction) quand ils ont essayé de mener à bien des projets analogues à ceux étudiés ici. La faible taille de notre échantillon prouvait déjà le très faible taux de succès de telles initiatives, ce que confortent les témoignages de solitude que nous avons recueillis [1].

1. En revanche, si l'enquête avait été menée un an plus tard, cet échantillon aurait été plus conséquent : des professeurs de lycée construisent des sites d'une qualité remarquable, ce qui accessoirement pose de redoutables problèmes de concurrence aux universités qui s'engagent en 2001 dans la publication électronique. Mais en retour, l'explicitation des conflits vécus aux premières heures du développement du web en sciences humaines aurait peut-être souffert de cette situation nouvelle, où le web était déjà passé dans les mœurs.

3.4 L'informatique accusée

Quelques enquêtés détaillent bien en quoi les résistances de leurs collègues vis-à-vis de l'internet sont aussi rationnelles, témoignant d'un rejet d'une informatisation aussi peu pensée que prétendument indolore[2]. Il s'ensuit deux types de peurs, non pas fantasmées mais bien légitimes, et ces inquiétudes (ou lassitudes) mettent en évidence le fait que, pour la majorité des universitaires l'internet est bien un élément de l'informatique.

Déjà, les réseaux tentaculaires ont mauvaise réputation : « crainte que le résultat n'engendre de nouvelles difficultés dans le travail (exemples de grands projets informatiques 'foireux' et médiatisés, comme celui de la BnF...) ». De telles défiances apparaissent aussi dans le domaine de la santé : « méconnaissance ou résistance face à l'informatique, mais également expérience des dysfonctionnements pouvant créer des situations très critiques en pratique médicale — je fais là plutôt allusion à la mise en réseau et à l'informatisation des services hospitaliers parfois bien laborieuse, avec des écueils tout de suite évidents et des bénéfices longtemps attendus ».

Ces témoignages s'inscrivent dans une mémoire collective qui va bientôt avoir 30 ans : à la fin des années 1970, la CNIL[3] a été fondée suite au débat engagé par les risques de liaison entre fichiers de la police et de la Sécurité sociale. Déjà l'informatique en réseau inquiétait. Ensuite, ce risque a été partiellement oublié[4], mais d'autres fiascos ont fait parler d'eux, comme par exemple le système de billetterie de la SNCF Socrate[5].

2. Cf. Le paragraphe 4.3 page 168.

3. Commission nationale de l'informatique et des libertés.

4. Et, dans les années 1990, on peut s'étonner de voir la CNIL rester trop fidèle à sa mission, à savoir surveiller l'État (et donc empêcher les fonctionnaires que sont les chercheurs de croiser des données individuelles, même quand leur déontologie est manifeste, cf. le rapport Silberman : http://www.education.gouv.fr/rapport/silberman/default.htm), et en même temps faire preuve d'un laxisme déconcertant dès que des entreprises privées croisent et recoupent des fichiers pour multiplier les agressions à la vie privée, notamment sous forme de harcèlement publicitaire ciblé.

5. [Nau93].

Ainsi, la méfiance des universitaires face à l'internet est-elle liée à la crainte de voir apparaître une nouvelle « usine à gaz ». Les premiers articles de la « presse sérieuse » au sujet de l'internet étaient très orientés, voire stupides[6]. Or, les universitaires susceptibles d'avoir un pouvoir de décision n'avaient que la presse pour se faire une opinion sur l'internet : les spécialistes étaient rares, sinon peu écoutés. C'est ainsi que le réseau des réseaux pouvait être imaginé comme un (petit) réseau du type de ceux que l'on apprenait à exécrer en France.

À cette attitude dubitative face aux grands projets réticulaires, s'ajoutent les constats d'échec devant l'ordinateur individuel. Un enquêté rappelle l'état d'esprit de ses collègues : « 'exalté' pour les passionnés d'informatique et beaucoup de méfiance et de résistance face à l'outil informatique souvent vécu comme peu fiable, contraignant, une perte de temps chez les autres ». Ce qui renvoie aux conditions anarchiques de la diffusion des ordinateurs dans les années 1990 : pendant longtemps, les budgets de recherche permettaient d'acheter des machines, mais pas d'embaucher du personnel ; les effets de mode, puissamment orchestrés par une publicité tapageuse, ont été déterminants pour l'achat de machines et de logiciels coûteux et inefficaces.

On comprend alors mieux pourquoi des personnes énervées par les pannes et *bugs* du système de la BNF, mais aussi par le fonctionnement aberrant de leurs machines, aient pu émettre des doutes face aux discours prétendant que ces ordinateurs et leur mise en réseau puissent transformer structurellement leurs méthodes de travail.

3.5 Force de la tradition

Mais les errances de l'informatique et des grands projets n'excusent pas complétement l'attitude de nombreux universitaires face à l'internet. Les personnes interrogées donnent deux types d'explications à l'immobilisme généralisé auquel elles ont été confrontées : d'une part, la passion du pouvoir et de

6. Cf. le paragraphe 2.1 de l'annexe (page 397).

la propriété, voire de la rente, d'autre part, une série de peurs, cette fois-ci moins rationnelles, très liées à l'incompétence informatique et au refus de changer ses habitudes.

3.5.1 Routine

Au chapitre de la propriété, trois témoignages sont sans appel : « méconnaissance et le jeu habituel de pouvoir dans une petite institution », « jalousie ordinaire » et « rivalités collégiales et incompréhension des motifs fondamentaux qui vont à l'encontre du sentiment de possession d'un secteur qui semble inéluctablement attaché en sciences humaines et lettres à la maîtrise — réelle ou illusoire ! — d'un domaine de recherche ». Ce partage des domaines, des territoires, pourrait-on dire, s'accorde avec un respect des monopoles d'autrui, et se double de pratiques grégaires : « facteur générationel ; incompétence ; manque d'élan vers la nouveauté ; désir de rester intégré aux pratiques de la communauté existante » dit l'un, conscient malgré tout de l'esprit opportuniste de ses collègues : « en tout cas prêts à sauter dessus si et quand cela commencera à fonctionner ». Tradition et pouvoir vont de pair : « l'émergence de nouvelles techniques signifie généralement une remise en cause de savoirs anciens acquis ; peu ont le goût du changement et beaucoup craignent de voir leur pouvoir menacé. Les réactions hostiles ont souvent pour origine la peur. Je dérangeais ».

Cette peur du changement, dans le quotidien comme dans les règles de la concurrence, apparaît fréquemment : « ignorance et une certaine forme de conservatisme » ; « crainte de ne pas savoir faire évoluer des compétences acquises depuis longtemps et des pratiques devenues routinières, donc rassurantes aussi » ; « rupture trop grande avec leurs habitudes de travail traditionnelles. Impression que ces projets étaient purement opportunistes ou formels, et sans enjeux véritables ». On comprend alors l'attrait de l'immobilisme : « pour un grand nombre d'individus, il fallait fournir un effort conséquent pour comprendre la situation. Rien ne semblait à leurs yeux pouvoir justifier un tel effort ». On mesure là les effets d'inertie produits par l'acquisition d'un outillage mental spécifique, qu'il convient, par esprit d'économie, d'actualiser le moins possible.

3.5.2 Valeurs

Évidemment, un tel conservatisme est intimement lié à une claire perception des valeurs propres au monde universitaire[7]. Celles-ci s'échelonnent suivant un axe qui va de la manipulation technique, décriée, à la publication, valorisée. L'adoption simpliste de la loi du « *publish or perish* » qui structure le paysage économique de l'université, associée à une surenchère théorique négligeant l'importance du laboratoire du chercheur, expliquent en grande partie le manque de discernement des universitaires face à l'internet : peu ont compris que le renouvellement de leur science passait par une modernisation — ou au moins par une adaptation — de leurs méthodes de travail et de recherche, et que l'informatique n'était pas qu'une technique visant à rabaisser le chercheur au niveau de la secrétaire, mais s'inscrivait dans la panoplie de ces méthodes.

On comprend alors que les disciplines les plus traditionalistes, comme les sciences de l'antiquité et les études littéraires[8], aient été les plus fermées aux sirènes de la modernité, qu'elle soit prônée par des universitaires ou les publicitaires : « incompréhension du médium, opposition devant tout changement technologique », « préjugés sur l'incompatibilité entre technologie et littérature ». Ainsi, certains des enquêtés étaient obligés de faire appel à une constante pédagogie, qui ravirait les enseignants du primaire : « on retrouve la séparation entre les activités intellectuelles et techniques. En ma qualité de X, mes activités informatiques passaient pour être hors-sujet. D'où ma parade constante : montrer qu'Internet (ou telle autre activité informatique) sont au service de [ma discipline] ». Un historien analyse ces réactions technophobes : « en large partie une totale absence de culture technique, très répandue, qui empêchait de percevoir les possibilités ouvertes par le réseau.

7. Cf. le chapitre 1 de la partie II (page 99).
8. Cf. le tableau 4.1 (page 172).

S'y ajoutait aussi une méfiance pour l'univers de la science et de la technique, facilement assimilé à de la cuisine sans grand intérêt, soit un univers de valeurs sans doute pas très différent de celui des critiques d'art ou des critiques dramatiques voyant arriver la photographie ou le cinéma [...] beaucoup considèrent encore sans doute la production scientifique comme une création exprimant sans l'aide de moyens mécaniques une individualité forcément exceptionnelle »[9]. Dans cette opposition primaire entre technique et esprit pur, l'idée que les processus de la pensée puissent être altérés par une modification de l'écriture était (et reste souvent) inconcevable.

3.5.3 Pouvoirs de papier

Ce mépris pour la technique va bien sûr de pair avec une forte inquiétude quant à la pérennité du système traditionnel de valorisation du chercheur : la publication d'un article dans une revue savante imprimée.

Cela peut être expliqué de façon ingénue : « [...] d'autres trouvaient ça sympathique, gentil, bref ils croyaient que je perdais mon temps (il est vrai qu'Internet 'rame' à certaines heures) avec des choses qui ne pouvaient pas être scientifiques, puisqu'elles n'étaient pas imprimées et publiées sur du papier. Comment citer une source virtuelle, inexistante ? » ; L'économiste est plus direct : « beaucoup de scepticisme. Craintes de pillage des travaux ainsi diffusés (pb de *copyright*) » ; cette relation à la norme, et à l'absence d'audace est la règle : « bof ! Et puis, tout le monde est obsédé par la liste d'articles 'sérieux' à publier ; pas question de mettre une idée neuve (si on en a !) en ligne — par peur de se la faire 'piquer' ». Ainsi, la peur devient multiforme. Elle touche la crainte du vol (d'une idée, d'un texte), la crainte de l'illégitimité (abscence de reconnaissance, de validation des travaux), voire de la perte de confidentialité[10]. Et la peur paralyse : « il y avait, il y a toujours,

9. On pourra remarquer qu'une telle attitude n'est pas propre aux sciences humaines, puisque la valorisation de l'esprit pur, qui n'utilise que le papier et le crayon pour exprimer son génie, était aussi attestée chez les mathématiciens dans les années 1980.

10. Cf. le paragraphe 3.2 de l'annexe (page 403).

surtout une méconnaissance totale d'Internet comme outil de publication, et ce à tout point de vue. Les gens ne savent pas ce qu'il est possible de publier mais ils s'interrogent aussi sur beaucoup de questions qui se posent d'ailleurs (mais qui les paralysent souvent) : pour qui écrit-on ? ne risque-t-on pas d'être pillé ? etc. ».

Du coup, ce fameux scepticisme, souvent teinté de mépris, peut s'afficher : « mes articles sur Internet étaient accueillis avec intérêt par ceux qui avaient envie de s'y mettre, et avec indifférence par les autres » ; « soyons clairs. Bien que notre premier numéro ait été salué par le corps professoral de l'X et d'ailleurs, ainsi que par nos pairs, il n'y avait guère de monde qui pensait qu'un numéro deux et encore moins que des numéros trois et quatre seraient mis en ligne. Or, c'est fait » ; « [notre initiative] était sans doute assez largement vue comme un perte de temps, ou un hobby ».

Dans cette compétition entre revue papier et revue électronique, les opportunités offertes par l'internet, notamment en matière de possibilité de mise à jour ou de correction d'articles, étaient encore moins perçues, ou, quand elles l'étaient, jugées encore plus menaçantes pour le statut de l'auteur, comme en témoigne une des personnes les plus actives : « enthousiasme pour beaucoup, scepticisme pour quelques-uns. Peu pressentaient les changements apportés par l'édition électronique. Encore aujourd'hui, certains sont réticents à l'idée que des modifications ou des corrections puissent être apportées après publication (la notion de version originale du texte persiste, l'archivage est mis en question). Les possibilités d'hypertexte ne sont guère employées. Les questions techniques et juridiques de *copyright* ne sont pas encore élucidées, cela constitue un frein pour certains ; beaucoup pensent que ce mode de diffusion n'est pas reconnu par les instances d'évaluation au même niveau que le sont les publications traditionnelles. C'est un frein encore plus important ».

De façon générale, les détracteurs de l'internet apparaissent « très sceptiques sur le succès possible d'une revue électronique, qui manquait d'après eux de légitimité. Depuis, là aussi, les choses ont tendance à changer, même si bcp considèrent encore que, dans la hiérarchie des publications, une publication électronique 'vaut' moins qu'une publication papier, en dépit du sérieux reconnu du comité de rédaction et de la qualité des articles ». Mais l'auteur de cette remarque met aussi en évidence l'incohérence ce tels raisonnements : « crainte de la concurrence du support électronique par rapport aux publications 'classiques'. De toutes façons, de nombreuses revues sont concurrentes entre elles... ». Les chercheurs méconnaîtraient-ils cette concurrence entre revues traditionnelles ? Connaissant la hiérarchie implicite de ces revues, on en doute. On sait que leurs positions relatives font l'objet d'incessants enjeux de pouvoir, et on imagine que les universitaires, déjà bien occupés à multiplier leurs interventions politiques dans un domaine qu'ils maîtrisent bien, n'ont guère envie de voir de nouveaux objets comme les nouvelles revues électroniques perturber les règles d'un jeu déjà complexe : le seul registre sur lequel fonctionnent ces revues est bien celui de l'économie symbolique[11].

On peut donc en conclure, comme le suggèrent plusieurs enquêtés, que l'internet était refusé principalement parce qu'un engagement dans l'édition électronique malmenait, tout en la mettant à jour, l'architecture du pouvoir symbolique universitaire.

3.6 Une lente reconnaissance extérieure

On a vu que les pionniers ne pouvaient trouver le soutien escompté au sein de leurs collègues proches : mépris, mises à l'écart, et bien sûr refus de promotions étaient la norme.

Toutefois, quelques-uns ont réussi à valoriser leur marginalité : en réponse à la question « au contraire, en avez-vous tiré rapidement des bénéfices ? Si oui

11. Rappelons que les arguments sollicités dans le domaine de l'édition non-savante (rémunération de l'auteur, profit de l'éditeur), ne valent pas ici : les revues savantes imprimées en sciences humaines ne paient pas leurs auteurs ; elles sont coûteuses, systématiquement subventionnées. Leurs éditeurs sont rarement des entreprises multinationales disposant d'un monopole les incitant à appliquer de façon optimale les « lois du marché », comme cela peut se produire pour l'édition scientifique.

lesquels ? », un enquêté pense que « participer à la création d'une revue électronique a plutôt confirmé l'image de chercheur innovant — également due à d'autres activités —, ce qui m'a je crois aidé par la suite à obtenir certaines reconnaissances. (ex. : promotion à la 1$^{\text{re}}$ classe des Maîtres de Conférences dès que je l'ai demandée au CNU ; nomination — sur projet transdisciplinaire — à l'Institut Universitaire de France comme membre junior) ». Deux autres expriment une même impression positive : « [la participation à l'équipe] X et au réseau Y m'ont au contraire permis de pouvoir trouver un autre poste plus intéressant », rappelle la personne auparavant « mise à l'écart » ; « sûrement dans les responsabilités qui m'ont été confiées à l'X (dept Y et maintenant institut Z) ».

Mais souvent, les bénéfices ne sont que personnels ou symboliques. Le précédent témoignage est aussi complété par les propos suivants : « [mais je n'ai reçu] aucune rétribution supplémentaire. [Sinon], la possibilité de récupérer des budgets de recherches et, plus encore, l'élargissement très rapide de mon collège invisible perso et de mes ressources documentaires ».

Ces acquis — construction d'un réseau de collègues, capacité à tirer parti des ressources documentaires de l'internet, acquisition de nouvelles compétences professionnelles — sont fréquemment cités : « je crois que j'ai été un peu plus vite au courant de certaines mutations dans mon domaine de recherche, mais pour en tirer bénéfice il faut le publier, le faire savoir, ce qui est un autre travail. Plus profondément, Internet m'a vraiment élargi les idées, et rendu curieuse » ; « difficile à évaluer actuellement. Mais je gère les sites de deux sociétés savantes, ce qui permet d'être au point techniquement, aux dépens de mes réelles compétences » ; « à part l'acquisition de compétences, et un fonctionnement plus efficace de certaines activités de recherche il ne me semble pas [en avoir tiré de bénéfices] » ; « bénéfices actuellement essentiellement dans le domaine de la formation continue, des échanges de données ».

La personne qui évoquait sa relégation au sein de son université témoigne clairement des effets de réseau : « en revanche, à l'extérieur de l'université française, il semble que ce travail m'ait permis de toucher un public large intéressé par les questions touchant à l'histoire de la langue et des représentations du langage en secteur français [...]. Ce travail a permis en particulier de nouer des relations et de créer des connexions inattendues, en France (à X) et à l'étranger (à Y) ».

Même dans le cas d'un engagement dans l'internet suite à un choix institutionnel, la reconnaissance vient lentement : « la période d'incertitude quant à la volonté politique de réalisation a été un peu pénible comme toujours quand on travaille sans savoir si une réalisation verra le jour [...]. Depuis que le projet est accepté, il y a une véritable reconnaissance du travail accompli les années précédentes et du travail exceptionnel mis en œuvre pour la réalisation. La reconnaissance reste symbolique (entendre : aucune rémunération supplémentaire pour les heures en plus, mais nous sommes tous dans le même cas, alors, on finit par avoir presque honte de songer au manque à gagner). C'est déjà ça. Je pense aussi avoir acquis en peu de temps une expérience multiple : non seulement une meilleure connaissance dans le domaine, mais aussi sur les plans administratif, financier, qui accompagnent nécessairement la confection de tels dossiers, laquelle expérience pourra m'ouvrir d'autres horizons professionnels ».

Parfois, les pionniers doivent se satisfaire de la notoriété : « une autonomie créative. Une certaine publicité dans les salons où l'on cause » ; « bénéfices financiers nullissimes. Une certaine reconnaissance institutionnelle et éditoriale » ; « être coordinateur de X lors de son développement a été un pari au départ car cela m'a pris beaucoup de temps mais je pense que cela m'a servi au cours de mon doctorat pour 'me faire un nom' auprès de personnes qui apprécient beaucoup le travail que nous faisons. C'est d'ailleurs pourquoi je ne vais pas tarder à démissionner de mon poste car si la figure du coordinateur n'est pas toujours au centre de la revue, elle est quand même favorisée. Chacun son tour d'en profiter ! ;-) [Mes bénéfices se situent donc] au plan du capital symbolique essentiellement. Certainement pas financièrement ! ». Et le chercheur bascule parfois dans le groupe des personnalités publiques : « cela n'a été valorisé que tardivement (maintenant cela fait vraiment chic). [...] Ah oui, j'ai aussi fait une couverture de magazine professionnel d'informatique, et on m'a classé parmi 'les gourous de l'Internet' pendant un petit temps. Donc, du point de vue symbolique c'est pas mal. Disons que cela rembourse au symbolique les frais d'achat informatique... On peut faire des piges aussi, il n'y a pas grand monde qui ait quelque chose d'original à raconter là-dessus. Mais après, si l'on veut être sérieux, on passe plus de temps sur sa machine qu'à valoriser ce qu'on fait avec ».

Mais la reconnaissance a aussi ses travers professionnels. Les collègues d'un pionnier savent que son investissement « limite [son] activite de recherche » ; cela se traduit par « un très fort gain en notoriété, mais pas forcément en considération » et peut donc nuire. Dans le monde universitaire, la gestion de la réputation médiatique obéit toujours à des lois complexes.

Restent les personnes qui n'ont acquis qu'une reconnaissance très localisée, sans autre forme de compensation : « symboliques : on passe par moi pour mettre de l'information » ; « j'en ai tiré, je crois, une certaine visibilité au sein du laboratoire X. Pour le moment, je crois que c'est le seul bénéfice direct que j'ai pu retirer ».

Car toutes ne réussissent pas à négocier cette forme de capital typique du monde universitaire : « des bénéfices sur le plan symbolique. Mes compétences sont reconnues, proclamées, utilisées, mais non rétribuées par un meilleur poste ou une perspective d'évolution de carrière ». D'autres s'en sortent mieux : « sur le plan symbolique, et plus tard comme compétence demandée sur le plan professionnel » ; « tout cela a été pris en compte avec beaucoup de retard » ; « aucun bénéfice financier. Bénéfice symbolique important, au sein de mon institution comme ailleurs (notamment en France) ».

Et naturellement, la relation à la publication imprimée revient en force : « je crois que je pourrais en titrer des bénéfices dans l'avenir. Mais pour l'instant cela a été surtout une perte de temps. Si au lieu de m'occuper d'Internet je m'était consacrée davantage à la publication de livres, aujourd'hui ma réputation auprès de mes confrères serait établie de facon plus solide. Heureusement que j'avais publié avant de m'intéresser à Internet et que j'ai continué de le faire, autrement personne ne me prendrait en considération » ; « beaucoup pensent que ce mode de diffusion n'est pas reconnu par les instances d'évaluation au même niveau que le sont les publications traditionnelles. C'est un frein encore plus important ».

Quelques arguments peuvent expliquer l'ostracisme qu'ont subi les premiers « éditeurs électroniques » en sciences humaines : par exemple, l'inconsistance du web francophone et les désorganisations qu'induisait l'informatique, avec ou sans réseaux. Mais ces constats étaient à leurs yeux autant d'excuses pour ne pas perturber le fonctionnement d'un monde universitaire régulé par un mépris pour la technique et la valorisation de la propriété privée, fût-elle intellectuelle. Les pionniers, en s'écartant de pratiques grégaires, ont pris de gros risques, et on ne s'étonne pas que les avantages obtenus à persister dans la publication intensive sur le web ont été avant tout d'ordres privé et intellectuel, et non pas professionnel. Cependant, en constituant un réseau de collègues avec lesquels ils échangeaient des informations scientifiques, partageaient des expériences, construisaient des sites web savants et des programmes de recherche, la plupart des innovateurs ont fini par acquérir une forme de reconnaissance, et parfois une réelle renommée, médiatique ou scientifique. En s'impliquant dans les modes d'écriture contemporains, ils sont devenus plus au fait de la recherche (internationale) dans leur discipline en même temps qu'ils ont gagné une meilleure lecture du fonctionnement de leur monde professionnel.

Chapitre 4

En l'an 2000

On pourrait penser qu'à l'automne 2000, les témoignages précédents des enquêtés renvoient à un passé désormais révolu : on ne peut plus passer devant une affiche publicitaire ou un kiosque à journaux sans voir de multiples mentions à l'internet ; en multipliant les financement de recherches, les politiques publiques françaises comme européennes accordent aussi une grande place à cet objet, tantôt vu comme médium (enseignement à distance, préservation des patrimoines culturels, etc.), tantôt comme technique industrielle. Et pourtant, les commentaires actualisés des enquêtés prouvent que ces multiples formes d'injonction technologique ne sont pas suivies d'effet ; on va le voir, la situation dans le monde des sciences humaines a très peu évolué.

4.1 Des activités prolongées et amplifiées

En réponse à la question « quelles sont aujourd'hui vos activités liées à Internet ? Sont-elles dans le prolongement direct de votre initiative ? », continuité et développement sont les maîtres mots des enquêtés ; 20 personnes le rappellent. Pour elles, cela signifie tout d'abord passer beaucoup de temps à réaliser des tâches *a priori* insignifiantes et routinières[1] : s'informer, en

1. Comme on peut s'en douter, l'utilisation de l'internet à des fins scientifiques n'implique pas pour autant la disparition des activités besogneuses typiques de la recherche.

consultant les moteurs de recherche et en s'abonnant aux listes de discussion pertinentes, échanger, en lisant son courrier électronique et en y répondant : « maintenant Internet c'est : d'abord le mail, pour les messages quotidiens et pour quelques listes de discussion scientifiques qui marchent bien et auxquelles je suis fidèle. Ensuite la gestion de mes signets, qui me renvoient à des sites essentiels pour moi. Enfin, des recherches 'à l'aveugle' quand il me manque une information » ; « [je joue] de plus en plus le rôle de relais d'infos diverses entre le monde scientifique anglo-saxon et mes collègues et étudiants français » ; « très forte activité de recherche de contenus sur l'Internet [...] Correspondance : j'y passe des heures... » ; « réalisation en cours de ce projet + consultation des listes de discussion liées à cette activité ».

Ces activités de consultation et de communication arrêtent souvent les chercheurs qui découvrent l'internet. À plusieurs reprises, des enseignants de l'ENS ont formulé la remarque suivante : « je ne réponds pas à mon mail, car sinon, j'y passe des heures et je ne peux plus travailler ». Il est toujours difficile de comprendre pourquoi de tels universitaires excluent le courrier électronique de leurs activités professionnelles, ce qu'ils ne font pas pour leurs communications téléphoniques[2].

Restent les activités de production. Les enquêtés sont désormais responsables, non pas d'un, mais de plusieurs sites : « je suis cowebmaster du réseau X, webmaster du laboratoire Y, de l'école doctorale de Z » ; « j'ai mentionné ci-dessus mon activité principale mais j'aurais pu ajouter que j'ai collaboré à la création de plusieurs sites (association X, doctorat conjoint des universités Z, T, U). Par ailleurs, [...], je peux mentionner à nouveau la création de plusieurs listes de discussion [...dont] une dans le cadre d'un cours effectué par vidéo-conférence entre les universités A, B et C » ; « oui, j'ai développé ce qui m'intéressait et notamment des sites[3] [...] je fais une revue avec mes étudiants (http://www.x.fr) ».

2. Ceci met en évidence, non pas la différence entre un échange oral traditionnel et un échange écrit, plus rare (la lettre est un élément essentiel de la relation entre universitaires), mais la légitimité sociale accordée à un mode d'échange.

3. Trois sites personnels et un d'enseignement.

Ces activités d'enseignement autour de l'internet prennent de plus en plus d'importance : « par ailleurs, je pousse mes étudiants à employer Internet pour leur faire chercher et trier de l'information, tester un sujet de maîtrise. J'ai fabriqué et testé un module de licence au printemps 2000 » ; « [mes activités] se situent dans un prolongement naturel. À la fois sur le plan scientifique [...] et à un niveau de vulgarisation et de diffusion de connaissance » ; « je fais également des formations à l'Internet pour les X » ; « maintien de pages personnelles avec travaux et cours, + liens » ;

Mais tout cela reste parfois précaire et fragile, notamment quand les pionniers sont en thèse et n'ont donc pas encore un poste stable. Il s'agit alors de songer à la survie du projet : « Je continue à gérer les listes électroniques et le site web, mais de loin seulement. Mon but est de réussir la transition (c'est-à-dire que mon initiative me survive) » ; « assurer la pérennité de la revue, ce qui n'est jamais sûr lorsqu'elle est effectuée par des étudiants, ce statut étant précaire par nature ».

4.2 Les pionniers et leurs institutions

Le temps des franches résistances et des conflits directs est enfin révolu. Mais, à de rares exceptions près, déjà repérées page 200, il semble que les institutions continuent de privilégier l'inertie comme pièce maîtresse de leur stratégie face à des enjeux qui les dérangent ou qui les dépassent.

Signe de cette paralysie générale, les enquêtés ont répondu aux questions « votre institution ou d'autres (françaises ou étrangères) soutiennent-elles aujourd'hui votre projet ? » et « au contraire, si votre institution émet des résistances, quelles sont-elles ? » en donnant un sens très large au mot *institution*, qui de fait dépasse les frontières de leur université ou de leur laboratoire... Ceci s'interprète par un besoin de trouver des collaborations lointaines, en l'absence de soutien local.

4.2.1 Conflits politiques

Quand l'institution d'appartenance des enquêtés manifeste un intérêt pour l'internet, c'est rarement pour faciliter la tâche des pionniers. Tout d'abord, apparaissent des conflits d'intérêt entre collègues : « dans le cadre de X, je suis libre et je travaille véritablement en réseau. A l'Y, [...] c'est parfois la course d'obstacles car je ne suis pas webmaster en titre. Cette dénomination semble devenir au sein de certaines institutions un titre honorifique et là encore les questions de pouvoir commencent à se faire sentir. Il y a ceux qui font et ceux qui causent ». Ensuite, les directions peuvent tenter de « court-circuiter » les experts pour déléguer à des entreprises la construction de sites web institutionnels, en privilégiant la forme sur le fond, comme cela se fait déjà dans de nombreuses grandes écoles[4] : « en revanche, l'institution (X) a soigneusement évité de me confier des responsabilités dans le domaine (de manière générale, au moins à l'X la mise en œuvre se fait soit par le haut — le directeur — soit par l'extérieur — les consultants — jamais en interne par remontée de projet. C'est sensible pour tout ce qui est informatique mais nullement spécifique). Je suppose que c'est la démarche administrative par excellence (toujours se méfier des troupes) ». Et l'obéissance aux effets de mode peut conduire certains directeurs à entrer clairement en conflit avec leurs experts si ces derniers s'intéressent avant tout aux contenus, comme l'explique un des enquêtés, condamné à venir tous les jours passer huit heures dans un bureau sans ordinateur, et à qui toute initiative a été explicitement interdite : « [je suis] chef de projet multimédia (sic), provisoirement à l'Y, [en fait] en quête de poste et de projet intéressant ». Il explique que son conflit avec la direction découle de « choix politiques de systèmes informatiques centralisés (Lotus Domino Serveur) sous-tendus par des convictions politiques

4. D'autres témoignages, extérieurs à l'enquête, provenant de la Sorbonne et d'écoles d'ingénieurs, mettent en évidence un souci de s'approprier l'internet à des fins d'image : les sites institutionnels sont pris en charge par les cellules de communication, qui entrent souvent en conflit avec les chercheurs soucieux d'une diffusion scientifique. Les travaux de ces derniers sont souvent relégués au fin fond d'arborescences Unix.

qui n'ont pas évoluées depuis le début des années 70 ». On retrouve là le lien entre informatique de service et pouvoir, avec la preuve qu'en l'an 2000, certains responsables d'institutions ont une conscience claire des enjeux de l'internet, même si les solutions qu'ils choisissent sont sujettes à caution.

4.2.2 Inerties

Sinon, quand ces enjeux ne sont pas encore perçus, on constate une amélioration sur le plan matériel, comme le remarque la personne qui amenait à son bureau le modem acheté sur son budget personnel : « pas d'implication générale, mais une bonne volonté certaine pour m'acheter un matériel correct depuis 1998. (D'un autre côté, il ne faut pas rêver, je n'ai même pas 128 Mo de mémoire RAM) ». Cette tendance semble générale : « depuis 1999, fourniture d'ordinateurs et d'accès ». Et l'obtention de telles ressources finit par faciliter les initiatives collégiales : « mon institution a créé un petit groupe d'enseignants qui veulent travailler sur la pédagogie, avec le soutien des services de la documentation. Donc les gens motivés viennent, et le soutien est lié à une certaine émulation et à la possibilité d'utilisation libre d'appareils ». Ainsi, ces dotations matérielles ne sont pas réellement prolongées d'une réflexion : « aujourd'hui on en est à proposer aux gens de bons moyens techniques (partage de réseau, forums etc.) ».

Et, le plus souvent, l'inaction est de rigueur. Neuf personnes soulignent une absence d'engagement institutionnel (au sens large, comme précisé page 227) : « ça piétine toujours [, pour cause d'] indifférence, absence de coordination dans les projets de ce type au niveau de l'institution » ; « pas de réticences. Simplement, il n'y a pas de soutien financier » ; « non pas résistance, mais indifférence » ; « le projet est de peu d'envergure, mais les étudiants qui entretiennent le site du X le font par pur volontariat ». Aux critiques directes ont succédé des promesses très virtuelles : « les soutiens à X sont multiples en France comme en Europe mais uniquement moraux » ; « soutien moral à différents niveaux de mon institution. Aucun autre soutien institutionnel » ; « aides financières attendues »... Au point que certains peuvent en devenir moroses : « je ne suis pas demandeur de soutiens », rappelle un chercheur, lui aussi correctement doté en matériel.

Sinon, la bureaucratie reste le meilleur rempart contre les initiatives : « dans notre évaluation très récente de l'équipe de recherche que je dirige désormais (agréée [...] en juin 2000) aucun compte n'a été tenu de nos ambitions informatiques : la dotation correspond au fonctionnement basique d'une équipe littéraire au sens le plus traditionnel (ex : 15 KF en équipement par an !) ». Suite à de violentes résistances locales, une autre personne a fini par faire héberger son site au Canada. Il ne désespère pas de le relocaliser un jour dans son institution, mais exprime son désabusement : « une demande est en cours ; mais comme la décision doit être d'abord du ressort du gouvernement de l'université avant d'être concrétisée par les informaticiens compétents, il est impossible de dire si cette demande aboutira ».

Ces fonctionnement sont tels que deux autres personnes ne croient plus à la possibilité de monter des projets audacieux dans une institution publique : « difficulté de faire aboutir un projet web dans une institution publique » ; « mais les projets intéressants dans l'administration peuvent-ils exister aujourd'hui ? J'apprends il y a une semaine que l'Académie des Sciences avec qui j'avais pris contact remet son projet d'archives numériques à dans 2 ans. Cela fait rêver... Leurs collègues US eux ont déjà un superbe site ».

En fait, ces formes multiples d'inertie traduisent une inflexion obligée (suite aux injonctions ministérielles) de la précédente politique de rejet total de l'internet. La méfiance reste dominante, comme le rappelle un professeur d'université : « [ma direction] suspecte effet négatif sur productivité en termes de recherche, même si [elle] reconnait utilité en matière d'enseignement ».

4.2.3 Audaces

C'est en fait un exceptionnel esprit entrepreneurial qui anime les pionniers. La philosophe qui avait réussi à impliquer huit institutions européennes dans son programme reçoit aujourd'hui d'autres formes de soutien : « projet

APN[5] du CNRS, financement de missions de jeunes doctorants... ». Dans cette course au « montage » impliquant plusieurs institutions, les provinciaux ne sont pas les plus mal placés : « Oui, l'X [m'aide] par son infrastructure et sa politique générale, la Région Y par un programme d'action prioritaire, le Ministère de la recherche (idem). D'autres demandes sont en cours ». Mais pour ces chercheurs-entrepreneurs, qui doivent compléter leurs cultures initiale et informatique par des connaissances administratives (en matière de gestion publique) comme juridiques (droit d'auteur, d'éditeur et *copyright*), et passer beaucoup de temps à constituer un réseau dense de relations nationales ou internationales, rien n'est définitivement acquis.

Déjà, les résistances réapparaissent de façon claire quand on aborde l'organisation des disciplines ou de nouveaux thèmes de recherche, et quand on bouscule le dogme de l'imprimé. Même dans une discipline nouvelle, aux thématiques transversales, le scepticisme règne : « par contre, le projet de revues scientifiques en SHS produites et diffusées de manière électronique se heurte à de plus grandes résistances. Dans le domaine de la recherche (en Info-Com), la prise en compte de l'internet et des réseaux électronique comme réalité tangible, comme objet d'études de la discipline n'est pas encore admise. cela reste très 'exotique' ».

Un chercheur tente d'expliquer la paralysie de ses collègues : « Internet est encore largement inconnu, si ce n'est en termes d'utilisation, du moins en termes de production ». Et pourtant, celui-ci tire son expérience du Canada, pays que l'on aurait pu croire plus « avancé » que la France. La publication sur le web n'est toujours pas évaluée : « [il n'y a] plus de résistances mais encore une mauvaise prise en compte du travail effectué. Les travaux sur le web ne bénéficient toujours pas de la reconnaissance officielle du moins en terme de carrière » ; les personnes qui avaient obtenu des premiers financements ne les voient pas transformés quand ils en demandent pour leurs revues savantes : « nous avons demandé de l'aide comme pour une publication normale mais nous attendons les résultats de ces demandes ».

5. Aide aux projets nouveaux.

Cette situation n'est pas propre à la France, comme certains témoignages canadiens l'ont rappelé : dès qu'il s'agit de raisonner en termes de contenus, de réfléchir sur l'apparition de nouvelles pratiques, tant intellectuelles que sociales, de penser l'actualisation des critères d'évaluation et de transmission des connaissances, la paralysie des universitaires semble étendue au monde entier. Le témoignage suivant met en valeur à la fois toute l'énergie dont doit faire preuve un chercheur quand il s'agit d'infléchir des habitudes ancrées dans le monde de la recherche, et la force de ces habitudes au niveau international : « nous avons rencontré nos plus graves difficultés de la part de l'institution américaine du *Social Science Quotation Index*. Alors que nous remplissions tous les critères (notamment système de double *referee* avec comité de lecture international), notre inscription a été refusée, au prétexte que nous n'étions pas suffisamment cités dans les revues internationales ! On nous maintient délibérément dans ce cercle vicieux. Cela est très grave car nombre de nos collègues anglo-saxons font état de ce prétexte pour ne pas nous envoyer d'articles, ou refuser que nous publiions leur article s'il s'agit d'actes de colloques ».

Même si l'on peut espérer que dans un futur proche, les revues électroniques se multiplieront, cette toute-puissance des revues papier et des organismes chargés de les référencer et d'en organiser la promotion peut singulièrement inquiéter, d'autant que les auteurs en sciences humaines[6] peu au fait de l'internet peuvent arguer de telles pratiques américaines pour maintenir un *statu-quo* qui se traduit par la valorisation exclusive des revues imprimées.

En 2000, on est donc passé d'un rejet total de l'internet et de ses potentialités éditoriales à une acceptation sans engouement de l'outil, sans qu'il y ait réflexion sur les usages possibles, ni sur les besoins, notamment en matière de culture informatique. Il est symptomatique qu'aucun des enquêtés ne signale l'existence de journées de formation au sein de son institution. Dans l'ensemble, tant que les enjeux éditoriaux du web ne sont pas compris, on ne remarque plus de franche résistance institutionnelle. Ce qui n'empêche pas les

6. Souvent *editors* ou membres de comités de rédaction.

pionniers de rester livrés à eux-mêmes : solitaires, ou en petits groupe, c'est à eux et à eux seuls de continuer à se former, à trouver des financements, en impliquant des institutions lointaines, tout en prolongeant leurs recherches dans leur domaine d'excellence et en construisant des revues savantes, en animant des groupes de discussion, etc. Autant d'activités qui ne sont toujours pas valorisées : l'internet reste perçu comme un réseau technique qui, dans le meilleur des cas, légitime l'achat de matériel. Le seul étalon de l'activité intellectuelle reste l'imprimé, même pour ce qu'on pourrait appeler l'institut international des poids et mesures de la renommée scientifique. On comprend alors que les chercheurs qui se soucient avant tout de l'accroissement de leur capital symbolique et qui supposent que le cadre économique qui le légitime est immuable — quitte à oublier sa relation avec la production scientifique — ne s'associeront ni aux actes, ni aux questions des pionniers.

4.3 Les collègues

4.3.1 Conformismes

À la question « que pensent aujourd'hui vos collègues de votre projet initial et de son évolution », les réponses sont brèves ou détaillées, mais rarement enthousiastes.

Deux personnes, dont le désarroi était déjà sensible, sont très directes. La première remet en question l'esprit de curiosité des chercheurs : « question métaphysique : mes collègues pensent-ils ? » ; la seconde insiste plus sur les rivalités mesquines entre des personnes qui n'ont qu'un pouvoir symbolique : « de la jalousie à l'indifférence ou à la dénégation de toute valeur et de tout intérêt, le spectre est large, et chacun y trouve sa place ! ». D'autres réponses, plus lapidaires, ne contredisent pas cette image d'un milieu sclérosé : « rien. Ils s'en foutent ! » ; « pas grand chose ». Les jalousies précédemment évoquées se prolongent dans l'attentisme : « [ils pensent] que probablement je n'avais pas tort, mais cela n'est pas encore dit... » ; « 'wait and see' pour la plupart ».

Quelques enquêtés s'efforcent d'être plus optimistes : « je suppose qu'ils apprécient... ? » ; « c'est une réussite qui les étonne » ; mais dans l'ensemble, les choses évoluent lentement : « un début de reconnaissance chez ceux qui étaient fermés et toujours le soutien des anciens » ; « l'indifférence a fait place à une certaine curiosité voire à un engouement qui reste encore un peu naïf » ; « bien plus d'intérêt et de reconnaissance. Pratiquement plus d'oppositions, indifférences qui ne disparaîtront jamais dans certains cas ».

Il est vrai que le travail nécessaire (apprentissage, construction d'un réseau collégial étendu) et le coût en temps et en moyens matériels, conjugués à la faible valorisation des activités sur l'internet, peuvent freiner les ardeurs. Un chercheur rappelle à quel point la consultation du web apparaît complexe aux personnes qui n'y ont pas été formées : « Maintenant les gens trouvent qu'il y a trop d'informations [sur le web, que] personne ne sait interroger » ; les chercheurs qui en sont à découvrir l'intérêt d'un site web gardent une attitude de simples consommateurs : « ils apprécient la disponibilité des informations. Je pense qu'ils n'ont encore que peu d'idées sur ce qu'ils peuvent faire du site internet. Comme mon projet est surtout la mise sous forme électronique des informations, je ne suis pas créateur du contenu. En ce sens, je suis limité par l'imagination des usagers du site ».

En fait, les pionniers sont toujours contraints par les logiques de rentabilité et d'investissement : « les réactions sont beaucoup plus positives mais il est toujours aussi difficile de trouver des personnes prêtes à investir du temps »; « le projet est bien perçu mais la 'rémunération' institutionnelle d'une participation à son développement n'est pas assez forte pour inciter des bonnes volontés nombreuses à se manifester ». Certes, l'analyse des rapports de force politiques plaide en faveur des pionniers : « la modification des rapports de force au sein de l'institution les oblige à envisager la production et la diffusion électronique sous un autre angle. Mais tout cela reste bien tiède ».

Au final, les innovateurs ont surtout réussi à sortir du guêpier dans lequel ils s'étaient fourrés. Une partie d'entre eux sait maintenant jongler avec les appels d'offres, les financements extérieurs, les contrats, et cette indépendance financière leur donne une autonomie relative — et une perception accrue du fonctionnement du monde universitaire. Mais ils n'en tirent pas pour autant des avantages professionnels : aucune ne mentionne de nouvelles responsabilités au sein d'une équipe constituée pour élargir leurs initiatives.

4.3.2 Frémissements

De telles tendances à l'inertie et à la morosité sont confirmées par les réponses à la question « de nouveaux collègues se sont-ils associés à votre projet ? Ou êtes-vous moins nombreux qu'auparavant ? » Certes, aucun enquêté ne répond par l'affirmative à la seconde question. Mais dans l'ensemble, on peut considérer que le nombre de chercheurs associés au projet initial n'a fait que doubler en moyenne, et que cet accroissement est surtout dû à la capacité des pionniers à enrôler leurs doctorants.

De façon générale, le nombre de collègues associés au projet a augmenté : « oui, quand même ! » ; il est en « augmentation progressive », souvent parce que « le projet est en expansion ». Pour certains, il reste « stable. Reconnaissance croissante de l'utilité ». La personne qui a conçu un module de licence travaille au sein d'un petit groupe : « on est quatre ou cinq ». Un autre enquêté a moins de chance : « je travaille seulement avec une collègue.... de X ! ». Une telle situation, vécue par un professeur, est encore plus délicate quand le pionnier est étudiant : « je suis dans la phase où j'essaie d'inclure de nouveaux étudiants pour entretenir et développer le site X. Les compétences techniques sont un frein majeur, à mon opinion ». Dans un tel contexte, on conçoit que le principal enjeu soit en fait de maintenir une taille critique : « [nous sommes] à peu près le même nombre depuis le numéro deux [de la revue électronique], six permanents et une bonne vingtaine de collaborateurs et de collaboratrices ».

En termes statistiques, l'évolution semble lente : « sur 50 chercheurs du labo, 2 ont accepté de [faire] mettre leurs travaux sur le web. Mais 10 à 15 suivent bien la mise à jour des informations personnelles (chgt de statut, financements) », indique un chercheur qui excuse ses collègues en rappelant que ce faible taux est peut-être lié à la reconnaissance de ses compétences : « non [il n'y a pas de nouvelle personne qui s'implique sur le site] ; mais tout le monde trouve ça bien que je le fasse ». Certes, on sent des frémissements : « [on est] plus nombreux » ; « oui, c'est un début » ; « [je constate] un intérêt un peu élargi » ; « le réseau de mes collègues grossit mais toujours sur le même principe des individus capables d'initiatives qui n'attendent pas la bénédiction de leurs responsables ».

On peut alors se demander si l'arrivée de nouveaux collaborateurs n'est pas le fait d'un militantisme de proximité. En effet, les experts offrent volontiers leur assistance : « surtout, de nouveaux projets naissent, et je m'y associe volontiers ». On se doute que les dynamiques fédératrices soient plus liées au savoir-faire social des pionniers qu'aux « propriétés intrinsèques du médium » comme on l'entend souvent : « notre comité de rédaction s'est élargi et diversifié » ; « nous sommes plutôt plus nombreux et les collègues encouragent désormais plus souvent leurs doctorants ou jeunes docteurs à publier dans X » ; « oui, plusieurs nouveaux collègues, doctorants... » ; « oui, de jeunes doctorants surtout ». Dans ce dernier cas, l'auteur est un maître de conférences, non habilité à diriger des thèses. Il organise par ailleurs un séminaire qui ne se tient pas à l'université, car il n'en a pas le droit. Il édite néanmoins une revue savante électronique et incite des doctorants à y publier. On voit ici à quel point l'investissement sur le web et le non respect des normes universitaires sont liés.

4.3.3 Impuissances

Pour autant, la place de l'internet dans la technique d'écriture n'est toujours pas comprise. Quand les enquêtés répondent à la question « pensez-vous que les représentations de vos collègues vis-à-vis d'internet se sont modifiées depuis votre initiative », leurs réponses sont plutôt négatives. Suivant les réponses, cet ensemble de collègues est pris au sens large (l'ensemble des universitaires) ou en un sens plus restreint (les collègues proches). Dans le premier cas, les répondants insistent surtout sur le rôle des médias : « non pas depuis mon initiative, mais depuis que tout le monde parle d'internet » ; « pas mal de gens incompétents bavardent sur le sujet. Ca ne me semble pas nouveau. Quant aux évolutions, non, je crois que je n'y suis pour rien » ; « modifiées suite aux discours des médias : intérêt suite au battage publicitaire » ; « j'en doute, pour la génération en place. J'en suis sûr pour la population de nouveaux chercheurs »[7]. Et l'usage généraliste de l'internet est maintenant passé dans les mœurs : « à coup sûr, ils ont tous un abonnement Internet ».

Ainsi, les chercheurs dans leur ensemble sont, comme d'autres groupes sociaux, sensibles aux discours des médias et à la surenchère publicitaire. Ces facteurs semblent plus déterminants pour aiguiser leur curiosité vis-à-vis de l'internet que le travail pionnier de leurs collègues. C'est ainsi qu'il faut entendre les réponses qui renvoient à un réseau collégial plus restreint : « depuis notre initiative, je ne sais pas — plus généralement, oui, sans aucun doute » ; « il y a peu de chances » ; « peut-être... » ;« je ne sais pas » ; « oui, mais je ne suis pas certain que ce soit seulement l'initiative en question qui ait provoqué ce virage » ; ; « oui, mais pas du fait de mes initiatives (sauf marginalement à X) ».

Certaines réponses sont moins diplomatiques, allant d'un « les bûches restent les bûches » à un énergique « pas assez ! ! » quand il s'agit d'expliquer le faible intérêt des collègues proches.

Les quelques réponses optimistes et argumentées concernent encore et toujours une poignée de personnes sensibilisées par le pionnier : « pour quelques collègues, oui. C'est une source de grande satisfaction » ; « oui, mais faire ne suffit pas ; il faut expliquer, montrer, démontrer. Les formations à l'Internet sont indispensables et plus encore les formations critiques ».

Et les représentations de l'internet au sein du monde universitaire évoluent lentement, ce qui est normal si l'on raisonne en termes d'outillage mental et non pas d'adoption d'un objet technique : « oui, certainement mais

7. Ce témoignage, souvent entendu, est à comprendre, non pas comme une opposition entre générations (les enfants manipuleraient des ordinateurs que leurs parents ne savent pas allumer), mais comme une intégration accélérée de l'outillage mental contemporain auprès de jeunes chercheurs.

c'est très lent et je le répète, cela n'implique pas forcément une implication dans la production sur le 'Net'. C'est d'ailleurs pourquoi je pense que les inégalités resteront considérables [...] entre la minorité des producteurs et la majorité des consommateurs ». Cette analyse canadienne rappelle, d'une part, que l'utilisation du web à des fins pédagogiques et éditoriales entre dans une classe de pratiques marginale, et que d'autre part, elle conserve ce statut marginal au sein même des chercheurs, *a priori* plus sensibilisés à la production auctoriale que l'ensemble d'une société.

4.4 Évaluation du web aujourd'hui

Pourtant, de telles inerties collégiales n'ont plus de raison d'être. Les enquêtés estiment que le contenu du web dans leur discipline est maintenant d'une grande qualité, même si plusieurs rappellent son caractère inégal et déstructuré : « ça va dans tous les sens, je suis perdu, malgré quelques balises » ; « foisonnant, de niveau très inégal » ; « encore embryonnaire, et très inégal en qualité (pages émises par de doux dingues jusqu'aux choses les plus pointues) » ; « toujours très disparate [...mais] des sites intéressants ».

Mais dans l'ensemble, ils sont satisfaits de l'évolution rapide des contenus : « aucune comparaison avec la situation d'il y a cinq ans, évidemment » ; « tout a fait intéressant, mais manquant encore de cohérence, d'outils de validation et d'orientation » ; « il semble s'améliorer régulièrement et propose aujourd'hui un ensemble de ressources fort éclatées ou disparates, mais qu'il est possible d'homogénéiser sous certaines préoccupations de requêtes ». On le voit ici, la « pléthore » chère à Paul Mathias[8] ne gêne pas les chercheurs qui ont appris à raffiner leurs recherches à force d'expérimentation et de navigation systématique. Certains tirent même un réel plaisir de cette prolifération désordonnée : « passionnant et chaotique » ; D'autres, plus mesurés, évoquent aussi cette transformation : « plutôt bon, en progrès rapide en tout cas » ; « meilleur » ; « il commence à y avoir des choses pas mal ».

8. [Mat01].

Au-delà de cette impression générale, les témoignages recueillis mettent en évidence des spécificités qui transforment les pratiques professionnelles. C'est le cas du médecin, qui montre comment le partage d'expérience peut améliorer de façon significative l'analyse de phénomènes rares et complexes : « très bon, en particulier dans le domaine de l'échographie en diagnostic prénatal, l'échange d'iconographies et de cas cliniques (souvent rares) est très appréciable ».

Les quelques réticences portent sur le côté « vitrine » des sites web d'institutions françaises : « toujours insatisfaisant en histoire, comme je l'ai déjà dit, mais les grandes coquilles vides tendent à disparaître et l'option 'vitrine et annuaire' semble condamnée à plus ou moins long terme. Espoir, donc ! » ; « encore trop maigre, et pas assez rationalisé. Toutefois, je suis en histoire X, où beaucoup de choses sont faites, déjà » ; « plus riche et plus fourni qu'il y a 4 ou 5 ans, mais peu aisé à décrire parce que résultant de multitudes d'initiatives locales souvent isolées. Quelques grosses institutions mettent cependant en place des serveurs très riches » ; « en France, gagnerait à être développé. C'est toujours très faible en Y, et en Z aussi. Aux États-Unis en revanche... C'est paradise land » ; « il y a encore un contenu assez faible, pour ne pas dire très faible, sur la Toile francophone en matière de recherches en communication. Ne serait-ce qu'en France, tous les laboratoires n'ont pas un site et si c'est le cas, le contenu est souvent pauvre. Il n'y a par exemple pas de texte mis en ligne, seulement des informations d'ordre administratif qui datent d'ailleurs parfois d'un certain nombre de mois ». Un autre chercheur fait une remarque analogue, mais pour les États-Unis : « toujours peu de choses en X. Assez faible. Les universités américaines se contentent de publier leurs programmes d'enseignement ». Mais si l'on aborde le web comme un tout, indépendamment de la localisation et de la langue des sites, l'impression de complétude se confirme : « le contenu du web pour les X est l'un des meilleurs parmi les disciplines littéraires. Il est extrêmement structuré : il existe plusieurs portails, et des sites universitaires qui ont choisi de se spécialiser pour offrir un contenu de haut niveau (bibliographies, collections, corpus [...]) ».

Les plus critiques sont en fait les plus exigeants, ceux qui voient clairement où pourrait en être le web aujourd'hui : « pour la géographie proprement dite, encore insuffisant ». Ce qu'explicite un de ses collègues : « on pourrait faire plus et mieux, en termes de bases d'information, de bases de données cartographiques [...], et surtout de logiciels didactiques [9]. La pratique de la discussion interactive est encore trop peu développée ». Une autre visionnaire s'inquiète, elle, non pas des contenus mais de leur accessibilité : « en bonne évolution, mais menacé par les sites payants ».

En l'espace de quelques années, le web francophone en sciences humaines est donc passé de la vacuité totale à l'état de bibliothèque prometteuse. Cependant, tous les enquêtés sauf un rappellent la richesse des informations sur les sites nord-américains et se plaignent du manque d'homogénéité des sites français en matière de recherche. Certains, conscients des potentialités du web en matière de logiciels en ligne, d'applications pédagogiques, ou de multiplicité de niveaux de lecture, s'estiment même frustrés. En fait, les pionniers restent pionniers, au sens où l'amélioration des contenus en France est largement due à leurs efforts, et au sens où leur expérience nourrit leur perception de l'étendue des possibles en matière éditoriale. Mais on le sait, l'inertie des institutions et des collègues, décrites auparavant, rendent difficile toute évolution rapide.

9. Cet enquêté et un autre sont les seuls à faire allusion à des logiciels intégralement *on-line*, appliqués à des méthodes, ou plus généralement à fonction pédagogique.

Chapitre 5

Bilan et enjeux

5.1 Culture technique et réseaux d'érudits

On comprend que les personnes interrogées aient la même appréciation quant à leur besoin de culture informatique en 2000 qu'au moment où elles ont décidé de s'impliquer sur le web. Aussi, comme on l'a vu au chapitre 2.2 (page 195), le nombre d'enquêtés affirmant ne pas « avoir ressenti le besoin de mettre à jour leur culture technique » reste quasi-constant : elles sont sept au lieu de six[1] ; on se souvient que, bien souvent, elles peuvent solliciter des proches (collègues du laboratoire, époux, etc.) en cas de besoin. Et certaines d'entre elles n'ont tout simplement pas besoin de mettre à jour leur culture informatique, car celle-ci est déjà fort développée, comme le signale le philosophe auteur de quatre sites web : « non, mais quand j'aurais le temps, je monterai une machine Linux ».

De façon générale, on constate un intérêt accru pour l'automatisation intégrale de sites web, pour la programmation et pour les protocoles structurés. On passera donc sur les réponses insistant sur les besoins de connaissance des langages permettant la réalisation de scripts (perl, php, etc.), ou plus évolués (java), et sur le désir de se familiariser avec des codages structurants comme XML (cité deux fois). Ainsi, même après quelques années de pratique inten-

1. En fait les six du départ, plus une.

sive, les trois-quarts des interviewés ressentent le besoin de compléter leurs lacunes en matière informatique.

Mais ils se distinguent clairement de la majorité de leurs collègues en ce sens que, dans leurs réponses, cette prétendue culture technique se confond avec la culture du chercheur : ils désirent approfondir leurs connaissances pour offrir un confort maximum à leurs lecteurs, et transmettre de façon optimale leur savoir-faire. Ils veulent comprendre les arcanes de la « boîte noire » qu'est le web, et ne se satisfont pas d'une pratique consumériste. Ces questions les amènent à remodeler leurs réseaux de sociabilité, leurs pratiques de lectures, et leurs centres d'intérêt, ce qu'ils détaillent avec précision.

Tout d'abord, la perception politique des enjeux économiques et éditoriaux est présente, que ce soit dans le but de rester lisible par tous, ou de rester maître de ses outils de production : « je n'ai aucune compétence en java, ni XML, ce qui s'avèrera un frein rapidement, même si notre parti pris est de permettre au plus grand nombre de butineurs de pouvoir nous lire sans encombre » ; « désir accru de maîtriser le monde Unix et de s'affranchir de Microsoft ». Ici, la compréhension de l'informatique comme culture complexe et la perception des enjeux autour des formats de fichiers sont patentes.

Et divers pionniers explicitent clairement en quoi leur engagement stimule leur curiosité, et les incite à s'intéresser à des domaines qui pourraient ne pas avoir de rapports avec leur discipline : « non [... mais] je serais intéressé en revanche de connaître les moteurs de recherche, comment se fait le cheminement vers un site » ; « oui, c'est en continu. Actuellement sur l'informatique linguistique et sur les réseaux [...] » ; « oui, mais dans le domaine précis de la bibliothéconomie ; je suis un stage [...] sur le catalogage informatisé et les normes d'échange des données bibliographiques ». L'approfondissement des protocoles informatiques facilite aussi la conceptualisation : « la mise à jour est en fait permanente, l'apparition dans le domaine qui est le mien (édition électronique) de nouveaux procédés et de nouveaux standards empêche de s'endormir. La tentative de construire un discours de recherche en Info-com sur les réseaux oblige à une remise en cause permanente ».

Bien sûr, ces analyses se développent dans un cadre collectif — mettant une fois de plus en évidence l'importance des effets de réseau —, et sollicitent l'acquisition d'une culture générale : « tous le jours. je le fais en lisant, en essayant, et en discutant avec des amis passionnés d'informatique ou informaticiens de profession », affirme une philosophe ; « ma culture est quasi-nulle. Je la mets à jour par mes relations avec les informaticiens », complète un autre. Le médecin insiste, lui, sur les soirées thématiques : « soirées de formation mensuelles au sein du club médical depuis le début, lecture quotidienne (rapide) des news spécialisées ». Il en est de même pour un autre enquêté : « acquisition de nouveaux progiciels, lecture d'ouvrages spécialisés, sessions de formation ».

Reste un dernier témoignage, à la fois alarmant, quant à l'état larvaire de cette culture au sein des sciences humaines, et optimiste, en termes de modernisation de la pédagogie : « certes. De toutes les manières possibles, notamment en organisant des séminaires de formation au sein de l'école doctorale et en assurant moi-même des cours d'initiation ». Il vient d'une spécialiste de littérature, dont on aurait pu penser qu'elle n'était pas obligée de se transformer en enseignante d'informatique. Mais *a contrario*, il manifeste, d'une part, cette nécessité que les chercheurs les plus audacieux ressentent, à savoir entreprendre systématiquement, d'autre part, cette vision claire de l'intégration de l'internet dans l'outillage mental du chercheur : la recherche documentaire sur l'internet et la manipulation de protocoles techniques deviennent, aux yeux de cette érudite, les outils de base de l'étudiant littéraire qui se destine à la recherche.

On réalise alors à quel point cet approfondissement de la culture *a priori* technique favorise l'acquisition d'une culture générale, infléchit les relations sociales des chercheurs, réoriente leur curiosité et donc leurs centres d'intérêt vers des questions théoriques, et stimule l'organisation d'enseignements nouveaux. Et les réponses des enquêtés prouvent une fois de plus l'incohérence du discours dominant chez les universitaires qui distinguent recherches théoriques et applications techniques : les dispositifs de recherche de connaissances, de mise en forme de celles-ci, et de publication des synthèses organisées doivent être enseignés, et, en retour, ils réorganisent la distribution et l'importance relative de ces connaissances.

Prétendre que ce nouvel ensemble de compétences n'est que « technique » revient à refuser que la mise en réseau de milliers de travaux de recherche touche au métier, au savoir-faire, à la compétence des chercheurs. Ne pas en avoir conscience consiste à refuser de s'interroger sur la culture implicite du chercheur patenté, faite de « petits riens », de « trucs de cuisine », d'intuition, mais aussi de culture érudite et de réseaux sociaux. Tout comme la science du milieu du XX^{e} siècle, et la manière de la faire, n'ont plus de rapport avec le joli catalogue qui décore la grande salle de la Sorbonne, la pratique scientifique d'aujourd'hui, avec les laboratoires, les méthodes, les axes de recherche contemporains, s'écartent de ce qui avait valeur de norme il y a 30 ans. À partir d'une transformation de l'outillage intellectuel du chercheur, on arrive logiquement à une transformation de ses pratiques intellectuelles[2]. Et bien sûr, plus, tard, à une évolution de ses thèmes de recherche.

5.2 Bénéfices intellectuels

5.2.1 Libertés

On a vu au chapitre 3.6 (page 219) que les bénéfices acquis par les pionniers étaient essentiellement symboliques, après une longue période où ils avaient subi mépris ou rejet de la part de leurs collègues. On ne reviendra donc pas sur la faible reconnaissance professionnelle obtenue, ni sur l'absence de retombées financières suite à leur engagement. De même, on passera rapi-

2. On préférera l'expression « pratiques intellectuelles », qui renvoie à toute la panoplie des activités du chercheur associées à l'élaboration de sa pensée, et d'ailleurs aussi à la construction de pratiques sociales spécifiques pour l'optimiser — même si Pierre Bourdieu décrit parfois un monde universitaire qui ne pense plus, qui ne s'occupe que d'accroître son pouvoir —, à une expression de plus en plus *à la mode* comme « pratiques cognitives ».

dement sur l'analyse de leur bilan[3] par les acteurs quand ceux-ci multiplient les « positif » (ou « très positif »), qui apparaissent 12 fois, sans compter les « si c'était à refaire, je le referais ! », ni les témoignages de réel plaisir. Au total, ce sont 20 réponses sur 24 à cette question qui témoignent de la double satisfaction des pionniers à s'être engagés dans l'internet : ils ont aujourd'hui la preuve que leur intuition d'il y a 3 à 6 ans était bonne, et leur expérience leur donne un recul qu'ils jugent « inestimable ».

Cette satisfaction est complétée de détails qui montrent comment l'incursion dans l'informatique génère des lieux étroits avec la culture classique, et s'appuie sur un laboratoire qui devient un espace d'amitié, tout en développant une éthique de la production intellectuelle : « je suis heureuse d'avoir osé me perdre dans les arborescences de Gopher. j'ai découvert bien des savoirs anciens en construisant des pages web et je suis contente de participer à quelques équipes vivantes qui sont l'occasion de rencontrer des personnes de formation et d'horizon différents » ; « je suis fatiguée, mais contente. J'ai plus d'amis qu'auparavant. Nous allons peut-être construire quelque chose ensemble » ; « 'il faut y croire' pour passer le cap des débuts souvent laborieux, quoiqu'on en dise, quand on n'est pas formé à l'informatique — avoir des amis ou confrères passionnés est très important pour avancer ».

De telles remarques sont encore plus explicites dans les réponses à la question « comment évaluez-vous les coûts et bénéfices de votre investissement ? ». En filigrane, apparaît de façon permanente le caractère fructueux de l'élargissement du réseau de compétence, souvent internationalisé.

La bonne connaissance du web et des réseaux de compétence associés (avec le courrier électronique et les listes de discussion) aide les chercheurs à optimiser leur travail : « ils m'ont permis et me permettent encore de fonctionner plus efficacement et de gagner beaucoup de temps » ; « gain de temps, communication inter chercheurs accélérée [...] ». Mais les plus impliqués dans la construction de sites n'oublient pas les jours passés à transformer le web

3. Lors de la question : « avec le recul, quel bilan feriez-vous de cette période comprise entre votre investissement premier et aujourd'hui ? »

francophone, auparavant désertique, en un instrument de recherche spécialisée : « très positifs sur tous les plans, mais au prix d'un travail excessif, à cause de la nécessité de tout faire soi-même ». « Il s'agit là d'un investissement humain en temps et en énergie considérable mais c'est très enrichissant, ne serait-ce que d'un point de vue personnel »

On ne s'étonnera pas de découvrir que les pionniers arrivent sans encombre à garantir leur autonomie, comme le montre le témoignage suivant : « le coût est lourd en temps mais le GROS gain est le sentiment d'être libre de choisir les projets de travail. Toute création 'web' demande beaucoup de travail et d'investissement personnel. Aussi, même les responsables peu conscients de la 'chose' saisissent qu'il faut, faute de reconnaissance, donner un peu de champ libre au 'faiseur' ».

L'implication dans des projets éditoriaux apporte aussi des bénéfices imprévus, en fait liés à l'accroissement de la notoriété du chercheur : « sur le plan symbolique, j'ai reçu une reconnaissance au sein du laboratoire (ce que je ne recherchais pas, mais qui me donne la possibilité de donner mon avis sur un certain nombre de choses) » ; « franchement, sur le plan personnel, je me suis amusé. C'est bon pour le développement moral. Je peux dire que j'ai été dans l'une des premières cohortes d'analystes français (cad loin derrière les autres) en sciences sociales d'un phénomène passionnant. C'est bien pour l'ego, qui est toujours étendu chez les universitaires. C'est un domaine où l'on se fait interviewer souvent, et où on peut jouer les experts (cela m'arrive très rarement). Symboliquement, cela compte. En revanche, cela coûte en temps, n'est pas reconnu par l'institution, qui ne soutient pas les projets informatiques, préférant imposer les siens ».

5.2.2 Actualisation de la recherche

Mais outre ces avantages, les enquêtés insistent sur les apports intellectuels de leur activité : « sur le plan intellectuel, j'ai trouvé la construction d'un site web passionnante », complète l'historien qui a maintenant le droit d'émettre un avis au sein de son laboratoire ; un littéraire précise les avantages de la programmation : « je suis personnellement satisfait d'avoir intellectuellement et matériellement investi dans l'informatique et internet. Cela constitue un plus dans mon profil professionnel, sans compter que la nécessité de programmer est une excellente école de rigueur intellectuelle ». Il « regrette seulement que ce genre de compétence soit si difficile à faire reconnaître officiellement dans [sa] discipline [...] » mais exprime de façon lucide en quoi l'informatique et l'internet réorganisent ses méthodes de travail et améliorent la qualité de sa production : « j'ai le sentiment d'être en phase avec le monde dans lequel je vis et, en tant que chercheur, de pouvoir mobiliser toute la palette des outils disponibles pour rédiger les meilleurs articles possibles dans mon domaine » ; un autre insiste aussi sur la transformation des modalités de son activité mentale : « énorme sur le plan intellectuel, malgré l'abandon d'autres formes de travail » ; et les explicitations de ces bénéfices sont parfois détaillées : « les coûts sont intégrés mais réels du moins sur le plan individuel : un investissement sans commune mesure avec la vie d'un enseignant-chercheur moyen, pour un résultat sans aucun rapport. Sur le plan symbolique, le bénéfice est plus net : le sentiment d'avoir, dans des secteurs bien particuliers, contribué au renouvellement de la pensée et de la pratique des universitaires, sans que cela soit généralisé, et la satisfaction d'avoir noué plus de contacts ces dix dernières années que les vingt années précédentes, avec une qualité de la relation et une permanence plus forte qu'avant » ;

Ces perceptions d'un redéploiement de l'outillage mental et de son actualisation par le laboratoire deviennent récurrentes : « dans un rapport très favorable. Dans la mesure où j'ai pris tôt le tournant, mon évolution suit celle des outils et ne demande pas trop d'efforts d'adaptation. Inversement les compétences acquises concernant la compréhension des phénomènes et le réseau d'experts sont inestimables » ; « l'ensemble m'est personnellement très positif. Je dirais même que cet investissement technique fournit une bonne base de réflexion pour mieux envisager l'épistémologie de ma discipline et pour mieux comprendre les enjeux et les intérêts philosophiques de la démarche de recherche en sciences humaines et lettres ».

Renouvellement de la pensée, explicitation des avantages des réseaux sociaux dans la production intellectuelle, et même apports épistémologiques font clairement partie des acquis des chercheurs qui ont fait le pari de l'édition électronique savante et de l'acquisition des techniques d'écriture associées.

Le pari d'une transformation de l'écriture a été gagné par ces pionniers ; ils se sont engagés dans l'exploration des formes contemporaines d'écriture en ayant conscience — dès le début, et sinon très vite à la lueur de l'expérimentation — que cela leur offrirait tous les plaisirs que peut attendre un érudit d'une exploitation raisonnée de son premier outil de travail, incluant l'accroissement des capacités mentales, un éclairage réflexif sur ces capacités, et une prise de conscience de leur statut proprement collectif. Ce qui les incite à s'offrir une lecture sociologique de leur environnement d'enseignants et de chercheurs.

Tout d'abord, ces analyses s'appliquent évidemment au monde universitaire lui-même : « ce travail pionnier était nécessaire, pour l'institutionnaliser, il reste à l'insérer dans la pratique scientifique reconnue, tant par les institutions françaises (comités d'évaluation des travaux des chercheurs et des publications) qu'étrangères (faire sauter le verrou de l'inscription dans les listes de citation). [...] Sur le plan intellectuel, j'espère que nous aidons une génération de jeunes à entrer dans l'ère Internet avec des moyens d'expression adaptés ».

Et cette institutionnalisation n'est pas gagnée : à la question « quels sont vos projets actuels ? », les personnes qui n'ont pas de statut, comme l'étudiant en histoire, ou qui attendent une promotion, comme un maître de conférences, n'oublient pas les contraintes professionnelles : « je travaille à ma thèse » ; « m'absorber dans mon agrégation[4] ». L'ingénieur chef de projet multimédia qui a eu tant de conflits avec ses deux hiérarchies successives, répond, lui : « trouver du boulot :-) ». Deux spécialistes des bibliothèques et de l'édition, espèrent, eux, finaliser leur projet : « arriver à réaliser enfin ce projet qui traîne depuis bientôt deux ans » ; « difficile en ce moment de voir au-delà du

4. Du supérieur.

démarrage tant attendu : il y aura forcément des ratés, des choses à modifier, à améliorer, un temps de réaction aux événements de cette mise en route ».

Les personnes moins fragiles statutairement ont l'intention de « continuer », tant dans le sens d'une meilleure maîtrise des outils que dans celui d'une amélioration de la production éditoriale : « améliorer ma page web, progresser en cartographie automatique, trouver des outils pour chercher plus vite et plus efficacement sur Internet », etc.

Le potentiel d'écriture des réseaux est clairement perçu, que ce soit en matière de travail collaboratif, ou de conception logicielle : « continuer dans le sens d'un développement des possibilités de fonctionnement en réseau du web » ; « améliorer l'aspect 'liens' de mon site, et créer des pages où interviennent les internautes de passage (exercices, réactions au contenu, demandes) »[5] ; « recherche-action sur l'édition des revues scientifiques, notamment en SHS » ; « créer un forum pour les doctorants ».

Avec ce regard sur leur propre monde, la sensibilité aux formes contemporaines d'enseignement et à l'élargissement du public s'affirme : « On parle peu de ce qui revient de cette mise à disposition, alors que ce n'est pas nul. Des gens du monde entier nous écrivent », remarque un historien ; « fait partie d'une réflexion sur les formes modernes de communication et de pédagogie. [Y] compris certains inconvénients », rappelle l'économiste ; « coûts élevés (en temps, en argent, en persévérance) ; bénéfices en cours : en reconnaissance, en espoir d'offrir bientôt un outil de qualité et des services améliorés, c-à-d aussi induisant de nouvelles relations avec le public » commente un spécialiste des bibliothèques ; un historien complète cette analyse : « incontestablement un gros investissement à titre personnel, et des bénéfices qui à titre personnel là aussi sont de l'ordre de la formation. Il me semble aussi que l'expérience prouve qu'existe un public en dehors du monde savant au sens étroit qui est demandeur des ressources fournies, ainsi d'ailleurs que parmi les jeunes chercheurs ».

5. Seconde manifestation de l'intérêt pour des logiciels en ligne. Cf. la note 9 de la page 240.

Ainsi, ces pionniers sont amenés à se poser concrètement des questions sur l'enseignement à distance d'une part, sur la relation entre le chercheur et son public d'autre part. Ils ne proposent pas de réponses explicites, mais leur engagement leur donne les moyens d'appréhender ces problématiques complexes de façon plus fine que ne le font la majorité des discours relatifs au *e-learning* ou à l'*edutainment*.

Mais, encore une fois, les enquêtés savent que rien n'est acquis, au contraire. Leurs projets sont encore fragiles : « assurer la pérennité de la revue ». Ils sont acculés à un esprit d'entreprise permanent : « consolider la position du vacataire qui assure le secrétariat de rédaction, trouver des moyens financiers pour aider à la maintenance des services offerts sur le site, assurer une meilleure visibilité internationale, résoudre les problèmes juridiques ».

De tels projets scientifiques sont mus par une éthique particulière, qui ne se satisfait pas de l'idéologie actuelle : « il reste énormément à faire, alors que les usages d'Internet se répandent, pour construire usages et institutions où les forces du marché sont maintenues à leur place, importante mais aucunement première. Bagarres terribles en vue sur l'enseignement [...] » ; « extension de l'édition en ligne, gratuite, numérisation à grande échelle, perfectionnement de l'outil de gestion automatique des hétérographes, association éventuelle et circonspecte avec des partenaires privés » ; « résister à une sorte de dissolution dans le commerce et dans la perte des contenus » ; « continuer à participer à la mise en place d'un réseau local de professionnels pour l'instant indépendant tout en restant vigilant sur la mise en place de la réglementation des pratiques dans le domaine médical ».

Pris entre des institutions qui ont le temps pour elles[6] et des entreprises multinationales qui aimeraient conquérir des parts du marché éducatif, les pionniers savent qu'ils sont condamnés à la vigilance, alors même qu'ils ont

6. « Les conservateurs veulent abolir le temps, éterniser l'état actuel du champ, [...] alors que les novateurs, sans même se soucier de concurrence qui que ce soit, introduisent, par leur propre intervention, le changement et font la temporalité spécifique du champ ». [Bou01], p. 127.

conscience de l'évolution de la structure et des motivations des apprenants. Il est coûteux d'être lucide.

5.3 Conclusion

Nous espérons, à la lueur de ces témoignages détaillés, avoir prouvé l'efficacité des concepts d'outillage mental et de laboratoire, mais aussi la faiblesse de l'acuité du monde universitaire face à de tels paradigmes.

Notre trentaine de pionniers découvre l'internet assez vite, entre 1991 et 1997. Très vite, pour eux, cet ensemble de protocoles est plus qu'un système de communication ou qu'un objet de consommation. C'est un instrument de travail, qui nécessite un savoir-faire, et à ce titre, ils ressentent le besoin d'acquérir une solide culture informatique, ou, à défaut, de s'entourer d'informaticiens.

Ils découvrent notamment assez tôt les potentialités du web en matière de production scientifique — historiquement, c'est à de telles fins que le WWW a été conçu — et la richesse des sites nord-américains en sciences humaines. Ils décident alors de participer à la fondation de son équivalent francophone.

C'est là qu'ils prennent le plus de risques : l'apprentissage de l'informatique leur coûte du temps, tout comme le travail de recherche sur les réseaux (web, listes de discussion, etc.) et celui de publication. Ce qui réduit leur participation à l'économie universitaire traditionnelle, dont ils ne respectent plus les règles, en négligeant d'accroître leur capital d'articles imprimés et en proposant d'autres modèles. Leur engagement leur coûte aussi de l'argent : achats d'ordinateurs, de modems, frais téléphoniques, etc. sont souvent engagés sur leur budget personnel. En effet, le plus souvent, leurs institutions ne les suivent pas, quand elles n'exercent pas de farouches résistances qui peuvent menacer la carrière des innovateurs.

On aurait pu penser que les réactions négatives de la hiérarchie résulteraient d'une attitude autoritaire pour casser une dynamique de plus en plus importante chez les universitaires. Mais cette vision simpliste ne s'accorde pas avec la sociologie du monde universitaire, telle qu'elle a été décrite précédemment [7] : l'écrasante majorité de leurs collègues méprisait ou ne comprenait pas les projets des pionniers.

Alors que ces derniers développaient alors une réflexion sur le fonctionnement de l'enseignement et de la recherche, renouvellaient les thèmes et les méthodes de leurs disciplines suite à leur compréhension des potentialités de l'informatique et de l'internet, se lançaient dans des activités entrepreneuriales — tant pour institutionnaliser leurs pratiques que pour les financer, dans un cadre juridique et économique *ad hoc* —, et découvraient l'importance des défis pédagogiques à relever — élargissement de la base des « apprenants » et irruption des entreprises dans le marché éducatif —, leurs collègues en étaient à se familiariser avec le courrier électronique [8].

En fait, les innovateurs restent condamnés à la solitude, principalement parce qu'ils se distinguent de leurs pairs et supérieurs sur un point : ils ont choisi d'adopter une attitude réflexive par rapport à leur outillage mental.

Nous ne chercherons pas à détailler les conditions sociales et historiques d'une telle prise de conscience [9], même si on peut se douter que le « face à face » avec une technique d'écriture en construction y est pour beaucoup.

7. Chapitre 1 de notre seconde partie (page 99 et suivantes).

8. Situation encore valable en 2002. Le témoignage suivant garde parfois une valeur actuelle : « il est encore très tôt dans le domaine qui est le mien pour parler de modification des thèmes de recherche ou des formes de la recherche induites par Internet. L'outil le plus fréquemment utilisé reste pour le moment le courrier électronique, le web l'est *a priori* encore peu », rappelle un historien en 2000 (cette citation et les suivantes sont tirées des commentaires des enquêtés à la fin du questionnaire).

9. Les pionniers étaient-ils marginalisés dès le départ, socialement comme professionnellement — voire culturellement, si l'on s'étonne de leur intérêt pour l'informatique ? Pour répondre à une telle question, il faudrait alors réaliser une ethnographie de leurs positions antérieures et de leurs parcours, ce qui serait l'occasion d'un autre travail, que pourraient, par exemple, réaliser les intéressés, comme Pierre Bourdieu l'a fait dans son « esquisse pour une auto-analyse » ([Bou01], pages 184 à 220). Convenons simplement que le statut de certains enquêtés (directeurs de recherche au CNRS ou professeurs d'université) n'incite pas à les imaginer comme des trublions de la recherche.

En effet, les enquêtés ne conçoivent pas l'informatique comme une technique méprisable ou utilitaire, mais comme une partie intégrante de la technologie de l'intellect [10]. On se souvient qu'ils utilisent quotidiennement les ordinateurs pour optimiser leur production scientifique comme leur enseignement. Mais ils ne le font pas sans recul : au contraire, celui-ci s'accroît au fil de leur usage des outils d'écriture [11].

La découverte des capacités offertes par l'écriture contemporaine les invite à un regard critique sur les comportements de leurs collègues, qui reproduisent des méthodes classiques sans s'interroger sur ce qu'elles conditionnent en matière culturelle [12].

Face aux routines et aux résistances de leurs collègues, ils sont alors conduits — peut-être pour se défendre face aux conventions de leur milieu, qui les marginalisent — à construire une sociologie de leur entourage [13]. Ce qui

10. Ce qui les conduit parfois à des réactions d'humour qui pourraient choquer les traditionalistes : « la médiocrité des bibliothèques françaises devient tolérable grâce à l'Internet ».

11. « Le Net n'a pas modifié, mais est en train de modifier cette logique, en redisposant dans un espace plus vaste de recherche les outils livresques ou institutionnels traditionnels ».

12. Par exemple, en évoquant l'apport des études sur les auteurs dits mineurs : « La notion d'exhaustivité revient en force : avec les méthodes manuelles, on fait des sondages dans les corpus littéraires, avec les CD-ROM et les moteurs de recherche sur Internet, on peut avoir le listing complet des occurrences de mots ou parties de mots dans des corpus librement définissables. Cet accès facile aux corpus (pourvu qu'ils soient disponibles) remet dans le circuit des auteurs ou des textes moins connus : puisque le moteur de recherche balaie toute une base de données avec la même facilité, la distinction s'efface entre auteurs connus et auteurs secondaires ».

13. Tant dans la description avisée des blocages, comme le prouve la première des deux citations suivantes, que dans la façon dont ces blocages sont étroitement liés à une profonde incompétence — technique ? —, particulièrement nuisible aux tenants d'un outillage mental actualisé, ce que montre la seconde citation : « plus un personnage a de l'autorité et une position acquise, moins il consent à se mettre à internet. La nouveauté est rejetée dans la mesure où elle semble permettre de remettre en question les hiérarchies établies. *A contrario*, les jeunes, non encore entrés dans la carrière, et les personnes plus âgées mais en position marginale sont réceptifs à cette nouveauté » ; « quand j'ai montré au directeur

leur est d'autant plus aisé qu'ils s'en éloignent spontanément : leurs activités scientifiques les incitent à réorganiser leur propre réseau social, notamment en multipliant les contacts avec les informaticiens.

L'analyse sociologique du monde universitaire apparaît donc comme un résultat naturel de ce travail réflexif sur les instruments de l'intellect. Mais elle n'est qu'un des éléments de cette prise de conscience généralisée, qui, rappelons-le, ne se développe pas en quelques semaines [14].

C'est maintenant la technologie de l'intellect dans son ensemble qui est l'objet de leurs réflexions.

Bien sûr, ils restent lucides et s'inquiètent d'une adoption sans recul des « nouvelles technologies » en leur monde : effets doublement pervers de la publication tous azimuts, qui dévalorise la production scientifique [15] en même temps qu'elle risque de la paralyser [16] ; croyances naïves en les capacités d'outils encore bien primitifs et d'autant plus contraignants [17]. Au vu du faible

des archives X l'image d'un manuscrit numérisé qui se trouvait sur mon disque dur, il a cru que par le simple fait d'être numérisée elle était disponible *ipso facto* sur internet... voilà le degré de la culture informatique de quelques uns de nos interlocuteurs ».

14. Inutile d'imaginer pour autant que les enquêtés aient acquis une avance incompressible sur leurs collègues dans le domaine de la publication électronique : les injonctions ministérielles et européennes, mais aussi l'arrivée de nouveaux acteurs facilitent d'autant plus les conditions de débats fructueux et de partage d'expériences qu'aujourd'hui, le web francophone en sciences humaines regorge d'initiatives convaincantes.

15. « Il me semble que tout cela n'a fait qu'accentuer un certain phénomène de 'poubellication' sauvage dans lequel on peut trouver son compte ».

16. Un collègue s'inquiétait récemment que la tendance à l'exhaustivité dans la citation des sources dans les normes (américaines) de publication tue la recherche : « s'il faut lire toutes les pages web avant que de rédiger un article sur un sujet donné, on ne peut plus inventer » (communication personnelle).

17. « Il y a cependant un inconvénient, qui tient, non pas à internet, mais aux moteurs de recherche : toute question qui ne peut se formuler par mots-clés est presque impossible à traiter par les moyens informatiques ; ainsi, des questions d'ordre littéraire sur les métaphores, les allusions, les circonlocutions risquent de devenir moins attractives dans les années qui viennent. Internet, et plus généralement tout moyen informatique exploitant des banques de données, génère une nouvelle forme de médiocrité auprès d'étudiants ou de chercheurs peu doués : croire qu'une compilation de renseignements glanés sur les serveurs

recul dont témoignent les universitaires contemporains au sujet de leur outil de travail, il y a lieu de s'inquiéter de la façon dont les vertus de cette réflexivité seront évoquées dans les écoles et les lycées quand l'internet (ou son équivalent du moment) sera massivement socialisé.

À partir de là, se posent les questions de la diffusion de l'outillage mental et du statut des promoteurs de cette diffusion. Celles-ci ne trouvent plus leurs réponses dans un cadre national : déjà, la situation française n'est pas singulière[18]. Ensuite, l'intégration de l'internet dans l'enseignement n'est pas pensée, malgré la multitude d'excellents cours en ligne. Ici encore, l'Amérique du Nord fait figure de pilote[19].

Ce qui, au final, pose la question du statut futur des enseignants. Si ceux-ci profitaient jusque là peu ou prou d'une alliance objective avec les éditeurs, l'arrivée de nouveaux acteurs internationaux risque fort de changer la donne. Les enquêtés sont sensibles aux monopoles de l'écrit qui se construisent et cherchent à les circonvenir. Ils s'inquiètent des nouvelles lois commerciales qui se mettent en place, plus adaptées au commerce qu'à la science[20]. Ils

tient lieu de science, et aussi considérer comme canoniques les textes trouvés sur support électronique, sans les comparer avec ceux présents sur d'autres supports (les éditions critiques imprimées, par exemple) ».

18. « La comparaison que je peux faire avec les collègues américains, italiens et suisses me permet de dire que les problèmes sont équivalents ailleurs et extrêmement variables selon les situations locales », affirme une géographe.

19. « Nombre [d'enseignants d'Amérique du Nord] associent déjà cours en amphi et mise en ligne de leur enseignement, avec fonctionnement interactif avec les étudiants » ; « de plus en plus de contenus de cours sont en ligne. Mes étudiants commencent tout travail et toute interrogation en regardant le site du cours avant de m'écrire par courriel pour poser leurs questions éventuelles ; le site leur permet de travailler à distance, quels que soient le lieu ou l'heure ».

20. « La recherche universitaire a un manque chronique de moyens financiers : quand les serveurs internet mettent en ligne des documents qu'ils n'ont pas eux-mêmes produits, ils se heurtent au problème du *copyright* : ils le contournent en offrant des textes, des cartes d'atlas, etc. tombés dans le domaine public, ce qui oblige alors l'universitaire internaute à travailler sur des documents dépassés... Régression paradoxale due à une législation commerciale inadaptée à l'avancée de la technique ».

ont compris que celles-ci sont avant tout destinées à faciliter la création en Europe d'un marché de l'éducation, convoité par les éditeurs commerciaux de produits pédagogiques.

Ainsi, leur implication dans une recherche actualisée, qui tire au mieux profit des instruments d'écriture contemporains, les incite à dynamiser leurs disciplines, à acquérir un regard lucide sur le fonctionnement de leurs institutions, et à expliciter un grand nombre d'enjeux sociaux, économiques et politiques liés à l'écriture.

Quatrième partie

Écritures de l'Homme et des machines

Chapitre 1

Problématique

Nous avons montré dans les parties précédentes que l'informatique participait d'une transformation de l'écriture, puis que l'étude de son appropriation ou de son rejet permettait d'expliciter les pratiques scientifiques, et par suite, l'organisation sociale du monde universitaire : la lucidité scientifique est intimement liée à la capacité de comprendre que l'écriture est une technique, et donc à l'explicitation de ses mécanismes. Le texte, qui est toujours très dépendant des instruments (matériels, cognitifs, etc.) qui permettent de l'appréhender, voit son statut fortement transformé du fait que l'on sait maintenant le traiter avec des machines. Et nos enquêtes ont prouvé le gain intellectuel que l'on pouvait tirer de ce constat, même si le coût d'adoption des techniques contemporaines d'écriture est élevé, socialement comme intellectuellement. Pour autant, il nous semblait essentiel de donner aussi une preuve *interne* de la réflexivité de l'écriture, en montrant comment le texte et ses instruments s'infléchissent réciproquement, et comment ils génèrent une étrange production de sens, qui, à son tour invite à reconsidérer le statut du texte et à repenser les outils qui le donnent à voir. Notre espoir étant que le caractère récursif de ces influences prouve de façon définitive la technicité et la réflexivité de l'écriture, ou, en d'autres termes, que l'assertion « l'écriture est une technique réflexive » ne soit plus considérée comme un postulat que l'on pourrait toujours refuser, mais comme une proposition irréfutable.

Avant d'étudier de façon approfondie un texte électronique particulier (les *access_log* d'un moteur de recherche), il nous semble utile de rappeler brièvement la variété des types d'archives électroniques dont nous pouvons maintenant disposer, les débats que leurs traitements sont susceptibles d'engendrer, et bien sûr les travaux scientifiques qu'ils ont déjà suscités.

1.1 Types d'archives électroniques

Au-delà de tout texte, de toute image réalisés par un individu sur son ordinateur personnel, un nombre croissant de bases de données (c'est-à-dire de textes et de procédures associées), exploitées ou potentiellement exploitables, existent de nos jours sur les serveurs.

Les plus connus de ces serveurs sont les sites web, qui hébergent des types très variés de fichiers : textes (articles scientifiques par exemple), images, logiciels, etc. Ces serveurs sont eux-aussi archivés à leur manière par les moteurs de recherche. Mais un individu peut aussi se constituer assez aisément de telles archives, avec des « aspirateurs de sites » comme wget. De tels outils, communément utilisés pour construire des sites miroirs, peuvent être détournés, par exemple pour télécharger les échanges d'une liste de discussion. Ainsi des textes, des débats, publiés sur un site distant, peuvent être transférés sur l'ordinateur du chercheur qui désire les étudier à tête reposée, et qui profite là d'une opportunité inattendue : le texte numérique est disponible, présent ; la saisie en a été déléguée aux utilisateurs eux-mêmes[1].

Nous sommes là dans le registre des archives sémantiquement cohérentes, dans la mesure où les textes obtenus ont en général un sens propre.

Il existe aussi un autre type d'archives, moins connues, mais répandues sur tout type de serveur et plus généralement sur toute machine servant de

1. Cf. le « Forum sur le recrutement et la gestion des universitaires et des chercheurs » organisé par le sénateur Yves Fréville au printemps 2001 : http://www.senat.fr/Vforum/13/forum.html. Mais l'abonnement à une simple liste de discussion diffusée par le courrier électronique permet aussi de construire et de traiter de tels types d'archives.

nœud dans un réseau électronique : par exemple, un ordinateur Unix (ou Linux) conserve l'ensemble des commandes produites par ses utilisateurs ou par les robots[2], mais aussi les en-têtes des courriers électroniques transitant par le serveur, donc émis ou reçus par ses utilisateurs, les accès web ou ftp[3], etc. De même, les routeurs enregistrent les flux IP qui transitent par eux[4]. Ces modes de constitution d'archives sont paramétrables[5], mais les fichiers produits sont peu utilisés du fait de leur faible valeur informative : ils servent avant tout à la sécurité informatique.

Au carrefour de ces deux types d'archives, nous rencontrons les *access_log* des moteurs de recherche eux-mêmes, qui vont solliciter toute notre attention. Le traitement de ces bases incite à développer des outils spécifiques d'appréhension du texte[6] et à comprendre en profondeur la logique de leur production, puis à imaginer des nouvelles modalités d'écriture, notamment pour produire une synthèse des résultats obtenus.

1.2 Archives et surveillance

Au-delà de leur intérêt pour l'anthropologie de l'écriture, de telles archives automatiques peuvent inquiéter. On sait maintenant retrouver des assassins grâce aux traces électroniques des téléphones portables. Sans entrer dans le domaine du droit, où nous avouons nos faibles compétences, il nous semble nécessaire de clarifier sur ce point la position et les moyens du chercheur. En l'occurrence, certaines archives comme les *access_log* sont considérées comme susceptibles d'être nominatives, dans la mesure où certains numéros IP peuvent (après enquête) être associés à des ordinateurs individuels. Mais, outre le fait que, bien souvent un tel numéro est en fait celui du modem d'un

2. Commande *lastcomm*, par exemple.

3. Fichiers *syslog xferlog* et *access_log*.

4. C'est-à-dire entre les réseaux qu'ils interfacent.

5. Par exemple avec le fichier *httpd.conf* pour les archives d'un serveur web.

6. D'autant que leur taille est rarement compatible avec les logiciels et systèmes d'exploitation grand public.

fournisseur d'accès à l'internet (FAI), son intérêt est faible si l'information qu'il procure ne peut être recoupée, complétée par une autre. En l'occurrence, pour connaître toutes les activités d'un individu sur l'internet, il faudrait disposer de tous les *access_log* des sites qu'il a visités. Chose irréaliste, sauf dans quelques cas particuliers : à l'intérieur d'une institution, ou d'un FAI, le principe commun aux deux exemples étant qu'à la sortie d'un réseau fermé[7], il est toujours possible de poser un mouchard intégral. Encore faudrait-il que la personne disposant de telles informations ait réellement l'ordre ou le désir de mener une enquête policière (ou marketing) ciblée.

Nous savons hélas que ce genre de choses arrive. Non seulement sur commission rogatoire, mais dans un cadre dépourvu de légalité. Avant les attentats du 11 septembre 2001, nous connaissions l'existence du programme *Echelon*[8], et ne pouvons que supputer celle de programmes (secrets) analogues développés en Europe. Après ces attentats, nous avons vu avec quelle célérité les fournisseurs d'accès américains se sont empressés de livrer à la CIA leurs archives[9]. Et nous savons que, depuis, leurs équivalents européens ont l'ordre de garder durant un an l'intégralité des traces des activités de leurs clients[10], sur lesquels ils disposent par ailleurs d'informations très sensibles, comme l'âge, l'adresse, etc.

Ainsi, nos libertés individuelles sont grandement menacées par des formes de surveillance qu'imposerait un prétendu intérêt général. En conclusion, il nous semble important que les lois déterminent précisément ce que l'on a le droit de faire avec de telles archives, et que les États s'y conforment. Pour ce qui nous concerne, la CNIL demande que soient déclarés les fichiers informatiques comportant des données nominatives, et l'entreprise qui nous a fourni les *access_log* a effectivement déclaré ces fichiers. Mais il apparaît qu'il ne suffit pas de respecter la loi : dans le cadre de tout travail sur des fichiers relatifs à des individus, même si les données que l'on manipule ne sont pas

7. Ou sur une ligne téléphonique ou ethernet.
8. http://www.europarl.eu.int/tempcom/echelon/mandate_fr.htm
9. http://interactif.lemonde.fr/article/0,5611,3324–223491-0,FF.html
10. http://www.iris.sgdg.org/actions/lsi/evol/art14.html

nominatives, l'honnêteté intellectuelle et l'exigence scientifique nous semblent essentielles, et vont de pair. Par exemple, les conclusions dépréciatives relatives aux internautes semblent typiques des analystes pressés qui cherchent leurs faiblesses pour mieux profiter de leurs tendances consuméristes. Ceci dit, nous restons étonné par le fait que de nombreuses personnes, aussi peu au fait des lois que de l'informatique, s'inquiètent des menaces que feraient peser sur la vie privée l'étude des *access_log* alors qu'elles ne s'émeuvent pas de la multiplication des programmes de surveillance intégrale. En revanche, le fait que dans un proche avenir, nous pourrions nous faire refuser l'analyse d'archives plus complètes [11] au motif qu'elles contiendraient des informations quasi-nominatives, sans tenir compte du projet scientifique qui nous anime, alors que de telles contraintes ne s'appliquent pas pour une entreprise ni pour une unité de services secrets, qui, bien sûr ne chercheront jamais à publier les résultats obtenus, nous choque.

1.3 Bref panorama des travaux effectués

En matière d'analyse des *access_log*, la quasi-totalité des travaux sont le fait des *Web searching studies*, aussi appelées *Web information retrieval*. De tels travaux sont aussi recensés sous la rubrique *Web Connectivity Analysis*. Jansen et Pooch [JP00] présentent de façon détaillée l'historique et l'état de l'art de telles recherches.

Force est de reconnaître que ces travaux sont essentiellement le fait d'équipes d'informaticiens, qui souvent sont motivés par le déterminisme technique (« le web a un impact majeur sur la société » [12]), et finissent par avoir une vision très centrée sur le fonctionnement des moteurs de recherche. Sinon, il énoncent deux types de résultats : les internautes sont mal à l'aise face au fonctionnement de ces machines, et ces dernières ne sont pas aussi efficaces qu'on pourrait l'espérer ; d'où le besoin d'« aider » l'internaute dans ses re-

11. Nous pensons par exemple aux traces laissées par des caches.
12. [JP00]), citant dans l'introduction Lesk (1997) et Lynch (1997).

cherches [MB00]. Souvent, ces mesures mènent à la production d'indicateurs complexes.

D'autres informaticiens proposent des travaux plus descriptifs, même si leur finalité reste la même : l'optimisation du fonctionnement du moteur. La référence à ce jour reste sûrement l'article déjà ancien « Analysis of a very large AltaVista query log » [S+98]. Le thème des usages s'affirme alors et des analyses de type sociologique commencent à apparaître. Mais le caractère daté de l'enquête (lié principalement aux limites du moteur d'alors) fait qu'une grande insistance est accordée à la présence ou à l'absence d'opérateurs booléens dans les requêtes — différence entre *advance querying* et *simple querying*, ce qui donne à penser que les 20 % de personnes[13] qui les utilisent sont plus « avancées » que les 80 % d'autres. Et malgré la multiplication d'indicateurs statistiques à chaque étape, les auteurs ne peuvent s'empêcher de donner la liste des 25 requêtes les plus populaires, donnant l'impression que les requêtes frivoles ou pornographiques ont une grande importance[14]. Sinon, cet article propose d'autres résultats que nous retrouverons en partie : deux-tiers des requêtes n'apparaissent qu'une fois en 6 semaines, 64 % des sessions consistent en une seule requête sans réplication. L'analyse devient nettement plus critiquable quand elle s'aventure dans l'étude des corrélations entre les 10 000 premiers mots : la focalisation sur le rang permet peut-être d'optimiser l'algorithme du moteur, mais n'informe en rien sur les faits, gestes et demandes des internautes, qui apparaissent en fait comme des « masses sans cesse recomposées ».

Il apparaît donc que le lien étroit qu'ont les chercheurs avec le monde des entreprises, et plus précisement celui des *start-up*, et leur culture de statisticiens les incitent à insister sur les requêtes majoritaires — même s'ils reconnaissent qu'elles n'ont qu'un faible intérêt — et sur l'utilisation des opé-

13. Résultats proposés par [S+98]. Dans [JP00], les travaux sur les *access_log* des moteurs de recherche évaluent entre 2 % et 8 % le taux de requêtes avec opérateurs booléens. Dans [BA02], le taux d'utilisateurs rédigeant de telles requêtes varie entre 3 % et 9 %.

14. Elles représentent en poids 1,5 % du total des requêtes, mais 0,00000016 % des requêtes distinctes. [S+98] pp. 7–8.

rateurs booléens. L'ennui est qu'ils ont créé ainsi une *dynamique* de recherche dont il sera difficile de se départir [15].

Bien sûr, à partir de telles incursions dans la sociologie, on ne pouvait faire marche arrière. La *Transaction log analysis* (TLA) espère comprendre les « actions de l'utilisateur », ses « interactions avec le moteur » (ou le serveur), et la façon dont il « évalue les résultats obtenus » ([JP00]). Les limites de cette démarche, déjà ancienne, sont par ailleurs reconnues. Mais elle sert de tremplin aux études sur les *access_log* et à leur couplage avec des études à caractère plus socio-démographique (allant jusqu'à l'observation participante ou les études en laboratoires, avec des panels spécifiques).

Il s'ensuit une description des internautes et de leurs pratiques « à la louche », qui finit par en donner une image relativement dépréciée : ils écrivent *en moyenne* deux mots par requête, s'intéressent à la pornographie (entre 5 % et 8 % des requêtes), n'utilisent pas d'opérateurs booléens, ne consultent pas plus d'une page de résultats [16], voire ne « savent pas lire plus de 4 lignes sur un écran » [17]. Pour aller plus loin dans l'analyse (par exemple pour rattacher de telles pratiques à des catégories socio-professionnelles), on doit alors faire confiance aux grandes entreprises américaines de sondages, qui évaluent de façon grossière le nombre des internautes, et parfois leur profil sociologique [18].

Cette prolifération de travaux généralistes en langue anglaise ne doit pas inciter à sous-estimer ceux en langue française. Par exemple, des informaticiens, comme Brigitte Trousse [19] et Mahieddine Djoudi [Djo01], conscients des difficultés à consulter le web, se sont attachés à développer des outils d'aide

15. Voir [Jan00] par exemple.
16. Résultats tirés des articles cités.
17. http://www.text-e.org/conf/index.cfm ?ConfText_ID=10
18. Le site http://www.pewinternet.org/reports du *Pew Internet Project*, l'un des moins mauvais du genre, raisonne cependant avec des catégories simplistes (âge, revenu moyen, etc.) qui ne sont jamais croisées.
19. Voir http://www-sop.inria.fr/aid/personnel/Brigitte.Trousse/publications.html pour une bibliographie exhaustive.

à la navigation. D'autres, comme Patrice Abry [20] ou François Baccelli [21], expliquent le fonctionnement du web avec des lois statistiques non classiques, et commencent à essayer de les corréler avec des comportements humains. Jacques Lajoie, grâce à sa culture de psychologue, sait éviter les écueils dépréciatifs des informaticiens américains et, le premier, a montré la richesse des requêtes des internautes, le foisonnement de leurs exigences [Laj98, Laj01]. Le laboratoire *Usages, Créativité, Ergonomie* [22], après avoir longtemps travaillé sur les *access_log* tout en réalisant par ailleurs des recherches de type ethnographique [23], a récemment réussi à combiner les deux : il a étudié durant un an l'intégralité des traces d'un millier d'internautes représentatifs de la population française. Les résultats sont éloquents, car on suit alors les individus un à un plutôt que d'essayer de reconstituer leurs pratiques à partir de leur passage par un point donné. La méthode employée permet un premier calage sociologique de ces pratiques, malgré le biais introduit par le fait que celles-ci s'opèrent exclusivement à domicile [BA02, VB$^+$02]. Citons enfin les travaux de Stat Québec [24], qui offre depuis plusieurs années une description assez complète des pratiques des Québécois en matière d'internet et de leur évolution.

Les recherches à caractère sociologique utilisant les archives électroniques sont surtout promues par les ingénieurs et les informaticiens, et le très petit nombre de sociologues français travaillant dans ce domaine étonne. Cette absence est-elle le fait du mépris pour la technique d'une discipline qui voulait fédérer les sciences humaines ? De la difficulté certaine à obtenir des sources et à en réaliser le traitement ? Paradoxalement, cette trop faible présence empêche l'épanouissement d'un réel débat scientifique sur le thème des usages : souvent les analyses pêchent par excès de préjugés sur les pratiques sociales. Pour remédier à cette situation, pour garantir une bonne analyse de leurs

20. http://www.ens-lyon.fr/~pabry
21. http://www.di.ens.fr/~baccelli
22. France Télécom R & D.
23. Voir par exemple [BV01].
24. http://diff1.stat.gouv.qc.ca/savoir/indicateurs/tic et http://diff1.stat.gouv.qc.ca/savoir/publications/rapports

données, les ingénieurs font appel à d'autres spécialistes. C'est ainsi que les économistes, les linguistes, les spécialistes de l'information et de la communication, mais aussi de l'érudition, sont de plus en plus nombreux à travailler, de façon directe ou indirecte, sur les archives électroniques et sur les usages de l'internet.

Il n'en demeure pas moins que l'intérêt pour les requêtes des internautes aux moteurs de recherche se répand. Par exemple, le site Metaspy[25] propose un instantané des 15 dernières requêtes apparues sur le moteur de recherche MetaCrawler. Diverses entreprises, proposent des analyses, gratuites[26], ou payantes[27]. Mais faute d'approche rigoureuse, les auteurs de ces travaux multiplient les erreurs méthodologiques, sans doute parce qu'ils corrèlent sans précautions des pratiques humaines supposées et des statistiques issues des archives électroniques. Leurs résultats semblent ne pas avoir d'autre fonction que de satisfaire leurs commanditaires : les internautes maîtriseraient tous les moteurs de recherche. Ils seraient satisfaits des résultats obtenus et leurs préoccupations seraient avant tout consuméristes[28]. De telles conclusions sont aussi stupides que peu crédibles, principalement du fait que leurs auteurs n'ont pas compris que ces traces particulières ne pouvaient être appréhendées que dans un contexte d'outillage mental (pour analyser les intentions des internautes, mesurer leurs pratiques, et construire les outils qui permettent de tirer le meilleur parti de telles sources). C'est aussi pour rappeler ce fait que nous avons désiré clore notre travail sur ce type d'archives.

25. http://www.metaspy.com

26. Exemple : *le Capteur de MSN*, http://communities.msn.com/lecapteur

27. Le site www.abondance.com vend 700 Euros son rapport de 105 pages intitulé « La recherche d'information sur le Web francophone : état des lieux ».

28. « 88 % des internautes estiment qu'ils trouvent 'toujours' ou 'souvent' l'information recherchée grâce aux outils de recherche ». Bien sûr, la pornographie et les loisirs viennent aux premiers et seconds rangs, dans ce rapport de mars 2002 à 700 Euros (http://etudes.abondance.com/recherche-res.html).

Chapitre 2

Traces d'un moteur de recherche

Pour montrer comment la liste produit du sens et que celui-ci ne s'obtient qu'en inventant des procédures de lecture et d'écriture, nous avons délibérément choisi une archive parmi les plus contemporaines, les *access_log* d'un moteur de recherche. Ces fichiers contiennent les questions que les internautes transmettent au moteur pour être aiguillés sur des pages web supposées pouvoir leur offrir des réponses. Parmi leurs auteurs, nous comptons des millions de personnes, qui savent qu'elles s'adressent à des machines, aussi leur syntaxe est très particulière. Mais une telle liste doit aussi beaucoup de son existence à ces mêmes machines, auxquelles l'homme délègue de plus en plus de travaux d'écriture, principalement administrative.

Notre objet, parmi les plus classiques — voire le plus vieux ? — de l'écriture, se distingue déjà grandement des listes habituelles. Malgré tout, notre démarche se rapprochera parfois plus de celle de l'archéologue que de celle du sociologue ou de l'informaticien : nos fichiers témoignent d'une activité parcellaire et fugitive d'individus, puisque la fonction d'un moteur de recherche est d'inciter l'internaute à le quitter ; de plus, nous n'avons pas la moindre indication socio-démographique relative aux auteurs de ces pratiques enregistrées. Aussi, leur interprétation s'annonce d'emblée délicate, ce qui fait à notre avis tout l'intérêt de ces masses textuelles : nous pourrons montrer en quoi de nombreuses opérations *a priori* simples — tris, comptages, classifications, etc. — génèrent toute une série de questionnements propres aux sciences humaines, lesquels seront à l'origine des meilleurs résultats que l'on puisse obtenir.

Les résultats suivants proviennent d'un premier travail personnel, et s'intègrent dans le cadre plus vaste d'une recherche labellisée par le RNRT (Réseau National de la Recherche en Télécommunications), réalisée par l'équipe *Réseaux, Savoirs & Territoires* de l'ENS en collaboration avec le laboratoire DMI/GRI Gestion et Recherche d'Informations (France Télécom Recherche & Développement). Les résultats des autres participants à ce projet d'étude relatif aux préoccupations des internautes ne seront pas exposés ici[1].

Le moteur de recherche, appelé Goosta par la suite, était sollicité environ 500 000 fois par jour au printemps 2000 et 700 000 fois un an plus tard. Les *access_log* du premier trimestre 2001 sur lesquels nous allons travailler comprennent donc près de 62 millions de lignes, qui, une fois nettoyés, dépassent quatre giga-octets ; leur traitement aura généré environ 100 Go de fichiers intermédiaires produits pour les besoins de l'analyse.

2.1 Structure des archives

Celles-ci, telles que nous les avons reçues, et une fois nettoyées par l'entreprise[2], sont stockées sur des fichiers quotidiens dont les lignes ont l'allure suivante :

142.149.110.99	P0-97795710-26	[1/1/2001 :0 :0 :11]	booster la cnnection a internet
193.250.163.23	P0-97834398-107	[1/1/2001 :9 :56 :38]	"moteurs de recherche"
152.162.12.144	P0-97267473-17	[1/1/2001 :10 :56 :28]	"gabriel fauré"
205.188.98.157	P0-97759141-325	[1/1/2001 :13 :17 :29]	louer une maison, bail ?

La première colonne correspond au numéro IP de la machine de la personne qui se connecte à Goosta. Dans les faits, c'est souvent le numéro IP du

1. D'une part car l'enquête ne fait que commencer ; d'autre part, pour des raisons déontologiques évidentes : nous n'imaginons pas publier dans le cadre d'une recherche personnelle les résultats de nos étudiants et collaborateurs.

2. En fait par Sylvie Tricot, que nous remercions ici.

modem du fournisseur d'accès et les « reroutages[3] », fréquents, complexifient la recherche de la localisation géographique de l'émetteur — déjà délicate dans les cas favorables.

La seconde correspond à un *cookie*, accepté ou refusé — qui peut aider à singulariser (et non à authentifier) l'auteur de la requête, au moins dans le cadre d'une session — mais qui parfois n'est attaché qu'à une machine collective.

La troisième colonne est évidemment la date.

La quatrième va retenir la majeure partie de notre attention : c'est l'expression de la requête de l'internaute. Les exemples précédents mettent déjà en évidence la présence de signes de tous types (guillemets, ponctuation...) et la présence de fautes de frappe ou d'orthographe. Malgré tout, la variété des requêtes (intégrant même des caractères de contrôle) est telle qu'un nettoyage draconien s'impose avant de pouvoir en opérer un premier traitement.

2.2 Réduction des requêtes

Nous avons étudié les requêtes du premier trimestre 2001, ce qui correspond à 88 jours d'archives, du 1er janvier au 31 mars 2001 — les données des 8 janvier et 22 février étant perdues. Nous obtenons alors 11 millions de requêtes distinctes (pour environ 62 millions d'occurrences), composées à partir de 3,4 millions de formes graphiques différentes.

2.2.1 Première procédure

Il nous faut donc simplifier ces trop nombreuses interrogations, tout en altérant le moins possible leur sens et la démarche de leurs auteurs.

La variété de ces requêtes est à peine imaginable : certaines contiennent plus de... 1700 caractères, d'autres intègrent des tabulations, des passages à

3. Opération qui consiste, pour le propriétaire d'un réseau, à délester une partie de ses modems sur ceux d'une autre région : ainsi une personne de Caen apparaît connectée à un modem de Toulouse.

la ligne, des successions de dièses ou de points, etc. Le nombre de mots qui les composent est colossal. Aussi, avons-nous réalisé un premier toilettage : les caractères de contrôle[4], les symboles &, @, les accolades et les signes de ponctuation (sauf les points) ont été remplacés par des espaces. La répétition de certains caractères a été éliminée : ainsi +++ et - - deviennent simplement + et - (opération appliquée aux caractères suivants : +, -, %, ., #). Ensuite, les parenthèses, apostrophes, signes =, * et $, ainsi que les guillemets (auparavant réduits au type le plus simple), ont été isolés par des espaces. Cette opération n'a pas été appliquée au point pour éviter de scinder des URL (www.x.y) et l'a été partiellement pour le tiret, de façon à conserver les mots composés[5]. Les majuscules ont été réduites en minuscules, et tous les caractères ont perdu leurs diacritiques (le ñ comme le ó ou le ç). Le ß et le æ ont aussi été tranformés en ss et en ae. Pour finir, les espaces précédant et terminant les requêtes ont été supprimés, et les espaces intermédiaires consécutifs — multipliés par les opérations précédentes — réduits à un seul.

Nous appelons **Nett1** cette première procédure de réduction[6]. Elle permet de réduire la taille du vocabulaire de 45 % : on passe de 3 348 586 mots distincts (après élimination des caractères de contrôle) à 1 842 768 ; de façon prévisible, elle réduit peu le nombre de requêtes distinctes : la baisse n'est que de 8 % (10 103 055 requêtes distinctes au lieu de 10 957 253).

Le tableau 2.1 détaille les résultats obtenus avec cette première procédure. Le total des fréquences des mots se monte à 143 276 089 pour 88 jours.

4. Imprévus mais présents : codes ascii 0 à 31, 127 à 160, et 255.

5. Ainsi *Franche-Comté* restera en l'état quand la requête *France -Paris* sera décomposée en *France - Paris*.

6. A partir de maintenant, les formes et requêtes citées le seront le plus souvent sous forme simplifiée : *chomage* sans accent, *gracq* sans majuscule, etc.

période	requêtes distinctes	mots distincts	*Rappel du total des accès*
1 jour (1er jan)	85 050	42 663	206 070
1 jour (4 jan)	200 828	83 663	690 338
30 jours	3 799 976	817 658	20 095 168
88 jours	10 103 055	1 842 768	61 812 583

TAB. 2.1 *Distribution des mots et des requêtes. Procédure* ***Nett1****. La taille du vocabulaire est ainsi réduite de 45 %.*

2.2.2 Programmation et interprétation

De tels nettoyages appliqués à trois mois d'archives prennent environ six heures de travail sur un ordinateur puissant[7]. Le résultat en est une simplification des requêtes, une réduction conséquente du nombre de « mots », ces derniers étant toujours définis comme une succession de caractères séparés par des espaces.

Il nous faut insister sur la complexité d'un tel travail. Nous verrons par exemple qu'un échantillonnage trop précoce biaise fortement l'analyse que l'on pourrrait faire de telles listes de requêtes. Les programmes doivent être exhaustifs (il faut penser à tous les cas de figure possibles) et surtout, doivent s'enchaîner de façon efficace : chaque programme et fichier résultat s'intègrent dans une procédure[8], et il arrive hélas fréquemment que l'on doive la relancer à partir du début, ce qui peut prendre plusieurs nuits de calcul[9], voire plus si l'on arrive en limite des possibilités de la machine[10], et l'on n'est jamais à

7. En l'occurrence un Sun Sparc 5, doté de 500 Mo de mémoire vive et de 46 Go de disque. Les programmes sont en général écrits en Perl, avec bien sûr un usage intensif de commandes Unix quand cela est possible. Nous remercions Jacques Beigbeder ainsi que Michael Totschnig, dont les conseils nous ont permis de diviser par six le temps machine, qui dépassait initialement une journée entière.

8. Le programme *k* produit un fichier résultat *k+1* à partir du fichier *k*, lui-même résultat du programme précédent...

9. Lire les archives quotidiennes sans les décompacter, les nettoyer, les compter, leur appliquer des opérations de recherche ou de réordonnancement, etc.

10. En termes de mémoire vive, si le nombre de variables constituées est trop important, ce qui arrive facilement, ou de mémoire de stockage, à l'occasion d'opérations de tri par

l'abri d'une erreur : dans de tels cas, où la lecture exhaustive des fichiers par l'homme est irréalisable, mieux vaut trop vérifier ses sources que pas assez.

Par exemple, un mois après avoir entamé de tels travaux, nous avons découvert que certaines lignes des archives étaient incomplètes : pour 2174 requêtes du mois de janvier 2001, il manquait au moins l'un des champs précédemment décrits, quand la ligne n'était pas intégralement vide.

À ce stade de la préparation des données, il convient de n'avoir aucun préjugé. Par exemple, on aurait pu se débarrasser de requêtes rares et peu lisibles, comme la requête « *# cat kjbible | tr cs a-za-z 012 | tr a-z a-z | sort | uniq* » qui n'apparaît que trois fois et qui semble difficile à catégoriser en fonction des mots qui la composent, bien qu'elle puisse être compréhensible pour un informaticien. Sinon, l'incompréhension peut être totale, comme avec cette requête : « *#1575 #1604 #1602 #1590 #1575 #1569 #1601 #1610 #1605 #1589 #1585* ». D'autres cas intermédiaires peuvent apparaître, comme pour la requête commençant par « *%28courbes near alg%e9briques%29 and %28%28repr%e9sentation* », dont le sens est clair. Il ne faut pour autant imaginer que les codages seront la cause principale de la difficulté à classer et à interpréter les requêtes ou les mots. Certes, certaines requêtes ne seront pas utilisables, mais il apparaîtra que les requêtes les plus simples et les plus *lisibles* seront parfois les plus délicates à analyser, souvent parce qu'elles sont composées d'un seul mot, polysémique suivant son contexte (*paris*, *carte*, etc.).

Ainsi, une première approche de ces données passe par leur description détaillée, sans aucune forme de préjugé. Ce n'est qu'après un temps assez long qu'émergent, au fil des programmes, des profils variés avec lesquels on finit par se familiariser.

---

exemple : il vaut alors mieux expliciter la partition sur laquelle on compte inscrire les fichiers temporaires sous peine de recevoir un *mail* avertissant que l'opération n'a pu être effectuée par manque de place (cf. l'option *-T* de la commande *sort*).

2.2.3 Seconde procédure

Pour faciliter la description générale des requêtes, et notamment pour en mesurer de façon plus simple le contenu et les agrégations, nous avons appliqué la procédure **Nett2** suivante : tout d'abord, les formes graphiques les plus courantes isolées par la procédure **Nett1** — et ayant souvent un statut spécifique — ont été supprimées. Ce sont : « le », « la », « les », « l », « de », « du », « des », « d », « un », « 2 », « et », « ou », « en », « a », « au », « sur », « pour », « x », « the », « par », « and », « of », « 1 » et « - » ; ensuite, il en a été de même pour les formes de ponctuation et les signes courants qui n'étaient pas encore éliminés avec la procédure **Nett1** : « " », « + », « * », « = », « $ », « (», «) », « $<$ » et « $>$ ». Le tableau 2.2 donne la liste de ces formes éliminées et de leur fréquence, et montre aussi leur grande importance dans le corpus issu de **Nett1**.

Nous appelons « mots-outils » ou « connecteurs » ces formes graphiques : en effet, si l'on considère le graphe dont les mots des requêtes sont les sommets et dont les arcs sont définis par la cooccurrence de deux mots dans une même requête, il apparaît intuitif que ces « connecteurs » garantissent l'existence d'un chemin entre des mots sémantiquement lointains[11] (bien sûr, une telle définition ne correspond pas à celle des grammairiens).

Ce second nettoyage nous fait perdre une partie de l'intention de l'émetteur (par exemple les signes « + », « - » et les guillemets disparaissent), mais rend le corpus plus aisé à traiter. Le nombre de requêtes distinctes se réduit alors à 8 616 504 et le nombre de mots distincts à 1 701 876[12]. Le nombre total de mots n'est plus que de 118 750 102, pour 61 805 286 requêtes. Autrement dit, la suppression de seulement 33 mots, signes ou caractères réduit la fréquence totale des mots de plus de 23 millions[13] et le vocabulaire de

11. Et ce, sans préjuger de la forme exacte du graphe. Pour se faire une idée de la fonction des connecteurs, les exemples suivants devraient suffire : on ne voit pas trop comment relier « excel » à « arcachon ». Or la forme « du » remplit ce rôle puisque l'on trouve les requêtes « ergonomie du logiciel excel » et « geographie du bassin d ' arcachon ».

12. Soit des gains respectifs de 15 % et de 5,5 %.

13. La différence entre le total des fréquences des formes éliminées (25,8 millions) et le

rang	fréquence	forme
1	7 095 156	+
2	4 647 379	de
3	3 698 633	"
4	1 238 572	d
5	1 142 901	la
6	908 149	et
7	805 368	des
8	789 994	du
9	749 225	le
10	734 240	a
12	680 930	en
13	659 555	l
14	610 726	les
21	335 486	sur
23	318 999	pour
42	174 312	au
52	157 130	un
57	152 126	2
65	143 801	-
67	141 663	x
89	115 039	(
93	113 340	)
105	103 487	the
179	73 394	par
180	72 998	and
192	68 321	of
275	52 528	1
289	50 385	*
582	28 909	ou
1951	10 020	=
3835	4816	$
19 815	542	>
45 277	152	<
Total	25 878 276	

TAB. 2.2 *Formes et caractères éliminés avec la procédure* ***Nett2****. Voir le paragraphe 2.2.3 pour savoir si la forme éliminée est isolée ou pas.*

140 000 termes. Bien sûr le nombre de requêtes reste quasi-constant, à l'exception des 7297 requêtes éliminées indépendamment de la procédure pour cause d'incomplétude (cf. page 274).

2.2.4 Précautions

Avant tout chose, nous rappelons le caractère exploratoire d'un tel travail, même s'il existe désormais une littérature abondante sur le sujet[14]. Notre but est de montrer que les outils d'appropriation que nous construisons pour étudier cette longue liste sont essentiels pour lui faire produire du sens autre qu'une succession de préjugés. Par exemple, nous ne connaissons rien des utilisateurs de Goosta, même si nous pouvons penser qu'ils se recrutent principalement parmi un « grand public »[15] aux contours bien vagues : nous savons aussi — grâce à nos enquêtes précédentes — qu'il n'y a pas de corrélation directe entre le niveau culturel et la compétence en matière d'internet documentaire. Enfin, si nous pouvons imaginer que les centres d'intérêt des internautes sont fortement contraints par les médias (télévision, presse, etc.) et par les normes de consommation, nous sommes d'autant moins intéressés à prouver une telle assertion qu'elle nous semble découler, non pas d'un conformisme de la population étudiée, mais de celui des chercheurs (et de leurs méthodes et outils) qui sont à l'origine de ce type de proposition.

Les descriptions qui suivent sont donc sommaires, et tant la structure du corpus que la volonté d'en dégager les traits saillants ne permettront pas, dans le cadre de ce travail, d'utiliser des méthodes syntaxiques raffinées, et encore moins des outils d'analyse sémantique. L'important est de *défricher* et d'explorer un domaine de recherche, d'en montrer les intérêts et les limites,

total des fréquences des mots disparus (23 millions) s'explique par le fait que certaines de ces formes étaient incluses dans des mots. Exemple : « *hot<hack* » donne maintenant deux mots avec la nouvelle procédure, puisque le symbole < est remplacé par un espace. Les symboles « _ » et « % » n'ont pas été supprimés.

14. Cf. le paragraphe 1.3 page 263.

15. Du fait de leur présence sur ce moteur : nos collègues universitaires utilisent en majorité Google, après s'être surtout servi d'AltaVista.

l'intuition qui nous porte étant qu'un tel travail est fructueux, non pas malgré son coût apparent, mais *du fait* de ce coût : il permettra de montrer ce que les sciences sociales — en usant au mieux du potentiel réflexif de l'écriture — peuvent apporter à ce type de recherche, aujourd'hui monopolisée par les informaticiens et les spécialistes du marketing.

2.3 Vocabulaire des requêtes

2.3.1 Richesse lexicale

Dans l'introduction à la seconde édition du *Petit Robert*, il est précisé que le nombre total d'entrées dépasse les 50 000. Nous pouvons concevoir qu'en intégrant les diverses flexions des noms communs et des verbes[16], les mots étrangers les plus courants, les néologismes et surtout les noms « propres » — étendus aux marques, aux produits et aux « célébrités » —, le dictionnaire « courant » se monte alors à quelques centaines de milliers de mots. Or, ici, ce nombre s'élève à près de deux millions.

Plusieurs facteurs peuvent expliquer cette hypertrophie lexicale.

— Tout d'abord, les fautes de frappe (très certainement liées à la trop petite fenêtre de la page du moteur, qui empêche de visualiser plus de 10 caractères consécutifs), d'orthographe et les constructions de mots composés (notamment avec le tiret et l'espace souligné, voire sans rien — l'espace entre les mots étant parfois absent).

— l'usage délibéré d'un français libéré des contraintes orthographiques, par exemple phonétique (*kdo*) ou humoristique (*windaube*), de façon à cibler plus efficacement les pages recherchées.

— Au-delà des noms « propres », la présence de mots relatifs aux objets techniques de consommation, qui sont déjà des hétérographes, du fait de la présence variable de l'espace inter-« mots » : *F801*[17], *900XJ*, *Photoshop5.0*.

16. Mais ces derniers sont rares dans un tel corpus.

17. A la requête « *F 801* », Google propose 726 000 réponses, et conseille de la corriger en « *F801* », qui n'en renvoie plus que 6240 (21 janvier 2002).

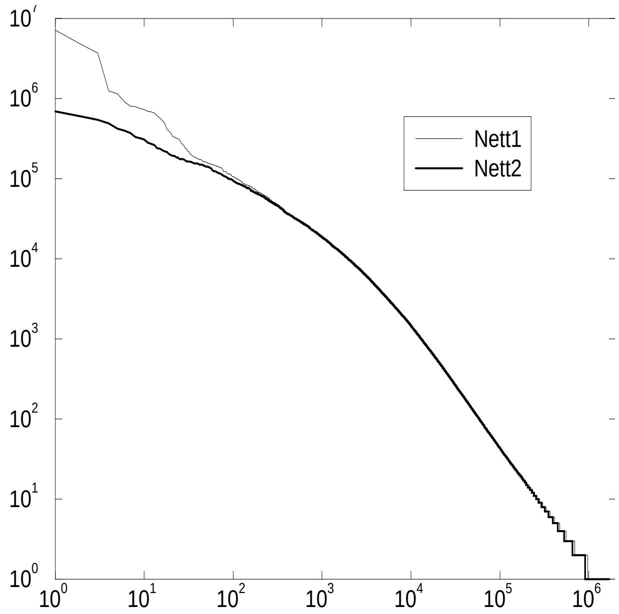

FIG. 2.1 *Graphe rang-fréquence des mots avec les deux procédures. En abscisse, le rang des mots, en ordonnée, leur nombre d'occurrences.*

Ce type de mots est virtuellement infini, puisque tout nombre peut avoir le statut de mot (pensons à 2001, 1914, 3615, 1515, etc.).

2.3.2 Distribution statistique

Avant de nous intéresser à certains éléments spécifiques du lexique, nous devons compléter — sous forme synthétique — les résultats déjà présentés dans les tableaux 2.1 et 2.2 (pages 273 et 276) : on se souvient qu'avec la procédure **Nett1**, les mots les plus fréquents étaient des connecteurs. La grande variation des fréquences des mots incite à mettre celles-ci en correspondance avec le rang de ces mots, tout en choisissant des échelles logarithmiques (figure 2.1) : sinon, le graphe, en forme d'équerre, est illisible.

Ce graphique montre qu'une fois passés les mots les plus fréquents, les deux procédures donnent des résultats identiques, qui par ailleurs mettent en évidence une assez bonne conformité du vocabulaire avec la loi de Zipf[18],

18. Qui dit que le produit de la fréquence d'un mot par son rang est constant.

N premiers mots	fréq. cumulées	% cum.	poids dans le lexique	fréq. dernier mot
10	21 809 617	15,2 %	0 %	734 240
100	40 293 662	28,3 %	0,005 %	106 135
1000	74 316 076	52,1	0,05 %	18 695
10 000	116 148 996	81,5	0,54	1455
100 000	136 275 312	95,6	5,4	44
1 000 000	142 433 321	99,9	54	1
1 842 768 (total)	143 276 089	100	100	1

TAB. 2.3 *Distribution des mots du dictionnaire de Goosta classés par fréquence décroissante. Procédure* ***Nett1****.*

surtout pour les mots à fréquence intermédiaire (entre 10 et 10 000 occurrences).

L'homogénéité de cette classe de mots, et le souvenir de l'intérêt porté par les lexicométriciens aux mots à fréquences rares nous inciteront à ne pas nous focaliser sur les mots à fréquences majoritaires, même si ces derniers ont un poids considérable dans le corpus, comme le montrent les tableaux 2.3 et 2.4.

En fait, la présentation par rangs décroissants proposée par ces tableaux est dangereuse : déjà, les premières fréquences se cumulent ; par exemple, les dix premiers mots obtenus avec la procédure **Nett1**, qui sont tous des connecteurs, accroissent le poids des 100 premiers, etc[19]. Mais surtout, les chiffres n'incitent qu'à étudier les 1000 premiers mots : suivant la méthode utilisée, ils épuisent 42 % ou 52 % des fréquences calculées, alors qu'ils ne constituent que 0,06 % du vocabulaire utilisé (les personnes toutefois soucieuses de connaître ces mille premiers mots et requêtes pourront consulter les paragraphes 4.1 et 4.2de l'annexe (p. 415 et 424). Or, la très grande majorité des analystes d'*access_log* tombent dans ce piège[20].

Considérons par exemple le tableau 2.5, qui présente la liste des 15 premiers mots, obtenue avec la méthode **Nett2**. Il ne nous apprend rien, si ce

19. La procédure **Nett2** corrige en partie cette illusion (voir tableau 2.4).

20. Par exemple, le rapport à 700 Euros précité (cf. note 27 page 267) étudiera les requêtes des internautes à partir des 550 mots les plus courants, réutilisés comme mots-clés.

N premiers mots	fréq. cumulées	% cum.	poids dans le lexique	fréq. dernier mot
10	4 456 598	3,75 %	0 %	309 901
100	17 314 889	14,6 %	0,006 %	94 410
1000	50 049 680	42,1	0,06 %	18 581
10 000	92 140 140	77,6	0,6	1461
100 000	112 255 915	94,5	5,9	43
1 000 000	118 074 638	99,4	58,76	1
1 701 876 (total)	118 750 102	100	100	1

TAB. 2.4 *Distribution des mots du dictionnaire de Goosta classés par fréquence décroissante. Procédure* ***Nett2****.*

n'est que les mots les plus utilisés sont polysémiques. Dans cette liste, le mot « *carte* » renvoie-t-il à « *carte bleue* », « *carte routière* », ou à « *carte de vœux* » ? Que penser d'une personne qui a saisi une requête contenant le mot « *sexe* » ? Recherche-t-elle des images obsènes ou vérifie-t-elle que son enfant n'y accèdera pas ? Dans quel contexte se comprend le mot « *location* » ? A-t-on un propriétaire, un futur locataire derrière l'écran ? Ces locations sont-elles saisonnières ou non ?

Face à une liste de ce type, la construction de regroupements — *sexe* avec *sex* et *nue* (rang 20), *photo* avec *photos*, *mp3* avec *musique* (rang 19), etc. — donne une illusion de sérieux et de rigueur, alors qu'elle risque fort de ne rien apporter, puisqu'elle s'appuie sur une méthode faussée dès l'origine.

Après avoir évité ce premier piège, qui consiste à se focaliser sur les mots les plus fréquents, nous retiendrons que la moyenne d'usage des mots est élevée, comprise entre 70 et 78 suivant la méthode choisie, alors que le taux d'hapax ne s'écarte pas trop des normes « littéraires » : 48 % dans les deux cas. Et le graphique 2.1 rappelle que seulement 250 000 mots — soit moins de 15 %, avec l'une ou l'autre méthode — apparaissent plus de 10 fois en 88 jours. En revanche, l'agrégation que proposent le tableau 2.6 et le graphique associé 2.2, met bien en évidence le très petit nombre de mots à fréquences élevées, et l'importance en taille comme en poids des mots aux fréquences faibles (par exemple entre 2 et 10 000).

fréquence	mot
688 278	sexe
576 184	france
529 294	paris
487 042	gratuit
419 503	location
393 489	photo
369 153	photos
330 797	carte
316 871	jeux
306 657	video
277 851	emploi
269 857	immobilier
260 077	sex
238 580	hotel
236 779	mp3

TAB. 2.5 *Les 15 mots les plus fréquents. Procédure **Nett2**. Voir aussi la liste des 1000 premiers page 424 de l'annexe (paragraphe 4.2).*

Classe de fréquences	Nombre de mots	Total des occurrences
1	798 095	798 095
2 à 10	649 653	2 516 234
11 à 100	195 793	5 929 300
101 à 1000	45 322	13 756 738
1001 à 10 000	11 053	32 747 893
10 001 à 100 000	1872	46 886 421
plus de 100 000	88	16 141 833

TAB. 2.6 *Classes de fréquences des mots. Procédure **Nett2**. 45 322 mots ont une fréquence comprise entre 101 et 1000, et leur poids total dans le corpus est de 13 756 738 (cf. le graphique 2.2).*

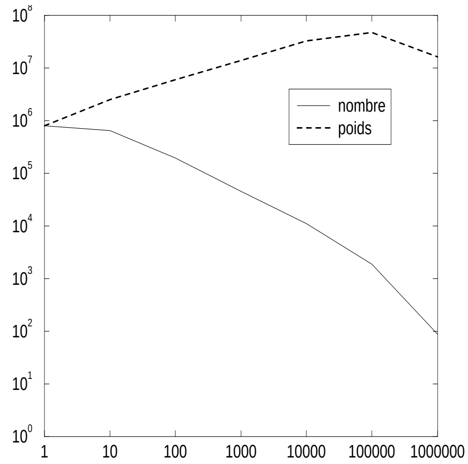

FIG. 2.2 *Nombre et poids des groupes de mots triés par fréquence. Lecture : 649 653 mots ont une fréquence comprise entre 2 et 10 (inclus). 88 mots ont une fréquence comprise entre 100 001 et 1 000 000. Le total d'occurrences des 649 653 mots de la seconde classe est de 2 516 234. Celui des 88 mots les plus fréquents est de 16 141 833 (cf. tableau 2.6). Procédure **Nett2**.*

2.3.3 Importance des mots rares

En effet, 86 % des mots apparaissent entre une et dix fois dans cette longue période.

Certes, nombreux sont les mots mal orthographiés (souvent des mots composés sans séparateur, comme « *galleriephoto* » ou « *comparaisonsautos* »), Or, si ceux-ci risquent fort de renvoyer un message d'erreur de la part du moteur de recherche, ils apportent néanmoins du sens et témoignent d'une variété des pratiques plus instructive que la simple présence de mots banals.

Par exemple, pour les données des 88 jours du printemps 2001, on a dénombré 4820 mots contenant la graphie « photo » (*photoaerienne*, *photoshop*, *photo_arbre_hevea*, *www.hp-photoworld.com*, *pphoto*, etc.) ; 433 mots contenant la graphie « peugeot », 274 contenant la graphie « apple », et 186 la graphie « microsoft » (Procédure **Nett1**). On pourrait regrouper ces formes avec la forme attendue, qui reste largement majoritaire : on dénombre 60 493 occurrences pour le mot exact « peugeot », quand la seconde forme est bien plus rare — *www.peugeot.fr*, avec 290 occurrences[21]. Mais nous savons qu'une telle démarche est hâtive. En revanche, les formes rares dévoilent donc un sens précis. Considérons par exemple le mot « chomage » : il apparaît 16 885 fois, et ses formes dérivées (au nombre de 56) apparaissent au total 224 fois. Parmi les formes graphiques les plus rares, on recense *emploi-chomage* (12 occurrences), *couverture-chomage* (4 occ.), ainsi que *chomage2000*, *chomage.com*, *assurance-chomage* (4 occ. aussi), et enfin les hapax *denonciatiochomage* et *definitionduchomageenfrance*. Paradoxalement, ces dernières formes font sens pour la personne qui analyse les requêtes, alors que d'autres, peut-être pertinentes pour le moteur, seront plus difficiles à interpréter.

Parmi ces mots rares, nous découvrons aussi des mots de notre vocabulaire courant : en trois mois, *heterodoxe* apparaît 10 fois, comme *horlogeries* (qui apparaît 2755 fois au singulier), *hydroxydes* est repéré 6 fois, comme *ineligibles* (qui apparaît 4 fois au singulier), *inquietante* ou *insatiablement*. De même, *analytiquement* n'apparaît que 4 fois — et *psychanalytiquement* une seule —, *scintillographie* 3 fois, *soliloquer* et *soliloque* une fois (procédure **Nett2**). Autrement dit, éliminer les mots à faibles fréquences consiste à réduire notre langue sans discernement.

Ce fait est particulièrement flagrant dans le cas des personnes. Des auteurs comme Pierre Bourdieu ou Julien Gracq[22] ne sont cités que respectivement 1679 fois (rang 9050) et 290 fois (rang 29 759). C'est faible face à Molière[23]. Le géographe Marcel Roncayolo a bien moins de succès avec 24 occurrences, et un rang de 148 568, soit autant que Jean-Claude Chamboredon (rang 146 118), à peine plus que Gérard Noiriel (23 occ., rang 152 429) ou Jack Goody (21 occ., rang 161 065), mais déjà loin devant Philippe Descola (12 occ., rang 230 696).

21. Les autres sont : *peugeot.fr* (220 occ.), *clicpeugeot* (78), *peugeot307* (66), *peugeot.com* (59).

22. En recherchant le mot exact *bourdieu*, puis *gracq*, et en évitant les formes qui les contiennent, comme *dubourdieu*, etc. (Procédure **Nett2**).

23. La première apparition concerne le nom seul du dramaturge : 18 325 occurrences, rang 993.

Ces exemples prouvent à quel point il est dangereux de négliger les mots rares : non seulement, on fonde l'analyse sur les mots les plus courants, souvent les plus polysémiques et les moins explicites, mais on risque de perdre des pans entiers de la culture maîtrisée — ou sollicitée — par les internautes.

Cette insistance sur les mots rares répond aux quelques études disponibles sur les requêtes des moteurs. Celles-ci ont une fâcheuse tendance à valoriser les « têtes de listes » que sont les mots à fréquences majoritaires, en les décontextualisant de façon artificielle : beaucoup sont des fragments de syntagmes nominaux ; et l'on verra que ces mots courants apparaissent fréquemment comme des requêtes complètes. Dans une telle situation, on ne peut rien déduire de l'intention de l'émetteur.

Nous comprenons donc qu'il est nécessaire, pour mener une étude efficace, d'oublier les mots fréquents, ou, au moins, de ne pas se focaliser sur leur fréquence.

2.3.4 Typologie formelle des mots

Pour nous persuader de la grande variété des pratiques, voire de leur étrangeté, nous pouvons préciser la distribution de la longueur des mots et leur homogénéité interne, définie de façon purement formelle comme suit : c'est le nombre de caractères alphabétiques du mot — c'est-à-dire ne comprenant que les symboles a à z —, divisé par sa longueur.

Tout d'abord, le graphique 2.3 décrit la distribution des mots (distincts ou cumulés) en fonction de leur longueur : si l'on rencontre un mot de 1099 caractères et un autre de 1456, les mots les plus fréquemment cités ont entre 3 et 12 caractères, la grande majorité des mots distincts en a entre 5 et 11[24] (procédure **Nett2**). On remarque aussi que ce groupe de mots entre 5 et 11 caractères est le plus représenté, puisqu'il correspond à 87 539 884 occurrences, soit encore 73 % du total des fréquences. La superposition des deux courbes permet d'avoir une idée de la fréquence d'apparition des mots

24. 73 % du vocabulaire, le maximum (226 631) étant atteint pour les mots de 7 caractères, les autres classes (mots de longueur 5 à 11) comprenant au moins 100 000 mots.

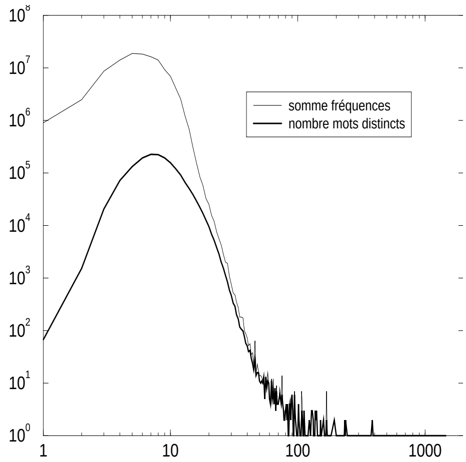

FIG. 2.3 *Nombre de mots d'une longueur donnée. En abscisse, le nombre de caractères. La courbe épaisse du bas traduit le nombre de mots distincts ayant une longueur donnée (en abscisse), l'autre décrit leur poids. Procédure* ***Nett2****.*

d'une longueur donnée : autant on conçoit que les courbes soient très proches pour les grandes longueurs (de tels mots apparaissent rarement plus de deux fois), autant les fortes répétitions de mots courts surprennent ; 68 mots de longueur 1 génèrent à eux seuls 900 420 occurrences, et 1524 mots de longueur 2 en génèrent 2 468 056. Certes, on sait qu'un seul mot de longueur 4 en génèrera près de 700 000, mais que *pèse*-t-il face à ces 1590 mots brefs ?

Reste à comprendre la forme de ces mots. Le graphique 2.4 donne le pourcentage de mots exclusivement composés de lettres de l'alphabet en fonction de leur longueur. Est rappelée, en bas du graphique, la fréquence de ces mots. On constate ici que les mots qui ont le plus de chance d'apparaître dans un dictionnaire[25] sont les mots de longueurs 5 à 13, puisque leur taux d'homogénéité interne dépasse 99 %. Et ce graphe permet de mieux comprendre la forme des mots courts : ce sont surtout des abréviations ou des combinaisons de chiffres et de lettres[26].

25. Mots composés exclus.

26. Voici quelques exemples de mot fréquents d'un ou deux caractères : *cd* (139 673 occ.), *st* (81 569 occ.), *pc* (84 671 occ.), *tv* (82 546 occ.), *3d* (37 709 occ.), *35* (19 462 occ.).

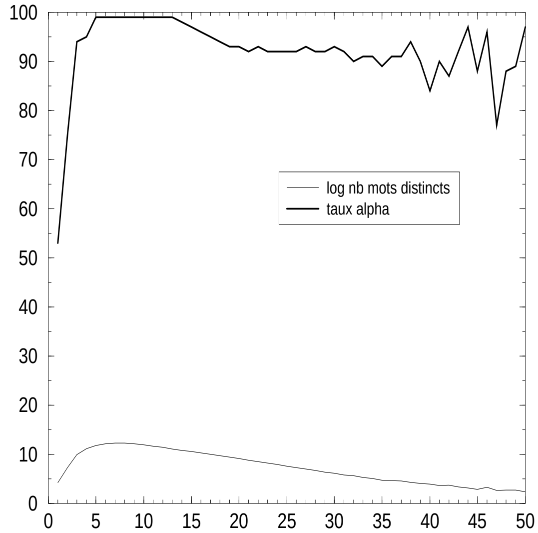

FIG. 2.4 *Homogénéité interne des mots : taux de caractères alphabétiques des mots distincts d'une longueur donnée. En abscisse, le nombre de caractères. Graphe limité aux mots de longueur inférieure à 50. Procédure* ***Nett2****.*

Le tableau 2.7 met en évidence cette singularité lexicale, tout en rappelant que la grande majorité des mots est, comme on l'a déjà vu, composée exclusivement de caractères alphabétiques.

Ainsi, de nombreux mots n'appartiennent pas à des dictionnaires : ils sont 118 314 à posséder un taux d'homogénéité interne inférieur à 80 %. Et ils apparaissent 1 357 865 fois en tout. Ceci incite à évaluer les mots composés de chiffres seuls, et de chiffres et lettres seuls. Par exemple, *2001* apparaît 124 902 fois et *2000* 110 893 fois. Au total, 179 mots parmi les 10 000 premiers

Taux de car. alpha	nb occ. (poids) du groupe	nb mots distincts
entre 0 et 49 %	211 085	42 724
entre 50 et 79 %	1 146 780	75 590
entre 80 et 89 %	741 872	148 562
entre 90 et 99 %	404 943	121 308
100 %	114 345 654	1 313 691

TAB. 2.7 *Taux d'homogénéité interne des mots. Procédure* ***Nett2****.*

sont composés de chiffres seuls, pour un poids total de 1 244 342 occurrences, et 286 ne comportent que des chiffres et des lettres, mais au moins un chiffre, pour un poids total de 1 938 809. Sur le total des « mots », on en dénombre en fait 11 293 qui ne sont que des nombres, et leur fréquence cumulée est de 1 554 055.

De tels regroupements rappellent ceux relatifs aux connecteurs de la première procédure. Ils mettent en évidence des usages qui nous semblent plus pertinents que ceux que décrivent les conclusions relatives aux mots les plus fréquents. Par exemple, parmi ces « mots-nombres », on dénombre 877 numéros de téléphone[27], pour 1252 occurrences. Cela signifie que certains utilisateurs imaginent que Goosta offre aussi les fonctionnalités d'un annuaire inverse[28] — d'ailleurs très demandé (rang 21, pour 44 681 occurrences).

*

L'étude des mots privés de leur contexte (la requête) a peu d'intérêt, tout comme la focalisation sur leurs rangs : la haute fréquence d'un mot ne nous apprend rien sur le nombre d'utilisateurs l'ayant cité, ni sur leurs intentions. Pourquoi conclure des pratiques à partir des mots *sexe* ou *mp3*, alors qu'on ne conclut rien des mots *de* ou *la* ? En revanche, les mots rares, trop vite oubliés, nous rappellent que les internautes ne sont pas aussi conformistes qu'on pourrait l'imaginer lors d'une étude pressée. Accessoirement, le choix de ces mots rares témoigne de leur curiosité : la saisie d'un ou plusieurs mots banals sur un moteur de recherche renvoie avant tout du bruit.

Nous découvrons aussi à quel point une étude approfondie est coûteuse : elle nécessite du temps (temps de l'homme, temps de la machine), et sollicite diverses disciplines. Pourtant, notre travail ne fait que commencer : le besoin se fait sentir d'étudier les requêtes elles-mêmes, les sessions, et enfin de tenter de cerner leurs auteurs.

27. Mots d'exactement dix chiffres, commençant par un zéro.

28. Etendue aux requêtes composées exclusivement de numéros de téléphone, une étude analogue montrera qu'il y a 1753 telles requêtes distinctes, pour un poids total de 4495 occurrences (intégrant donc les symboles espace, tiret et point)

Enfin, nous découvrons un étrange dictionnaire : comme les mots de *tous les jours*, il renvoie à une culture grand public, avec des mots éphémères (starlettes et starletons[29], mots d'actualité, etc.), et les fautes de frappe ou d'orthographe sont communes. Mais il témoigne aussi d'une explosion du vocabulaire, fruit de l'industrialisation (médicaments, noms de marques, pièces techniques, etc.) et il est difficile de savoir si cette « explosion » de ces mots exclus de la *littérature* a un lien ou pas avec la transformation des techniques d'appropriation et de recomposition de l'écriture.

Cette variété du vocabulaire aura eu le mérite de nous faire prendre conscience que, même si nous écrivons avec peu de mots, un bien plus grand nombre nous encombre l'esprit. Nous y reviendrons.

2.4 Typologie des requêtes

2.4.1 Distribution générale

On rappelle que l'on dénombre un total d'environ 62 millions de requêtes sur 3 mois (cf. tableau 2.1 p. 273), dont dix millions de requêtes distinctes pour la procédure **Nett1** et 8,6 millions pour la procédure **Nett2**.

Les deux procédures donnent les mêmes résultats en matière de distribution rang-fréquence des requêtes. Aussi le graphique 2.5 se limite-t-il aux résultats de la procédure **Nett2**, et pour les requêtes dont les rangs sont compris entre 10 et 4 500 000[30].

Ce graphique, qui met en évidence une exceptionnelle conformité avec la loi de Zipf, appelle divers commentaires. Tout d'abord, les requêtes semblent distribuées suivant trois régimes : Les 1000 premières, signalées par une portion de graphe en zig-zag dans la représentation log-log. Leur fréquence s'étale entre 284 462 (*sexe*, seule à dépasser 100 000 occurrences) et 4042 (*dijon*). Celles de rang supérieur à 300 000, signalées par une portion de graphe en

29. On pourrait appeler ainsi les célébrités d'un jour de sexe masculin.

30. De façon à mettre en évidence les premières des 4 millions de requêtes n'apparaissant qu'une fois.

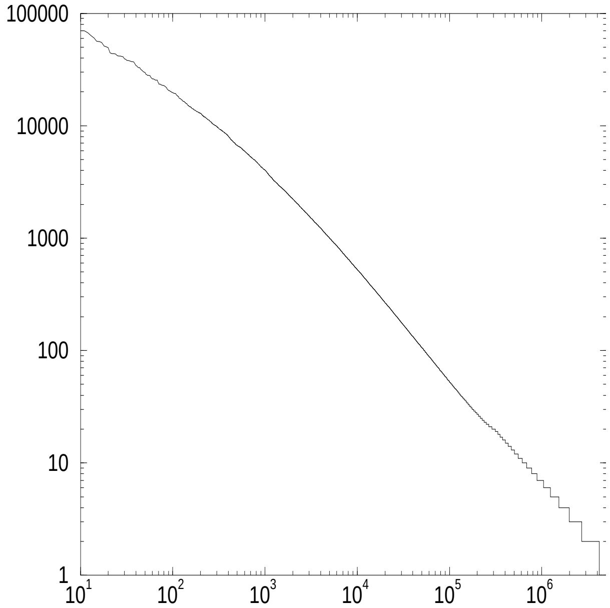

FIG. 2.5 *Graphe rang-fréquence des requêtes avec la procédure* ***Nett2*** *(rangs compris entre 10 et 4 500 000).*

escalier, preuve qu'elles sont très nombreuses à apparaître peu fréquemment (entre une et 20 occurrences, la 300 001^e étant « *hypnose contre harcelement* »). Mais ces requêtes constituent plus de 96 % des requêtes distinctes et pèsent 37 % des occurrences du total[31]. Ces importantes proportions prouvent, encore plus qu'avec les mots, que les évènements rares ne peuvent être éliminés de l'étude, même si leur masse les rendra particulièrement difficiles à analyser.

Enfin, le régime intermédiaire des requêtes de rangs compris entre 1001 et 300 000, qui se distingue par une remarquable linéarité de la courbe. Nous remarquons que la pente des requêtes de ce second groupe est la même que celle des requêtes de fréquences très faibles. Sa mesure précise amène à proposer les deux équations symétriques suivantes :

31. 8 329 469 requêtes ont une fréquence comprise entre 1 et 20 pour un poids total de 22 267 710 d'occurrences. De façon analogue, les 3 566 854 requêtes apparaissant entre deux et dix fois apparaissent 13 087 637 fois et il y a 4 436 513 requêtes singulières, soit 51,4 % du total des requêtes distinctes.

$$freq = 3 * 10^6 * rang^{-0.95} \tag{2.1}$$

et

$$rang = 6{,}6 * 10^6 * freq^{-1{,}05} \tag{2.2}$$

équations valables pour $rang \geq 1000$ et pour $freq \leq 5000$.

Ainsi, les requêtes, certes simplifiées, ont un comportement inattendu : que ce soit avec la première ou la seconde procédure, elles obéissent à des lois de puissance mieux que ne le font les mots du lexique[32].

La comparaison des deux tableaux 2.8 et 2.9 rappelle que les deux procédures donnent des résultats équivalents, contrairement aux mots, et on remarque un réel tassement de la distribution cumulée par rapport à celle du vocabulaire : les 1000 premières requêtes ne « pèsent » qu'environ 17 % du total. Et, quelle que soit la méthode employée, il faut atteindre le rang 100 000 pour obtenir des résultats analogues aux mille premiers mots. Par exemple, avec la procédure **Nett2**, les requêtes apparaissant au plus 20 fois (soit en moyenne une fois tous les quatre jours[33], ont un poids total conséquent, qui dépasse celui des 10 000 premières requêtes : ici, nous ne pouvons nier l'importance des requêtes rares.

Le graphique 2.6 synthétise ces résultats en mettant en évidence le problème méthodologique que l'on pressent : comment créer un échantillon significatif de ces requêtes, puisque les plages intermédiaires ont toutes des poids analogues ? La sagesse inciterait à accepter toutes les requêtes apparaissant entre 2 et 10 000 fois, ce qui correspond hélas à 4 179 698 de requêtes distinctes, pour un poids total de 50 909 962. Se limiter aux requêtes apparaissant entre 11 et 10 000 fois réduit le nombre de requêtes distinctes à 612 844 et leur poids à 37 822 325.

32. Cette conformité à une telle loi signifie, pour les physiciens, qu'aucun phénomène n'est vraiment négligeable (communication de Marc Barthélémy).

33. Et de rang élevé : supérieur à 300 000.

N premières requêtes	fréq. cumulées	% cum.	fréq. dernière req.
10	996 496	1,6 %	69 927
100	3 754 713	6	18 535
1000	10 213 110	16,5	3787
10 000	20 375 617	33	487
100 000	31 648 608	51,2	49
1 000 000	44 992 358	72,8	7
10 103 055	61 812 583	100	1

Tab. 2.8 *Distribution des requêtes classées par fréquence décroissante. Procédure* ***Nett1****.*

N premières requêtes	fréq. cumulées	% cum.	fréq. dernière req.
10	1 003 910	1,6 %	70 122
100	3 876 178	6,3	19 755
1000	10 726 894	17,4	4042
10 000	21 565 130	34,9	525
100 000	33 815 934	54,7	53
1 000 000	47 493 428	76,8	7
8 616 504	61 803 155	100	1

Tab. 2.9 *Distribution des requêtes classées par fréquence décroissante. Procédure* ***Nett2****.*

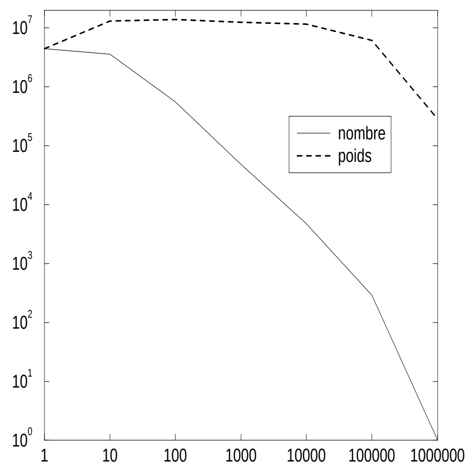

FIG. 2.6 *Nombre et poids des classes de requêtes triées par fréquence. Les graphes se lisent ainsi : 560 170 requêtes ont une fréquence comprise entre 11 et 100 (inclus). 291 ont une fréquence comprise entre 10 001 et 100 000. Le nombre total d'occurrences des 560 170 requêtes est de 13 828 486. Celui des 291 est de 6 172 218. Procédure* ***Nett2****.*

2.4.2 Complexité des requêtes

En fait, les requêtes majoritaires sont d'autant moins intéressantes qu'elles sont souvent composées de simples mots, et qu'elles participent faiblement à l'accroissement du vocabulaire : le graphique 2.7, relatif aux quatre millions de premières requêtes, montre que les mille[34] premières requêtes ne contiennent en moyenne qu'1,3 mots, et qu'il faut attendre les requêtes de rang 150 000 pour atteindre une moyenne de deux mots par requête.

Ce nombre de mots des requêtes s'accroît jusqu'à celles de rang 300 000 (2,8 mots en moyenne) puis « chute momentanément » et reprend son accroissement tout en restant souvent au dessus de 2. Cependant, à partir des requêtes de rang 600 000, on constate de grandes variations de la moyenne, qui parfois descend à 1,2 mots et peut s'élever jusqu'à 4 ou 5, et à partir du rang 2 000 000, cette moyenne dépasse souvent 3 mots.

34. Ce résultat a été testé sur les blocs de requêtes par paquets de 100 et de 500 avant d'être repris pour des paquets de 1000 : les résultats sont identiques.

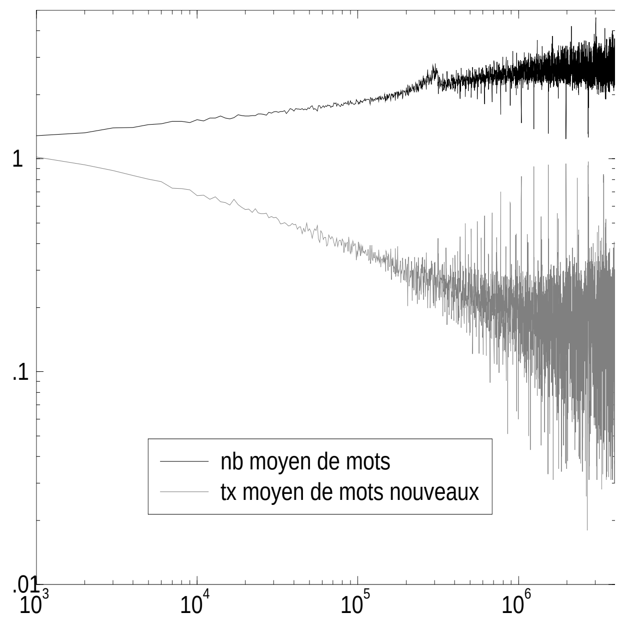

FIG. 2.7 *Richesse lexicale des requêtes (rang ≤ 4 000 000). En noir, le nombre moyen de mots pour des paquets de mille requêtes, ordonnées par fréquences décroissantes. En gris, le taux moyen de mots nouveaux. Procédure* ***Nett2****.*

On mesure bien la fiabilité de cette mesure moyenne en considérant l'écart-type (graphique 2.8), particulièrement faible pour les requêtes majoritaires (0,47), dépassant souvent 1 à partir des requêtes de rang supérieur à 300 000 : nous sommes assurés de ne rencontrer que des requêtes d'un ou deux mots parmi les 50 000 premières requêtes[35], alors que les requêtes de 4 mots ou plus sont communes pour les rangs dépassant 300 000. À ce seuil, nous retrouvons la brutale variation remarquée dans le graphique 2.7 ; à partir du rang 1 000 000, moyennes et écarts-types varient grandement.

35. Rang à partir duquel l'écart-type est supérieur à 0,7 et la moyenne dépasse 1,7.

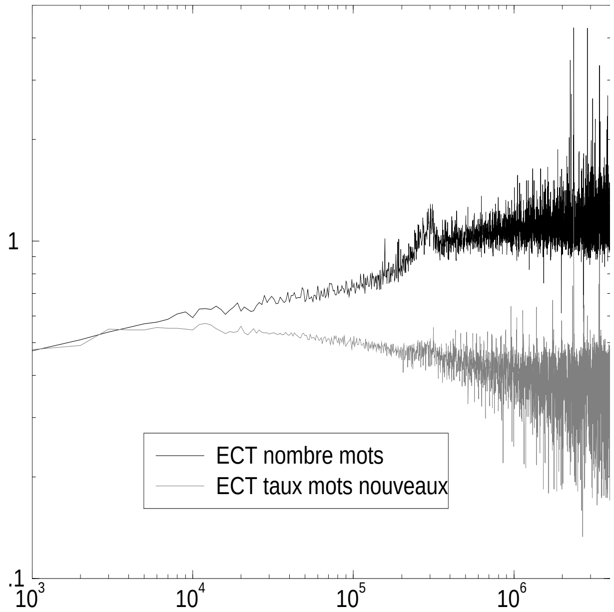

FIG. 2.8 *Écarts-types du nombre de mots et du taux de mots nouveaux pour les requêtes de rang ≤ 4 000 000, pour des groupes de 1000 requêtes (suite du graphique 2.7). Procédure* ***Nett2****.*

De même, le taux de mots nouveaux[36] suit une décroissance quasi logarithmique au début, et retrouve parfois une croissance rapide pour les rangs dépassant 250 000 (environ 27 % de mots nouveaux pour ce rang), en s'approchant parfois de 100 % pour les requêtes de rang supérieur à 1 000 000. Comme pour le nombre moyen de mots, la courbe de l'écart-type permet de confirmer ces analyses réalisées à partir de la moyenne : au-delà du rang 1 000 000, les variations de taille des requêtes peuvent être très importantes.

Nous pressentons donc que les requêtes rares sont celles qui génèrent la richesse lexicale du vocabulaire, celle-ci devant être entendue comme com-

36. Celui-ci est calculé ainsi : pour chaque groupe de mille requêtes, on compte le total des fréquences des mots, et le nombre de mots qui n'apparaissaient pas dans les groupes précédents. Le taux de mots nouveaux est le rapport de ces deux nombres. Il est évident que cet indicateur ne peut que décroître, puisqu'il tient compte de la totalité du vocabulaire qui a déjà été rencontré dans les précédentes requêtes et puisque la longueur des requêtes tend à s'accroître.

binaison de mots mal orthographiés, de noms propres et de mots rares qui ne renvoient pas au vocabulaire de la consommation courante, comme par exemple ceux qui sont utilisés dans les requêtes à caractère intellectuel.

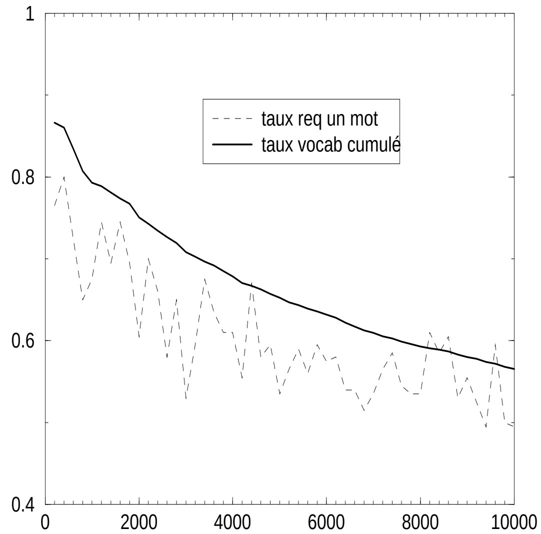

FIG. 2.9 *Requêtes d'un seul mot et vocabulaire cumulé des 10 000 premières requêtes. Les requêtes sont ordonnées par fréquence décroissante puis regroupées par blocs de 200. Le graphique se lit ainsi : les 200 premières requêtes sont à 77 % des requêtes d'un seul mot, et les 200 suivantes le sont à 80 %. 87 % des mots des 200 premières requêtes sont distincts, comme le sont 56 % des mots des 10 000 premières. Procédure* ***Nett2****.*

L'indigence du vocabulaire des requêtes majoritaires surprend, mais est confirmée par le graphique 2.9, qui donne, pour les 10 000 premières requêtes regroupées par paquets de 200, le pourcentage de requêtes d'un seul mot, et la taille du vocabulaire cumulé, divisée par le nombre de mots des requêtes qui le génèrent. Nous voyons combien pèsent les requêtes d'un seul mot (toujours plus de 50 % des requêtes, et souvent plus de 70 % pour les 2000 premières), et surtout comment ces requêtes majoritaires contribuent peu à l'enrichissement du vocabulaire : les 10 000 premières, composées au total de 14 335 mots (soit donc 1,4 mots par requête en moyenne), n'en contiennent que 8105 distincts (ce qui explique le taux de 56,5 %, pour ce rang 10 000). Or, ce vocabulaire réduit pèse lourd dans le total des occurrences des mots, puisqu'un mot dans une requête apparaît au moins autant de fois qu'elle[37]. En fait, les 1285 mots des 1000 premières requêtes réalisent un total d'occurrences de 13 246 600, impressionnant pour le publicitaire pressé[38], alors que l'internaute qui les adresse au moteur de recherche est quasiment assuré de ne pas obtenir les informations qu'il espère, d'une part parce que sa requête n'est pas assez précise, d'autre part parce que le moteur propose trop de réponses.

En fait, les requêtes majoritaires n'ont pas plus de pertinence que les mots majoritaires, car elles sont très souvent composées d'un seul mot, nécessairement banal. Elles sont tout peu informatives, pour l'analyste comme pour l'internaute : ce dernier risque fort d'avoir du mal à faire son choix parmi les réponses offertes si sa requête se limite à des expressions du type *horoscope* (rg. 9, 73 071 occ.), *corse* (rg. 80, 22 682 occ.), *islam* (rg. 415, 7786 occ.), *maison* (rg. 550, 6349 occ.), *broderie* (rg. 671, 5468 occ.), « *loterie gratuite* » (rg. 910, 4299 occ.), ou encore *restaurants* (rg. 992, 4059 occ.).

En consultant la liste des 1000 premières requêtes (paragraphe 4.1 de l'annexe, p. 415), on se convaincra de la quasi impossibilité d'interpréter ces requêtes majoritaires. À la limite, elles témoignent d'une incompréhension du fonctionnement des moteurs de recherche — ou d'un usage ludique, non motivé de ces moteurs — par une partie de la population, qu'on ne peut évaluer, puisque ces statistiques proposent des agrégations brutales : une requête n'est qu'un fragment de session. Là encore, tout raisonnement fondé sur des énonciations agrégées et qui apparaissent majoritaires, sans qu'on se soit donné la peine de vérifier qui les produisait ni comment elles étaient reformulées, n'a aucune validité.

37. Ainsi, chaque mot des 10 000 premières requêtes apparaît au moins 525 fois, comme chacun des 1285 mots des 1000 premières apparaît au moins 4042 fois.

38. Et les 6820 autres mots, complétant le vocabulaire de ces 10 000 premières requêtes, apparaissent 11 258 698 de fois. Le total des occurrences des mots intervenant dans les 10 000 premières requêtes étant de 28 558 610, on en déduit que les 1285 mots précités ont un poids complémentaire de 4 053 312 dans les requêtes de rangs 1001 à 10 000.

2.4.3 Mots des requêtes

En revanche, nous pouvons à nouveau tenter d'estimer la variété des termes employés à partir de ces requêtes. La faible présence de mots courants, comme *flexibilité*, ou *refus*[39] est étonnante. Pour le premier, la fréquence de la première requête le contenant est 343 (rang 15 553). Pour *refus*, il faut attendre la requête « *refus vente* », de rang 94 705 (56 occ.). Pour des mots comme *comestible*, *désarmement*, *malfaçon*, *perpendiculaire*, dont la fréquence dans le lexique est déjà rare, de l'ordre de 100, la fréquence de la première requête les contenant varie entre 46 (« *fleurs comestibles* », rang 114 361) et 13 (« *ethique desarmement* », rang 412 185). D'autres mots, comme *sophistiqué/e*, apparaissent en tout 13 fois (« *maquillage sophistique* », 4 occ., rang 1 790 689).

Ces exemples confirment l'intuition suivant laquelle le vocabulaire se construit avec les requêtes à faibles fréquences : à se focaliser sur les requêtes majoritaires, on construit une représentation simplifiée, voire méprisante, des internautes.

Pour vérifier cette impression, nous avons réalisé un test à partir du dictionnaire des 39 501 mots courants (ni verbes, ni noms propres) du dictionnaire orthographique de LaTeX constitué par Christophe Pythoud[40], qui a été légèrement modifié pour, d'une part, intégrer les mots masculins (la vedette de ce dictionnaire est le féminin), et, d'autre part, obtenir une graphie conforme à la procédure **Nett2**. Ce dictionnaire contient en définitive 43 422 mots, dont les noms communs et les adjectifs au singulier (masculin et féminin), ainsi qu'au pluriel si ceux-ci sont complexes[41].

Le graphique 2.10 montre comment l'ensemble de ces mots est épuisé par les requêtes. Nous montrons que les requêtes majoritaires utilisent une infime fraction du vocabulaire de la langue française traditionnelle. Par exemple, les 10 000 premières requêtes, dont on sait qu'elle sont composées de 8105 mots

39. Mots apparaissant respectivement 1260 et 845 fois dans le vocable (procédure **Nett2**).
40. http://www.gutenberg.eu.org
41. Comme « *abri-sous-roche* » et son pluriel « *abris-sous-roche* ».

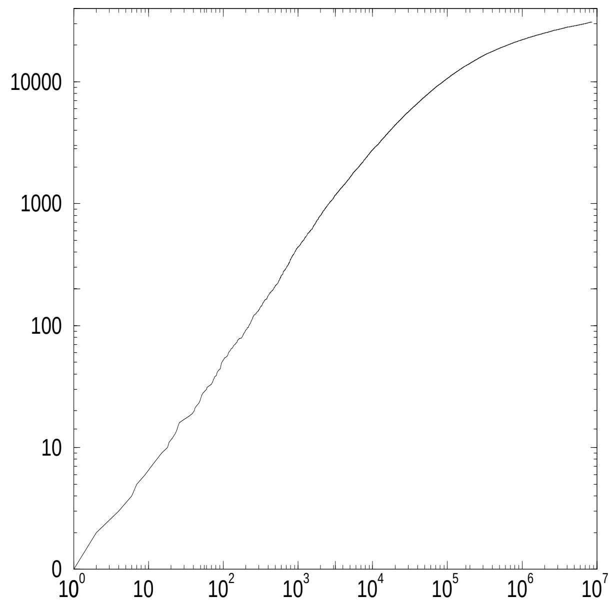

FIG. 2.10 *Utilisation du vocabulaire de la langue française au fil des requêtes. En abscisse, le rang n de la requête. En ordonnée, le nombre de mots courants du dictionnaire de LATEX repérés dans les requêtes de rang inférieur ou égal à n. Procédure* ***Nett2****.*

distincts, n'en contiennent que 2000 du dictionnaire choisi. Mais s'il faut attendre la requête de rang 100 000 pour épuiser le quart de ce dictionnaire (10 000 mots), après, la croissance se maintient, même si elle devient logarithmique : les requêtes très rares, de rang supérieur à 637 104, font apparaître autant de mots du dictionnaire, puisque le total cumulé des mots de son lexique passe de 20 000 à 30 927.

Par exemple, voici quelques mots apparaissant dans les requêtes de rang 100 000 à 101 000 : *diffuseur* (648 occ.) [42], *elaboration* (1235), *empirisme* (79), *epiphyse* (65), *ermite* (153), *felibrige* (55), *brin* (271), *filetage* (303), *fontanelle* (61), *funerarium* (91), *garrot* (115), *glissade* (167), *histamine* (103), *huchet* (145), *irrationnel* (145), *jais* (114), *kidnapping* (78), *lithosphere* (102),

42. Nous donnons ici le nombre total d'apparitions du mot exact. Si donc *diffuseur* apparaît 648 fois, ce ne peut être que dans des requêtes de rangs compris entre 100 000 et 8 000 000.

marinade (220), *metastase* (367), *milice* (138), *mistigri* (67), *nougatine* (93), *pessimisme* (436), *mineralogique* (198), *pneumocoque* (85), *funebre* (372).

Les requêtes de rang 1 000 000 à 1 001 000 ne sont pas en reste : grâce à elles sont introduits des mots comme : *poigne* (27 occ.), *exclamation* (20), *point-virgule* (14), *cyclostome* (11), *etatique* (46).

Et les requêtes de rang supérieur à 8 000 000 — qui n'apparaissent donc qu'une fois — contribuent elles-aussi à l'accroissement du vocabulaire. On passera sur les mots peu connus comme *yttrifere* ou *zoosemiotique* pour insister sur les suivants : *anarcho-syndicalisme*, *stakhanoviste*, *exterieurement*, *suburbain*, *suffisance*, *implicitement*, *zygomatique*. De tels mots sont des hapax puisqu'ils n'émergent qu'avec des requêtes singulières.

Nous avons donc doublement la preuve (avec le test dictionnairique comme avec ces exemples) que sans les requêtes peu fréquentes ou rares, toute une série de mots de la langue disparaîtraient. Or, ces mots rares (par exemple aux fréquences comprises entre 1 et 100 en trois mois) sont essentiels pour comprendre la diversité des questionnements des internautes. Mais leur coût d'extraction dans un contexte qui fait sens est particulièrement élevé, puisque le nombre de requêtes distinctes est cinq fois plus important que la taille du lexique. Le paradoxe que nous presentions se confirme : pour obtenir des résultats pertinents, il faudra oser oublier les requêtes majoritaires et ne s'intéresser qu'aux requêtes rares, complexes, et difficiles à extraire.

2.4.4 Quel *Français fondamental* ?

Après avoir montré que les requêtes majoritaires contiennent peu de mots présents dans un dictionnaire des noms communs, nous pouvons préciser la spécificité de leur lexique. Au sein du vocabulaire des 10 000 premières requêtes, on rencontre bien sûr des noms propres : tout d'abord, des noms de villes. *Paris* apparaissait déjà en bonne position (rang 3, 529 294 occ.)[43],

43. Les mots des listes suivantes ont été repérés parmi ceux des 10 000 premières requêtes ; par souci de cohérence, est rappelé entre parenthèses non pas leur nombre d'occurrences dans ces requêtes majoritaires, mais dans le total du corpus.

on peut y rajouter *toulouse* (102 097 occ. au total), *marseille* (89 091), *nantes* (63 566), *montpellier* (57 283), *nancy* (28 703), *venise* (24 532), *angers* (21 888), etc. On repère aisément des noms de régions ou de pays : *france* (rang 2, 576 184 occ.), bien sûr, mais aussi *corse* (82 903 occ.), *maroc* (72 191), *europe* (59 476), *cuba* (14 448), *vietnam* 12 570), etc. De personnes célèbres : *zola* (23 464), *maupassant* (18 614), *moliere* (18 325), *copernic* (16 644), *baudelaire* (6076) ou *shakespeare* (5253)...

Des acronymes d'institutions ou de formations françaises, comme *anpe* (83 937 occ. au total), *bts* (76 076), *dess* (50 173), *iut* (39 028), *iup* (32 125), *apec* (27 908), *insee* (19 548), *iufm* (18 773), *snes* (17 422), *cpam* (12 481), *dea* (12 034), *hlm* (10 293), *deug* (9023), *afpa* (8938), *capes* (8651), *smic* (5524), *cadremploi* (5160), *chu* (5138). On découvre de multiples autres abréviations (*pmu*, *btp*, *sci*, *tv*, *fm*, *4x4*, *tgv*, *mst*, *sarl*, etc.), d'autres connecteurs (*dans*, *aux*, *and*, *sous*, etc.), des prénoms, des mots étrangers (*sex*, *free*, *new*, etc.) ; des noms d'entreprises, françaises ou étrangères : *sncf* (85 628 occ.), *ratp* (30 303), *tf1* (37 498), *fnac* (37 401), *sagem* (36 121), *alcatel* (31 658), *michelin* (31 497), *degriftour* (25 976), *bnp* (13 539), *danone* (5810), *oreal* (5387), *aom* (5059), *axa* (5047). Et parmi ceux-ci, nombre de marques et de noms d'automobiles : *renault* (58 930), *bmw* (30 740), *yamaha* (27 614), *ferrari* (18 778), *clio* (11949), *volkswagen* (10 511), *206* (10 062), etc.

Bien sûr des noms d'« artistes » contemporains, et un nombre important de mots touchant à l'informatique et à l'internet. Nous oublierons ici les noms de logiciels, d'entreprises informatiques, de composants téléphoniques, pour ne conserver que les plus connus des protocoles et des formats de fichiers : *web* (94 278 occ. au total), *warez* (75 377), *download* (55 381), *http* (53 075), *linux* (46 338), *gif* (42 578), *napster* (33 512), *javascript* (26 976), *freeware* (26 429), *mpeg* (24 955), *html* (23 801), *hack* (23 155), *ftp* (21 820), *ip* (19 460), *php* (17 279), *asp* (15 170), *unix* (14 385), *usb* (13 314), *zip* (11 750), *icq* (11 434), *sql* (11 230), *password* (10 458), *pdf* (10 044), *gnutella* (8422), *irc* (8031), *intranet* (7676), *firewall* (7082), *proxy* (6832), *routeur* (6263), *mp4* (5445), *scsi* (5268), etc.

Notre but n'est pas de faire une liste à la Prévert, mais de mettre en évidence la variété d'un vocabulaire fort usité : ces mots, acronymes, etc., d'origine française ou non, sont souvent connus de tous, et il est fort possible qu'en quelques décennies, notre vocabulaire courant ait doublé du fait de leur multiplication. Mais cela signifie que les références classiques au sujet du vocabulaire (les 22 000 mots de l'œuvre de Victor Hugo, les 1500 mots du *Français fondamental*) sous-estiment le nombre de mots que nous avons en tête, même si on les retrouve peu dans notre langue écrite. Et pourtant ! Il est clair que si nous faisions l'étude de toutes les documentations imprimées, nous découvririons un lexique très important, de l'ordre de plusieurs centaines de milliers de mots.

L'étude du lexique, abordée par celle des requêtes, met donc en évidence un *Français contemporain* bien plus souple et plus varié que celui qui sert de référence dans notre langue écrite. Ce sera un des aspects positifs d'un tel travail que de rappeler ce fait : la multiplication des mots techniques et la grande extension des « noms propres » pose surtout des questions de type linguistique. Le seul intérêt de ces requêtes fréquemment saisies est de nous faire découvrir une nouvelle langue vivante, et pourtant écrite.

2.4.5 De l'amalgame à l'errance statistique

Nous avons démontré que les requêtes fréquentes contiennent des mots qui ne sont pas dans les dictionnaires, et que si on recherche de tels mots, il faut alors fouiller les requêtes rares. Entreprise longue et coûteuse. Par ailleurs, il apparaît que les requêtes fréquentes, souvent composées d'un seul mot, banal et polysémique, ne sont pas interprétables. Par exemple, la requête « *paris* » apparaît 25 496 fois. C'est beaucoup, même si certains pourraient prétendre que cela reste peu au regard du total d'occurrences du mot *paris* (529 294) ; pour la requête « *bmw* », c'est l'inverse qui se produit : recensée 10 478 fois (rang 269), cette brève requête constitue à elle seule 30 % des cas d'apparition du mot *bmw*. Pour savoir si les internautes s'intéressent aux automobiles ou aux motocyclettes, il faut repérer la seconde requête contenant ce mot (rang

5732 : « *bmw m3* »), et la troisième (rang 11 061 : « *moto bmw* »). Si l'on se limite aux 10 000 premières requêtes, on perd donc la référence aux deux-roues. De même pour la requête « *firewall* », qui apparaît 3201 fois (rang 1279), quand dans 45 % des cas, le mot et la requête se confondent. Or, la seconde apparition de ce mot apparaît dans la requête de rang... 36 274 (« *firewall linux* », 147 occ.), ce qui peut donner à penser soudain que le monde Linux est peut-être moins marginalisé qu'il n'y paraît au sein des visiteurs de Goosta. Mais l'enquêteur qui désire obtenir un panorama des pratiques et des centres d'intérêt des internautes à l'aide des seules 1000 ou 10 000 premières requêtes, doit témoigner, non pas de compétences en sciences humaines, mais d'un réel sens de la divination.

Ce n'est que grâce à une série de raccourcis, d'analogies, faisant appel à une sous-culture linguistique, sociologique et psychologique que l'on peut prétendre affirmer quoi que ce soit sur les « internautes » à partir des requêtes majoritaires. Or, la multiplication de faux raisonnements, et d'implicites est proprement effarante chez la majorité des personnes manifestant un quelconque intérêt pour les requêtes des moteurs de recherche. Ces analyses expéditives ne se rencontrent pas seulement dans les discours, privés ou médiatisés, mais aussi dans des articles à vocation scientifique. Pour mesurer à quel point un mélange de marketing, de pseudo-sociologie, et bien sûr, de présumée cognition tient lieu de démarche scientifique chez les consultants, on pourra se référer à l'article de Stefana Broadbent et Francesco Cara publié à l'occasion du colloque virtuel organisé par la bibliothèque publique d'information (BPI) du musée Georges Pompidou[44].

Reste un autre errement méthodologique, relatif à une autre discipline mal comprise : la statistique. Tout d'abord, l'idée — fortement répandue — que toute distribution statistique suit une loi de Gauss. Quand une telle situation se produit, le phénomène majoritaire correspond au « phénomène moyen », et les autres, s'ils ont un minimum d'importance, se distribuent autour de ce phénomène central. Il est fort possible que l'intérêt pour les premiers mots ou

44. http://www.text-e.org/conf/index.cfm ?ConfText_ID=10. Voir aussi [BC01].

requêtes soit induit par l'idée (fausse, nous l'avons vu) que les distributions de ces objets linguistiques obéiraient à la loi de Gauss : le mode définirait la moyenne, donc la norme.

Dans le même registre, on pourrait aussi expliquer le désintérêt pour les requêtes marginales. Un corollaire de notre loi « universelle » est que la somme des phénomènes rares est négligeable (courte queue de distribution). Dans le cas d'une variable qui suit une loi normale centrée réduite, le taux d'individus présentant une valeur supérieure à 3 est ridicule : 0,3 % environ. Il s'ensuit que l'on peut alors oublier les phénomènes dont les fréquences sont inférieures à 100 fois la fréquence maximale[45] (cf. le double graphique 2.11).

En adaptant ce principe aux requêtes[46], sachant que la première apparaît 284 462 fois, on pourrait négliger sans risque aucun les requêtes apparaissant moins de 2850 fois, et donc celles de rang supérieur à 1500. Nous avons mis en évidence les dangers induits par de tels raisonnements[47].

Mais ce n'est pas tout. La dernière erreur consiste à penser que tout échantillon d'une variable aléatoire quelconque finit par avoir un comportement gaussien (théorème de limite centrale). Et donc, même s'il est faux dans son acception stricte, le raisonnement précédent finirait par être valide. Or, ce théorème ne peut s'appliquer ici, car la loi théorique n'a pas de variance[48].

45. En effet, un modèle discret tiré de la gaussienne montre que, dans le cas d'une représentation rang-fréquence, une fréquence égale au centième de la fréquence de rang 1 correspond à une valeur égale à trois fois l'écart-type.

46. Certes, on pourrait en fait critiquer ce raisonnement en remarquant que les requêtes ne devraient pas être étudiées telles quelles, mais réorganisées de façon alphabétique : par exemple, la requête « *internet ecriture* » n'apparaît qu'une fois, alors que la requête symétrique apparaît six fois. Et le total des occurrences de requêtes contenant ces deux mots (« *internet* », « *écriture* ») se monte à 34. Mais un tel argument ne vaut pas pour les requêtes majoritaires, souvent composées d'un seul mot, et n'empêche pas de reprendre les raisonnement statistiques que nous critiquons.

47. Et nous pouvons imaginer sans effort à quel point la description de nos échanges nous ferait apparaître comme des rustres si l'on ne considérait que les 1500 phrases les plus fréquemment énoncées par la population française un jour donné.

48. Il faudrait que l'exposant de la loi de puissance soit inférieur à $1/2$. Or, on a vu qu'il est proche de 1 (cf. l'équation 2.1 page 291). Pour être précis, si on a affaire à une densité

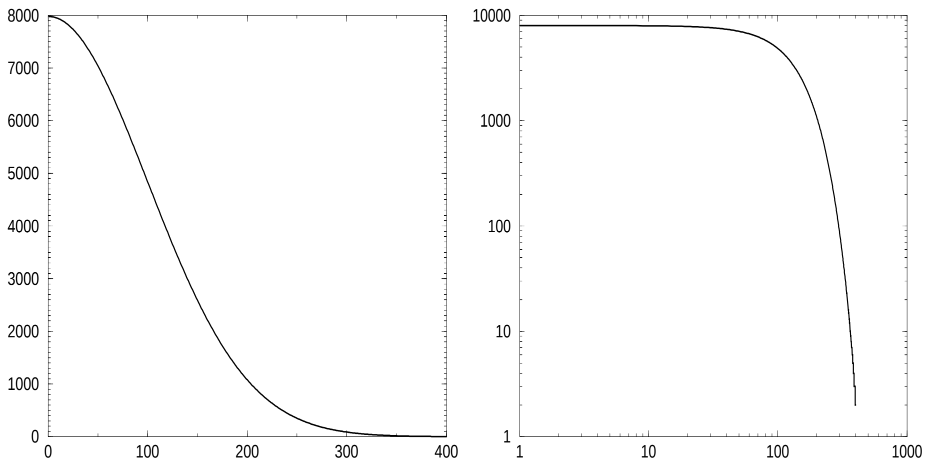

FIG. 2.11 *Graphe rang-fréquence d'une distribution gaussienne (centrée réduite). Simulation réalisée à partir d'une population d'un million de personnes. On estime le nombre d'individus (fréquence) vérifiant* $x < X \leq x + 1/100$ *(et idem pour* $-X$*). Ainsi, le rang vaut* $100 * x$*. À gauche, la représentation standard. On obtient une demi-gaussienne. À droite, la représentation log-log. On obtient une courbe quadratique, longtemps très plate, et non pas une droite, comme on l'a vu pour les distributions de lois de puissance. Accessoirement, on peut remarquer qu'un rang 300, correspondant à* $X \simeq \pm 3$*, donne une fréquence égale au centième de celle de rang 1.*

Ces résultats mettent bien en valeur les vices et les vertus de l'interdisciplinarité : il est dangereux d'appliquer des connaissances mal assimilées en linguistique, en sociologie, et en statistique : les pratiques linguistiques et intellectuelles ne se mesurent pas comme la taille ou le poids. En allant trop vite en besogne, on sous-estime la curiosité, l'intelligence et la diversité culturelle des personnes que l'on étudie. En revanche, on comprend bien comment des comptages sur des mots et des lignes, complétés par des tableaux et des graphiques qui les synthétisent, invitent à décrire une liste, à élaborer d'autres

de type $x^{-\mu-1}$, on obtient une variance finie dans le cas où $\mu > 2$; on peut montrer que, dans l'équation 2.1, l'exposant (0,95), vaut $1/\mu$. On peut traduire ce fait en disant qu'une distribution a une variance si elle met en évidence des phénomènes relativement serrés, et donc si la fréquence dépend peu du rang.

outils pour en affiner l'analyse, en même temps qu'ils font voler en éclats les préjugés consécutifs au refus de penser cette liste-texte comme un objet d'écriture avant toute chose.

2.5 Autres surprises statistiques

Cette étude est *a priori* source de désarroi : les représentations conventionnelles, les échantillonnages habituels ne fonctionnent pas. Mais elle est aussi très instructive sur le plan méthodologique : les phénomènes majoritaires sont peu pertinents, quand les plus informatifs sont rares. Mais d'autres faits bousculent encore nos préjugés.

2.5.1 Nombre de mots des requêtes

On lit souvent qu'une requête est composée de deux mots en moyenne [49]. Bien sûr, les deux procédures donnent des résultats distincts puisque la seconde réduit le nombre de mots (cf. le graphique 2.12) : avec **Nett2**, les requêtes de deux ou trois mots sont déjà clairement majoritaires dans l'ensemble des requêtes distinctes (67,5 %), alors qu'avec la procédure **Nett1**, il faut intégrer les requêtes de quatre mots pour obtenir un seuil analogue (66 %). Ensuite, les graphes apparaissent quelque peu translatés, même si pour les requêtes d'une dizaine de mots, le décalage est plutôt de trois mots. En terme de poids dans le total des requêtes, la seconde procédure favorise les requêtes de deux mots (aussi présentes que celles d'un seul, et qui constituent le profil moyen des requêtes) quand la première manifeste une décroissance forte dès le premier mot.

Au-delà de ces nuances, un invariant s'impose : quelle que soit la méthode utilisée, les requêtes composées d'un seul mot constituent 40 % du total en nombre d'occurrences [50]. Ici encore, les requêtes qui, dans une vision statis-

49. Par exemple, [S$^+$98] rappelle que la moyenne et l'écart-type de la longueur des requêtes valent respectivement 2,35 et 1,74.

50. Ces requêtes constituent 12 à 15 % des requêtes distinctes, ce qui prouve qu'elles

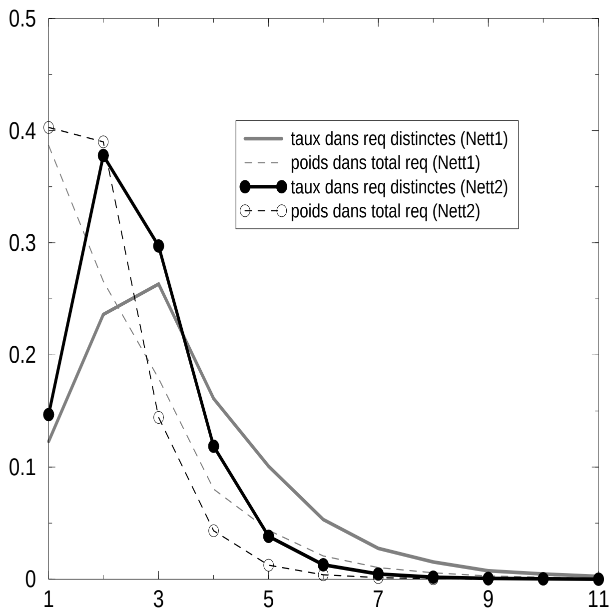

FIG. 2.12 *Distribution des requêtes en fonction de leur nombre de mots. Procédures* ***Nett1*** *et* ***Nett2****.*

tique primaire, sont les plus remarquables, sont aussi les moins instructives : pour faire retour à ce que nous avons déjà démontré, on ne dénombre que 7 requêtes de trois mots ou plus dans les 1000 premières, ce qui est fort peu. Pour les 9000 suivantes, ce type de requête ne s'élève qu'à 338.

Qu'obtenons-nous si nous oublions ces requêtes aussi pesantes que vides? Un simple argument de combinatoire permet de comprendre que, plus une requête est longue, plus sa fréquence est faible. Ainsi, l'intérêt des requêtes minoritaires, et parmi elles les requêtes longues, est triple : déjà, ce sont elles qui enrichissent le vocabulaire. Ensuite, la probabilité pour que l'internaute tire satisfaction de sa question au moteur est plus grande que pour une requête courte[51]. Enfin, son intention est compréhensible pour l'analyste.

participent malgré tout à l'accroissement du vocabulaire, puisqu'il y a au moins 8 millions de requêtes distinctes.

51. Sauf pour certaines requêtes très spécifiques comme « *ps2pcl* ». Mais celles-ci ne sont pas non plus fréquentes ; en l'occurrence, l'exemple que nous évoquons n'apparaît... jamais en trois mois.

nb occurrences	requête	nb mots
704	tony hawk s pro skater	5
643	pere noel est une ordure	5
383	sims ca vous change vie	5
314	c est arrive pres chez vous	6
300	credit agricole sud rhone alpes	5
285	on ne badine pas avec amour	6
197	eclipse lune 9 janvier 2001	5
180	centre gestion fonction publique territoriale	5
175	cartes voeux electroniques gratuites bonne annee 2001	7
169	vie est long fleuve tranquille	5
165	peine mort aux etats unis	5
162	dernier jour condamne victor hugo	5
156	bts assistant gestion pme pmi	5
155	guerre troie n aura pas lieu	6
142	crack paint shop pro 7	5
137	impact droit europeen droit interne	5
128	musee art moderne ville paris	5
124	forces faiblesses commerce exterieur francais	5
123	victor hugo dernier jour condamne	5
120	comment gagner argent dans courses chevaux	6

TAB. 2.10 *Quelques unes des 30 premières requêtes de 5 à 10 mots. Procédure **Nett2**.*

À titre indicatif, le tableau 2.10 donne les premières requêtes composées de 5 à 10 mots (procédure **Nett2**), à comparer avec les premières requêtes (voir le tableau 4.1 page 415). Malgré la suppression des connecteurs, celles-ci sont *lisibles* et interprétables, et la constitution de regroupements thématiques apparaît plus légitime qu'avec des requêtes d'un mot ou deux.

Il en est de même pour les requêtes de plus de 10 mots, dont voici le détail des cinq premières : « *tableau recapitulatif tous microprocesseurs amd intel actuel venir ainsi que leurs caracteristiques* » (48 occ.), « *marie elise eugenie crepin nee 28 mars 1871 forest montiers somme* » (40), « *est ce que echanges entre partenaires sociaux modifient concretement aide qu animateur apporte jeunes difficulte elaborer realiser leur projet vie* » (34), « *cours universitairse informatique droit aes sciences politiques ressources* » (30), et « *redaction une* *note synthese mise reseau ordinateurs routeurs administrateur reseau entreprise* » (29).

D'une part, le sérieux apparaît : point de requêtes frivoles ou naïves ici. D'autre part, nous comprenons mieux comment se construisent les sessions : la probabilité est grande que ces requêtes de plus de 10 mots soient chacune des éléments d'une même session. Outre la spécificité du vocabulaire (« *Mme Crépin* », etc.), la présence de fautes de frappe corrobore cette intuition (*universitairse*).

Reste, avant de donner quelques exemples, puis de décrire les sessions, à montrer comment cette richesse lexicale se construit avec le temps.

2.5.2 Importance du temps long

La lecture du tableau 2.1 (page 273) a mis en évidence un fait troublant : le total des mots d'une journée n'est pas représentatif du vocabulaire recensé sur de plus longues périodes. En effet, il est d'environ 80 000 pour une journée d'activité normale (comme le 4 janvier 2001), de 817 658 pour un mois (30 jours de janvier 2001), et de 1 842 768 pour les 88 jours du printemps 2001[52]. Le même tableau montre que le nombre de requêtes distinctes croît aussi avec le temps.

Ce résultat, qui peut sembler déroutant, est synthétisé dans le graphique 2.13 et appelle tout de suite une remarque méthodologique. En effet, nous pouvions penser que le nombre de requêtes distinctes fût estimé aisément à partir d'un sondage réduit du corpus, en limitant par exemple l'échantillonnage à une semaine. Il n'en est rien : étonnamment, ce nombre semble proportionnel à la durée. Même à la fin de l'échantillon temporel, entre les 55^{e} et 88^{e} jours, le nombre de nouvelles requêtes croît encore d'environ 97 000 par jour. Et ces nouvelles requêtes ne sont comptées qu'une fois[53]. Ainsi, même après

52. Et pour un total de 61 812 583 requêtes, dont 10 103 055 distinctes après application de la procédure **Nett1**.

53. Évaluation réalisée à partir de la courbe du cumul des nouvelles requêtes (courbe la plus ascendante du graphique 2.13). Si une nouvelle requête apparaît plusieurs fois un jour

une simplification drastique (ici la procédure **Nett2**), le taux de nouvelles requêtes par jour compose environ un septième du total des requêtes (alors que ces dernières sont comptées autant de fois qu'elles apparaissent). La lecture du graphique 2.14 montre qu'en fait, en période de semaine, les nouvelles requêtes (toujours comptabilisées une seule fois) constituent au moins 40 % des requêtes distinctes de chaque jour. On pourrait alors interpréter ce phénomène comme une conséquence d'une activité combinatoire des internautes à partir d'un lexique donné, fixe au bout d'un certain temps. La croissance du nombre de mots avec la taille de l'échantillon montre que cette hypothèse est fausse. Celle-ci est aussi linéaire avec le temps (en moyenne 20 000 nouveaux mots par jour, cf. graphique 2.13), et ces nouveaux mots constituent au minimum 15 % des mots distincts de chaque jour (cf. graphique 2.14).

*

À travailler sur ces listes textuelles réalisées par des millions de personnes, nous découvrons un univers statistique inattendu. Il nous faut négliger l'évidence pour trouver la richesse (linguistique, culturelle) à la marge. Nous devons aussi nous débarrasser de tous nos préjugés, d'autant, qu'au-delà des comptages, nous savons que le couple {*requête, fréquence*} est, jusqu'à nouvel ordre, sociologiquement arbitraire en ce sens qu'il s'obtient en additionnant des « fragments de pratiques » chez les internautes ; il découle d'une agrégation, d'une collection qui n'a pas de rapport avec des pratiques individuelles.

2.6 Approches thématiques

Ces précautions prises, nous pouvons nous intéresser non pas au sens des requêtes, mais à certains des thèmes qu'elles pourraient évoquer. Pour cela, nous proposons des outils lexicométriques simples, mais robustes.

Commençons par un constat, qui rappelle l'importance du temps long. Les requêtes contenant exactement le mot *gracq* apparaissent presque tous

donné, on ne la compte qu'une fois, et elle n'est plus comptée les jours suivants, même si elle réapparaît.

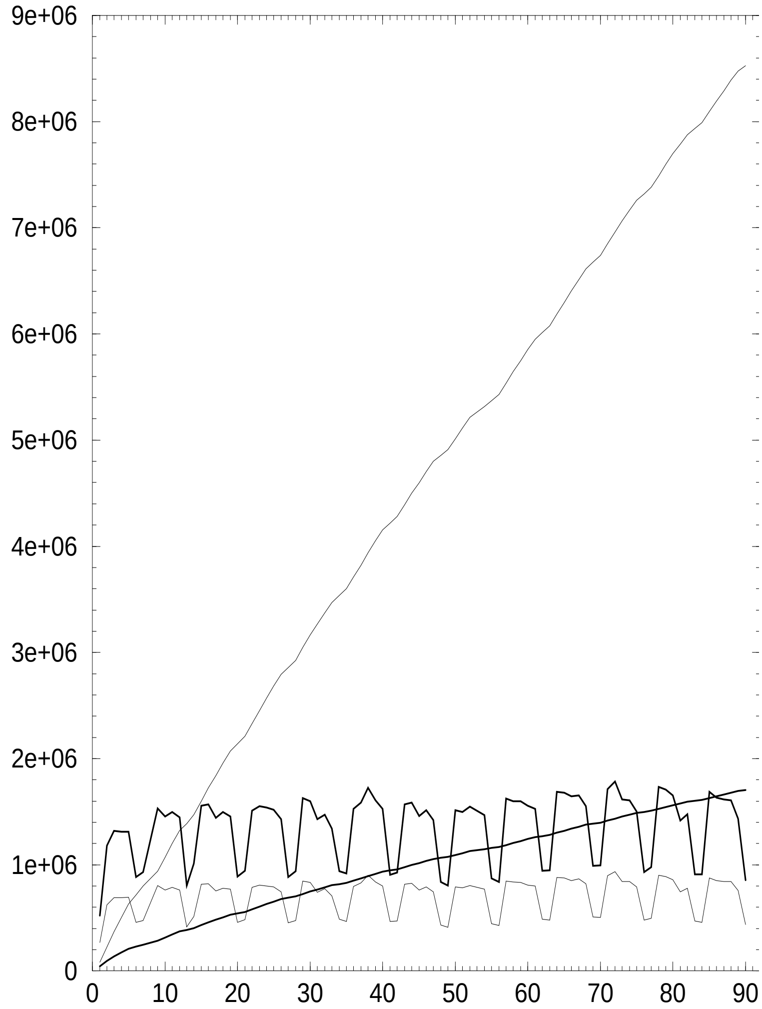

FIG. 2.13 *Accroissement du vocabulaire et des nouvelles requêtes au fil des jours. En abscisse, les jours. Les courbes épaisses correspondent aux mots, les autres aux requêtes. Les deux courbes à croissance variable du bas rappellent les nombres de requêtes et de mots de chaque jour : les fléchissements signalent les week-end Procédure **Nett2**.*

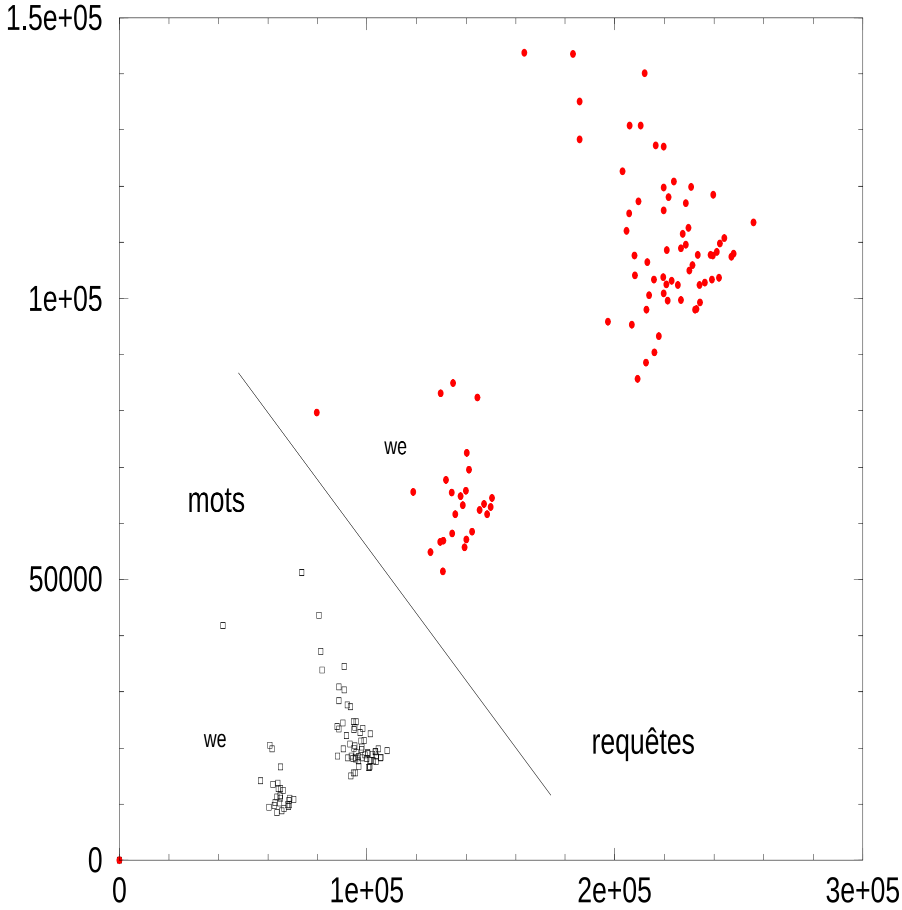

FIG. 2.14 *Importance de la nouveauté. Chaque point représente un jour.* Cercles *: En abscisse le nombre de requêtes distinctes d'un jour donné. En ordonnée le nombre de requêtes nouvelles du jour (c'est-à-dire non apparues les jours précédents).* Carrés *: même chose pour les mots. On distingue clairement, pour les mots comme pour les requêtes, deux groupes : les points représentatifs des week-end, et ceux des jours de la semaine. Les points sur la première diagonale sont ceux du premier jour. Procédure* ***Nett2****.*

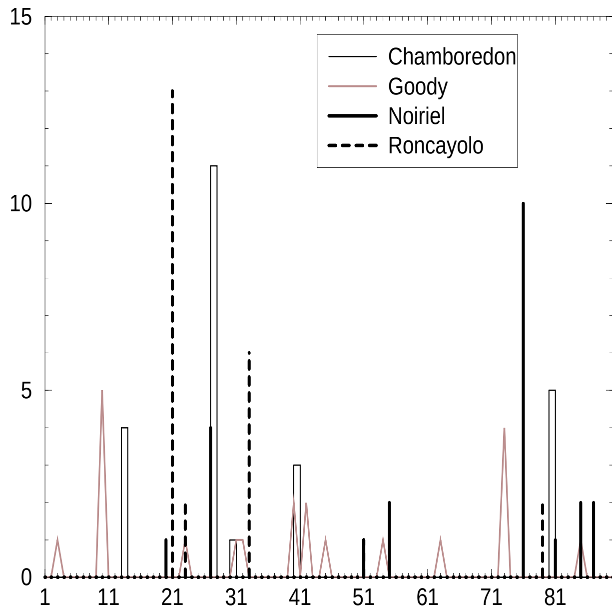

FIG. 2.15 *Date et fréquence des requêtes contenant les noms des chercheurs suivants : Chamboredon, Goody, Noiriel et Roncayolo.*

les jours, mais rarement plus de dix fois par jour. En revanche, les références au mot *roncayolo* sont rares : 13 occurrences le 21^{e} jour, 6 le 33^{e}, une le 51^{e} et 2 le 79^{e}. Il en est de même pour *goody*, mieux distribué dans le temps, mais n'apparaissant jamais plus de 5 fois dans une journée. Le graphique 2.15 rappelle la distribution de ces noms d'érudits rarement cités.

Mais cette démarche, qui consiste à chercher un mot dans une liste de requêtes, peut-être améliorée. Puisque nous avons remarqué de grandes oppositions entre des requêtes brèves et vagues, mais nombreuses, et d'autres, précises et rares, nous pouvons tenter de corréler les thèmes des requêtes à leurs caractéristiques lexicales.

2.6.1 Méthode

L'idée est simple : nous définissons une liste de mots susceptibles d'avoir le même profil (comme *anpe* et *emploi*) et repérons les mots qui leur sont associés dans l'ensemble des requêtes ; puis nous calculons le « taux de cooccurrence » de chaque mot trouvé avec les mots de la liste initiale. Par exemple, *offre* apparaît 26 357 avec *emploi*, et 43 728 fois en tout. Il a donc un taux de cooccurrence de 60 % avec *emploi*. Autrement dit, il n'y a que quatre chances sur dix pour que *offre* n'apparaisse pas avec *emploi* à ses côtés.

Quelques sondages montrent qu'un taux de cooccurrence de 30 % est déjà fort élevé. Avec cette méthode lexicale, on retrouve des synonymes, des segments répétés (même si dans le cadre d'une requête, ceux-ci peuvent être inversés), et l'on se constitue au fur et à mesure une liste de base de plus en plus consistante, qui finit par être composée des parties suivantes :

– fragments graphiques associés au thème sans ambiguité. Par exemple, *employ* entre dans le registre de l'emploi, car *employé*, *employeur* le contiennent. Les mots mal orthographiés, mais pertinents, comme *proffessionel*, sont acceptés ;
– mots ou expressions acceptés : *continue* (très liée à *formation*, mais non pas la forme *continu* — pour éviter « continuité », « discontinu(e) » ; de même, dans cette catégorie, nous sélectionnons *pare*[54] ce qui exclut « parent ». Les expressions composées de plusieurs mots s'intègrent dans cette classe ;
– fragments refusés : *location*, si on a accepté *saisonnier*, très lié à *emploi*, alors qu'on ne veut pas de « *location saisonniere* » ; *mode* pour éviter de tomber sur les nombreuses recherches de modes d'emploi ; *mex*, *lingerie*, etc. si l'on recherche des requêtes relatives à TEX et à LATEX, mais qu'on ne veut ni de restaurants mexicains, ni de fantasmes vestimentaires. Dans ce registre, on pourra aussi éviter la forme *.html* (et donc les URL), si l'on se limite à l'informatique professionnelle.

À partir de là, un programme bref (57 lignes) mais nécessitant quelques heures-machine, sélectionne les requêtes (procédure **Nett2**) qui satisfont à nos contraintes. Celles-ci sont présentées avec leur nombre d'occurrences, et, à titre de vérification, avec le mot à l'origine de leur sélection.

54. Plan d'Aide au Retour à l'Emploi.

	formes acceptées	porno, hardcore, perversion, irotisme, erotisme, erotiq, hexibitionniste, exhibition, exibition, lingerie, sexy, boutikue, monikue, esthele, femminine, charme, chrame, jarretelle, voissa, fellation, sodomi, bondage, masturbation, clito, vagin, partouze, voyeuris
sexe	formes exactes acc.	sex, sexe, sexi, soutien-gorge, guepiere, bas nylon, cul, seins, pamela anderson, nues, lesbiennes, salopes, libertine, sperme, voyeur
	formes refusées	mycose, infection, prelevement, psycho, chirurgie
	formes acc.	job, recrutement, annualisation, employ, salaire, salari, emploi, anpe, stage, interim, temporaire, employabilite, assedic, precaire, licenciment, licenciement, remunerateur, remunere, indemnite, allocation, precaire, apprenti, demandeur, diplomant, diploment, professionel, proffessionel
emploi	formes exactes acc.	boulot, pare, conge, conges, formation, formations, apec, unedic, reinsertion, cheque, cdd, continue, chomage, travail, apprentissage, rtt, 35h
	formes refusées	location, mode, sexe
	formes acc.	proxy, firewall, password, passwd, unix, javascript, linux, ascii, unicode, tutoriel, emacs, postscript, sprintf, debian, redhat, kppp, html, php
inform.	formes exactes acc.	ftp, tcp, telnet, ip, tcpip, cgi, cgi-bin, sql, mysql, routeur, scsi, perl, latex, lisp, gnu, xml, svg, css, dtd, crypt, eof, goto, require, dns, smtp, hub, script, ssh
	formes refusées	sex, star-system, avery, pantalon, cuir, vetement, fetichisme, vinyl, bondage, lingerie, .html, www.

TAB. 2.11 *Sexe, emploi et informatique.*

On dispose ainsi d'un sous-corpus thématique que l'on peut étudier plus précisément. Dans une version améliorée, le programme calcule, pour chaque mot trouvé, son taux de cooccurrence avec la totalité du lexique de base.

2.6.2 Application à quelques thèmes

Cette méthode a été testée en sélectionnant quatre thèmes : la pornographie — parce qu'elle fait couler beaucoup d'encre —, l'emploi, l'informatique professionnelle, puis l'informatique au sens large. Le tableau 2.11 rappelle tout d'abord les lexiques de chacun des trois premiers thèmes — quasi-exhaustifs, car construits par itérations successives.

Appliquée au sexe, l'étude fait apparaître 2 145 012 requêtes, dont 152 071 distinctes. Les requêtes à connotation pornographique ne dépassent donc pas 3,5 % du total des requêtes. Au vu de la façon dont l'internet permet une exploration intime secrète, sans médiation (du vendeur de journaux pornographiques par exemple) ni contrainte juridique (on peut chercher à l'étranger ce qui est interdit en France), ce taux apparaît assez faible.

La seconde étude, destinée à mettre en évidence les préoccupations professionnelles des internautes, montre que le total des requêtes sur ce thème est de 1 169 990, soit 2 %. C'est plus que ce qu'on attendait, surtout si l'on garde en mémoire les discours relatifs à la morale, ou à l'opposé, à la publicité, qui valorise un internaute consumériste et sans soucis. D'autant que les requêtes distinctes sont ici plus nombreuses que dans le thème précédent : 161 628.

La troisième est réservée à une informatique professionnelle, spécifique aux programmeurs et aux adeptes de l'écriture informatique : c'est pourquoi des mots de « consommation courante », comme *napster*, *mp3* ou *windows* n'ont pas été retenus, et que les requêtes relatives à des recherches d'URL ont été délibérément rejetées. La méthode est donc stricte, et les résultats surprenants : on dénombre ainsi 61 038 requêtes distinctes, pour un total d'occurrences de 293 848.

L'étude des sessions nous fera découvrir d'autres résultats, mais d'ores et déjà, on est étonné par le nombre de requêtes simples dans le registre pornographique : les requêtes d'un mot sont 889 960, soit 41 % du total, quand ce taux baisse à 29 % pour les requêtes relatives aux préoccupations professionnelles (337 025 requêtes) et à 27 % pour la thématique informatique (79 959 requêtes). On découvre aussi un autre résultat très significatif : pour le premier thème, le poids des requêtes de deux mots est inférieur à celui des requêtes d'un seul (795 113), alors que c'est le contraire (465 276 et 118 137), pour celles des deux autres thèmes.

Dernier point, les requêtes à caractère pornographique composent un vocabulaire de 32 900 mots distincts quand celles du second thème sont plus

formes acc.	3com, .gif, .zip, adsl, altavista, antivirus, applescript, appleshare, applet, appletalk, ascii, booster, bsd, cgi, cisco, compilateur, compression, cookie, crontab, debian, defragm, dell, dll, dns, download, emacs, faq, filesystem, firewall, forward, freeware, ftp, gamez, gcc, gnutella, google, hack, hacker, hewlett, hqx, html, http, hypertext, icq, iis, illustrator, imprimante, internet, intranet, javascript, jpeg, jpg, kde, linux, localhost, logiciel, macintosh, mail, makefile, modem, mp3, mp4, mpeg, mpg, napster, netscape, olitec, outlook, passwd, password, pcmcia, photoshop, php, plug-in, pop3, popup, postscript, ppp, programming, proxy, reboot, redhat, robotics, scanner, scsi, servlet, shareware, shutdown, smtp, software, sprintf, sql, ssh, ssl, stylesheet, svg, sybase, tableur, tcp, telecharg, telnet, traceroute, tru64, tutoriel, unicode, unix, url, usb, warez, wget, win2, win32, win9, windows2, windows9, word2, word9, www, wysiwyg, xml, yahoo
formes exactes acc.	ip, routeur, perl, latex, gnu, hub, script, gifs, irc, web, system, excel, nt, configure, bios, macros, explorer, novell, spam, fetch, lisp, sparc, utf
formes refusées	sex, star-system, avery, pantalon, cuir, vetement, fetichisme, vinyl, bondage, lingerie, yvette, appartement, thermocopression, maill, tre-mail

TAB. 2.12 *Informatique large*

riches : 40 210 mots distincts (et celles du troisième contiennent 18 708 mots). Déjà, ces quelques indicateurs montrent comment on peut à la fois cerner de façon exhaustive et précise (notamment grâce aux listes de mots rejetés) un ensemble de requêtes liées à un thème, et mesurer sa popularité autrement qu'en faisant de simples additions de fréquences de mots souvent polysémiques s'ils sont pris seuls.

Le quatrième thème est celui de l'informatique au sens large, qui intègre les pratiques de consommation — incluant la recherche de biens culturels comme les logiciels et les extraits musicaux —, et des pratiques plus intellectuelles, qui vont du désir de bien faire fonctionner son ordinateur (quitte à apprendre le mode d'emploi d'un logiciel ou d'un modem) à la recherche d'un anti-virus ou d'un *firewall* pour protéger ses données. Il englobe l'informatique des spécialistes. Le tableau 2.12 décrit la façon dont il est construit (et montre la difficulté à repérer sans erreur l'intérêt pour les produits Windows).

Nous obtenons là 2 648 764 requêtes, dont 521 900 distinctes. Ce thème risque fort de traduire la préoccupation première des internautes, avec 4,3 %

Thème	Tt. rq.	Tt. rq. dist.	Taille lx.	pds rq. 1 m.	pds rq. 2 m.	pds rq. > 2 m.
info large	2 648 764	521 900	209 140	968 195	875 510	805 059
sexe	2 145 012	152 071	32 900	889 960	795 113	459 939
emploi	1 169 990	161 628	40 210	337 025	465 276	367 689
info spéc.	293 848	61 038	18 708	79 959	118 137	95 752

TAB. 2.13 *Caractéristiques des quatre thèmes étudiés. Abréviations : Tt. : total ; pds : poids ; rq. : requêtes ; dist. : distinctes ; lx. : lexique ; m. : mots.*

du total des requêtes, et 6 % des requêtes distinctes. Son lexique est varié, puisqu'il est composé de 209 140 mots distincts. Le poids des requêtes d'un seul mot est de 36,5 %, et celui des requêtes de deux mots est légèrement inférieur aux premières (875 510 contre 968 195) ; ainsi, ce thème a un profil intermédiaire entre celui de la pornographie, et ceux, plus « exigants », de l'emploi et de l'informatique spécialisée.

Le tableau 2.13 récapitule l'ensemble de ces résultats, tout en introduisant d'autres indicateurs.

Ainsi, nous obtenons plusieurs moyens susceptibles de discriminer les préoccupations des internautes : en nous souvenant que nous travaillons sur des requêtes, et donc sur des agrégations de pratiques, et non pas sur des individus, et en évitant tout jugement moral, nous avons mis en évidence quelques indicateurs plus pertinents que le poids total des requêtes d'un thème donné. En l'occurrence, après avoir utilisé le nombre de requêtes distinctes, et le taux de requêtes d'un mot, il apparaît que le taux de requêtes distinctes, la taille du lexique divisée par le nombre de requêtes distinctes (richesse lexicale), et le poids des requêtes de trois mots ou plus sont des indicateurs pertinents pour mesurer l'importance d'un thème aux yeux des utilisateurs (cf. tableau 2.14) : les requêtes à caractère pornographique, non négligeables, sont communes, brèves et pauvres ; leurs auteurs apparaissent peu motivés. Ils le sont déjà beaucoup plus quand ils sont confrontés au mauvais fonctionnement de leur machine, et quand ils sont en quête d'outils leur permettant d'apprendre l'informatique ou d'acquérir des biens culturels : un vocabulaire très étendu est sollicité, et les requêtes longues ne sont pas rares. La problématique de

Thème	poids brut	tx req. dist.	indic. richesse lex.	pds req. > 2 mots
sexe	3,5 %	7 %	21 %	21,4 %
info large	4,3 %	20 %	40 %	30,4 %
emploi	2 %	14 %	25 %	31,4 %
info spéc.	0,4 %	21 %	31 %	32,6 %

TAB. 2.14 *Indicateurs des quatre thèmes étudiés. L'indicateur de richesse lexicale est donné par la taille du lexique divisée par le nombre de requêtes distinctes.*

l'emploi témoigne d'efforts du même ordre (en termes de longueurs des requêtes, par exemple), même si le vocabulaire est moins étendu que dans le cas de l'informatique large. Mais il est aussi plus normalisé et en proportion, plus important que dans le premier cas. Enfin, les requêtes touchant à une informatique de type Unix, si elles sont peu nombreuses, témoignent *a priori* d'une culture spécifique, sollicitant néanmoins un grand nombre de mots, et les plus forts taux de requêtes longues et distinctes.

Et donc, plutôt que de plaquer des interprétations (morales, consuméristes, intellectuelles) sur les comportements des internautes en fonction de l'agrégation de leurs pratiques, nous aimerions proposer deux axes de lecture : la motivation et la compétence. Les divers thèmes étudiés ici (au moins les trois premiers) n'ayant aucune raison d'être rattachés à des comportements spécifiques[55], nous constatons que notre population est assez passive face au thème de la lubricité : il y a de fortes chances qu'en ce domaine, l'offre soit supérieure à la demande, voire qu'elle la crée. Situation non surprenante, où les concepteurs de sites ont une grande influence dans le « formatage des usages » — ce qu'on mesure par exemple très clairement avec la longueur de la fenêtre destinée aux requêtes, et avec le fouillis de nombreux portails. L'expression de la compétence est plus délicate, puisqu'elle renvoie à une culture spécifique à l'informatique et à l'internet, et à une culture traditionnelle de

55. Nous espérons même qu'une personne intéressée par une thématique ne se contraint pas qu'à celle-ci.

la documentation, elle-même biaisée par le degré d'expertise de la personne dans le domaine de sa recherche (un spécialiste de statistiques pouvant être fort démuni et donc « naïf » s'il doit rechercher des textes du Moyen-Âge).

Ces quelques exemples, représentant déjà 9 % du total des requêtes, auront rappelé le danger des interprétations rapides, et mis en évidence, grâce à quelques indicateurs simples, la variété des besoins des internautes, et la relation — heureusement prévisible, mais rarement citée — entre leurs besoins (culturels, financiers, etc.) et leurs motivations.

*

Ces études montrent qu'il est aisé, quitte à construire les outils *ad hoc*, de mettre en évidence des caractéristiques propres aux archives électroniques provenant d'un moteur de recherche, mais aussi combien il est difficile d'analyser le comportement de ses utilisateurs, dont nous savons peu de choses, si ce n'est que leurs niveaux culturels et leurs préoccupations sont très variables (d'une personne à l'autre, et d'un moment à l'autre pour une même personne).

Nous sommes en effet impressionné par le caractère pluri-dimensionnel des usages : le type des mots et requêtes rencontrés, leurs fréquences, leurs emplois invitent avant tout à la réserve, en même temps qu'ils soulèvent des questions informatiques et méthodologiques aussi essentielles que passionnantes.

Ces premières descriptions devraient maintenant nous permettre de prolonger l'étude jusqu'aux auteurs de notre étrange liste.

Chapitre 3

Pratiques individuelles

Avant de pouvoir décrire les comportements de personnes, il nous faut passer par un objet intermédiaire, la session. Celle-ci correspond à une séquence d'interactions d'un individu avec le moteur. La mise en évidence des sessions est assez délicate, comme on va le voir.

3.1 Sessions

Une session est une série de questions d'un « même » utilisateur au moteur, qui doivent se succéder dans un intervalle de temps « raisonnable », limité ici à 20 minutes. Nous ne pouvons savoir si, durant une session, l'utilisateur reste systématiquement sur le moteur de recherche ou pas : il peut par exemple aller et venir entre le site du moteur et les pages des réponses qui lui sont proposées ; ce qui importe, c'est que s'il émet ou répète[1] plusieurs requêtes, chacune apparaisse moins de 20 minutes après la précédente.

Le moyen idéal pour repérer une session est le cookie, identifiant unique. Mais certaines personnes refusent les cookies. Dans ce cas, un nouveau cookie

1. Si une personne adresse une requête et si, après avoir visualisé les 10 premières réponses, elle demande à voir celles de la « page » suivante, sa requête sera à nouveau archivée dans les *access_ log*, et ainsi de suite. Ainsi, le verbe « émettre » est-il ici synonyme de « répéter ». Pour un rappel de l'allure de nos archives, cf. le paragraphe 2.1 p. 270 et le tableau 3.1 p. 323.

leur est attribué, même en cas de répétition de la requête : ainsi une personne qui refuse les cookies donne l'impression d'effectuer plusieurs sessions, et il faudra corriger ce type de biais. Par la suite, le cookie servira aussi à repérer plusieurs sessions d'un même utilisateur, ce qui nous aidera à décrire des profils d'usages.

La réalisation des programmes permettant de reconstruire les sessions est un vrai casse-tête chinois ; elle nécessite des centaines d'heures de travail — et parfois autant d'attente pour obtenir les premiers résultats. C'est pourquoi les résultats de cette partie se limitent à un échantillon réduit du corpus tiré de la procédure **Nett2** : la semaine du 1er au 7 mars 2001 (incluant un week-end).

3.1.1 Reconstitution des sessions

Il est assez aisé de repérer les sessions des personnes acceptant les cookies si ces dernières ont émis au moins deux requêtes ou si, à partir d'un même numéro IP, elles ont réalisé au moins deux sessions (de longueur variable) : il suffit de trier le fichier *access_log* par numéro IP[2], puis par cookie et par date ; si, pour un même numéro IP, un cookie apparaît deux fois consécutives, cela signifie que l'« utilisateur[3] » les accepte. Nous pouvons alors, eu égard aux contraintes temporelles que nous avons choisies, définir précisément la ou les sessions de cette « personne ». Ce premier travail nous permet de classer en sessions 90 % des requêtes. Les autres sont archivées dans un second fichier, qui contient les sessions avec refus de cookie et certaines sessions brèves.

Comme chaque cookie intègre sa date de création, les second et troisième champs de ce nouveau fichier sont redondants. Nous commençons alors par tenter de reconstituer les sessions des personnes refusant les cookies : pour un numéro IP donné, nous définissons comme « proches » deux requêtes consécutives (dans un délai réduit, avec les contraintes évoquées précédemment)

2. Durant une connexion, un utilisateur garde toujours le même numéro IP.

3. Même si, dans le cas de machines collectives (familiales, estudiantines, etc.), un même cookie peut référer à plusieurs utilisateurs.

num IP	cookie (variable)	date en sec.	requête	date
12.x.y.z	983793834-2529516	5490234	rer paris parc	5/3 :13 :3
12.x.y.z	983793885-2445840	5490285	rer paris parc	5/3 :13 :4
12.x.y.z	983793972-2473830	5490372	rer paris places parc	5/3 :13 :6
12.x.y.z	983794439-2500759	5490839	metro paris places parc	5/3 :13 :13

TAB. 3.1 *Exemple de reconstitution d'une session.*

si on retrouve au moins la moitié des mots de la plus courte dans la plus longue[4]. Les deux requêtes sont alors intégrées dans une même session. Il suffit ensuite d'entamer une démarche itérative.

Le tableau 3.1 explique une telle situation : au début, la requête est répétée, ensuite, elle se complexifie, mais on retrouve toujours plus de la moitié des mots de la requête la plus courte dans la plus longue.

Notre méthode a tendance à augmenter légèrement le nombre de sessions. Par exemple, si une « personne » a consulté dix pages de réponses après avoir saisi la requête « *yves montand* », puis sélectionne quelques minutes plus tard la requête « *chansons* » suivie de « *chansons yves montand* », deux sessions apparaissent. Pourtant, il conviendrait de les regrouper. Mais la méthode a aussi des avantages : elle fait mieux que réunir dans une même session les requêtes identiques[5].

Avant d'être finalisé sous la forme que nous expliquons, l'algorithme a été testé sur plusieurs cas de figure, et affiné. Sa fiabilité est accrue du fait que nos requêtes sont fortement nettoyées. Il permet de reconstruire 26 493 « sessions-individus » pour 114 343 requêtes, soit une moyenne de 4,32 requêtes pour une

4. Après suppression des derniers connecteurs non isolés source d'ambiguité et des autres mots d'un seul caractère ; bien sûr, si une des deux requêtes apparaît vide après cette procédure, l'algorithme refuse la proximité.

5. Dans les faits, les archives complètes du moteur permettent de suivre les personnes qui refusent les cookies, et qui prolongent l'interrogation du moteur à partir d'une même requête (c'est-à-dire les personnes qui consultent plusieurs pages consécutives du moteur relatives à une requête donnée). Nous ne disposons pas de ces informations, mais notre méthode les reconstitue.

telle unité quand cette moyenne est de 4,42 pour les sessions avec acceptation de cookies : les écarts sont faibles.

Après cette reconstitution des sessions avec refus de cookie et de longueur supérieure à deux, nous nous intéressons maintenant aux sessions de longueur 1. Pour éviter celles avec refus de cookie qui n'ont pu être traitées par l'algorithme, nous repérons, parmi les requêtes restantes, celles dont le cookie associé date d'avant mars 2001. On en dénombre 245 054.

Au total, on obtient 5 034 395 requêtes pour 1 328 309 sessions, soit une moyenne de 3,79 requêtes par session. Le reliquat, oublié, correspond à 5,6 % du total des requêtes : un millier dû aux sessions reconstituées mais rejetées à cause d'incohérence des dates des cookies [6], diverses sessions que nous n'avons pas réussi à reconstituer, des sessions de longueur 1 de nouveaux utilisateurs [7], etc.

Nous pouvons alors estimer le taux maximal de personnes refusant les cookies : les requêtes rejetées sont au nombre de 296 614. Elles s'intègrent dans des sessions avec refus de cookie, mais non reconstituables, et dans des sessions de longueur 1 réalisées par des nouveaux arrivants. En négligeant ces derniers et en appliquant les résultats obtenus avec les requêtes conservées pour l'étude, nous estimons à environ 78 000 les sessions avec refus de cookies qui ne sont pas prises en compte. En les ajoutant aux 26 493 reconstituées, cela nous donne environ 100 000 sessions avec refus de cookie. Sachant qu'on obtient une « moyenne » de deux sessions par individu (on sait que cette notion n'a de sens que pour des estimations grossières), on peut évaluer à 7 % le taux de personnes refusant les cookies [8].

6. Exactement 364 sessions correspondant à 1136 requêtes.

7. Nous aurions pu essayer de réperer les requêtes dont la date d'émission était strictement supérieure à la date du cookie, plutôt que de rejeter ces requêtes si la date du cookie était postérieure au 1er mars...

8. 100 000 / (1 328 309 + 78 000). D'après les gestionnaires du moteur, le nombre de personnes refusant les cookies ne dépasse pas 10 %. Dans [S$^+$98], ce taux était évalué à 4 %.

3.1.2 Choix de la procédure

Initialement, un tel travail avait été opéré en respectant les deux procédures **Nett1** et **Nett2**. La taille des fichiers étudiés se monte à 325 Mo (pour une semaine) dans chaque cas. La seconde procédure s'avère plus avantageuse, notamment pour reconstituer les sessions avec refus de cookies ; mais elle supprime de nombreux connecteurs.

Notre choix mérite donc une explication, car certains analystes partent du principe que la syntaxe de l'utilisateur témoigne de sa capacité à maîtriser la logique du moteur de recherche[9]. Ce dont nous doutons, pour plusieurs raisons : comme le proposait Google à ses débuts (printemps 2000), il est inutile de raffiner une requête avec des « + », des « - », des « et », des « ou », des guillemets si le moteur est efficace. A charge pour l'utilisateur de compléter sa requête avec d'autres mots-clés si le nombre de réponses est trop élevé. Ce que font les chercheurs expérimentés[10], que l'on devrait classer parmi les personnes à faible compétence documentaire si l'on acceptait le préjugé que nous critiquons ; d'autre part, la syntaxe du moteur étudié est mal connue des utilisateurs. À l'occasion de nos nettoyages des requêtes, nous avions remarqué de nombreuses expressions commençant par des guillemets, mais sans guillemets finaux, dotées de symboles « ++ » ou « && » ; enfin, les réponses du moteur sont ambiguës : par exemple, la requête « *"atelier internet"* » renvoie de façon très approximative aux pages contenant cette expression, puisque dans les dix premières réponses, nous avons trouvé une page ne contenant nulle part l'expression recherchée[11], mais l'expression « *atelier sur internet* ». De même, la requête « *pamela -anderson* » permet bien de retrouver des pages sans le mot *anderson*, chose impossible avec la requête « *pamela - anderson* ».

9. Cf. le paragraphe 1.3 page 263, et notamment la note 13 de ce paragraphe.

10. Enquête réalisée auprès d'une trentaine de collègues. Avec Google, on est dans une situation inverse de boucle entre conception et appropriation : autant, dans le cas de Goosta, l'internaute est contraint par la largeur de la fenêtre de requête, autant, dans le cas de Google, les concepteurs ont été contraints par les utilisateurs de rajouter une « recherche avancée », alors même qu'elle est peu utile.

11. Y compris dans les *metatags*.

Aussi, le raffinement de requêtes avec des opérateurs booléens et des expressions complexes ne peut-il être la preuve que d'une maîtrise relative, mais assurément non complète, de l'informatique documentaire. Et il suffit d'une erreur (de l'utilisateur, ou de l'algorithme du moteur[12]) pour que ces requêtes dites complexes ne soient pas efficaces. Cette insistance des analystes des *logs* sur les opérateurs booléens témoigne avant tout du respect pour une tradition de recherche qui n'a plus d'intérêt à nos yeux : les travaux de Silverstein datent de 1998, à l'époque où AltaVista — employeur du chercheur — s'imposait comme le meilleur moteur, grâce à cette possibilité d'insérer des opérateurs booléens.

3.1.3 Indicateurs des sessions

Les sessions de la semaine sont au nombre de 1 328 309.

Tout d'abord, ont été repérés les mots et requêtes fréquents, communs ou rares : mots apparaissant plus de 1500 fois (les 10 000 premiers[13]), entre 1500 et 100 fois (seuil correspondant à une apparition par jour en moyenne et au rang 60 000[14]), et les... 97 % d'autres ; requêtes apparues plus de 500 fois (environ les 10 000 premières, couvrant assez bien les 10 000 premiers mots), entre 500 et 20 fois (ce qui mène au rang 300 000[15]), et autres.

Nous calculons alors, pour chaque session, les nombres de mots *distincts* fréquents, communs ou rares de la session[16], et faisons de même pour les

12. Gardons en mémoire le fait que les requêtes composées d'un URL donnent aussi des réponses aléatoires.

13. Exactement les 9767 premiers, couvrant environ 80 % des occurrences de mots. Nous tirons ici parti des études précédentes sur les mots et requêtes.

14. Exactement 58 643.

15. Rangs exacts : 10 572 et 314 676. Rappelons que les 300 000 premières requêtes réalisent 63 % du total des occurrences des requêtes, rang à partir duquel on rencontre une brusque variation de leur nombre de mots. Une étude analogue a été réalisée en prenant comme rang limite 150 000 (requêtes apparaissant plus de 36 fois en trois mois, et fréquemment composées de deux mots). Les résultats changent alors légèrement, mais pas de façon critique (quelques pour cent).

16. Autrement dit, un mot rare apparaissant deux fois ne l'est compté qu'une, et la valeur

requêtes. S'en déduisent les indicateurs *typmot* et *typreq*, qui ont les valeurs -1 (rare), 0 (commun) ou 1 (fréquent), définies par la présence d'un mot ou d'une requête de plus bas rang[17], et la taille du lexique de la session (nombre de mots distincts).

Nous dénombrons aussi le total des occurrences de mots (avec répétition), le nombre maximal de mots des requêtes, et bien sûr la longueur de la session (nombre de requêtes). Ensuite, le nombre de répétitions des requêtes[18], qui permet de se faire une idée du nombre de requêtes distinctes d'une même session. Enfin, la présence ou absence de requêtes dites « simplistes » dans la session (requêtes exclusivement composées d'un numéro de téléphone, ou d'une URL) et la durée de cette dernière.

Bien sûr, certains indicateurs s'avèrent corrélés à d'autres (comme le total des occurrences de mots, s'accroissant automatiquement quand une requête est répétée). Mais, au vu des travaux du chapitre précédent, on ne s'étonnera pas de rencontrer une grande variété de « types » de sessions, même si l'on simplifie de façon draconienne[19] les indicateurs ainsi construits, et bien qu'aucune spécificité thématique n'ait été introduite : à partir des sessions étudiées, nous obtenons 150 050 profils différents ; si le premier réunit 168 480 sessions, 70 % des profils n'apparaissent qu'une fois, ce qui signifie

à 2 de l'indicateur *motrare* signifie que deux mots rares distincts ont été repérés une ou plusieurs fois dans la session.

17. Si toutes les requêtes d'une session sont fréquentes (*typreq*=1), sauf une, de rang intermédiaire, *typreq* sera abaissé à cette valeur 0. Bien sûr, cette méthode privilégie la moindre forme de complexité ; mais elle permet au moins de bien repérer les sessions composées de requêtes et de mots très fréquents.

18. Si deux requêtes consécutives d'une même session sont identiques, ce nombre est augmenté de 1.

19. En limitant la plupart des valeurs à 5 : nombre maximal de mots dans une requête, nombre de mots et requêtes distincts fréquents, communs et rares ; pour les autres indicateurs : si la longueur de la session dépasse 5, elle est réduite à deux classes, de 5 à 10, et plus de 10 ; il en est de même pour le nombre total de mots et le nombre de mots distincts ; le taux de répétition est réduit à des classes de largeur 5 s'il dépasse 5 ; quant à la durée de la session, elle est comptée par groupes de 5 minutes si elle est inférieure à 10 minutes, par groupes de 10 mn sinon, avec comme limite supérieure une heure ou plus.

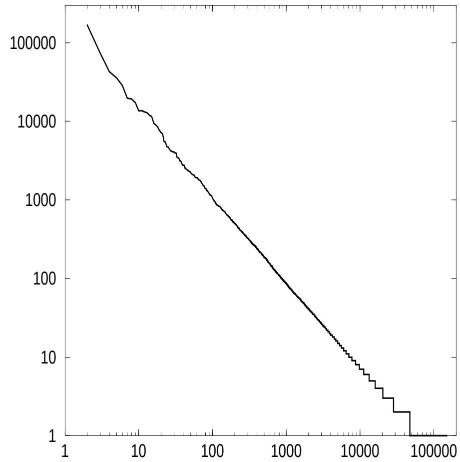

FIG. 3.1 *Représentation log-log des types de sessions constitués classés par rang et par fréquence.*

que 8 % des sessions sont absolument hors-normes. Le graphique 3.1 met en évidence cette structure, qui manifeste l'importance des profils de sessions atypiques et qui montre que les types de sessions se distribuent eux aussi selon une loi de puissance.

3.1.4 Description des sessions

L'objet étudié ici, la session, est bien plus homogène que le mot ou la requête. Ceci-dit, nous n'étudions encore que des « micro-phénomènes », qui ne correspondent pas vraiment à des réalités sociologiques : ce n'est pas en comptant le nombre de fois qu'une personne entre dans une église qu'on peut en déduire qu'elle est pratiquante ; un questionnaire, où elle répond par oui ou par non à la question posée, est autrement plus révélateur. Même quand nous regrouperons les sessions pour tenter de décrire des pratiques humaines, il conviendra de se remémorer cette situation : comme lors de l'étude des mots et requêtes, nous devrons rester particulièrement vigilant quant à l'interprétation des résultats statistiques relatifs aux sessions. Et ces précautions nous permettront de proposer quelques descriptions qui, à notre connaissance, n'ont pas été réalisées à ce jour.

3.1.4.1 Tris à plat

La majorité des sessions sont brèves : 883 363 durent moins de 5 minutes, soit les deux-tiers. 323 142 entre 5 et 20 minutes, mais 55 367 plus d'une demi-heure (9 % des sessions durent plus de 20 minutes). Très certainement, la durée permettrait de distinguer les internautes à domicile et dotés d'une connexion lente et coûteuse (qui privilégieraient les sessions courtes), des autres, disposant de l'internet sur leur lieu de travail, ou d'une connexion rapide et forfaitaire chez eux.

38 % des sessions sont de longueur 1 et 69 % d'entre elles sont de longueur au plus 3. À l'opposé, 23 % des sessions ont une longueur supérieure ou égale à 5. Ainsi, les sessions courtes sont majoritaires. De façon analogue, 19 % des sessions ont un total d'occurrences de mots réduit à... 1, quand ce dernier dépasse 10 pour 20 % des sessions.

Pour la grande majorité des sessions, les requêtes aussi sont courtes : 35 % des sessions sont composées de requêtes d'un mot. 75 % des sessions sont composées de requêtes d'un ou deux mots. À l'opposé, seulement 8 % des sessions contiennent une requête d'au moins 4 mots.

Le lexique est le nombre de mots distincts d'une session. 29 % des sessions ont un lexique de taille 1 ; ce qui signifie que près d'une session sur trois est composée d'une seule requête d'un seul mot, éventuellement répétée. 59 % des sessions ont un lexique de taille inférieure ou égale à 2 et 75 % en ont un de taille inférieure ou égale à 3. Elles ne sont que 15 % à disposer d'un lexique de taille supérieure ou égale à 5.

34 % des sessions ne contiennent que des requêtes fréquentes (et 53 % au moins une requête fréquente) ; 58 % des sessions ne contiennent aucune requête rare quand 16 % contiennent au moins 2 requêtes rares. De même, 55 % des sessions ne contiennent que des mots fréquents et 79 % ne contiennent aucun un mot rare. 5 % contiennent au moins deux mots rares.

Le nombre de répétitions témoigne d'une certaine insistance de l'utilisateur. Par exemple, une session avec quatre répétitions peut être composée d'une même requête lancée 4 fois (ce qui signifie que l'internaute a jeté un œil sur les 50 premières réponses), ou de quatre requêtes lancées chacune deux fois[20]. 54 % des sessions ne contiennent aucune répétition, quand 11 % en contiennent plus de 5 (et 4 % plus de 10). Cela signifie que dans la moitié des sessions, les utilisateurs ne cherchent pas à consulter la seconde page de résultats du moteur. Sont-ils satisfaits des premières réponses, reformulent-ils leur question, ou partent-ils visiter un autre moteur ? Nous ne pourrons pas répondre à ces questions, sauf partiellement à la seconde.

Par ailleurs, on dénombre 8758 sessions « simplistes », c'est-à-dire contenant au moins une requête composée exclusivement d'un URL ou d'un numéro de téléphone, témoignant certainement d'un grand malaise de l'utilisateur devant le moteur, qui va de la confusion entre moteur de recherche et annuaire téléphonique inverse à l'incapacité à distinguer fenêtre d'adresse du navigateur et fenêtre d'interrogation du moteur[21].

Il y a deux façons d'analyser ces premiers résultats. Ou l'on considère que « les » internautes posent avant tout des requêtes banales et se satisfont tout de suite des réponses du moteur : les sessions sont très brèves, en temps comme en longueur ; pour la moitié d'entre elles, la seconde page de résultats n'est jamais consultée ; seulement un quart des sessions ont un lexique de taille supérieure à 4. Cette première hypothèse tend à suggérer que, ou le moteur est très performant, ou les internautes se satisfont de peu. Elle ne nous convient pas : si nous nous souvenons que l'internet renvoie à de nouvelles pratiques d'écriture, de recherche (et de lecture sélective), la première idée qui nous vient à l'esprit est que la grande majorité des internautes est déçue ou

20. Bien sûr, d'autres cas de figure sont possibles entre ces deux extrêmes. À noter que trois sessions sont de longueur comprise entre 1520 et 1590, et composées d'à peu près autant de répétitions. Dans ces derniers cas, nous sommes donc assuré d'avoir affaire à des robots pilotés par des pirates.

21. En plus, le moteur ne propose même pas la redirection attendue (test effectué avec la requête « *http://barthes.ens.fr* » le 11 février 2001). Cependant, on peut imaginer que des documentalistes vérifient ainsi le référencement de leur site. Mais on doute que ce soient des employés des entreprises aux URL suivants qui aient saisi 2080 fois la requête « *www.caramail.com* », 1763 fois « *www.yahoo.fr* » ou 1375 fois « *www.tf1.fr* ».

désemparée. A l'opposé, se dégage un noyau de 10 à 20 % des « utilisateurs-sessions » qui, malgré ces difficultés, ont tendance à se montrer insistants : ils cherchent à parcourir un grand nombre de pages de réponses, n'hésitent pas à rédiger des requêtes non communes, voire complexes, ni à les remanier. Ainsi se dessinent des tendances, des types d'usages.

Mais il ne faut pas pour autant confondre sessions majoritaires et pratiques des internautes. D'une part, nous n'étudions pas encore des « individus » ; d'autre part, prétendre que 80 % des sessions ont un profil donné sans rappeler la présence des 20 % restantes reviendrait à retomber dans les errances déjà dénoncées auparavant : les salaires mensuels des Français ne sont pas tous inférieurs à 10 000 F, tout comme leur culture ne se limite pas au contenu du *Quizz*.

3.1.4.2 Quelques tendances

Les requêtes de moins de deux mots sont systématiquement majoritaires, quelle que soit la longueur de la session. Bien sûr, plus une session est longue, plus la probabilité de voir apparaître une requête de trois mots au moins s'accroît : elle passe de 13 % pour les sessions de longueur 1 et augmente d'environ 7 % à chaque classe de longueur pour atteindre les 50 % pour les sessions de longueur supérieure à 10. De façon analogue, plus le nombre maximal de mots des requêtes — noté *maxmots* — est élevé, plus la probabilité que la session soit longue augmente : 75 % des sessions dont les requêtes ne contiennent qu'un mot sont de longueur inférieure ou égale à 2. Elles ne sont que 24 % des sessions pour lesquelles *maxmots* vaut 5 ou plus (mais il n'y a que 37 616 telles sessions, soit 2,8 %).

De même, plus *maxmots* augmente, plus la session risque de contenir des répétitions : par exemple, deux sessions sur trois, parmi celles dont *maxmots* est supérieur ou égal à 5, contiennent au moins une répétition. Mais 64 % des sessions dont la ou les requêtes ne contiennent qu'un mot sont sans répétition (et 53 % des sessions dont *maxmots* vaut 2). On peut s'étonner de cette pratique, habituellement réservée aux personnes qui ont l'habitude d'adresser des requêtes très ciblées, quitte à les reformuler si la première page de réponses ne les satisfait pas.

En fait, 55 % des sessions composées de requêtes d'un seul mot sont de longueur 1. Il n'y a donc même pas de réénonciation, ni de poursuite de la lecture des résultats. Ce taux d'abandon (ou de « dérive ») instantané baisse à 35 % pour les sessions où *maxmots* vaut 2, et à 22 % pour celles il vaut 3 : dans l'ensemble, ce taux semble particulièrement élevé.

20 % des sessions composées de requêtes d'un seul mot ont un total d'occurrences de mots égal à 2. Traduit, et combiné avec les résultats précédents, cela signifie que quand un internaute saisit une requête d'un mot, il y a trois chances sur quatre (20 % + 55 %) pour qu'il ait une des trois réactions suivantes :

— il ne dépose aucune autre requête et ne cherche pas à consulter la seconde page de résultats[22] ;

— il regarde la seconde page et sa session sur Goosta est alors terminée ;

— il émet une autre requête d'un mot et là aussi, cela signifie la fin de la session.

Même si on peut admettre que l'internaute utilise le moteur à des fins ludiques (ou didactiques), on est malgré tout amené à penser que dans de nombreux cas, l'usage du moteur conduit plutôt à une forme de « désespoir » : n'oublions pas que 35 % des sessions ne contiennent pas de requêtes de plus d'un mot, ce qui déjà, témoigne d'une faible connaissance du fonctionnement d'un moteur de recherche. Bien sûr, ces sessions sont banales : 60 % d'entre elles ne sont composées que d'un mot fréquent. Le fait que dans 75 % des cas, on retrouve la pratique évoquée ci-dessus donne à penser que si l'internaute n'utilise pas le moteur à des fins de démarrage d'une navigation, il arrête bien vite sa session. D'autant qu'il ne reformule pas sa requête : 81 % des sessions d'un mot (*maxmots*=1) ne sont pas modifiées. Ces effets d'uniformité se traduisent aussi avec la faible présence des mots rares (de rang supérieur à 60 000) : 79 % des sessions n'en contiennent aucun ; pour les sessions à

22. Bien sûr, il peut néanmoins consulter une à une les dix premières réponses.

requête(s) d'un mot, comme pour celles contenant une requête de deux mots, ce taux est de 81 %. Il baisse légèrement (75 % ... 3 mots) mais reste tout de même à 58 % pour les sessions dont une requête comprend au moins 5 mots. On découvre donc que même ces sessions, dont on pouvait penser qu'elles fussent complexes, contiennent peu de mots rares.

On peut vérifier que la durée d'une session est plus corrélée à un type de matériel qu'à des pratiques intellectuelles. Par exemple, si 81 % des sessions de moins de 5 minutes ne contiennent que des requêtes d'au plus deux mots, ce dernier profil constitue encore 56 % des sessions de 20 à 30 minutes, et 47 % des sessions de plus d'une demi-heure. Les sessions de longue durée ne favorisent donc pas l'expression de requêtes d'au moins trois mots. On ne s'étonnera pas du fait que plus une session est longue en nombre de requêtes, plus elle a tendance à durer, même s'il faut attendre des sessions de plus de 10 requêtes pour qu'elles soient nombreuses à dépasser les 30 minutes (les sessions de longueur 5 à 9 constituent 27 % de ces sessions à longue durée, celles de longueur supérieure à 10, 69 %). De même, plus une session dure, plus la chance qu'apparaisse un mot ou une requête rare augmente. Mais, d'un autre point de vue, les sessions comprenant un total d'occurrences de mots supérieur à 10 se distribuent assez bien dans les classes de temps définies : chaque telle classe[23] comprend au moins 17 % de ces sessions. Certes, il y a là un effet d'écrasement de la distribution, mais on apprécie de ne pas voir de lien direct autre que prévisible entre la durée et le profil de la session.

Le croisement de la longueur de la session avec la taille du lexique donne en revanche des informations contrastées : comme prévu, les sessions de longueur 1 dont le lexique n'est composé que d'un mot sont très nombreuses (255 722) et celles de même longueur, mais avec un lexique de plus de 10 mots très rares : 94 exactement. Mais le graphique 3.2 montre aussi un autre phénomène : quelle que soit la taille du lexique, on rencontre un nombre non négligeable de sessions longues, de plus de 10 requêtes (entre 7722 et 41 157 — ceci pour les sessions à lexiques de 5 à 9 mots, non représentées ici[24]). Au-

23. Seuils à 5, 10, 20 et 30 minutes.

24. Pour une meilleure lecture, nous avons préféré ne représenter que trois tailles de

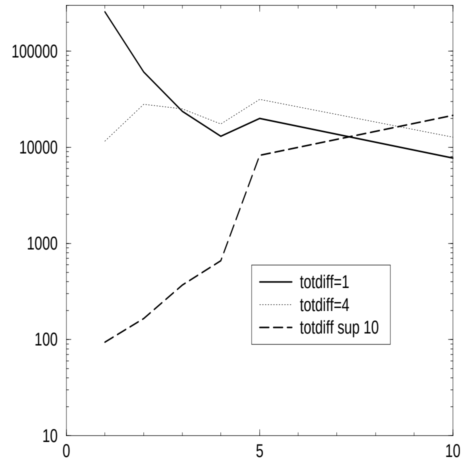

FIG. 3.2 *Nombre de sessions d'une longueur donnée (1, 2, 3, 4, 5 à 9, 10 ou plus) ayant un lexique (*totdiff*) de taille 1, 4 ou supérieure à 10.*

trement dit, il semble que se dégage une population de curieux, même chez les personnes supposées émettre des requêtes banales. En fait, ce résultat donne à penser qu'on a là au moins deux catégories d'utilisateurs : des personnes qui utilisent le moteur à des fins professionnelles, ou avec une demande précise, et d'autres, plus passives, qui consultent néanmoins de nombreuses pages de réponses. En effet, 30 % des sessions de longueur supérieure à 10 correspondent à des requêtes ne contenant que des mots fréquents. Certes, il faudrait affiner ces analyses, en regardant si ces requêtes sont identiques ou pas. Mais ces 33 274 sessions empêchent de confondre tous les auteurs de requêtes composées de mots fréquents : une minorité parmi les utilisateurs d'un lexique banal a un mode de consultation qui rompt avec la norme (de façon analogue, 38 % des sessions avec plus de 10 répétitions — donc elles aussi de longueur supérieure à 10 — ne contiennent que des mots fréquents).

En revanche, le désir de cibler les requêtes est assez peu présent : si on se rappelle que 21 % des sessions contiennent au moins un mot rare, elles ne sont que 5 % à en contenir au moins deux. Cette proportion s'accroît logiquement avec la longueur des sessions, mais très modestement : 11 % des sessions de

lexique, les autres présentant *grosso modo* des profils intermédiaires à ceux qui sont décrits.

	typreq = - 1	typreq = 0	typreq = 1	Total
typmot = - 1	18 %	4	0	22
typmot = 0	11	10	2	23
typmot = 1	12	11	32	55
Total	41	25	34	100 %

TAB. 3.2 *Typologie des sessions par types de mots et requêtes (pourcentages calculés à partir des poids, à 0,5 % près).*

longueur 5 à 9 contiennent au moins deux mots rares, et 19 % des sessions de plus de 10 requêtes. La taille du lexique semble être la condition d'apparition de ces mots rares : pour les lexiques de taille 2, le taux de sessions avec au moins deux mots rares est de 2 %. Il double à peu près pour chaque unité supplémentaire, et vaut 42 % pour les 31 004 sessions à lexique de taille supérieure à dix[25].

3.1.4.3 Requêtes et sessions

Si l'on décompose l'ensemble des sessions par types de mots et de requêtes, on obtient le tableau 3.2 : 34 % des sessions sont composées de requêtes de type 1, ce qui semble conforme avec le poids des 10 000 premières requêtes (tableau 2.9 page 292).

Ce premier résultat pourrait donner à penser qu'étudier les sessions plutôt que les requêtes n'apporte rien. Mais on découvre que 18 % des sessions contiennent au moins une requête rare et un mot rare ; or, calculé à partir de la simple étude des requêtes, le taux de requêtes rares comprenant au moins un mot rare n'est que de 11,7 %. Ainsi, l'étude des sessions met en évidence des pratiques plus complexes que celle des requêtes seules.

Mais les sessions de longueur 1 nous apportent d'autres informations, récapitulées dans le tableau 3.3 : peu de sessions avec requêtes et mots rares, 52 % de sessions composées de requêtes fréquentes. Ces sessions de longueur

25. taille 3 : 4 % ; 4 : 7 % ; 5 : 15 %.

	typreq= - 1	typreq = 0	typreq = 1	Total
typmot = - 1	8 %	3	0	11
typmot = 0	6	11	3	20
typmot = 1	9	12	49	70
Total	22	26	52	100 %

TAB. 3.3 *Typologie des sessions de longueur 1.*

1 constituent, on le sait, 38 % des sessions et sont les plus difficiles à interpréter, traduisant certainement une « fuite » des internautes. Un simple calcul à partir de la comparaison des tableaux 3.2 et 3.3 montre qu'alors le taux de sessions de *typreq* = 1 parmi celles de longueur 2 ou plus n'est que de 23 % (au lieu de 34 %) quand celui des sessions de *typreq* = - 1 et *typmot* = - 1 est supérieur, s'élevant à 24 %.

Nous comprenons alors que l'intérêt des publicitaires pour les requêtes majoritaires n'est pas seulement sans fondement scientifique : il est aussi suicidaire, puisque celles-ci doivent en grande partie leur importance au comportement des internautes qui fuient au plus vite le moteur. Ces personnes semblent capables de nous proposer un sondage électoral en ne consultant que les personnes dont on sait qu'elles ne voteront pas...

3.1.5 Typologie des sessions

Nous disposons maintenant des moyens de proposer une catégorisation des sessions. On appellera SE (comme session élémentaire) une session vérifiant l'une des trois conditions suivantes :

— elle n'est composée que de requêtes fréquentes[26] (*typreq* = 1) ;

— elle contient des requêtes fréquentes ou communes d'au plus deux mots, ces derniers étant alors fréquents ou communs ;

— son lexique total ne dépasse pas 2 mots, fréquents ou communs.

26. Voir le paragraphe 3.1.3 page 326 pour les définitions des mots « fréquent » (aussi synonyme de « courant »), « commun », « rare ».

Les autres sessions sont appelées SNE.

Puisque l'on travaille sur des sessions, et donc sur des ensembles de requêtes, il est clair que la définition d'une SE mérite bien son nom : le seuil de complexité permettant de définir une SNE est on ne peut plus bas. Par exemple, une session composée de la requête « *musique reunionnaise* », répétée 11 fois, suivie de la requête « *zouk* » (4 fois) est une SNE. De même pour la session composée des requêtes suivantes : « *ucpa* », « *pep* », suivie de « *colonies vacances* » (2 fois). Ou encore « *crous ly* » suivie de « *crous lyon* ».

On dénombre 799 303 SE, soit 60 % des sessions, dont la moitié (51 %) sont de longueur 1, et 20 % de longueur 2. À l'opposé, 37 % des SNE sont de longueur 1 ou 2. Ainsi, la catégorie SE englobe-t-elle 79 % des sessions de longueur 1.

Par construction, les SE ont un lexique réduit : pour 88 % d'entre elles, la taille de ce dernier est inférieure à 2 (43 % de taille 1, 45 % de taille 2). C'est l'inverse pour les SNE : cette taille est supérieure ou égale à 3 pour 82 % d'entre elles (et dépasse 5 dans 34 % des cas).

Nous ne reprendrons pas le détail de la plupart des autres indicateurs : le lecteur sera convaincu, au vu des travaux précédents, que l'on construit là deux classes d'usages distinctes. Nous avons jugé intéressant d'en construire un nouveau, afin d'étudier le lien entre la longueur d'une session et le caractère systématique de ses répétitions : il suffit de repérer les sessions de longueur k composées de $k-1$ répétitions.

Tout d'abord, pour les SE, la probabilité d'abandon à chaque étape est très forte : 61 % au début[27], pour ne passer en dessous de $1/2$ que si la longueur dépasse 4. Mais il ne reste alors plus que 19 % des SE. Le phénomène est inverse pour les SNE : il y a plus de sessions de longueur 2 que de longueur 1, et les sessions de longueur 4 ou plus constituent encore 48 % des SNE ; de façon générale, pour les SNE, ce taux d'abandon à chaque étape est constant et faible : aux alentours de 20 %. 27 % des SE sont en fait composées d'une

27. Autrement dit, les SE de longueur supérieure ou égale à 2 constituent 39 % des SE.

seule requête adressée plusieurs fois (donc avec autant de consultations de pages de résultats), contre 11 % des SNE (et 21 % en moyenne). Cette différence, *a priori* faible, traduit de fortes oppositions entre SE et SNE, comme le montrent le tableau 3.4 et le graphique 3.3 (pages 339 et 340).

Quand l'auteur d'une SE rédige sa première requête, s'il n'abandonne pas sa session, il a plus d'une chance sur deux de consulter la page suivante de réponses, et ce, indépendamment du nombre de pages déjà consultées. Bien sûr, c'est l'inverse qui se produit avec une SNE : le taux de répétition systématique tombe très vite.

Nous voyons se profiler, non pas deux catégories de populations (un individu réalisant une SE peut très bien réaliser une SNE plus tard, et inversement), mais d'attitudes. Avec la première (SE), la requête, souvent banale, n'est pas modifiée, mais il y a recherche d'un lien intéressant, quitte à consulter plusieurs pages successives de réponses, jusqu'à abandon du moteur[28]. Il est clair que cette démarche n'est pas optimale, mais on ne peut que s'étonner devant une telle insistance : elle témoigne ou d'une certaine passivité, ou d'une curiosité risquant fort de mener à l'insatisfaction (les requêtes sont banales). La seconde privilégie la rédaction de requêtes plus complexes et leur reformulation en cas de non satisfaction.

Le graphique 3.3 met en évidence une autre situation : quand la longueur d'une session dépasse 10, la tentation de prolonger cette session est forte, de 80 %, et ce, indépendamment de son type. Il semble que, passé un effet de « surcharge cognitive », les attitudes de type SE se confondent avec les autres. Ce fait est à relier avec celui découvert avec le graphique 3.2 de la page 334 : il existe des modes de curiosité indépendants du type *a priori* de la session (banalité, taille du lexique). On remarque, aussi, pour les deux catégories, des comportements étranges lors des longues SE : la probabilité de consulter une page supplémentaire augmente quand la longueur dépasse 18, et il y a plus de sessions de longueur 20 que de longueur 19.

28. Ce qui n'empêche pas de consulter un ou plusieurs liens, puis de revenir au moteur et de demander à visualiser une nouvelle série de réponses.

Lg	Nb SNE	Tx répet. SNE	Ratio lg SNE	Nb SE	Tx répet. SE	Ratio lg SE
1	92 396			410811		
2	105 382	0.27	1.14	161 963	0.64	0.39
3	78 355	0.16	0.74	72 412	0.58	0.45
4	56 942	0.12	0.73	40 774	0.56	0.56
5	42 032	0.09	0.74	24 694	0.55	0.61
6	31 685	0.08	0.75	16 344	0.55	0.66
7	24 398	0.07	0.77	11 503	0.55	0.7
8	19 418	0.06	0.8	8309	0.55	0.72
9	15 006	0.05	0.77	5944	0.53	0.72
10	12 255	0.05	0.82	4622	0.54	0.78
11	10 130	0.05	0.83	3700	0.53	0.8
12	8260	0.04	0.82	2930	0.53	0.79
13	6853	0.04	0.83	2293	0.53	0.78
14	5651	0.04	0.82	1827	0.55	0.8
15	4772	0.03	0.84	1496	0.54	0.82
16	3974	0.03	0.83	1272	0.53	0.85
17	3492	0.04	0.88	1063	0.52	0.84
18	3007	0.03	0.86	890	0.54	0.84
19	2543	0.04	0.85	925	0.61	1.04
20	2282	0.09	0.9	1587	0.8	1.72
Total	528 833			799 303		
% étudié	100 %			97 %		

TAB. 3.4 *Typologie SE et SNE. Le tableau se lit ainsi : il y a 105 382 SNE de longueur (lg) 2, soit 1,14 fois plus que des SNE de longueur 1. 27 % des SNE de longueur 2 sont des répétitions de la requête initiale. On dénombre 3700 SE de longueur 11, soit 80 % des sessions de longueur 10. 53 % de ces 3700 sessions sont en fait des répétitions systématiques d'une même requête.*

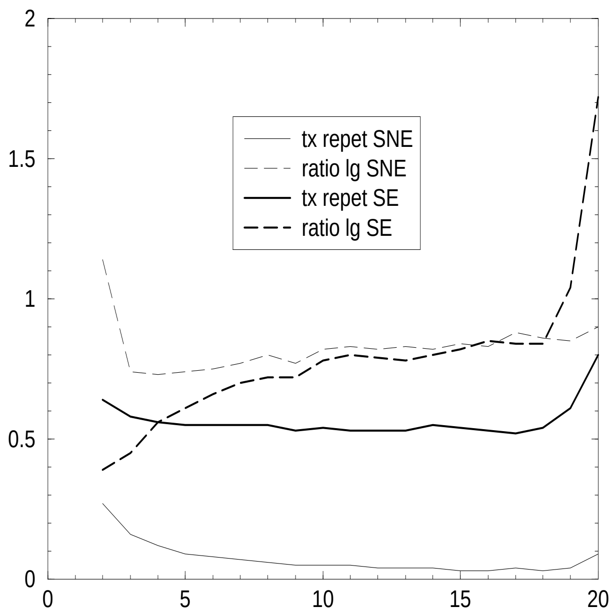

FIG. 3.3 *Répétitions dans les SE et SNE. Voir tableau 3.4 et commentaire. En abscisse, la longueur d'un type de session (on démarre à 2). Le taux de répétition des SE est étonnamment constant. Pour des longueurs supérieures à 18, on constate des phénomènes surprenants.*

3.1.6 Vers les usages

Ce travail sur les sessions était aride, mais indispensable. Pour décrire les 95 % des sessions réalisées durant une semaine, nous avons construit des indicateurs dérivés de notre étude précédente relative aux mots et requêtes. Diverses statistiques, complétées par la mise en évidence de certains comportements, nous ont prouvé que les sessions sont majoritairement simples et brèves.

Cette étude s'éloigne clairement des thèmes de sociologie-marketing communément rencontrés. Certes, nous avons montré à plusieurs reprises leur inconsistance, ce qui nous permet maintenant de mieux comprendre le grand nombre de faillites des *start-up* de l'internet. À condition de mener un travail rigoureux, de tester pas à pas des hypothèses avec des mesures exhaustives, de construire un raisonnement scientifique, il appert que le thème qui s'impose est celui du rapport de nos co-auteurs de listes à l'écriture contemporaine. Ils sont engagés dans un double processus d'écriture et de lecture sélective, qui n'a rien d'évident ni d'acquis. En l'occurrence, 60 % des sessions témoignent d'une très faible maîtrise de l'internet.

Mais ces résultats ne nous permettent pas encore de déduire des profils d'utilisateurs : d'une part, nous ne savons pas encore si certains auteurs d'un type donné de session (par exemple élementaire, de longueur 1, etc.) réalisent à d'autres moments des sessions d'un autre type (non élémentaires, etc.) ou pas. D'autre part, la distribution du profil des sessions — et les résultats précédents — nous rappellent la puissance, mais aussi la faiblesse de nos typologies.

Leur faiblesse est due au fait qu'il est aisé de définir précisément une classe (par exemple les SE), mais pas l'autre : des SNE, nous savons juste qu'elles sont des « non-SE ». Autant la notion de comportement « moyen » ou « majoritaire » a du sens pour la catégorie définie de façon objective, autant elle n'en a pas dans l'autre[29]. Et c'est clairement dans ce second groupe qu'on retrouvera le lot de pratiques variées (usages sophistiqués du vocabulaire et du moteur de recherche, etc.) qu'on a eu l'occasion de vérifier jusque là.

La force de notre méthode réside tout d'abord dans le fait qu'elle est itérative. Par exemple, on pourrait, au sein des SNE, définir précisément une sous-classe, puis le complémentaire de cette sous-classe, et ainsi de suite. Mais surtout, il nous apparaît que cette méthode est la seule possible. Au lieu de définir des « groupes » aux définitions vagues dans lesquels on insérera de force des comportements ou des personnes, pour respecter une idéologie ou pour obtenir coûte que coûte des phénomènes gaussiens, cette méthode par « dichotomies successives » semble être la seule qui respecte la structure statistique des objets que nous étudions. Puisque, parmi ceux-ci, ceux qui sont rares ont une grande importance, le seul moyen de les mettre en évidence

29. Ici, le langage courant et la terminologie statistique se confondent : puisqu'il n'y a souvent ni moyenne ni variance dans les phénomènes étudiés, une opération de dichotomie fera qu'une (au moins) des deux sous-classes obtenues n'en aura pas non plus.

consiste à les dégager, par une succession d'opérations, de la gangue des objets trop visibles qui les masquent.

3.2 Individus

Après la récolte de tant d'indices relatifs aux mots, requêtes, et sessions, nous pouvons maintenant opérer un regroupement des pratiques en fonction des « individus-cookies ». Un tel individu sera donc décrit par l'ensemble des sessions réalisées par un cookie particulier. Sa description sera nécessairement complexe[30], si l'on désire garder le maximum d'indicateurs pertinents relatifs à des « objets » qui s'emboîtent les uns dans les autres.

Chaque session est repérée par son type (SE ou SNE), son caractère (« simpliste »[31] ou pas), sa longueur, sa forme (répétition systématique ou pas), la taille de son lexique (nombre de mots distincts utilisés dans la session[32]). La durée n'a pas été conservée puisqu'elle n'apparaissait pas pertinente.

Puis sont décrits les thèmes de chaque session en fonction des types de requêtes rencontrées : informatique spécifique (E comme informatique *é*troite), informatique au sens large (L pour *l*arge), pornographie (S comme *s*exe) et emploi (W pour *w*ork). Deux autres thèmes ont été ajoutés : anthropologie

30. Le travail de programmation imposé par cette reconstruction nous a fait prendre conscience qu'une étude des usages invite au rapprochement avec la problématique de la conception des moteurs de recherche : tant sur le plan des méthodes lexicométriques et statistiques, que sur celui de la rapidité des calculs, et donc de l'efficacité des algorithmes utilisés, les problèmes et limites rencontrés sont du même type. Par exemple, cette partie n'aurait pu être réalisée si nous n'avions pas acquis une machine très performante (2 Go de mémoire vive, disques rapides), qui a pu réduire efficacement des temps de calculs qui duraient, sur le Sun, plusieurs heures. Nous espérons d'ailleurs que le profilage des usages réalisé ici pourra servir un jour à l'optimisation de ces moteurs.

31. Requêtes exclusivement composées d'un numéro de téléphone, ou d'une URL. Cf. page 327.

32. Cet indicateur sera utilisé pour créer des typologies fines des internautes.

et sociologie[33] (G comme *G*oody), dictionnaire et traduction[34] (D comme *d*ictionnaire). Une requête n'entrant dans aucune de ces rubriques était classifiée « autre » (A comme *a*utre).

Ce travail a repris, en la généralisant, la méthode exposée au paragraphe 2.6 (cf. page 310) ; cependant, puisque nous travaillions sur plusieurs thèmes à la fois, il a fallu choisir un ordre pour classifier les requêtes : le programme commence par vérifier si une requête renvoie au thème G, puis E, puis D[35], puis W, puis L[36], et enfin S[37].

Enfin, les sessions sont regroupées par cookie.

À partir de la longueur et de la taille du lexique d'une session, nous déduisons des indicateurs dérivés : longueurs minimale, maximale et moyenne des sessions[38] des internautes ; nombre de sessions de longueur 1, de longueur supérieure ou égale à 10 ; minimum et maximum des tailles de lexique des sessions de chaque personne.

Pour chaque personne, nous calculons enfin le nombre de requêtes d'un thème donné (A, D, E, etc.) dans l'ensemble de ses sessions.

Nous obtenons ainsi un fichier assez complexe, dont chaque ligne correspond à un internaute : on y retrouve son numéro de cookie, la description

33. Mots entiers : 'bourdieu', 'goody', 'mauss', 'ehess' ; expressions : 'ethnog', 'ethnolog', 'anthropolog', 'sociolog', 'durckheim', 'malinowski' ; dans notre échantillon, les auteurs Descola, Noiriel, Chamboredon et Lévi-Strauss n'apparaissaient pas (cf. le graphique 2.15 page 313) ; expressions rejetées : 'sex', 'grossiste', 'jean', 'pantalon', 'veste', 'caille' (ce dernier mot pour 'mauss'...).

34. Requêtes comprenant les formes 'dictionnaire' et 'traduction', sans exclusion d'aucune autre.

35. Car E ne contient pas la forme *telecharg*, mais L oui, et on rencontre fréquemment « *dictionnaire telecharger* ». Cette requête entre donc dans le thème D.

36. De façon que les recherches d'emploi en informatique apparaissent dans le thème W.

37. Ce qui ne pose pas de problème, puisque les thèmes précédents excluent les formes relatives à ce dernier. Mais c'est aussi pour cette raison que, dans le repérage du thème E, le mot LaTeX a été enlevé, comme les exclusions des termes « lingerie », etc.

38. La longueur moyenne des sessions est un peu tassée vers les valeurs inférieures dans la mesure où c'est la partie entière du nombre de requêtes divisé par le nombre de sessions. La longueur maximale est la plus grande des longueurs des sessions rencontrées.

synthétique de chacune de ses sessions[39], et les indicateurs numériques initiaux et dérivés[40], dont la valeur maximale vaut dix (la valeur 10 signifie donc « dix ou plus »).

3.2.1 Premiers résultats

Ces « utilisateurs » sont, au total, 640 885 . À lui seul, ce nombre légitime tous les efforts entrepris depuis le début de cette enquête, puisque nous arrivons à une population d'enquêtés dont la taille n'a plus aucune commune mesure avec une enquête sociologique traditionnelle.

Les résultats précédents donnent à penser qu'une « personne-cookie » réalise en moyenne deux sessions dans la semaine[41]. Dans les faits, 58 % des individus ainsi repérés ne font qu'une session, 20 % en font deux, 9 % trois, et donc 13 % quatre au plus.

La majorité des membres de notre panel utilise donc peu le moteur Goosta. Les taux trouvés sont conformes avec ceux que l'entreprise rencontre pour la totalité du portail qui héberge le moteur — et donc nécessairement plus consulté que le moteur lui-même : 42 % des « utilisateurs-cookies » réalisent une seule visite par semaine, 16 % en réalisent 2, 10 % 3. Ces faibles usages du moteur sont à mettre en regard des travaux de Valérie Beaudouin et de Houssem Assadi ([BA02]), qui montraient que les usagers à domicile consultaient un moteur de recherche 30 fois dans l'année en moyenne et donc,

39. Exemple : _selg=1AX2_snelg=1AX3_selg=1AX1_selg=1AX2_snereplg=3AAAX4. Cet internaute a donc réalisé 5 sessions, dont 3 SE. 4 sessions sont de longueur 1, une de longueur 3, cette dernière étant en fait à « répétition systématique ». Aucune requête n'est « simpliste ». Toutes renvoient au thème A. Le chiffre après la lettre X donne la taille du lexique de la session.

40. Nombre de sessions, de SE, de SNE, taille du lexique minimal, maximal, longueur maximale, minimale, moyenne, nombre de sessions de longueur 1, de longueur supérieure à dix, nombre de sessions avec répétition systématique, nombre de requêtes dont le thème est A, D, E, G, L, S, W.

41. Cf. page 324. Notre fichier comprend environ 5 millions de requêtes pour 1,3 millions de sessions.

moins d'une fois par semaine[42]. D'autres études ([BC01, Djo01]) rappellent la difficulté qu'ont les internautes à utiliser un moteur de recherche (mais n'offrent pas de réelles statistiques). Ainsi, même si nous reconnaissons les limites de notre échantillonnage — pour une étude plus poussée, il conviendrait de prendre une plage d'étude bien supérieure à la semaine, ce qui nécessiterait une machine autrement puissante —, nous espérons proposer une étude précise du comportement de ces 641 000 personnes.

Tout d'abord, nous évoquerons rapidement le cas des « profanes », ou nouveaux venus à l'internet. La constance de ces flux de néophytes expliquerait autant l'accroissement du nombre d'internautes que la lente évolution des usages de l'internet. Nous avons alors décidé de tester la pertinence de cet argument. Or, le cookie permet de connaître la première date de consultation du moteur par l'internaute. Il nous a permis de classer nos utilisateurs en trois groupes : personnes arrivées en mars 2001, en janvier ou février 2001, ou auparavant. On sait qu'un tel indicateur a deux biais : les personnes faisant une seule session de longueur 1 sont toutes venues sur Goosta au moins une fois avant mars (c'est ainsi qu'elles ont été repérées) ; inversement, parmi ces mêmes personnes ne réalisant qu'une session, mais de longueur supérieure à 2, les « nouvelles » sont légèrement sur-représentées, puisqu'elles intègrent les sessions avec refus de cookie que nous avons reconstituées. Les dates de première apparition du cookie sont assez bien distribuées : 39 % des personnes sont déjà venues sur Goosta avant 2001, 31 % en janvier ou février. Ceci dit, cette variable n'apporte pas d'informations probantes en matière de discrimation des pratiques. Il n'est donc pas sûr que les « anciens » aient des usages différents des nouveaux internautes qui découvrent Goosta.

La longueur moyenne des sessions vaut 1 pour 30 % des utilisateurs, 2 pour 24 % d'entre eux, 3 pour 14 % d'entre eux, et dépasse 5 pour 24 % d'entre eux. Il y a donc beaucoup de sessions courtes.

42. Toujours *en moyenne*, et en gardant en mémoire les précautions à prendre quand on utilise cet indicateur, un utilisateur émettait 8 requêtes par mois sur la totalité des moteurs, sachant que la moitié du panel est constituée de « faibles utilisateurs », auteurs de moins de quatre requêtes par mois ([BA02, VB^{+}02]).

69 % des utilisateurs ont réalisé au moins une SE. Ceci montre l'intérêt d'une étude sur les personnes : elle met en évidence une forme de panachage des pratiques, puisqu'on ne rencontrait que 60 % de SE. De façon analogue, 65 % des personnes n'ont jamais fait de session avec répétition systématique, alors que 79 % des sessions étaient sans répétition (systématique). On peut vérifier cette combinaison d'usages en considérant le pourcentage d'auteurs de sessions de longueur un [43] : il n'est que de 22 % quand on dénombrait 38 % de sessions de ce type ; une minorité d'internautes réalise plusieurs sessions de longueur 1. En revanche, 14 % des utilisateurs ont réalisé au moins une session de longueur supérieure ou égale à 10, et 35 % une de longueur supérieure ou égale à 5 (23 % des sessions étaient de longueur supérieure ou égale à 5).

Nous n'en déduisons pas pour autant que les utilisateurs réguliers ne font que des sessions courtes : au contraire, le pourcentage des personnes réalisant au moins une session de longueur supérieure ou égale à 5 croît avec leur nombre de sessions, passant de 25 % chez les auteurs d'une session à 86 % pour ceux d'au moins dix.

Les auteurs de sessions « simplistes » forment, quant à eux, une catégorie assez homogène : ils sont 7797, soit quasiment autant que le total de ces sessions (8758).

3.2.2 Une difficile appropriation

Une grande partie des utilisateurs manifeste une préférence pour les SE : 57 % des personnes n'ayant réalisé qu'une session ont fait une SE ; mais pour 78 % des personnes ayant réalisé deux sessions, au moins une est une SE. Ce taux va croissant : 99,2 % des auteurs d'au moins dix sessions ont réalisé au moins une SE. Cette augmentation invite à la modélisation : du tableau 3.5, il ressort qu'un utilisateur a une chance sur deux de réaliser une SE, quel que soit le nombre de sessions qu'il effectue. Cet important résultat prouve tout d'abord qu'une majorité d'utilisateurs de Goosta appartient au « grand

43. A distinguer d'une longueur moyenne valant 1.

Nb sessions	Au moins 1 SE	1- $(1/2)^{Nb\ sessions}$
1	0.57	0.5
2	0.78	0.75
3	0.87	0.88
4	0.92	0.94
5	0.95	0.97
6	0.96	0.98
7	0.97	0.99
8	0.97	1
9	0.98	1
10 ou plus	0.99	1
Moy	0.69	0.64

TAB. 3.5 *Modélisation de la production de SE. Pour chaque groupe d'utilisateurs ayant réalisé k sessions dans la semaine, on calcule la proportion de ceux qui ont réalisé au moins une SE, et on la compare au modèle théorique supposant qu'à chaque session, l'utilisateur a une chance sur deux de réaliser une SE.*

public »[44], mais montre surtout la difficulté qu'ont les internautes de France à exprimer des requêtes précises, si l'on accepte l'idée que notre panel est représentatif de ces derniers. Les croisements SE, SNE et nombre de sessions donnent des informations du même type : entre 70 et 80 % de ceux qui sont venus au moins deux fois sur le moteur ont réalisé autant de SE que de SNE (pourcentage bien sûr calculé sur les sessions paires).

Cette tendance à la simplicité se double d'une tendance à l'abandon : il suffit de considérer les sessions de longueur 1. On rencontre au moins une telle session chez 29 % des personnes venues une seule fois, chez 62 % des personnes ayant réalisé 2 sessions, chez 75 % de celles en ayant réalisé 3, chez 88 % des auteurs de 5 sessions, et... chez 98,5 % des auteurs d'au moins 10 sessions[45]. Certes, il est possible de retrouver du premier coup un site précis

44. « Le public ciblé par Goosta est constitué des 'bons pères de famille' » (Christian Licoppe, conférencier invité à l'Atelier Internet de l'ENS, séance du 12 février 2002).

45. En prolongeant le modèle du tableau 3.5, il semble qu'un utilisateur de Goosta a approximativement quatre chances sur dix de réaliser une session de longueur 1. Un tel modèle soulève des questionnements sociologiques. Mais il peut être efficace pour les concepteurs

en saisissant les mots-clés *ad hoc*, mais au vu des résultats précédents, nous sommes plus tenté d'attribuer la présence de ces sessions courtes à un échec qu'à un succès.

On pressent une corrélation entre l'émission de requêtes trop générales et cette forme d'abandon instantané du moteur. Pour la mettre en évidence, nous explicitons les relations entre les 4 variables suivantes : nombres de SE et de SNE, nombre de sessions de longueur 1, et (suite au travail précédent) nombre de sessions à répétition systématique. Cette évaluation est un peu délicate car les variables sont liées, mais non comparables quand le nombre de sessions varie[46].

Pour chaque tranche $SE + SNE = k$, on remarque que le nombre de sessions de longueur 1 et celui de sessions à répétition systématique croissent directement avec le nombre de SE.

À titre d'exemple, le tableau 3.6 donne les taux de sessions de longueur 1 rencontrés chez les personnes ayant fait 2, 3 ou 4 sessions. Nous savions déjà que le taux de sessions de longueur 1 croît avec le nombre de sessions. À nombre de sessions constant, le taux de sessions de longueur 1 croît avec le nombre de SE[47] : ce taux de sessions de longueur 1 devient très important[48] pour les sessions à majorité SE (qui sont d'ailleurs de plus en plus majoritaires quand le nombre de sessions augmente).

Nous obtenons des résultats analogues, bien que moins spectaculaires, pour les répétitions (cf. tableau 3.7). Aussi sommes-nous tenté de « combiner »

---

de moteurs de recherche.

46. Par exemple, si une session est constituée d'une répétition systématique, elle ne peut être de longueur 1. En revanche, la somme de ces deux formes de sessions, comme la somme $SE + SNE$, ne peut dépasser le nombre total de sessions, et il est difficile de comparer l'auteur d'une seule session avec celui d'une dizaine.

47. Cf. le taux inverse, décrit dans la colonne *0 lg1* ; par exemple, parmi les auteurs de 3 sessions, ceux qui n'ont réalisé aucune SE sont 47 % à avoir réalisé au moins une session de longueur 1 ; ce pourcentage s'élève à 87 % pour les auteurs de 3 SE.

48. Exemple : parmi les 23 070 personnes ayant réalisé 4 sessions dont au moins 2 SE, entre 49 % et 74 % d'entre elles — en tout 14 174 personnes, soit 61 % de cet échantillon — sont auteurs d'au moins 2 sessions de longueur 1.

SE_SNE	0 lg1	1 lg1	2 lg1	3 lg1	4 lg1	Total individus
2 sessions						125025
0_2	66	28	7	-	-	27 077
1_1	39	50	11	-	-	46 448
2_0	24	41	35	-	-	51 500
3 sessions						57411
0_3	53	31	12	3	-	7323
1_2	32	47	18	4	-	14 308
2_1	20	39	34	7	-	19 021
3_0	13	28	36	23	-	16 759
4 sessions						30793
0_4	43	32	17	6	2	2480
1_3	27	42	23	8	1	5243
2_2	15	36	35	12	2	7795
3_1	11	26	35	24	4	8626
4_0	8	17	28	29	17	6649

TAB. 3.6 *Pourcentage de personnes ayant réalisé 2, 3 ou 4 sessions, dont k de longueur 1 (lg1). La dernière colonne rappelle le total des individus ayant réalisé un type donné de sessions.*

les deux indicateurs, afin de regrouper deux types de pratiques opposés dans la forme (répétition sytématique ou abandon direct), mais témoignant l'une comme l'autre d'absences de reformulation des requêtes. Nous appelons *comb* ce nouvel indicateur, qui vaut la somme des deux.

À titre d'exemple, le graphique 3.4 donne, pour les auteurs de 4 sessions, le nombre de ceux qui ont réalisé k SE et dont le comb vaut l. La relation entre le *comb* et le nombre de SE apparaît assez directe.

Les résultats du tableau 3.8 montrent que, même parmi les personnes n'ayant réalisé qu'une session, dont les pratiques sont difficiles à commenter, 79 % des auteurs d'une SE ont un *comb* valant 1 : environ 169 000 individus ont en fait réalisé une SE, de longueur 1, ou systématiquement répétée. Chez les auteurs de 4 SNE et d'aucune SE, 22 % d'entre eux ont un *comb* supérieur ou égal à 3. Ce pourcentage, pour les auteurs de 4 SE, se monte à 84 %.

Aussi un *comb* élevé (proche du nombre de sessions) apparait comme un bon indicateur d'une faible motivation ou d'une curiosité difficile à exprimer,

SE_SNE	0 rép.	1 rép.	2 rép.	3 rép.	4 rép.	Total individus
2 sessions						125025
0_2	78	19	3	-	-	27 077
1_1	68	29	3	-	-	46 448
2_0	57	33	10	-	-	51 500
3 sessions						57411
0_3	68	24	6	1	-	7323
1_2	62	32	6	1	-	14 308
2_1	54	34	10	1	-	19 021
3_0	47	34	15	5	-	16 759
4 sessions						30793
0_4	61	26	10	3	1	2480
1_3	57	32	9	2	0	5243
2_2	50	36	11	2	0	7795
3_1	44	35	15	4	0	8626
4_0	39	32	18	8	3	6649

TAB. 3.7 *Pourcentage de personnes ayant réalisé 2, 3 ou 4 sessions, dont k avec répétition systématique (rép).*

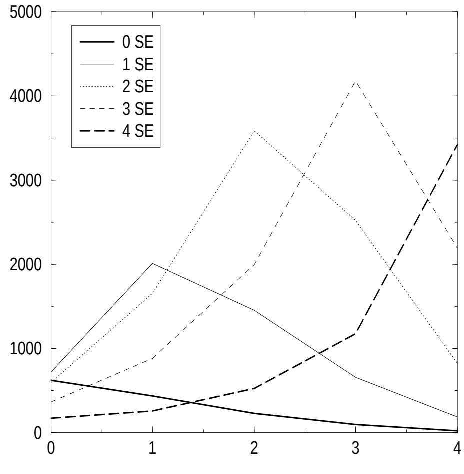

FIG. 3.4 *Auteurs de 4 sessions. En abscisse, le* comb. *En ordonnée, le nombre d'individus.*

SE_SNE	0 comb.	1 comb.	2 comb.	3 comb.	4 comb.	5 comb.	6 comb.	Total
1 session								373 128
0_1	73	27	-	-	-	-	-	159 841
1_0	21	79	-	-	-	-	-	213 287
2 sessions								125 025
0_2	49	35	16	-	-	-	-	27 077
1_1	15	63	21	-	-	-	-	46 448
2_0	4	27	69	-	-	-	-	51 500
3 sessions								57 411
0_3	35	34	22	10	-	-	-	7323
1_2	11	49	30	9	-	-	-	14 308
2_1	4	23	56	17	-	-	-	19 021
3_0	1	9	31	59	-	-	-	16 759
4 sessions								30 793
0_4	25	29	24	15	7	-	-	2480
1_3	8	38	32	17	5	-	-	5243
2_2	3	19	46	26	7	-	-	7795
3_1	1	8	29	48	14	-	-	8626
4_0	0	3	12	33	51	-	-	6649
5 sessions								18 026
0_5	20	24	21	18	9	7	-	967
1_4	6	30	31	22	8	3	-	2155
2_3	3	14	39	29	12	4	-	3456
3_2	1	7	25	42	20	5	-	4299
4_1	0	2	12	32	43	10	-	4081
5_0	0	1	5	15	36	44	-	3068
6 sessions								11 266
0_6	12	21	25	16	12	10	4	486
1_5	5	21	25	25	14	7	3	1002
2_4	1	13	30	30	17	7	2	1607
3_3	1	6	20	35	25	11	2	2217
4_2	0	2	10	29	38	16	3	2437
5_1	0	1	5	15	33	37	9	2108
6_0	0	0	2	6	18	34	40	1409

TAB. 3.8 *Pourcentage de personnes ayant réalisé 1 à 6 sessions, dont k avec répétition systématique ou de longueur 1 (comb).*

surtout dans le cas de sessions élémentaires. Il va nous aider à construire une première typologie des utilisateurs.

3.2.3 Première classification

3.2.3.1 Deux groupes

Pour renouer avec notre démarche itérative de construction de groupes, nous appelons *GR1* le groupe de personnes qui émet des requêtes banales ou qui semble désemparé en matière d'internet documentaire.

Plus précisément, *GR1* est défini comme suit : ce sont les personnes qui vérifient l'une ou l'autre des conditions suivantes :

— elles n'ont réalisé que des SE ;

— leur nombre de SE est strictement supérieur à leur nombre de SNE et leur *comb* est égal à leur nombre de sessions.

Les personnes n'appartenant pas à *GR1*, qui ne vérifent donc aucune de ces deux contraintes, seront dénommées *GR0* [49].

La deuxième condition ne change pas considérablement le profil des 296 885 auteurs exclusifs de SE : elle ne leur ajoute que 5852 personnes. On a donc 47,2 % des utilisateurs qui entrent dans la catégorie *GR1*.

Il nous semblait essentiel de vérifier que notre définition était compatible avec nos indicateurs initiaux relatifs aux mots et requêtes. Le *GR1* est co-responsable de la quasi totalité des requêtes courantes (il en manque 27 sur 10486) ; en termes de fréquence, cela fait 45,2 % de ces requêtes. Mais il n'est responsable que de 7,6 % des requêtes rares distinctes, dont le total vaut 880 792. En nombre d'occurrences, cela est équivalent : 7,7 %. Les mêmes calculs appliqués aux mots rares donnent des résultats analogues : 11,3 % du total des mots rares distincts, soit 9,9 % en poids. En revanche, le *GR0* est bien responsable de la quasi-totalité des requêtes rares. De même pour les

49. De façon générale, une catégorie de personnes ayant des usages « simples » sera dénommée X1 et l'autre X0 ; les personnes de X0 ayant les pratiques les plus communes seront des X01, les autres, des X00, etc.

mots. Le tableau 3.9 récapitule l'ensemble de ces résultats.

Au regard des autres données, les deux groupes sont bien distincts : peu de sessions multiples pour les *GR1* (70 % d'entre eux n'ont réalisé qu'une session, contre 47 % des *GR0*) ; tous les *GR0* ont réalisé au moins une SNE, contre seulement 2 % des *GR1*. La taille du lexique maximal renforce ces opposition : elle vaut 1 ou 2 pour 86 % des *GR1*, mais pour seulement 11 % de l'autre groupe[50].

79 % des auteurs *GR1* d'une seule session ont un *comb* qui vaut un. Un *comb* égal au nombre de sessions apparaît chez 69 % des auteurs *GR1* de deux sessions, chez 65 % des *GR1* auteurs de trois sessions, etc. Pour l'autre groupe, les taux démarrent bien plus bas et chutent très vite : 27 % des auteurs d'une session, 19 % de ceux de deux, et moins de 5 % ensuite.

Ainsi, avons-nous mis en évidence un groupe spécifique, qui émet surtout des requêtes banales : les SE sont quasi-systématiques, et leur lexique est fort réduit. Dans une session, la reformulation d'une requête est rare. Le groupe *GR1* semble donc particulièrement désarmé devant un moteur de recherche. Or ce groupe constitue la moitié de notre population.

Mais cela ne signifie pas pour autant que le *GR0* soit composé de personnes expertes *ès* internet documentaire. En effet, ce second groupe est défini en négatif. Nous savons qu'il contient les personnes qui ont réalisé des sessions sophistiquées, mais aussi d'autres, qui en ont réalisé de bien plus communes (au vu du grand nombre de requêtes fréquentes qui apparaissent dans le tableau 3.9). Aussi ce constat nous invitera-t-il à prolonger l'entreprise de dichotomie itérative que nous avions évoquée à la fin du chapitre 3.1.6 (page 341).

50. De même, la longueur maximale d'une session ne dépasse par 4 pour 82 % du *GR1*, alors que cette valeur 4 vaut médiane pour l'autre. Enfin, la longueur moyenne d'une session vaut 2,46 dans le premier cas, contre 4,05 dans le second. 5,6 % des personnes du *GR1* ont réalisé une session de longueur supérieure à 10, contre 21,5 % du *GR0*.

Classe	Nb distinct(e)s	Poids
Requêtes rares		
GR0	813 973	1 651 620
GR1	58 306	122 017
mixte	8513	36 583
mixte0		19 278
mixte1		17 305
Requêtes communes		
GR0	96 625	556 736
GR1	19 870	89 174
mixte	64 054	833 998
mixte0		514 126
mixte1		319 872
Requêtes fréquentes		
GR0	27	467
GR1	5	81
mixte	10 454	1 743 502
mixte0		955 838
mixte1		787 664
Mots rares		
GR0	237 575	639 885
GR1	16 550	38 076
mixte	13 772	118 517
mixte0		78 017
mixte1		40 500
Mots communs		
GR0	14 225	224 822
GR1	715	4540
mixte	32 308	1 247 802
mixte0		971 832
mixte1		275 970
Mots fréquents		
GR0	13	1371
GR1		0
mixte	9686	7 461 526
mixte0		5 855 987
mixte1		1 605 539

TAB. 3.9 *Récapitulatif des nombres de requêtes et mots rares, communs ou fréquents suivant le groupe. « Mixte » signifie qu'un ensemble de mots ou requêtes est utilisé par des personnes des deux groupes. Est rappelée alors en troisième colonne la contribution de chaque groupe au nombre d'occurrences de ces mots ou requêtes.*

3.2.3.2 Thèmes

De façon générale, le thème A (« autres »[51]) est — par construction — largement majoritaire : 95 % des internautes l'ont évoqué dans au moins une requête. Sinon, le pourcentage de personnes s'étant intéressé à un thème donné est environ le double du taux de requêtes sur ce même thème[52].

Le taux d'internautes émettant des requêtes dictionnairiques semble faible (près d'1 %) mais n'est pas négligeable, puisque le lexique constituant le thème D n'est composé que de deux mots[53]. Il en est de même pour le thème de l'écriture informatique et de la programmation (E) : il n'était pas acquis qu'une personne sur cent sollicite des mots-clés aussi complexes et spécifiques. Par suite, le taux de personnes ayant rédigé au moins une fois rédigé une requête à connotation pornographique apparaît modéré (6,2 %), entre les préoccupations professionnelles (4,1 %) et l'informatique « grand public » (L), qui motive plus de 9 % des utilisateurs. Nous ne nous étonnons pas de la faible apparition du thème G, qui renvoie à toute la variété des préoccupations des internautes, mais il n'est pas désagréable d'apprendre qu'une personne sur 1500 s'intéresse à la sociologie et à l'anthropologie.

Il était tentant de nuancer ces premiers résultats (voir tableau 3.10) en profitant de la décomposition en deux groupes de nos internautes ; bien sûr, les grandes différences des longueurs moyennes des sessions et de leur nombre

51. Rappelons la convention mnémotechnique choisie pour les abréviations des thèmes : A comme autres, D comme dictionnaire, E comme (informatique) étroite, L comme (informatique au sens) large, G comme Goody (anthropologie, etc.), S comme sexe, et W pour travail (work).

52. Cf. le tableau 3.10 et celui de la page 319.

53. La fascination sur les rangs et fréquences des mots ou requêtes tendrait à faire oublier ce thème au profit de la pornographie : le mot « sexe » est en première position avec 688 278 occurrences ; le mot « traduction » a pour rang 147 et pour fréquence 75 772 ; le mot « dictionnaire » a pour rang 165 et fréquence 69 329. La somme de ces deux dernières fréquences ne vaut que le cinquième de celle du mot « sexe ». Pourtant, 15 776 personnes ont saisi le mot « sexe » dans leurs requêtes, quand 5506 — soit le tiers — ont saisi au moins une fois « dictionnaire » ou « traduction » (démarche exclusive : les requêtes de type S et P n'ont pas été prises en compte).

Thème	Nb personnes GR1	Nb personnes GR0	% moyen
A	277 620	328 248	94,54
D	2259	3171	0,85
E	1873	4930	1,06
G	63	318	0,06
L	20 274	39 564	9,34
S	18 815	20 870	6,19
W	8680	17 633	4,11
Total	302 737	338 148	

TAB. 3.10 *Nombre de personnes de chaque groupe ayant émis au moins une requête d'un thème donné.*

vont certainement influer sur les nombres de personnes s'intéressant à un ou plusieurs thèmes. Déjà pour le thème A (autres), les statistiques divergent : 8 % des *GR1* ne l'ont jamais cité, contre seulement 3 % des *GR0*. 32 % des personnes du *GR0* l'ont évoqué plus de 10 fois, contre 8 % du GR1. Souvent, un thème est évoqué par deux fois plus de personnes du *GR0* que du *GR1*. Le cas extrême se produit pour le thème G, avec un rapport de un à cinq. Mais les thèmes E, L et W offrent aussi des résultats surprenants, puisque deux fois moins de *GR1* que prévu les évoquent.

Un thème ne suit pas cette loi, celui de la pornographie : le rapport vaut 1. Et il est le seul où l'abandon est moins systématique pour le *GR1* : comme le montre la classification 3.5, qui décrit le profil d'un thème (A exclu) en fonction du nombre de personnes l'ayant cité une fois, deux fois, etc. jusqu'à 10 fois ou plus, le *GR1* apparaît homogène : les distances entre les thèmes (E_GR1, W_GR1, D_GR1 et L_GR1) sont faibles ; mais le S_GR1 est présent au sein du profil *GR0*, nettement plus dispersé[54]. L'idée n'est pas que les personnes les plus désemparées face à l'internet seraient des obsédés sexuels, mais que leur difficulté à trouver des réponses conformes à leurs

54. La spécificité du thème D, lui aussi non « conforme », s'explique par les invitations de tous les moteurs à effectuer des traductions en ligne.

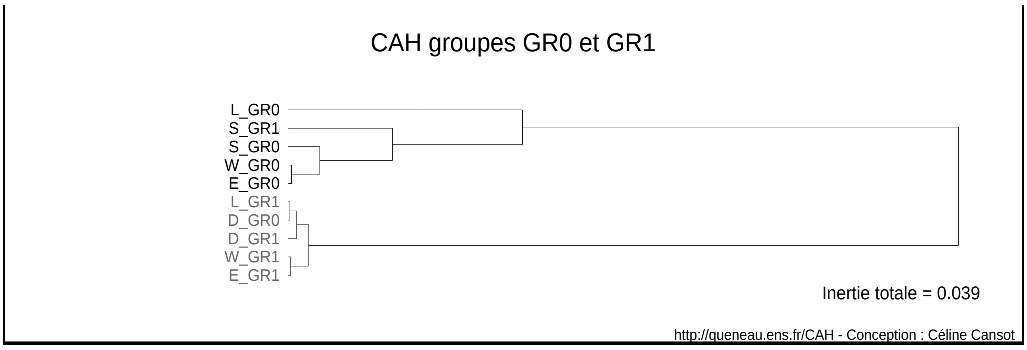

FIG. 3.5 *Classification des thèmes et groupes en fonction du nombre de personnes ayant émis 1, 2, 3..., 9 ou plus de dix requêtes d'un thème donné.*

réelles préoccupations les incite à n'utiliser les moteurs de recherche que dans un cadre ludique[55]. Forme d'exclusion intellectuelle, violente entre toutes.

*

Après avoir découvert qu'*en moyenne*, les utilisateurs de Goosta avaient une chance sur deux de réaliser, à tout moment, une SE, nous avons prouvé que ces 640 000 personnes se décomposent en fait en deux groupes de tailles équivalentes et aux pratiques bien distinctes : l'un n'effectue quasiment jamais de SNE, et chaque session est bien souvent composée d'une seule requête, très brève et rarement reformulée : l'utilisateur s'en va vite, dès l'obtention de la première série de réponses, sinon, après avoir parcouru une, deux, voire trois pages de réponses. On est dans une logique d'essais, timides et conformistes, rarement réitérés, puisque 70 % des GR1 ne sont venus qu'une fois dans la semaine. L'échec semble la règle. On comprend alors que seul le thème

55. Certes, les différences de nombre et de longueur de sessions accentuent le caractère « monothématique » des sessions *GR1*. Mais au final, le résultat reste le même. Par exemple, 5 % des *GR1* ayant émis des requêtes de type S en ont aussi émis de type L. Ce taux se monte à 15 % pour les *GR0*. Inversement, 4 % des *GR1* ayant émis des requêtes L en ont émis de type S (et 8 % des *GR0*). Autrement dit, quand les *GR0* prouvent leur capacité à témoigner de plusieurs centres d'intérêt, les *GR1* n'ont, quant à eux, pas les moyens de tirer parti des ressources culturelles de l'internet.

de la pornographie ait quelque succès pour cette population responsable de seulement 8 % des requêtes rares.

Nous pouvons maintenant prolonger notre démarche itérative.

3.2.4 Première généralisation

Nous pourrions réaliser ces itérations de façon quasi-automatique. Mais le fait que certains des indicateurs soient fort synthétiques nous incitera à la prudence, ce qui nous permettra de prouver la fiabilité de notre méthodologie, et d'opérer quelques vérifications, qui garantiront que les indicateurs que nous avons choisis au fil de ce travail sont robustes.

Nous allons donc commencer par réduire légèrement *GR0* et vérifier que les mots rares dans les deux nouveaux groupes obtenus se distribuent bien de la façon escomptée, puis étudier attentivement les auteurs d'une seule session, avant d'évaluer le taux de personnes familiarisées avec le fonctionnement des moteurs de recherche.

Nous décomposons le *GR0* en deux sous-groupes : *GR01*, et son complémentaire *GR00*.

GR01 comprend les personnes dont :

— le *comb* est égal à leur nombre de sessions

ou dont

— au moins les trois-quarts (strictement) de leurs sessions sont des SE.

Ainsi, nous sélectionnons ici encore des individus aux requêtes banales, qui réalisent des sessions brèves ou sans reformulation. *GR01* a donc un profil très proche de *GR1*.

La taille de *GR01* n'est pas négligeable : 70 593 personnes, soit 21 % de *GR0*, responsables de 152 921 sessions[56]. À eux deux, *GR1* et *GR01* réunissent 58,2 % des utilisateurs, et sont responsables de 17 % des requêtes rares distinctes[57]. Si nous agrégeons ces deux groupes (dont la réunion est

56. *GR00* comprend donc 267 555 personnes pour 661 564 sessions.

57. En tenant compte, ici encore, des requêtes rares qui sont dans un groupe *et* dans son

donc le complémentaire de *GR00*), les oppositions statistiques découvertes auparavant se maintiennent ou se renforcent ; par exemple 76 % des personnes du groupe *GR1* ∪ *GR01* ont un lexique maximum qui vaut 2, quand ce taux tombe à 5 % pour son complément *GR00* (au lieu de 86 % face à 11 %).

3.2.4.1 Mots rares

Nous désirons d'abord vérifier la pertinence des critères de rareté, qui sont définis de façon statistique. Si quelques sondages attestaient de leur qualité, il reste néanmoins possible que des mots rares (ou des requêtes rares) soient en fait des mots banals dans lesquels se seraient glissées des fautes de frappe. Dans ce cas, la rareté ne serait pas synonyme de sophistication.

Aussi, parmi les mots rares rencontrés dans les requêtes émises par des personnes *GR1* ∪ *GR01*, en avons-nous sélectionné un sur cent ; de même pour *GR00*. Ce qui a donné respectivement 455 et 1982 mots. Ces mots ont été classifiés en quatre groupes : *erreur* en cas de faute de frappe manifeste (*materiek*, *nostagie*, etc.), *doute* s'il était difficile de garantir l'erreur (noms inconnus, comme *lilootes*, *cedat*, *acopsante*, ou ressemblant à des mots connus, comme *auhan*, ou *maschio*), *web* (URL plus ou moins complètes, mais sans faute de frappe, comme *www.dad* — mais *www.education.gouvr.fr* rentre dans la catégorie *erreur*, à cause du *r* de *gouvr*), et enfin *autres*, *a priori* corrects, comme *lievremont*, *moho*, *narcotiques*, *presley*, ou *110ch*.

Dans chaque classe d'utilisateurs, le taux d'*erreur* est identique : un mot sur trois est mal écrit. En revanche, le taux de *doute* est plus élevé pour *GR1* ∪ *GR01* que pour *GR00* : 32,3 % contre 17 %. Il en est de même pour la catégorie *web* : 15,6 % contre 5,2 %. Au final, seul un mot rare sur cinq semble sans faute (ou susceptible d'exister dans un large dictionnaire) pour le groupe agrégé, face à un sur deux[58] pour *GR00*. Le tableau 3.11 synthétise ces résultats.

complément, ce qui, bien sûr, augmente les statistiques. Cf. le tableau 3.9 page 354. GR1 et GR01 réunis sont responsables de 26 % des mots rares distincts.

58. Au vu de la taille de l'échantillon, les marges d'erreur sont faibles : entre 4 % et 2 %.

Type de mots	% dans (GR1 ∪ GR01)	% dans GR00
erreur	31,4	33,3
doute	32,3	17
web	15,6	5,2
total	79,3	55,5
reste	20,7	44,5
Nb mots	455	1982

TAB. 3.11 *Taux et type de fautes dans les mots rares rencontrés dans le groupe GR1 ∪ GR01 et son complémentaire GR00. Échantillonnage : 1 mot sur 100.*

Aussi, les mots rares de *GR1 ∪ GR01* sont-ils fréquemment des mots susceptibles de n'être pas compris par le moteur, et on ne peut que se satisfaire de leur faible présence dans les requêtes de cette classe. *A fortiori*, si une personne de *GR1 ∪ GR01* saisit un mot rare, comme il y a 8 chances sur 10 pour que celui-ci soit incompris du moteur, sa réaction de rejet ou d'abandon risque d'être encore plus forte. Cela nous conforte dans l'idée que ce groupe *GR1 ∪ GR01* est particulièrement désarçonné par le fonctionnement des moteurs de recherche. À l'opposé, le fort taux de ces mots rares au sein de la classe complémentaire *GR00*, et le plus faible taux d'erreur prouvent que cette dernière émet effectivement des requêtes plus sophistiquées. Il est même possible de mesurer ce fait : une personne de *GR00* a *en moyenne* 10 fois plus de chances[59] de saisir un mot rare correct qu'une personne de *GR1 ∪ GR01*.

59. Le calcul s'établit ainsi : on dénombre 198 205 mots rares distincts propres aux 267 555 *GR00*, 45 509 propres aux 373 330 *GR1 ∪ GR01*, et 24 183 communs aux deux groupes. En tenant compte de cette « intersection », on en déduit qu'une personne de *GR00* a contribué à 0.83 mot rare, face à 0.19 pour le groupe complémentaire. Les ratios du tableau 3.11 permettent d'en déduire le taux de mot rare correct par personne de chaque groupe. Leur rapport vaut 9,57.

3.2.4.2 Biais éventuels

Avant d'aller plus loin, nous devons maintenant vérifier si certaines catégorisations ne créent pas de biais statistiques : nous travaillons sur des valeurs entières, comme le nombre de sessions, et il serait possible que des contraintes du genre *nombre de SE > 3 / 4 * nombre de sessions*, favoriseraient par exemple les auteurs d'au moins 5 sessions.

Le type de critère utilisé a effectivement une incidence sur les agrégations construites, mais celle-ci est limitée : pour montrer ce fait, considérons par exemple *GR00*, que nous scindons en deux groupes; CL1 est défini par les critères suivants :

— le nombre de SE est strictement supérieur au nombre de SNE
ou
— le lexique maximal est inférieur ou égal à 3.

CL0 est son complémentaire. CL1 réalise bien les sessions les plus communes[60].

Ici, deux personnes qui ont le même profil, mais dont l'une réalise deux sessions quand l'autre en réalise trois, ne seront pas classées dans le même sous-groupe[61].

Le graphique 3.6, montre que les deux sous-groupes ont une fréquentation analogue du moteur, ce qui est déjà une information riche en soi. Cependant, nous constatons des différences quand le nombre de sessions vaut deux ou trois : c'est bien là que le critère employé induit une rupture. Mais si l'on fait la somme, pour CL0 comme pour CL1, des auteurs de 2 ou 3 sessions,

60. Les 99 682 CL1 ont réalisé 263 095 sessions, les 167 873 CL0, 398 469 sessions. CL1 produit moins de requêtes rares que CL0 : 178 649 distinctes contre 579 823, soit environ 1,8 requêtes rare par personne contre 3,45 pour les CL0. 45 % des CL1 ont réalisé au moins une session de longueur 1, contre 31 % des CL0 ; pour 23 % des CL1 et 49 % des CL0, la longueur moyenne des sessions dépasse 5. 64 CL1 ont émis une requête de type G, contre 218 CL0.

61. Cas limite : une personne réalise une SE et une SNE ; l'autre la même SNE et deux fois la même SE. Ceci dit, on comprend que le « bénéfice du doute » dont profitera la première ne s'étendra pas à la seconde.

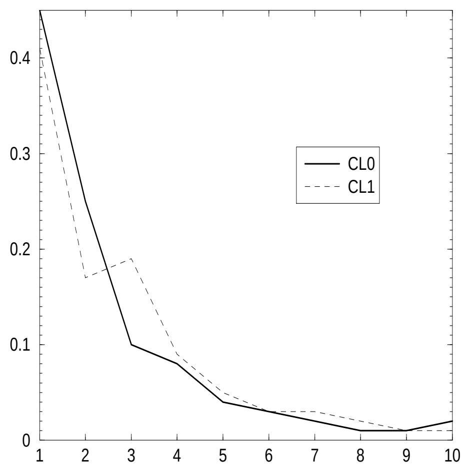

FIG. 3.6 *Pourcentages de CL1 et CL0 réalisant un nombre de sessions donné. En abscisse, le nombre de sessions.*

ces brutales variations s'annulent : 35 % au total pour les CL0, 36 % pour les CL1. D'autre part, ces biais, qui se compensent, donc, ne s'étendent pas quand le nombre de sessions dépasse le seuil choisi : à partir de 4 sessions, les résultats sont identiques pour les deux groupes.

Ainsi, les critères employés sont susceptibles d'introduire des biais aux abords du seuil choisi à la marge, mais ces biais sont peu importants.

3.2.5 Mesure de l'expertise

Nous savons que les personnes mal à l'aise consultent peu le moteur, quand les autres le font plus fréquemment [62]. Cependant, il n'est pas impossible que des utilisateurs familiers des moteurs se glissent parmi la population des personnes qui consultent rarement Goosta. Par exemple, les « experts » habitués à utiliser plusieurs moteurs à la fois ; il serait alors erroné de prétendre que le nombre de sessions d'une personne est en relation directe avec ses capacités documentaires sur l'internet. Aussi, pour mesurer le taux de personnes disposant d'une compétence certaine, devons-nous distinguer les auteurs d'une et de plusieurs sessions.

62. 70 % des *GR1* ont réalisé une seule session, contre 47 % des *GR0*.

3.2.5.1 Auteurs d'une session

Nous dénombrons 373 128 auteurs d'une seule session et 267 757 auteurs d'au moins deux. 69 % des premiers entrent dans la catégorie *GR1* $\cup$ *GR01*, face à 44 % des seconds[63].

Mais les 31 % d'autres personnes n'étant venues qu'une fois ont — par construction, puisqu'ils sont dans *GR00* — un comportement différent des *GR1* $\cup$ *GR01* : pour tous, la longueur de la session vaut au moins 2, et elle dépasse (au sens large) 5 dans 48 % des cas. Tous sollicitent un lexique de taille supérieure ou égale à 2. De plus, 42 % d'entre eux (48 911 personnes, soit 13 % des *GR00* n'ayant réalisé qu'une session) ont un lexique[64] de taille supérieure ou égale à 5. Un tel seuil semble témoigner d'une capacité certaine des utilisateurs à tirer parti d'un moteur de recherche.

Ces 48 911 personnes sont auteurs de 126 932 requêtes rares distinctes, quand les 67 921 autres (toujours *GR00* auteurs d'une seule session) ont émis 87 037 autres requêtes rares distinctes (en sus, 1452 sont communes aux deux sous-groupes). L'étude de quelques sessions réalisées par les personnes utilisant un lexique consistant donne à penser que leurs requêtes sont variées, reformulées, et efficaces[65].

Nous pouvons donc en déduire que 13 % des personnes utilisant rarement Goosta (venues une fois seulement dans notre fenêtre temporelle) sont certainement familières des moteurs de recherche. Nous découvrons une classe

63. Resp. 256 296 et 117 034 personnes. Voici par ailleurs les résultats relatifs aux mots et requêtes : Les *GR1* $\cup$ *GR01* n'ayant réalisé qu'une session sont responsables de 23 % des 276 401 requêtes rares distinctes émises par la totalité de ces visiteurs occasionnels (et 25 % en poids). Ils sont à l'origine de 33 % des 101 009 mots rares distincts rencontrés, mais ils sont responsables de 70 % des occurrences des requêtes courantes. Pour les auteurs d'au moins deux sessions, les résultats sont les suivants : 624 429 requêtes rares distinctes, dont 14 % rédigées par les *GR1* $\cup$ *GR01* ; 195 281 mots rares distincts, dont 21 % pour les *GR1* $\cup$ *GR01*.

64. Qui se confond avec *lexmax* et *lexmin* (nos internautes n'ont réalisé qu'une session).

65. Sinon, les thèmes abordés le sont en proportion assez proche des résultats tirés de l'ensemble du panel (A : 97,4 % ; D : 0,5 % ; E : 0,9 % ; G : 0,11 % (en hausse) ; L : 8,55 % ; S : 4,8 % (en baisse) ; W : 4,5 %.

d'usagers compétents, mais volatiles, qui ont un comportement en rupture complète avec la grande majorité des autres visiteurs épisodiques, pour lesquels nous avons dû constituer des discriminations du plus bas niveau possible afin de repérer un minimum de maîtrise du moteur.

3.2.5.2 Utilisateurs réguliers

Nous pouvons maintenant porter notre intérêt sur les auteurs d'au moins deux sessions : les autres laissent des traces fugitives, et seulement 13 % d'entre eux semblent maîtriser le fonctionnement de Goosta.

Parmi ces auteurs d'au moins deux sessions, presque la moitié (44 %) sont dans le groupe *GR1* $\cup$ *GR01*. Nous doutons que les 150 723 personnes restantes — de *GR00*, donc — soient toutes familiarisées avec l'informatique et les requêtes sur l'internet : en effet, 97 243 de ces personnes ont émis au moins une session de longueur 1. Un sondage rapide sur ces sessions de longueur 1, comme sur celles dont le lexique ne contient qu'un mot, montre qu'elles sont souvent très simples et peu évocatrices (« anpe », par exemple).

Sinon, 45 % de ces 150 723 personnes ont réalisé une session de longueur supérieure ou égale à 4. 59 % d'entre elles ont un *lexmax* supérieur ou égal à 5, ce qui commence à être important. Mais seulement 9 % ont un *lexmin* supérieur ou égal à 4, et 4 % un *lexmin* supérieur ou égal à 5.

Sachant que nous ne pouvons appliquer un critère aussi simple que précédemment, puisqu'ici, nos auteurs de multiples sessions peuvent alterner requêtes banales et complexes, nous nous proposons d'utiliser une fois de plus les indicateurs habituels pour opérer une coupure au sein de ce groupe de personnes venues au moins deux fois dans la semaine, et semblant un minimum familiarisées avec les moteurs de recherche. Nous regroupons d'une part les personnes remplissant l'une des conditions suivantes (groupe GR001), d'autre part son complémentaire[66] (dénommé GR000) :

66. Ces deux derniers groupes ne sont pas écrits en italiques pour rappeler que leur réunion ne donne pas *GR00* tout entier, puisqu'ils ne comprennent que les auteurs d'au moins deux sessions.

— le nombre de SE est supérieur ou égal au nombre de SNE ;

— les sessions de longueur 1 composent au moins la moitié du total des sessions ;

— le lexique minimum d'une session vaut 1 (il existe donc au moins une session composée d'un seul mot) ;

— le lexique maximal des sessions ne dépasse pas 3 ;

— la longueur moyenne des sessions vaut 1.

GR001 est bien plus important que GR000 : 115 885 personnes, responsables de 430 362 sessions, quand GR000 ne représente que 34 838 personnes (pour 114 370 sessions). Nous remarquons que GR000 a de fortes chances d'intégrer les personnes dont le *lexmax* est supérieur ou égal à 4.

Bien sûr, les deux groupes émettent de nombreuses requêtes rares : 542 539 à eux deux, dont 7307 communes à GR001 et à GR000. Mais chaque personne de GR000 est responsable d'environ 5,5 requêtes rares distinctes en moyenne, quand ce taux tombe à 3,1 pour GR001. Inversement, et toujours « en moyenne », une personne de GR001 est responsable de 4,8 requêtes fréquentes[67], et une personne de GR000 de 3,6. Enfin, les sessions des GR000 sont en moyenne de longueur 5. En matière de thèmes, les taux de GR000 sont plus élevés que pour la moyenne, ce qui est en partie logique, puisque qu'ils réalisent plusieurs sessions : 1,3 % d'entre eux ont émis des requêtes de type D, 9,7 % de type S et 8,2 % de type W. Les plus fortes augmentations se retrouvent pour les thèmes E (2,4 %), G (0,2 %) et L (18,2 %). Ce dernier pourcentage prouve que les personnes qui maîtrisent l'internet ont conscience du besoin de continuer à se cultiver dans le domaine de l'informatique.

Au vu de ces résultats, nous sommes conduit à penser que c'est parmi ces 34 838 personnes — soit 13 % des auteurs d'au moins deux sessions — que nous retrouverons les internautes disposant d'une maîtrise minimale de l'outillage intellectuel contemporain.

67. Cette fois en nombre d'occurrences, puisque ces requêtes ne sont que 10 000 environ.

*

Une majorité d'internautes (58 %) n'ont réalisé qu'une session dans la semaine. Nombre de ceux-ci sont particulièrement démunis face au fonctionnement du moteur. À l'opposé, 13 % semblent en avoir une bonne maîtrise. Restent alors les personnes venant deux fois ou plus, plus faciles à étudier. Là encore, seulement 13 % des utilisateurs attestent d'une telle maîtrise. Nous pouvons en déduire que 87 % de nos 640 000 internautes sont particulièrement désemparés face à de tels outils. Et pourtant, nos critères ont été d'une simplicité maximale.

Il faudrait vérifier ce résultat, avec des études approfondies, incluant des enquêtes ethnographiques, mais d'ores et déjà, les témoignages d'étudiants, évoquant leurs collègues, professeurs, amis et parents, nous donnent à penser qu'il est juste, même s'il déplaît.

3.2.6 Quelles requêtes fréquentes ?

En effet, nos 87 % de personnes inexpérimentées, mal à l'aise face à la technique comme face à l'internet documentaire, constituent l'« échantillon représentatif » de l'internaute français. C'est au travers de leurs mots et requêtes les plus fréquents que les analystes des *access_ log* vont déterminer des profils et des évolutions. Pour le plaisir de montrer comment leurs résultats sont bousculés par l'étude attentive d'une population précise, nous proposons d'étudier ce que sont les requêtes fréquentes au sein du groupe GR000, qui est à la fois représentatif des personnes sachant se servir de l'internet et des personnes venant régulièrement sur Goosta.

Notre raisonnement est le suivant : pour aller dans le sens des analyses que nous comptons critiquer, et pour réduire le poids des fautes de frappe, nous ne considérons que les requêtes fréquentes ou communes apparues dans les 3 mois, tous utilisateurs confondus. Nous sommes donc assuré que la fréquence totale de ces requêtes est supérieure à 20. Nous comptons combien de fois les membres de GR000 saisissent ces requêtes, puis comparons le rang général d'une requête avec celui qu'elle a dans GR000.

rang dans GR000	requête	fréq. dans GR000	rg initial	fréq. totale
1	sexe	531	1	284462
2	immobilier	345	2	93755
3	fond ecran	306	18	51067
4	mp3	296	5	79200
5	pokemon	281	22	43856
6	anpe	279	10	70122
7	sexe gratuit	278	11	69967
8	moteur recherche	275	52	28490
9	chat	256	4	84861
10	webcam	241	14	60406
22	meteo	186	6	77851
24	emploi	183	7	74660
41	caramail	154	8	73966
47	sex	147	3	91962
111	horoscope	91	9	73071

TAB. 3.12 *Requêtes apparaissant les dix premières dans GR000 ou dans le total.*

Pour à la fois expliciter cette méthode et montrer en quoi les variations sont faibles pour les requêtes majoritaires, le tableau 3.12 donne les fréquences et rangs respectifs des requêtes qui sont les dix premières dans GR000 ou dans le total.

Les deux premières requêtes du corpus de trois mois voient leurs rangs inchangés. Mais déjà apparaissent quelques différences. Par exemple « *moteur recherche* » voit son rang baisser[68] (8 au lieu de 52) quand celui de « *sex* » passe de 3 à 47. On est donc tenté de repérer les requêtes originelles qui voient leur rang croître dans GR000 (fort intérêt en moyenne, mais faible intérêt pour les GR000), et à l'inverse, les requêtes que nos spécialistes n'aurait pas remarquées, mais qui sont sollicitées par les auteurs GR000. Pour cela, on se fixe un seuil *limite* (2, 5 ou 10), et l'on inscrit dans un fichier les premières ({rang initial * *limite*} < {rang dans GR000}), et dans un autre les secondes ({rang dans GR000 * *limite*} < {rang initial}).

68. Le rang est synonyme de classement, mais les verbes associés sont opposés en français : si un objet passe du second au cinquième rang, celui-ci croît, mais son classement baisse.

Afin d'éviter une trop forte influence de l'échantillonnage sur le résultat, nous ne considérons que les requêtes fréquentes ou communes dont la fréquence dans GR000 est inférieure à la moitié de la fréquence dans le corpus entier[69]. Les résultats pour *limite* = 5, sont présentés dans l'annexe (tableaux 4.1 page 433 et 4.2 page 441).

Bien sûr, toutes les requêtes composées d'un mot simple, polysémique quand il est seul, voient leurs rangs s'accroître grandement : ainsi, les requêtes très vagues — mais trop faciles à interpréter ? — comme « *impot* », « *race* », « *parisien* », « *loto* », « *sexy* », « *loterie* », « *erotique* », « *pmu* », « *anal* », etc., ont des rangs entre 1000 et 5000 dans GR000 quand ce rang initial était compris entre 20 et 600. Et d'autres, comme « *barcelone* », « *golf* » ou « *gif anime* », se retrouvent au rang 11 000 alors qu'ils semblaient « importants », si l'on en croit leurs rangs initiaux : respectivement 737, 398, 647.

Inversement, des requêtes bien plus pertinentes se retrouvent dans les premiers rangs : « *fievre aphteuse* » (rang 11 au lieu de 419), « *geographie* » (124 au lieu de 764), « *seine maritime* » (149 au lieu de 12 667), « *louis 16* » (307 au lieu de 13 775), « *ubu roi* » (433 au lieu de 19 749), « *egalite homme femme* » (593 au lieu de 34 871), « *declaration droit homme citoyen* » (679 au lieu de 20 403), « *elections communales* » (877 au lieu de 45 389), « *phlebologie* » (1642 au lieu de 54 448).

Ces quelques exemples convaincront assurément le lecteur que les représentations d'autrui les plus primaires, aisées à réaliser si on effectue une lecture rapide et non raisonnée des rangs et fréquences des requêtes, s'effondrent lorsqu'on s'intéresse aux utilisateurs ayant un minimum de pratique des moteurs. Les préoccupations culturelles, géographiques et politiques des internautes apparaissent alors bien plus clairement.

69. En effet, on sait que certaines requêtes n'apparaissent qu'une seule journée. Cf. le graphique 2.13 de la page 311. Sont donc rejetées les requêtes exclusivement émises par des personnes de GR000 durant cette semaine.

Chapitre 4

En guise de synthèse

À partir d'une fonctionnalité spécifique des ordinateurs, la création d'archives textuelles instantanées, nous avons cherché à savoir ce que ces listes pouvaient apporter au chercheur contemporain. Pour cela, nous avons choisi des archives massives, textuelles, fruit de l'activité de millions de personnes.

4.1 Réflexivité

Déjà, il est apparu que ces documents produits par des machines ne pouvaient pas être analysés sans machines. Aucun traitement n'a pu être réalisé sans un travail préalable d'écriture, qui consistait à rédiger des lignes de programmation afin de produire d'autres listes et des graphiques. Ainsi, une nouvelle forme de l'écrit conduit à la création de nouveaux dispositifs d'écriture et de lecture, qui entrent en rupture avec ceux qui étaient auparavant à notre disposition.

Ces outils ont eu une incidence certaine sur notre raisonnement. Tout d'abord, ils nous ont incité à développer ou à réadapter des méthodes assez peu répandues dans le domaine des sciences sociales, linguistiques et lexicométriques. Ensuite, du fait que l'on travaillait sur des très grands nombres d'objets et de personnes, ils nous ont permis de mettre en évidence des lois statistiques rarement remarquées, les lois de puissance. Ce qui incite à réviser la plupart des axiomes et des méthodes de la sociologie : les agrégations brutales en classes sociales, en groupes de capital culturel donné, déterminant des normes ou des *habitus* nous apparaissent comme des extrapolations subjectives doublement appuyées sur la loi de Gauss — tant par la volonté de trouver des moyennes et des marges que par la batterie de tests développés pour mettre en évidence des écarts ou des indépendances — qui ne laissent pas de place à l'immense variété des pratiques individuelles. Dans notre étude, chaque personne a une très grande importance, et négliger un profil rare apparaît aussi dangereux que de supprimer les hapax de la littérature française. Nous avons montré à quel point les analystes pressés de telles archives multiplaient les erreurs de raisonnement parce qu'ils faisaient appel à une *doxa* sociologique, linguistique et économique lénifiante.

Le troisième effet est celui de l'infléchissement de notre problématique : un travail sur les listes produites par autrui incite à se pencher sur le rapport d'autrui à l'écrit avant même d'évaluer le sens de ce qu'il écrit. Ce point nous apparaît essentiel dans la mesure où l'objet d'étude est avant tout une trace de pratiques intellectuelles : on ne peut espérer déduire les préoccupations ou les préférences d'un individu à partir d'un texte si on ne sait comment il les exprime, quelles sont ses capacités à les énoncer.

Ceci induit à son tour un travail délicat de retour au texte, et aux outils qui permettent de l'appréhender. Mais ce rapport à l'écriture est mesurable, à partir de faits concrets, ce qui, une fois de plus, nous semble plus constructif pour témoigner de telles pratiques intellectuelles — et donc culturelles — que l'appel à toute une série de concepts préconstruits dont on ne peut savoir jusqu'à quel point ils ne sont pas créateurs des catégories qu'ils permettent de mettre en évidence.

Ainsi, plus qu'une méthodologie, c'est une épistémologie des sciences sociales — dont on n'oubliera pas à quel point elles se fondent sur l'écriture — qu'incite à repenser un tel travail sur les listes contemporaines.

4.2 Sens

Cette étude a aussi mis en évidence des faits sociaux.

Nous avons découvert que nos dictionnaires habituels sont trop réduits pour être représentatifs du nombre inimaginable de mots que nous avons en tête ; nous avons aussi montré que les internautes sont représentatifs des préoccupations de notre société, dans tous ses aspects : culturels, professionnels, ludiques, etc. Que leurs centres d'intérêt, leurs comportements sont mesurables, même s'il faut pour cela développer des outils sophistiqués. Et donc qu'une étude des usages, des pratiques à partir de telles archives est pertinente, à condition de s'en donner les moyens. Il y a là place pour une nouvelle approche scientifique. Proposer d'en faire une « discipline » semble cependant prématuré et dangereux. Prématuré car il n'est pas sûr que l'on dispose de la masse critique de spécialistes (de la langue, du nombre, de la société). Dangereux, car on s'expose à s'enfermer dans une tour d'ivoire où les échanges avec les autres « disciplines » se réduiraient comme peau de chagrin. Or, c'est bien parce qu'on travaille sur la technique commune à toutes les sciences qu'est l'écriture qu'une telle approche se doit de prendre le meilleur des unes et des autres.

Enfin, nous avons mesuré combien cet outillage mental contemporain se diffusait lentement : à partir d'un échantillon massif de 641 000 personnes, et en choisissant une série de critères minimalistes, nous sommes arrivé à la conclusion qu'au plus 13 % de ces personnes l'avaient plus ou moins assimilé, alors que les autres étaient condamnés à des comportements systématiques, plus proches de ceux d'un futur noyé sans planche ni bouée que de ceux du *surfeur* expérimenté que nous propose la publicité.

Ceci ne nous étonne pas. Une technique d'écriture en évolution, parce qu'elle est en relation directe avec notre pensée, notre perception du monde et d'autrui, ne peut être que violemment ségrégative.

Conclusion

L'internet témoigne de l'actualisation d'une des plus vieilles et des plus fidèles techniques de l'homme : l'écriture.

Aussi, l'importation d'une expression comme les « nouvelles technologies de l'information et de la communication » n'apporte rien si ce n'est un brouillard conceptuel qui gêne les analyses rigoureuses de l'internet. Pour se persuader que l'écriture est une technique, il suffit de se pencher sur la complexité des interactions qui mettent en correspondance le système graphique, le support et l'activité intellectuelle. Ces trois objets sont à prendre comme un tout, l'outillage mental ; celui-ci se développe et se réorganise suivant deux régimes temporels, l'un propre à l'individu, l'autre à la société. Le premier en définit le cadre local (usages, remaniements, apprentissages, activités des érudits...) ; le second témoigne de la façon dont l'institutionnalisation de pratiques et de statuts (l'auteur, le livre, la forme littéraire, etc.) conditionne ses modes de transmission et de valorisation : la maîtrise de l'écriture — étendue à tous ses instruments — sollicite bien des convoitises, intellectuelles, politiques, économiques. Les actes et choix qui en découlent ayant à leur tour une influence, parfois imprévue, souvent non dite, sur cette technique (stabilisation de la graphie française ou de la langue allemande, renommée des grands auteurs, etc.).

La notion de technique de l'intellect prend alors tout son sens, même si cette expression est d'un emploi difficile : « technologie de l'intellect » semble plus acceptable, d'une part, du fait du mépris des lettrés pour la technique, d'autre part — et ceci explique peut-être cela — suite à l'engouement récent pour la notion de technologie en France. En effet, la valorisation de la pensée « pure » va souvent de pair avec une critique de la technique, vite accusée d'assassiner une culture — par ailleurs jamais précisément définie. La notion d'outillage mental est souvent refusée par les spécialistes des sciences humaines, pourtant les premiers concernés par son emploi, tout comme est rejetée l'idée que son développement passe par l'existence de collectifs.

Dans un tel contexte, l'ordinateur risque fort d'être mal reçu, alors qu'il s'inscrit dans le renouvellement de notre technique de l'intellect. Il a même été inventé pour cela, et non pour améliorer la machine à écrire que nous avons abandonnée il y a vingt ou trente ans. Au-delà de l'évidence — processus de lecture, d'écriture et de réécriture permis par la machine — la séparation entre le code et son support visuel, la multiplicité des textes intermédiaires entre l'un et l'autre, et la proximité des programmes et codages avec notre système graphique classique témoignent des ruptures et des continuités induites par ces machines sur l'écriture. C'est en comprenant comment elles fonctionnent, comment elles sont *écrites*, que l'on mesure l'importance des transformations fastes ou néfastes qu'elles induisent sur notre outillage mental : bien utilisés, le tableau, l'expression régulière, le graphique et l'hypertexte accroissent singulièrement nos capacités. À maîtriser de tels instruments, dont la manipulation exige une série de connaissances hétérogènes et complexes, nous augmentons, de façon quantitative comme qualitative, notre aptitude à mesurer, à construire des raisonnements, à inventer. À l'opposé, le conformisme techniciste véhiculé par les publicitaires et partagé par une majorité d'universitaires prétend que les machines ne peuvent changer nos manières de travailler, de penser. Obéir à un tel mythe nourrit l'illettrisme (souvent doublé de formes très manuelles d'incompétence, ou d'une taylorisation de la gestuelle des travailleurs de l'intellect), sinon favorise l'acceptation d'un outillage mental de mauvaise qualité produit par quelques industriels soucieux de construire des rentes et des monopoles autour de l'écriture.

Le monde universitaire est le garant et le dépositaire légitime de l'outillage mental, qu'il est aussi chargé de le transmettre et de l'affiner. Mais il en est aussi dépendant, dans son fonctionnement : l'écriture est une technique réflexive, et l'acceptation de cette réflexivité est à la fois le moteur et le cadre explicatif de la science. On comprend alors que les personnes désireuses de maximiser leur pouvoir au sein du monde universitaire aient peu d'intérêt au dévoilement d'une évidence qui aboutit logiquement à la critique de leur appareillage intellectuel, souvent archaïsant, et donc de leur production scientifique. Mais ni l'arrogance, ni l'engagement dans la reproduction administrative d'une caste de lettrés ne suffisent à empêcher des réajustements scientifiques dans le monde universitaire ; ceux-ci sont souvent initiés par des chercheurs issus de disciplines « inférieures » — et souvent considérées comme techniques —, qui renouvellent l'outillage mental en usant de méthodes puisées à des sources hétérogènes et en profitant au mieux du travail collectif. Ainsi, l'écriture et la science sont-elles dans un état de perpétuelle interdépendance, même si cela est masqué.

On comprend alors les inquiétudes et les espoirs qu'a pu susciter l'internet en France au sein des sciences humaines. Son apparition en 1992 à la division « littéraire » de l'École normale supérieure (ENS) témoigne déjà de conflits de positions entre les diverses disciplines de cette institution chargée de reproduire l'élite de l'Université : les sociologues faisaient alliance avec les informaticiens. Les premiers étaient soucieux de leur image de théoriciens n'ayant pas peur des machines, mais ne connaissaient pas l'internet. Les seconds n'avaient que ce mot à la bouche, et favorisèrent l'acquisition d'odinateurs Unix. L'appropriation de ces machines et des réseaux n'alla pas de soi. Elle fut longue, difficile, et n'a pu se produire qu'au sein de petits réseaux qui ont constamment sollicité l'assistance des mêmes informaticiens. Ce sont aussi ces derniers qui ont évalué et publicisé les possibilités éditoriales du web, après avoir proposé une lecture de la légitimation scientifique au prisme de la publication imprimée et de son articulation avec les éditeurs.

Ces expériences restèrent marginales : quatre ans après l'introduction de l'internet, à peine 10 % des élèves de l'ENS littéraire et de son encadrement avaient découvert les réseaux, s'impliquaient dans l'acquisition de la culture informatique nécessaire à leur bon usage, et découvraient comment l'internet modifiait les méthodes de travail des chercheurs. Mais la faible légitimité des nouveaux techniciens de l'écriture et le caractère difficilement acceptable de leurs intuitions en terme d'outillage mental ont conduit les représentants d'une hiérarchie fondée sur les distinctions universitaires à mettre à l'écart ces expérimentateurs quand il s'est agit de combler le grand retard informatique de l'institution en 1997.

Les professeurs d'université, qui raisonnaient en termes d'échanges manuscrits ou de machines à écrire, ont alors facilité la diffusion d'une informatique bureaucratique, indigeste et impensée. Les élèves et les chercheurs furent laissés pour compte, la publication sur le web honnie. Jusqu'en 2001, l'ENS littéraire a vécu sous le règne des tenants de conceptions datant des années 1970, expurgées des notions d'outillage mental et de réflexivité : l'écriture ne pouvait en aucun cas être aussi l'affaire des informaticiens, mais devait rester la propriété des maîtres du beau langage, qui seul pouvait témoigner de la noblesse et de la pureté de leur pensée.

Cet exemple illustre comment réagit un pouvoir universitaire aux abois : à l'ENS, la recherche d'une adaptation raisonnée à une forme d'écriture évolutive n'a pas été la norme. Mais au moins, les conflits découlant de l'introduction de l'internet étaient-ils lisibles : la hiérarchie de la reproduction universitaire s'opposait, avec tous les moyens à sa disposition, aux chercheurs et aux inventeurs.

L'introduction de l'internet au sein des universités, aussi difficile et aussi conflictuelle, témoigne, quant à elle, d'un conformisme étendu à l'ensemble d'une profession — ce qu'on ne pouvait déceler à l'ENS du fait de sa paralysie électronique. Très peu de personnes se sont impliquées avant 1998 dans des projets éditoriaux d'envergure sur le web[1]. D'après les témoignages détaillés d'une trentaine de pionniers — aux statuts fort variés, puisque le professeur d'université côtoie le professeur de lycée, le directeur de recherche au CNRS l'étudiant — l'internet professionnel en sciences humaines ne peut se développer sans une longue pratique, sans l'acquisition d'une solide culture informatique. Mais on se marginalise très vite à s'engager dans ce type de production, coûteuse en temps comme en argent : l'incompréhension, le doute, voire le mépris et la relégation constituent la panoplie des réactions communes des collègues ou supérieurs, — à plus forte raison si l'on appartient à une université ou une institution parisienne. La notion de propriété, la peur de s'engager dans des activités non rentables, un profond mépris pour la tech-

1. Toujours en sciences humaines.

nique apparaissent constitutifs d'une Université qui témoigne de singulières capacités de résistance ; la défiance face à l'internet s'est maintenue au moins jusque la fin 2000.

Mais le fait d'expérimenter de façon approfondie l'internet offre aussi de réels bénéfices, certes symboliques, mais en fait propres au fonctionnement scientifique : tout d'abord, une lecture sociologique de son propre monde, en grande partie parce qu'un engagement concret dans l'édition électronique invite à comparer ses avantages avec ceux de l'édition imprimée : les deux ne se confondent pas, même si une telle comparaison met en lumière l'économie universitaire. Ensuite, une lecture de la dynamique des disciplines dans un cadre mondial. Enfin, des gains conceptuels, intimement liés à la compréhension du caractère technique de l'écriture, qui vont de l'approfondissement de méthodes passées et contemporaines à celui de l'épistémologie d'une science. Au-delà de ces acquis, s'affichent les enjeux réels de l'internet en sciences humaines et la façon dont l'outillage mental et l'idéologie du moment recomposent l'enseignement, aux plans national comme international : les logiques à l'œuvre en France se rencontrent aussi dans d'autres pays européens ; même le système de mesure américain de la notoriété des chercheurs reste inféodé aux normes mises en place par le pouvoir de l'imprimé.

Si le monde de l'enseignement et de la recherche résiste tant à une quelconque transformation de sa technique d'écriture, c'est bien parce qu'une telle mutation favorise une posture réflexive qui change ses règles du jeu, du seul fait que ces règles — étant celles de la science — restent implicitement fondées sur la consécration de celles et ceux qui réussissent à optimiser leur outillage mental.

Dans ses formes contemporaines, cet outillage n'est pas que fait d'outils informatiques que l'on adapterait à de « vieux » textes, numérisés et sauvegardés sur un support électronique. Il est aussi un tout : ordinateurs, programmes, et archives textuelles, que l'on peut glaner au travers du monde, ou qui témoignent de l'intérêt du monde entier pour des textes ou des procédures déposés sur un serveur spécifique. Tous ces textes (articles scientifiques publiés sur le web, archives de consultation d'un serveur, etc.) s'appréhendent comme autant de listes, indépendamment de leur statut (source primaire ou synthèse achevée). Travailler sur l'écriture dans sa triple acception *support*[2] / *texte*[3] / *outils conçus pour le manipuler* mène à des résultats à la fois surprenants et attendus. Attendus parce que c'est toute l'efficacité de la technique de l'intellect qui s'affiche au grand jour. Surprenants parce que les détours obligés, les méthodes à appliquer ou à inventer, les découvertes, et la vision scientifique que l'on acquiert suite à cette immersion dans les listes modernes offrent un plaisir et une lucidité toujours renouvelés. Caractéristiques qui rapprochent le chercheur et l'explorateur — l'un comme l'autre savent qu'ils feront des découvertes imprévues —, les sciences de la nature et l'anthropologie — l'intellect se lisant, s'organisant à la lueur de l'ensemble des outils, méthodes et mesures que l'homme a créés pour comprendre son propre monde.

Les archives électroniques sont souvent sales, comme le sont les tablettes d'argile abîmées ou couvertes de poussière. Avant même de tenter de les lire, il faut les nettoyer patiemment et délicatement, les multiples programmes dédiés à ce toilettage apparaissant comme autant de légers coups de pinceaux. Ce n'est qu'ensuite que le texte, auparavant succession illisible de caractères, apparaît comme tel. Mais souvent, il est encore trop massif. Pour l'appréhender, il faut alors le décomposer en mots, c'est-à-dire en extraire une liste élémentaire, qui n'a aucune valeur scientifique, mais qui crée une distanciation essentielle pour concevoir les futurs traitements du texte.

Les archéologues pressés qui ont étudié jusqu'alors les *access_log* des moteurs de recherche ont oublié cette étape intermédiaire : tant le profil statistique de la liste de mots que les coups de sonde que l'on peut donner en cherchant des mots communs, étranges, ou rares n'expliquent rien de la structure de l'archive, ni de l'intention des scribes qui l'ont créée, dont on sait seulement qu'ils sont très nombreux. Mais ces découvertes stimulent double-

2. Disque dur ou réseau.
3. Combinatoire tirée d'un alphabet.

ment le doute essentiel à la production scientifique. D'une part, l'univers que l'on découvre est proprement exotique : on pressent que nos grilles mentales habituelles seront peu opératoires. D'autre part, cet exotisme n'est pas nouveau : il est celui — permanent — que l'on ressent quand on est confronté à une langue et à ses éléments. Aussi comprend-on bien comment un outil créé pour construire une liste ouvre une problématique qui va inviter à reconsidérer ses hypothèses initiales et à reconstruire d'autres outils ; on comprend aussi que cette démarche sera récursive. À partir de là, on entre dans le territoire de la réflexivité de l'écriture : les instruments de l'informatique dédiés au texte (tris, comptages, expressions régulières, etc.) et à l'image, les méthodes, les questions, et les retours à l'archive interagissent de façon systématique et plurielle. La synthèse d'un raisonnement aboutissant au dessin d'une simple droite détermine la compréhension d'un phénomène collectif et les moyens de le décomposer. Par exemple, la population des auteurs des requêtes à un moteur de recherche n'est pas la réunion de quelques groupes qui auraient des caractéristiques explicites ; à un groupe, homogène, s'oppose toujours son complément, qui témoigne, lui, de la variété — proprement incommensurable, si l'on raisonne en termes statistiques — des pratiques d'écriture ; et ainsi de suite. Cette découverte nous apparaît d'autant plus importante qu'elle induit une méthode (itérative) de construction de catégories qui promet d'être très fructueuse en sociologie.

On remarque aussi une interdépendance prononcée entre la qualité de l'outil construit et la qualité de la question qui en est à l'origine ; naturellement, les deux ne s'émancipent pas des propriétés matérielles de l'instrument qui permet ce travail sur l'écriture, puisque certaines procédures ne sont pas réalisables si l'on ne dispose pas de moyens de calcul et de stockage conséquents. Certes, depuis plus d'un siècle, les méthodes statistiques d'échantillonnage répondent de façon très efficace à nos lacunes en matière de comptage et d'exhaustivité, et il est d'autant plus légitime d'en abuser que l'on dispose de données massives. À condition de garder en tête le fait que les estimateurs proposés sont souvent fondés sur des hypothèses fortes, pas toujours vérifiées ni même vérifiables, et que l'originalité de la pensée humaine est telle qu'un taux de sondage acceptable dans les cas d'école peut s'avérer très insuffisant quand on travaille sur la langue et ses usages.

On mesure ainsi comment l'étude d'un *texte-liste* informatique renvoie à l'explicitation de toutes les interactions entre la pensée, le système de signes et le support. Il n'y a là rien de bien nouveau, si ce n'est qu'on peut, à la lumière de notre technique de l'intellect, imaginer aisément les dynamiques que produisaient celles du passé. Bien sûr, le grand acquis de tout travail de ce type est méthodologique : l'étude formelle d'une liste mène non seulement à la compréhension de la dynamique réflexive de l'outillage mental, mais aussi à la reformulation épistémologique, à la mise en place de modèles et de *méta*-méthodes utilisables par quiconque désire explorer dans un cadre spécifique les produits de l'écriture contemporaine.

Cette mise en perspective détaillée de la pratique scientifique, à partir du questionnement d'un objet de l'écriture, offre aussi (preuve de son efficacité pratique) des résultats surprenants. On ne doute pas que la technique de l'intellect dont nous explorons les limites soit maîtrisée à des degrés fort variables. En l'occurrence, les archives électroniques que nous avons étudiées regorgent de cahiers d'écriture d'écoliers : 87 % des auteurs de ce long texte sont complètement démunis face à cette technique.

Si de nombreux universitaires ont tant de mal à s'approprier un outillage mental renouvelé, ce qui explique qu'il soit faiblement socialisé, on ne s'étonne pas que seulement 13 % de la population française en ait une maîtrise minimale. En revanche, le fait que la relation directe entre l'internet et l'écriture, entre l'informatique et la technique de l'intellect soit si rarement évoquée, nous surprend.

Nous n'aimerions pas être trop peu nombreux à comprendre ces faits, ni à pouvoir profiter pleinement de ces capacités de l'outillage mental.

Annexe

Chapitre 1

Enquêtes courrier électronique

Ces deux enquêtes s'appuient sur l'analyse des fichiers *syslog* de serveurs de courrier électronique.

Ces fichiers gardent les traces des en-têtes des messages émis et reçus par les possesseurs de comptes. Pour chaque message, on peut retrouver l'adresse de son expéditeur, la date et l'heure d'envoi, sa taille, et d'autres informations techniques (relayage des machines, émission par un robot, etc.). De même pour le destinataire. L'ennui étant que ces deux types d'informations, relatives à l'expéditeur et au destinataire, ne sont pas sur les mêmes lignes. Il s'ensuit donc un travail d'appariement qui peut devenir singulièrement complexe, mais qui ne sera pas évoqué ici.

Par choix déontologique[1], l'analyse s'est limitée à la description formelle des échanges, sans en étudier ni la taille, ni l'heure d'envoi ; par suite, les comptes ont été totalement anonymisés.

1. L'analyse des archives électroniques permettant de comprendre les faits et gestes des internautes est hélas de plus en plus développée au sein des « start-up » actuelles, qui voient ainsi le moyen d'attirer la manne publicitaire. Certaines vont jusqu'à installer des mouchards sur des ordinateurs ou des *palm-pilot* à l'indu de leurs propriétaires. Dès que ces derniers se connectent, la totalité de leurs activités électroniques sont alors transmises à l'entreprise.

1.1 École littéraire

1.1.1 Méthodologie

Au printemps 1996, l'internet était introduit au sein de l'ENS littéraire depuis quatre ans, et le serveur *elias* regroupait quasiment tous les comptes des enseignants de cette division tentés par le courrier électronique. À ce groupe s'ajoutaient des élèves (en majorité du département de sciences sociales), des chercheurs CNRS, des étrangers. Mais certains élèves avaient aussi un compte sur les machines qui leur étaient dédiées, et gérées par le SPI. Cette enquête de 50 jours consécutifs (mai et juin 1996) est exhaustive quant aux pratiques des enseignants et des chercheurs de l'École littéraire et des laboratoires (CNRS) associés. Elle est représentative des usages des élèves (en sciences humaines), comme le prouvera l'enquête suivante.

Il y a 103 possesseurs de comptes électroniques au moment de l'enquête : 38 élèves ou étudiants rattachés au département de sciences sociales, 13 personnes de l'encadrement (secrétaire, enseignants-chercheurs, etc.), 4 personnes de la cellule informatique littéraire, 9 enseignants-chercheurs et 13 élèves de l'École littéraire non rattachés au département (pensionnaires étrangers inclus), 13 anciens élèves et autant d'« invités » (personnes extérieures à l'ENS qui avaient des liens plus ou moins formels avec l'institution).

1.1.2 Résultats généraux

Le total des 10407 messages qui ont transité par le serveur en 50 jours correspond à :
— 1059 messages émis en interne (x@elias écrit à y@elias) et donc autant de reçus en interne ;
— 269 messages adressés à des personnes de l'ENS (i.e. d'un autre sous-réseau que celui des 103 utilisateurs) et 458 messages reçus provenant de personnes de l'ENS ;
— 2659 envoyés à l'extérieur et 5962 reçus de l'extérieur.

Chaque utilisateur du courrier aurait donc émis ou reçu 2 messages par jour en *moyenne*.

En fait 27 personnes n'ont envoyé aucun message et n'en ont reçu que deux. Pour 25 autres, le total « émission + réception » est compris entre 3 et 10, avec une seule personne qui en a émis 6, ce qui constitue le maximum pour ce groupe. Ces flux sont quasi-inexistants si on se rappelle la durée de l'enquête : 50 jours. Un premier frémissement apparaît chez les personnes pour lesquelles le total « émission + réception » est compris entre 12 messages et 50 inclus. Ce groupe de 20 personnes commence à rédiger des messages : environ autant qu'il en reçoit. Ceci témoigne d'un timide début de socialisation du courrier électronique auprès de ce petit groupe.

Enfin un groupe de 30 personnes a émis ou reçu entre 56 et 2499 messages pour la période considérée, ce qui correspond à un minimum d'un message émis ou reçu par jour. Nous appelons « utilisateur régulier » une personne qui a de telles pratiques, c'est-à-dire dont la somme des messages émis et reçus dépasse un par jour en moyenne[2]. Et encore, dans ce groupe, il y a deux personnes, peut-être à l'étranger, qui n'ont émis aucun message, et les membres de la cellule informatique littéraire, qui ont une fonction d'initiateurs des pratiques électroniques. Si nous nous intéressons à l'indicateur « messages émis », qui témoigne plus de pratiques volontaristes, il apparaît que sur ces 30 personnes, 21 ont émis plus d'un courrier par jour ouvré (soit plus de 35 messages émis en 50 jours), et 14 en ont émis plus d'un par jour. Au sein du pôle de socialisation de l'internet lui-même, 86 % des personnes les plus sensibilisés au courrier électronique ne s'en servaient donc pas.

Au sein du département de sciences sociales lui-même, dont 51 membres disposaient d'un compte, on dénombrait quatre élèves et cinq enseignants « utilisateurs réguliers ». Les autres usagers réguliers venaient chercher dans ce département le confort électronique dont ils ne pouvaient disposer dans le leur ou dans leur université (d'accueil ou d'origine). La majorité de ces der-

2. Et reconnaissons qu'un tel seuil prête à rire en 2002 : ce qui prouve à quel point le courrier électronique n'était pas entré dans les mœurs universitaires en France en 1996.

statut	H	F	Total	H rég.	F rég.	Total rég.	% rég
Élèves (et DEA) sc. soc.	25	13	38	2	2	4	10
Anciens élèves (litt.)	6	7	13	1	1	2	15
Élèves litt.	5	8	13	1	2	3	23
Encadr sc. soc.	9	4	13	5	0	5	38
Encadr info. litt.	2	2	4	2	2	4	100
Encadr litt.	5	4	9	1	2	3	33
Invités	8	5	13	6	3	9	69
Total	60	43	103	18	12	30	29

TAB. 1.1 *Profils agrégés des 103 possesseurs de comptes sur la machine « elias » au printemps 1996*

niers était en étroit contact avec les universités américaines, ce qui explique leur statut de « pionniers ».

Le tableau 1.1 récapitule la composition des utilisateurs occasionnels et réguliers ; il complète les catégories énoncées précédemment par une classification par sexe, afin de prouver que l'hypothèse de répartition sexuée des pratiques ne tient pas. La totalité des personnes ayant émis ou reçu plus de 51 courriers est comptabilisée ici dans les utilisateurs réguliers.

1.1.3 Variété des usages

Si le courrier émis en interne est encore faible (1059 messages sur un total de 3987 messages émis), ce qui témoigne que le fait de ne pas lire son courrier était admis, on discerne déjà quelques « grands » utilisateurs : deux personnes émettent ou reçoivent 40 à 50 messages par jour : ce sont deux chercheurs étrangers, l'un sans attache institutionnelle, résidant en France, l'autre, invité pour un an dans un laboratoire CNRS. Ces deux personnes ont des pratiques opposées : la première écrit peu et reçoit beaucoup de courriers provenant de listes de discussion (181 messages émis pour 2318 reçus), la seconde se constitue un réseau serré de relations sociales (1095 messages émis pour 855 reçus). Mais toutes deux témoignent du statut particulier des « invités » dans cet échantillon. Apparaît aussi un standard de toutes les études statistiques sur les usages informatiques : la distribution n'est absolument pas symétrique. En termes de « consommation », une grande partie de la population étudiée a des pratiques très faibles, quand un petit groupe monopolise les pratiques : par exemple les trois premiers utilisateurs sont responsables de 51 % du trafic et les trois premiers groupes de faibles utilisateurs définis précédemment (71 % de la population) ne participent aux échanges électroniques qu'à hauteur de 7 %.

S'il est évident qu'on ne peut corréler l'usage du courrier électronique avec une quelconque compétence disciplinaire (celui-ci peut servir à un usage ludique ou privé, et des chercheurs réputés peuvent, encore aujourd'hui, ne pas se servir du *courriel*), la typologie des listes de discussion d'alors peut expliquer en partie le peu d'engouement des chercheurs en sciences humaines pour ce mode de communication. Il est évident que ces listes ont une forte incidence sur les statistiques du courrier reçu : parmi les messages, 3303 proviennent de listes de discussions extérieures à l'ENS[3]. Ainsi plus de la moitié du courrier externe au réseau local provenait de listes de discussion, lesquelles ne touchaient qu'une quinzaine d'abonnés, qui, évidemment, font partie des utilisateurs réguliers. Ces listes, au nombre de 73, provenaient dans leur grande majorité d'universités américaines : 30 exactement, face à trois listes françaises et trois canadiennes (plus précisément, québécoises). Cette décomposition met d'ores et déjà en évidence l'absence d'offre d'information en sciences humaines en France à cette période : certes, le nombre d'utilisateurs réguliers du courrier électronique est fort réduit, mais en France les outils et les réseaux permettant de dynamiser les pratiques, même sur le plan extra-professionnel (car toutes les listes auxquelles sont abonnées nos personnes ne sont pas représentatives de groupes d'universitaires), étaient alors fort lacunaires dans l'espace francophone. En revanche, la forte présence de listes américaines, y compris pour des thèmes touchant aux sciences de l'antiquité, met en évidence le second emprunt indispensable à la socialisation

3. Trois listes, émises par le SPI, étaient adressées aux membres de la cellule informatique littéraire, ce qui correspond à un total de 108 messages.

réussie de l'internet : après avoir imité les « pratiques » des informaticiens, on importe celles des résidents outre-Atlantique.

Le timide usage du courrier électronique au sein de l'École littéraire peut s'expliquer au vu du faible nombre d'ordinateurs : la plupart des enseignants ne disposaient pas de machines sur leurs bureaux[4], et le nombre total des ordinateurs disponibles pour les élèves et enseignants (plus de mille personnes au total) ne dépassait pas soixante-dix, en incluant la trentaine de machines dédiées aux élèves. Pour beaucoup, l'ordinateur n'était que le prolongement d'une machine à écrire, dont la pratique était déjà complexe quand il s'agissait d'utiliser des polices de caractères « exotiques » ou de relire des fichiers au format inconnu.

Il n'en reste pas moins que seulement neuf personnes ayant un statut officiel au sein de l'ENS littéraire (incluant donc ses enseignants) ont émis plus d'un courrier par jour pendant la période de l'enquête[5].

1.1.4 Tableaux

1.1.4.1 Légende

Nous décrivons ici la légende commune aux deux tableaux.

S : sexe ;

T : type. R= régulier ; S= membre de la cellule informatique littéraire.

ST : statut. E= élève du département de sciences sociales ou étudiant de son DEA ; G= invité, enca= membre de l'encadrement au sein du département de sciences sociales ; enca I= membre de la cellule informatique littéraire ; enca L= membre de l'encadrement au sein d'un autre département littéraire ; AE= ancien élève de l'ENS, E G= élève d'un autre département littéraire ou pensionnaire étranger.

TTE : total des messages émis.

TTR : total des messages reçus.

TTT : total des messages émis et reçus.

RLENS : nombre de messages provenant de listes ENS (aucun n'a été envoyé à de telles listes).

4. Et s'ils en disposaient, rares étaient celles connectées à Internet.

5. Voir tableau 1.3, en considérant les utilisateurs ni AE ni G.

RLFR : nombre de messages provenant de listes françaises (aucun n'a été envoyé à de telles listes).

ELCA : nombre de messages adressés à des listes canadiennes.

RLCA : nombre de messages reçus de listes canadiennes.

ELEDU : nombre de messages adressés à des listes universitaires américaines.

RLEDU : nombre de messages reçus de listes universitaires américaines.

ELAU : nombre de messages adressés à des listes différentes des précitées.

RLAU : nombre de messages reçus de listes différentes des précitées.

EINT : nombre de messages émis à destination du réseau local.

RINT : nombre de messages reçus et provenant du réseau local.

EENS : nombre de messages émis à destination d'autres réseaux de l'ENS.

RENS : nombre de messages reçus et provenant d'autres réseaux de l'ENS.

EEXT : nombre de messages envoyés à l'extérieur de l'ENS.

REXT : nombre de messages reçus et ne provenant pas de l'ENS.

TTE : total des messages émis.

TTR : total des messages reçus.

TTT : somme des messages émis et reçus.

E/E+R : rapport des messages émis sur le total des messages émis et reçus.

REPART : cumul des flux.

1.1.4.2 Usage des listes

S	T	ST	TTE	TTR	TTT	RLENS	RLFR	ELCA	RLCA	ELEDU	RLEDU	ELAU	RLAU
F	R	E	26	30	56	0	0	0	0	0	0	0	0
H		G	0	60	60	0	0	0	0	0	0	0	0
H	R	E	43	18	61	0	0	0	0	0	0	0	0
H	R	G	24	41	65	0	0	0	0	0	0	0	0
H	R	enca	20	47	67	0	0	0	0	0	0	0	0
F	R	E	28	41	69	0	0	0	0	0	0	0	0
H	R	G	32	38	70	0	0	0	0	0	0	0	0
F	R	G	40	42	82	0	0	0	0	0	0	0	0
H	R	enca	36	52	88	0	0	0	0	0	0	0	0
F	R	enca L	49	42	91	0	0	0	0	0	0	0	0
H	R	AE	43	55	98	0	0	0	0	0	0	0	21
F		E G	0	100	100	0	0	0	0	0	0	0	93
H	R	enca L	50	82	132	0	0	0	0	0	0	0	0
F	R	E G	67	74	141	0	0	0	0	0	0	0	0
H	R	G	83	58	141	0	0	0	0	0	0	0	0
H	R	enca	5	143	148	0	0	0	0	0	76	0	53
F	R	enca L	76	74	150	0	0	0	0	0	0	0	0
F	R	G	83	90	173	0	0	0	0	0	0	0	0
F	S	enca I	70	118	188	20	0	0	0	0	0	0	0
H	R	enca	42	168	210	0	0	0	0	0	83	0	20
H	R	E	124	119	243	0	0	0	0	0	0	0	0
H	R	E G	136	128	264	0	0	0	0	0	0	0	0
H	R	G	98	178	276	0	0	0	0	0	0	3	69
H	R	enca	151	159	310	0	0	0	0	1	1	0	0
H	S	enca I	13	312	325	46	0	1	155	3	3	1	40
F	R	AE	153	318	471	0	0	0	0	4	82	0	46
F	S	enca I	77	705	782	0	431	1	209	0	0	0	0
H	S	enca I	859	486	1345	42	0	0	0	0	0	0	0
F	R	G	1095	855	1950	0	0	0	0	0	25	0	0
H	R	G	181	2318	2499	0	429	0	212	1	261	1	994
Total			3704	6951	10655	108	860	2	576	19	531	5	1336

TAB. 1.2 *Usage des listes chez les utilisateurs réguliers.*

1.1.4.3 Description détaillée des pratiques

TAB. 1.3 *Usage du courrier électronique chez les 103 utilisateurs de la machine « elias ».*

S	T	ST	EINT	RINT	EENS	RENS	EEXT	REXT	TTE	TTR	TTT	E/E+R	REPART
F		E	0	2	0	0	0	0	0	2	2	0	0
H		E	0	2	0	0	0	0	0	2	2	0	0
H		enca	0	2	0	0	0	0	0	2	2	0	0
F		AE	0	2	0	0	0	0	0	2	2	0	0
H		E	0	2	0	0	0	0	0	2	2	0	0
F		E G	0	2	0	0	0	0	0	2	2	0	0
H		E	0	2	0	0	0	0	0	2	2	0	0
F		E G	0	2	0	0	0	0	0	2	2	0	0
F		AE	0	2	0	0	0	0	0	2	2	0	0
H		AE	0	2	0	0	0	0	0	2	2	0	0
H		E	0	2	0	0	0	0	0	2	2	0	0
H		G	0	2	0	0	0	0	0	2	2	0	0
H		enca L	0	2	0	0	0	0	0	2	2	0	0
H		enca	0	2	0	0	0	0	0	2	2	0	0
F		E	0	2	0	0	0	0	0	2	2	0	0
H		E	0	2	0	0	0	0	0	2	2	0	0
F		G	0	2	0	0	0	0	0	2	2	0	0
H		E	0	2	0	0	0	0	0	2	2	0	0
H		E	0	2	0	0	0	0	0	2	2	0	0
F		enca	0	2	0	0	0	0	0	2	2	0	0
H		E	0	2	0	0	0	0	0	2	2	0	0
F		E	0	2	0	0	0	0	0	2	2	0	0
F		enca L	0	2	0	0	0	0	0	2	2	0	0
F		E	0	2	0	0	0	0	0	2	2	0	0
H		E	0	2	0	0	0	0	0	2	2	0	0
F		E G	0	2	0	0	0	0	0	2	2	0	0
F		enca	0	2	0	0	0	0	0	2	2	0	0
H		E G	0	2	0	1	0	0	0	3	3	0	0
F		E	0	3	0	0	0	0	0	3	3	0	0,01
H		E	0	2	0	0	0	1	0	3	3	0	0,01
F		E	0	2	0	0	1	0	1	2	3	0,33	0,01
F		AE	0	2	0	0	0	2	0	4	4	0	0,01
F		E G	0	2	0	0	0	2	0	4	4	0	0,01
H		AE	0	4	0	0	0	0	0	4	4	0	0,01
H		enca	0	2	0	0	1	1	1	3	4	0,25	0,01
H		E	0	2	0	0	1	1	1	3	4	0,25	0,01
H		E	0	2	0	0	1	1	1	3	4	0,25	0,01
H		AE	0	2	1	1	0	0	1	3	4	0,25	0,01
F		E	0	2	0	0	1	1	1	3	4	0,25	0,01
H		enca L	0	2	0	0	1	2	1	4	5	0,2	0,01
H		E	0	2	0	0	1	2	1	4	5	0,2	0,01
H		E	2	4	0	0	1	0	3	4	7	0,43	0,01
F		AE	0	2	0	0	5	0	5	2	7	0,71	0,01
H		AE	0	8	0	0	0	0	0	8	8	0	0,01
H		E	0	8	0	0	0	0	0	8	8	0	0,01
H		AE	0	2	1	1	1	4	2	7	9	0,22	0,01
F		E G	0	2	0	0	2	5	2	7	9	0,22	0,01
H		E	0	2	0	0	2	5	2	7	9	0,22	0,01
H		E G	0	9	0	0	0	1	0	10	10	0	0,02

suite sur la page suivante

suite de la page précédente													
S	T	ST	EINT	RINT	EENS	RENS	EEXT	REXT	TTE	TTR	TTT	E/E+R	REPART
H		E G	0	2	0	0	0	8	0	10	10	0	0,02
H		E	0	2	1	1	3	3	4	6	10	0,4	0,02
F		E G	0	0	0	0	6	4	6	4	10	0,6	0,02
H		E G	1	3	0	2	2	4	3	9	12	0,25	0,02
F		enca	0	2	0	0	7	3	7	5	12	0,58	0,02
H		E	3	10	0	0	0	0	3	10	13	0,23	0,02
H		E	0	2	0	0	3	8	3	10	13	0,23	0,02
H		E	0	2	0	0	5	7	5	9	14	0,36	0,02
F		E	0	2	0	0	7	5	7	7	14	0,5	0,02
F		AE	14	1	0	0	0	0	14	1	15	0,93	0,03
H		G	0	2	0	0	0	15	0	17	17	0	0,03
F		E	0	2	0	0	9	6	9	8	17	0,53	0,03
H		enca	0	9	0	0	7	4	7	13	20	0,35	0,03
F		enca L	0	2	0	0	8	10	8	12	20	0,4	0,03
H		E	10	9	0	0	1	0	11	9	20	0,55	0,03
H		E	1	3	0	1	17	19	18	23	41	0,44	0,04
F		E	0	2	0	0	19	23	19	25	44	0,43	0,04
F		enca	18	14	0	0	5	7	23	21	44	0,52	0,05
F		G	6	15	0	0	2	23	8	38	46	0,17	0,05
H		enca L	9	18	1	0	16	3	26	21	47	0,55	0,05
F		E	0	0	0	0	15	34	15	34	49	0,31	0,06
H		E	0	2	1	2	22	22	23	26	49	0,47	0,06
H		enca L	14	17	0	0	9	9	23	26	49	0,47	0,07
F		AE	4	6	3	3	12	22	19	31	50	0,38	0,07
F	R	E	6	9	3	1	17	20	26	30	56	0,46	0,08
H		G	0	2	0	0	0	58	0	60	60	0	0,08
H	R	E	4	5	0	0	39	13	43	18	61	0,7	0,09
H	R	G	23	39	1	2	0	0	24	41	65	0,37	0,09
H	R	enca	4	29	0	0	16	18	20	47	67	0,3	0,1
F	R	E	5	17	1	1	22	23	28	41	69	0,41	0,1
H	R	G	16	20	0	0	16	18	32	38	70	0,46	0,11
F	R	G	39	41	0	0	1	1	40	42	82	0,49	0,12
H	R	enca	8	23	0	0	28	29	36	52	88	0,41	0,12
F	R	enca L	2	3	5	4	42	35	49	42	91	0,54	0,13
H	R	AE	3	3	0	0	40	52	43	55	98	0,44	0,14
F		E G	0	4	0	0	0	96	0	100	100	0	0,15
H	R	enca L	9	16	5	16	36	50	50	82	132	0,38	0,16
F	R	E G	5	12	2	9	60	53	67	74	141	0,48	0,17
H	R	G	0	2	0	0	83	56	83	58	141	0,59	0,19
H	R	enca	0	4	0	0	5	139	5	143	148	0,03	0,2
F	R	enca L	10	6	0	0	66	68	76	74	150	0,51	0,21
F	R	G	13	25	0	0	70	65	83	90	173	0,48	0,23
F	S	enca I	18	32	50	85	2	1	70	118	188	0,37	0,24
H	R	enca	2	4	6	7	34	157	42	168	210	0,2	0,26
H	R	E	19	21	4	4	101	94	124	119	243	0,51	0,28
H	R	E G	50	50	0	0	86	78	136	128	264	0,52	0,31
H	R	G	33	40	0	0	65	138	98	178	276	0,36	0,33
H	R	enca	24	25	10	6	117	128	151	159	310	0,49	0,36
H	S	enca I	0	11	3	96	10	205	13	312	325	0,04	0,39
F	R	AE	47	67	7	12	99	239	153	318	471	0,32	0,43
F	S	enca I	22	19	34	29	21	657	77	705	782	0,1	0,49
H	S	enca I	587	237	130	169	142	80	859	486	1345	0,64	0,61
F	R	G	17	21	0	1	1078	833	1095	855	1950	0,56	0,78
H	R	G	11	23	0	4	170	2291	181	2318	2499	0,07	1
Total			1059	1059	269	458	2659	5962	3987	7479	11466	0,35	

1.2 Usages des élèves

Une seconde enquête, du même type, a été réalisée durant deux semaines à la fin de l'année 1996 à partir des archives du serveur de courriel des élèves. Le détail des résultats n'est pas développé ici car il est disponible à l'URL http://barthes.ens.fr/atelier/articles/guichard-mai-98.html.

L'enquête a concerné 851 personnes, pour un total d'environ 40 000 messages. Si l'on exclut les 222 élèves n'ayant rédigé aucun message en quinze jours[6], il en reste 629, dont 229 ont émis au moins un message par jour en moyenne, soit 36 % des comptes retenus. Ces taux d'usagers réguliers[7] varient légèrement suivant les disciplines[8], et l'on rencontre 20 % de littéraires ayant un tel usage régulier du courrier électronique. Mais le nombre d'élèves littéraires à disposer d'un compte sur les machines du SPI était alors réduit : 245 au total, pour environ 500 personnes. Ces usagers réguliers étant au nombre de 33, nous en déduisons que 6,6 % des élèves littéraires ayant un compte sur le serveur étudié se servaient du courrier électronique.

Les deux enquêtes offrent donc des conclusions analogues : ce n'est qu'une très faible proportion des élèves, ou des enseignants de l'École littéraire qui étaient un tant soit peu familiarisés avec le courrier électronique en 1996.

6. Car dans cette catégorie sont inclus des réels non utilisateurs, comme des personnes parties temporairement à l'étranger, et disposant d'un compte extérieur à l'ENS à partir duquel aucune statistique ne peut être produite.

7. Dans cette enquête, « régulier » signifie donc « qui rédige au moins un courrier par jour en moyenne ».

8. Et les nationalités, les élèves étrangers invités s'avérant les plus grands utilisateurs, avec un taux d'usagers réguliers de 52 %.

Chapitre 2

La presse et l'internet en 1996

2.1 L'internet diabolisé

Dans son supplément « télévision, radio, multimédia » daté des 17-18 novembre 1996, le quotidien le Monde publiait deux pages entières sur l'internet. Le titre est symptomatique de la façon dont les élites françaises concevaient alors les réseaux et leurs protocoles : « Internet [1], de l'enfer au paradis ». Le sur-titre n'est pas moins évocateur : « Le retour en grâce des réseaux ». Pour qui ne voulait pas lire le long et ennuyeux texte d'Annie Kahn, un encart résumait les dix aspects « diaboliques » et « angéliques » d'Internet (avec les icônes associées). Tous les poncifs étaient alors réunis, avec, comme premier point négatif : « Internet favorise la pédophilie » ; à l'opposé, l'intérêt de l'internet ne résidait pas dans un quelconque accès à la connaissance, mais dans la face opposée du premier travers : « Internet permet d'arrêter les pédophiles » [2].

Les autres points négatifs valaient le premier, et les points « positifs » ne faisaient que nier le préjugé en regard. Aussi, nous nous contenterons de rappeler les autres critiques : Internet facilite l'espionnage industriel (point 2) ; les fanas d'Internet s'intéressent surtout aux messageries roses (3) ; Inter-

1. Nous respectons ici la graphie du journal.
2. Avec un second encart à la clé sur le thème : « Grâce à Internet, au Royaume-Uni, la police a démantelé un réseau de pornographie enfantine ».

net spolie les auteurs (4) ; Internet accroît les inégalités entre info-riches et info-pauvres (5) ; Internet est un nouveau porte-voix pour le modèle culturel américain (6) ; Internet, c'est Big Brother (7) ; Internet va supprimer des emplois (8) ; Internet facilite les trafics en tout genre (9) ; Internet facilite la communication entre terroristes (10).

Entre les pédophiles, les Américains et les terroristes, la boucle était bouclée. Certes, Khaled Kelkal n'avait été tué qu'un an auparavant[3] et les médias avaient fait leurs gorges chaudes de l'affaire Dutrou durant tout l'été. Après de virulentes critiques contre le réseau internet, réservé aux poseurs de bombes et à la mafia (d'où les thématiques du Monde), les médias s'étaient lancés dans une attaque en règle contre « ce réseau de pédophiles » : une dépêche de l'AFP avait été jusqu'à annoncer qu'« un million d'images pornographiques et 40 millions de pages Internet seraient consacrées à la pornographie enfantine », si un démentait suivait le lendemain, le mal était fait : le 30 août, divers journaux, dont Libération reprenaient l'information initiale[4].

3. Auteur présumé d'attentats, il est abattu en septembre 1995.

4. Pour Libération : cf. l'article de Francis Mizio, « Pédophilie incontrôlable sur le Net ». Ces informations sont encore disponibles sur plusieurs pages, dont
http://www.netreporter.fr/netframe/chronique/html/airont.htm
et http://www.multimania.com/uzine/OURS/afp2.htm. Cette dernière cite le rectificatif de l'AFP :
« From : AFP MAIL
Subject : AFP-MAIL : tie.-social-media
Date : Thu, 29 Aug 96 20 :36 :38 METDST
Mailer : Elm [revision : 70.85.2.1]
tie.-social-media CORR
Conf.-Enfance/Internet : le visage "high-tech" de la pornographie enfantine de l'envoyée spéciale de l'AFP Bénédicte MANIER
RECTIFICATION
VOICI VERSION CORRIGEE de notre envoi de Stockholm diffusée mercredi et qui comportait de graves erreurs de chiffres dans le premier paragraphe.
STOCKHOLM, 29 août (AFP) - Un million d'images pornographiques sur 40 millions de pages Internet sont actuellement consacrées à la pornographie, y compris celles concernant des enfants 'qui sont facilement accessibles', estiment les experts présents au congrès de Stockholm sur l'exploitation sexuelle des enfants. ».

Mais le facteur le plus impressionnant de cette période, et là, le journal le Monde reproduit parfaitement les échos de l'idéologie du moment, est l'assignation de vices et de vertus à ce qui n'est pensé que comme une technique. Nous ne savons pas si la machine à laver et les moteurs à 16 soupapes avaient été accusés en leur temps des mêmes infamies. Mais l'on peut penser, à la suite de Serge Halimi[5] que le web, dont on avait oublié qu'il nous venait de Genève pour le traiter de média américain, effrayait suffisamment les pouvoirs organisés autour d'une diffusion de l'information à sens unique pour qu'on le diabolise[6].

2.2 Le virus informatique

Au moins un magazine a pris le contre-pied de cette idéologie : « Le virus informatique » (qui n'existe plus depuis juillet 2000). Dans son premier numéro, daté de février 1997, un dénommé Nick Larsen publie en page 20 l'article « Devenez pirate en toute simplicité ». Le bandeau est explicite : « On a souvent accusé Internet de tous les maux, d'être un nid de vipères et de pirates en tout genre. Avec un réseau aux ramifications internationales, il est en effet impossible de censurer des fauteurs de troubles situés dans des pays dénués de tout sens de la justice. Mais que dire lorsque ces mêmes informations sont regroupées sur des CD-Rom vendus en toute impunité dans bon nombre de boutiques françaises ? ». L'auteur explique que diverses informations sensibles, telles que comment réaliser un virus, fabriquer une bombe, ôter une protection logicielle ou téléphoner gratuitement, étaient en vente libre d'octobre 1995 à février 1996 ; le cédérom fut censuré suite à une plainte des sociétés éditrices de cartes bancaires... alors que d'après l'auteur, « les numéros de cartes proposés étaient déjà invalides ».

5. [Hal97].
6. Cet article fut publié au moment où le premier ministre Alain Juppé décida de s'intéresser à Internet (http://www.cnam.fr/educasup/internet.htm, disparue). On peut donc imaginer que dans l'article précité, Le Monde s'attachât en fait à « dédiaboliser » Internet, comme il le prétend dans son sur-titre.

Chapitre 3

Revues électroniques

3.1 Économie

Par *site savant*, nous entendons un site réalisé par un érudit, ou par un groupe d'experts (membres d'une équipe, laboratoire) qui contient des informations rares, donc précieuses (sont donc exclus les « sites vitrines » à faible contenu). Une *revue savante électronique* se distingue d'un site savant en ce sens qu'elle est constituée autour d'un comité de rédaction, qui définit une politique éditoriale, opère un travail de sélection scientifique des articles à publier, des ouvrages dont elle décide de rendre compte.

Dans leur forme électronique, la revue et le site savants s'éloignent beaucoup de l'imprimé : on n'est pas obligé d'attendre l'obtention d'un nombre minimum de pages pour publier un ou plusieurs articles. La diffusion en est quasi-instantanée. Dans sa forme actuelle, un article (ou un ouvrage) imprimé est soumis à de fortes contraintes : une taille à ne pas dépasser, un nombre limité d'illustrations, souvent en noir et blanc, obligent le chercheur à « aller à l'essentiel », à convaincre plus qu'il ne démontre, et l'empêchent de faire profiter, pour analyse critique ou pour prolongation de la recherche, ses lecteurs des données qu'il a souvent patiemment acquises et organisées.

Avec le faible coût de la publication électronique, des outils essentiels, et traditionnels, retrouvent donc leur place : sources primaires (iconographie, archives, tableaux statistiques, bases de données, notes explicatives, etc.) ou dossiers documentaires. Ces objets peuvent se construire au fur et à mesure (ex : bibliographie).

Les publications électroniques proposent aussi des outils nouveaux : liste de pointeurs, description des colloques et manifestations, comptes rendus de séminaires sont les plus classiques. L'insertion d'un moteur d'interrogation interne rend la revue totalement indexable [1] et permet au lecteur de la parcourir et de la reconstruire suivant ses propres centres d'intérêt [2].

Enfin, les concepteurs peuvent donner au lecteur la possibilité de travailler lui-même sur les sources proposées pour créer les objets scientifiques dont il a besoin : par exemple, à partir de tableaux statistiques relatifs à des données spatialisées, laisser le lecteur réaliser la carte de son choix, quand l'éditeur ou l'auteur n'aurait proposé qu'une sélection réduite de telles cartes. Cette production d'outils — qui n'est jamais qu'une forme de génération de pages dynamiques — peut offrir une autonomie totale au lecteur : si on lui propose de produire des statistiques à partir de données existantes sur le serveur, on peut lui permettre d'en produire à partir de ses propres données.

Réactivité plus grande à l'égard de l'état des lieux de la recherche, indexation quasi-automatique, argumentations renforcées, grâce à l'exhaustivité potentielle des sources et méthodes, mais aussi grâce à l'écriture hypertextuelle, liberté accrue du lecteur (en même temps que son profil peut-être connu avec l'analyse des archives électroniques), donnent des avantages décisifs à la revue électronique par rapport à son équivalent actuel imprimé.

Par ailleurs, le site savant — souvent production individuelle — n'a pas sa place dans l'imprimé [3].

1. Ce qui n'est pas nécessaire si les moteurs de recherche les plus connus indexent régulièrement le site : un outil comme Google est aussi un méta-index de toutes les revues en ligne.

2. On remarquera que l'unité de référence n'est plus le volume, mais l'article.

3. A l'exception peut-être des compilations d'articles d'un auteur en fin de carrière. Pour un exemple de site savant, cf. http://www.chass.utoronto.ca/~wulfric/

3.2 Confiance et piratage

Reste à s'assurer de la qualité des documents électroniques. On pourrait déjà répondre que cela ne concerne pas l'auteur, mais le lecteur. Sinon, pour organiser cette confiance, il suffit de reprendre les méthodes éprouvées par l'expérience : comité éditorial, scientifique, ou de lecture, par exemple. De façon générale, les constructions sociales extérieures à l'internet sont suffisantes : on l'a vu avec la publication de la conférence inauguraule de Roger Guesnerie au Collège de France[4]. Aussi, les questions habituelles sur la qualité de la confiance à accorder à un document ou à un auteur inconnu ne tiennent pas quand elles sont émises par des universitaires : ces derniers savent très bien que ce n'est pas l'imprimé qui certifie, mais son contenu (sinon le plus mauvais magazine de mode aurait autant d'importance qu'une revue scientifique). Mais l'expression de ces questions, qui discréditent en termes moraux une technique pour mieux nier le caractère technique de l'écriture, permet aussi de ne pas s'interroger sur la notion de qualité dans la production d'articles scientifiques : des économistes et des mathématiciens avouent que le tiers des articles de leurs disciplines sont entachés d'erreurs, allant de la bévue involontaire (mais parfois conséquente) à la falsification délibérée.

Reste, intimement liée à la question de la confiance positive, son pendant négatif : la peur de se faire piller. Celle-ci est en partie légitime : par exemple, ce travail ne sera pas publié sur le web avant sa validation par un jury et sa certification finale par l'institution. Mais, comme pour un article imprimé, la propriété intellectuelle de sa variante électronique tient à sa citation : il suffit que suffisamment de personnes puissent témoigner de l'antériorité d'une idée, d'une démonstration, pour que — dans un contexte à définir[5] — son auteur soit reconnu. Cette peur du pillage oublie par ailleurs deux réalités, l'une historique, l'autre sociologique. Commençons par la seconde : comment se fait-il que la majorité des mathématiciens, physiciens, etc. publient directe-

4. Cf. page 173.

5. Par exemple, la légitimité d'un inventeur n'est souvent que nationale, en vertu des logiques nationalistes de chaque pays. Cf. encore [Edg98].

ment leurs articles sur le serveur de Los Alamos sans validation ? La première raison est que ce serveur certifie à sa manière leur propriété intellectuelle. La seconde, moins connue, est qu'il vaut mieux que l'article soit d'une qualité scientifique irréprochable : sinon, ce ne seront pas deux ou trois relecteurs d'une revue qui seront au fait des défauts de l'article, mais des centaines de spécialistes, qui ne se priveront pas de conspuer l'auteur. Ici encore, la vérité scientifique profite au mieux de la logique de la concurrence inhérente au monde scientifique [6].

Reste à savoir si l'auteur veut pénétrer dans le marché de l'édition ou dans celui de la science : comme on peut le lire dans l'ouvrage précité, on ne dépose pas de brevet pour une découverte si on espère obtenir le prix Nobel [7]. Tout d'abord, un chercheur ne reçoit pas de droits d'auteur quand son article est publié. Il en est de même pour de nombreux auteurs d'ouvrage [8]. Parce qu'ils n'ont pas réussi à pénétrer le marché marchand de l'édition — publication de manuels, accès à des éditeurs disposant d'un solide réseau de diffusion —, ou parce qu'ils ne le désirent pas. A ce titre, de multiples auteurs du web — incluant de nombreux élèves des Écoles normales supérieures [9], mais aussi des professeurs du supérieur ou du second degré — s'engagent dans la publication sans en attendre des bénéfices financiers. Par désir de rendre ce qu'on a

6. [Bou01], p. 162.

7. [Bou01], p. 105.

8. Les éditions de l'Harmattan donnent un très bon exemple de ce type de situation.

9. Cf. les exercices corrigés de mathématiques réalisés par les élèves de l'ENS Lyon. De même, à l'ENS Ulm, peu importait aux premiers normaliens qui ont publié sur le web que leurs pages aient un statut d'article validé par les pairs, ou qu'il leur facilitât une promotion. La plupart étaient jeunes (rappelons-nous que le premier serveur web de l'ENS est celui des élèves), séduits par l'hypertextualité, par l'expérience de mise en réseau, par la construction d'un gisement d'informations de tous types, et souvent, l'humilité ou l'humour étaient de rigueur. Voir les expériences de double traduction automatique des poèmes de la Fontaine par Henri Desbois (hélas disparues suite à la panne du disque dur du serveur mercator.ens.fr, mais aisément reproductibles), les outils proposés sur le site Nouvelle Gallica Judaica (Cf. note 31 de la page 148), ou les premiers romans d'élèves (http://www.eleves.ens.fr :8080/home/coulmont/mac/meurtre.html).

reçu [10], on entre là dans une logique de contre-don, assimilable au « potlatch ». La pratique peut aussi être militante : désir de valoriser une discipline, un thème de recherche, de fédérer une dynamique. De telles motivationss sont évidemment à croiser avec le plaisir personnel de se faire connaître. Enfin, plusieurs auteurs, avant tout attachés à de sobres publications scientifiques, tirent beaucoup de satisfaction lorsque leurs pages web génèrent un effet retour [11].

Reste une autre forme de plaisir liée à la publication, peut-être trop peu évoquée au sujet de l'internet universitaire : celui de la création. Il ne s'agit pas de sombrer dans l'idéologie du « créateur » (d'entreprise, d'affiche publicitaire, etc.), souvent associée au « créatif » [12], mais de rappeler les formes de satisfaction liées à une production peu ou prou artistique, comme l'écriture d'une nouvelle, le tirage d'une photographie, l'interprétation d'une sonate, la réalisation d'un logiciel. A ce titre, la réalisation de pages web, simples ou sophistiquées, relève parfois d'un plaisir de conception d'une œuvre.

Quant à la référence à l'histoire, elle incite à se rappeler que la sensibilité au piratage varie suivant les époques : Jean-Pierre Vittu rappelle que l'*Encyclopédie* de Diderot était une « contrefaçon dans ses planches des manuscrits de la *Description des Arts et Métiers* et des mémoires techniques conservés par l'Académie des sciences » [13] Et même dans la situation la plus marchande, les avantages scientifiques et économiques du pillage peuvent être importants : Le même auteur affirme « qu'au XIX^{e} siècle, les exportations de livres français ne progressèrent que sur des marchés déjà ouverts par les pirates ». Ainsi, l'appel à la propriété intellectuelle masque-t-il avant tout une

10. Par exemple suite à la découverte d'un logiciel gratuit et aisé d'emploi pour écrire en japonais...

11. Même anonyme : par exemple, quand un inconnu leur adresse un message électronique leur demandant un complément d'information, une référence bibliographique, des conseils, voire une base de données entière, ou tout simplement les remercie pour leur contribution.

12. Le Petit Robert en donne la définition suivante : « créateur dans le domaine commercial ».

13. Pour cette référence et la suivante, cf. [Vit90].

volonté de ne rien faire ; ou du moins de ne rien faire qui fût gratuit, sur le plan financier comme symbolique, en acceptant une logique économique qui, parfois, est tellement désuète qu'elle ne s'applique plus, et donc dessert celui qui l'adopte.

3.3 Publication en 3 temps

Les potentialités de l'édition électronique renvoient une fois de plus à un savoir-faire qui n'est pas toujours à la portée de tous. Nous rappelons ici les trois temps de l'évolution de la publication chez les auteurs des deux enquêtes. Ce point a surtout un intérêt historique, dans la mesure où se multiplient les logiciels « clé en main » d'édition de revues savantes électroniques, et dans la mesure où l'on va peut-être, dans les prochaines années, assister à des spécialisations professionnelles comme celles que l'on rencontre dans l'imprimé. D'autres lectures de l'évolution des sites web peuvent être faites, mais celle-ci met en évidence le lien entre développement technique et questionnements sur les pratiques de recherche.

- La première phase, souvent assez timide, consiste à reproduire sur un serveur des documents qui auraient leur place dans un produit imprimé : présentation d'un département, d'une équipe de recherche, annonces et comptes rendus de séminaires, de débats, et même liste de pointeurs ou reproduction d'une revue imprimée existante vont dans ce sens : indépendamment du projet scientifique (article d'humeur ou réelle revue savante), le document mis à disposition du lecteur internaute est imprimable. Certes, de nombreux renvois hypertextuels jalonnent de tels documents, comme c'est le cas des listes de pointeurs, l'insertion d'images ou d'icônes est possible, mais la structure traditionnelle du texte n'est pas fondamentalement remise en cause, à l'exception de la « mise en réseau » rapide des documents.
- Un accroissement des compétences techniques ou de la connaissance des pratiques sur le web incite alors les auteurs à offrir plus de liberté

au lecteur, en lui proposant notamment des ressources interrogeables : base bibliographique ou catalogue de bibliothèque en constituent le premier type d'exemple. Plutôt que de télécharger un fichier conséquent (qu'on ne désire pas nécessairement « offrir » au premier venu), le lecteur pourra sélectionner des « fiches » correspondant à des requêtes données. Ce type de proposition peut s'accorder avec la possibilité de réaliser une recherche « plein-texte » sur tout ou partie du serveur. Techniquement, ce type d'outils fait appel à des compétences spécifiques en matière d'analyse textuelle : cela renvoie à la programmation attachée à la recherche de mots, d'expressions. Longtemps, le langage *Perl* et les *cgi-bin* associés furent utilisés. Aujourd'hui, le succès du langage *Php* tend à remplacer ce premier outil. Déjà, le rapport au texte traditionnel se complexifie : d'une part, celui-ci devient base de données, puisque des extraits du texte sont proposés à la demande de l'utilisateur.

D'autre part, ces possibilités d'indexation totale — longtemps dévalorisées par les littéraires, et qui, avec le développement des moteurs de recherche, connaissent un grand succès dans le monde industriel — remettent en cause les notions d'index traditionnels : tout mot est potentiellement indexé, et toute forme d'index est envisageable (ce qui incite à en multiplier les types). A ce sujet, nous ne pouvons résister au plaisir de citer Gérard Genette quand il présente son index des noms[14] : « Avec le lot habituel d'erreurs et d'omissions, cet index renvoie aux occurrences effectives des noms d'auteurs et à leurs occurrences implicites par mentions de titres. Un peu plus utile aurait été un index des titres (parfois plusieurs par œuvre) avec indication des noms (même remarque) et des dates (idem), mais on m'assure qu'il aurait été plus long que le livre. Tel qu'il est, comme de la plupart, sa véritable fonction est d'éviter à l'auteur la marque infamante : *no index.* ». On assiste alors à une première rupture avec tout un lot de pratiques imposée à l'auteur.

14. [Gen87], p. 381.

À ce stade d'« offre de services », les responsables du site prennent conscience des avantages de la publication électronique, notamment en termes de coût. Le serveur finit par être un outil qui centralise les travaux d'un groupe de chercheurs, et à cette publication sont souvent associés d'autres outils de communication : lettre d'information ou forum (liste de discussion) en sont les exemples les plus connus. Ainsi, la publication proprement dite se complète par la mise en place d'un réseau[15] ouvert ou semi-fermé[16]. C'est aussi le moment où les auteurs mesurent l'audience des pages et dossiers (avec les fichiers *access_log*). Cette phase est aussi celle d'une découverte, plus douloureuse : la constitution d'un réseau social de chercheurs nécessite des mises à jour fréquentes (annonce du prochain séminaire, publication de nouveaux documents, mais aussi vérification de la permanence des liens proposés). Ainsi, l'activité de gestion prend de plus en plus d'importance et renvoie à une organisation de plus en plus automatisable du serveur, tout en imposant de multiples vérifications manuelles. C'est à ce moment que le site trouve les moyens humains et financiers pour survivre, ou meurt.

- La troisième phase est celle de la rupture totale avec l'imprimé. Elle ne se manifeste pas tant par l'introduction de documents multimédia (il est aisé d'insérer un enregistrement audio ou vidéo dans une page web) que par le mode de production des divers formats de texte ou d'image : celui-ci peut devenir très automatisé, suite à un travail de programmation qui peut conférer aux outils utilisés le statut de réels logiciels et qui donne encore plus de liberté à l'utilisateur-lecteur. Certes, on souscrit peut-être alors aux sirènes de l'interactivité. Mais de tels objets sont assurément exclus de toute forme d'ouvrage imprimé, et certains,

15. À ce stade, différent du premier (on publie sur le web comme on jette une bouteille à la mer, espérant une réponse ou réaction), les personnes mise en réseau le sont suite à un accord explicite de leur part.

16. Par exemple, en proposant à un comité éditorial de consulter pour évaluation des articles proposés, qui sont déposés dans un dossier accessible par mot de passe.

comme les cartes animées, peuvent avoir un caractère fortement pédagogique. Dans ce même ordre d'idées, divers projets en cours incitent à mettre à disposition sur le web des logiciels complets, de statistiques, de cartographie, d'analyse textuelle, de travail collaboratif : plutôt que de lancer une application de ce type, à condition qu'elle réside sur son disque dur, on transmet par le réseau un fichier de données afin que la machine distante en organise le traitement et en renvoie le résultat. On est renvoyé dans de tels cas à la fabrication d'outils *ad hoc*, qui facilitent la recherche, et qui renvoient des résultats qui parfois, ne s'insérent plus du tout dans un imprimé.

Ainsi, un site web élaboré conteste les formes d'organisation de la publication aux marges du texte noble : en bousculant la notion d'index, si utile mais si fastidieux à réaliser, en transformant le rapport au texte, traduit comme une vulgaire base de données, en faisant de même pour la bibliographie, en rapprochant le document érudit de la somme des outils et méthodes si rarement enseignés qui en ont facilité la conception. L'internet met en valeur de nouvelles possibilités techniques, proprement informatiques, qui réhabilitent les aspects obscurs, non-dits, de la production scientifique que sont les outils et méthodes.

3.4 Le questionnaire

Voici la présentation du questionnaire correspondant à l'enquête de la troisième partie et la série des questions qu'il comportait.

Madame, Monsieur

Dans le cadre de ma thèse, j'étudie l'incidence d'Internet sur les pratiques des chercheurs en sciences humaines. Connaissant vos activités pionnières en matière de publication sur le Web, de fédération de listes de discussion et/ou de réflexion sur l'orientation de la recherche, j'aimerais que vos preniez le temps de répondre à ce bref questionnaire. Je m'intéresse particulièrement *à la genèse de vos activités, aux raisons de vos choix, au soutien dont vous disposiez alors, aux résistances que vous avez rencontrées ; je compte aussi sur vous pour faire un bilan de votre expérience et pour décrire comment celle-ci a modifié vos intuitions ou convictions d'alors concernant l'organisation de la recherche.*

Il n'y a pas de contrainte à la longueur de vos réponses.

Ce questionnaire est adressé individuellement à une trentaine de personnes ; il est tout à fait confidentiel et si certains de vos commentaires seront cités dans ma thèse, ce sera sous couvert de l'anonymat. Les informations personnelles que je vous demande (âge, sexe, statut) seront agrégées et ne permettront donc pas de rompre cet anonymat. Je prie les personnes que j'ai l'habitude de tutoyer d'excuser le caractère formel de ce questionnaire et je souhaite à tous une excellente fin d'été.

Je vous remercie de votre attention.

3.4.1 Historique et programme

Ia. En quelle année avez-vous découvert Internet (sous sa forme la plus large : mail, news, Web, etc.) ?

Ib. En quelle année avez-vous décidé de vous impliquer personnellement sur Internet ?

Ic. Sous quelle forme était-ce ? (forum, pages Web, etc.)

Id. Quelles étaient vos motivations ? (Demande hiérarchique, désir de dynamiser votre communauté scientifique, incitation venant d'amis ou collègues, etc.)

Ie. Votre démarche était-elle individuelle ou collective ?

3.4.2 Votre culture technique

IIa. Votre culture informatique vous semblait-elle un gage de succès de votre entreprise ?

IIb. Sinon, avez-vous ressenti le besoin de compléter cette culture technique, voire de l'acquérir entièrement pour mener à bien votre projet ?

3.4.3 Financement du projet

IIIa. De quelles ressources financières ou techniques disposiez-vous au départ ?

IIIb. Avez-vous été aidé(e) par votre institution dès le début de votre projet ? Si oui, sous quelle forme ? (budget de recherche, vacations, cession de machines, mise à disposition d'informaticiens etc.)

IIIc. Sinon, avez-vous été aidé(e) par une autre institution ? Si oui, sous quelle forme ?

IIId. Avez-vous dû faire appel à votre budget personnel pour mener à bien vos activités ? (achat de machines, logiciels, modems, etc.). Si oui, pouvez-vous estimer le montant de vos dépenses annuelles ?

IIIe. À combien évaluez-vous le temps passé à cette activité ? Ce temps était-il pris sur vos heures de travail ou sur vos loisirs ?

3.4.4 État de l'art et réception du projet

IVa. Quel était alors l'état des contenus sur Internet relativement à votre discipline ? (inconsistant, excellent, etc. Vous pouvez nuancer suivant les pays ou institutions)

IVb. Quel était l'état d'esprit de vos collègues (au sens large) vis-à-vis d'Internet ? (Toute classification est bienvenue)

IVc. Et vis-à-vis de votre initiative ?

IVd. Si vos collègues se désintéressaient de vos activités ou s'ils les critiquaient, à quoi, selon vous, était-ce dû ?

IVe. Vos activités vous ont-elles nui professionnellement ? (promotion retardée, mise à l'écart, etc.)

IVf. Au contraire, en avez-vous tiré rapidement des bénéfices ? Si oui lesquels ? (sur les plans financier et symbolique, professionnel comme extra-professionnel)

3.4.5 Aujourd'hui

Va. Quelles sont aujourd'hui vos activités liées à Internet ? Sont-elles dans le prolongement direct de votre initiative ?

Vb. Votre institution ou d'autres (françaises ou étrangères) soutiennent-elles aujourd'hui votre projet ? (Si oui, depuis quand et quelles sont les modalités de l'aide ?).

Vc. Au contraire, si votre institution émet des résistances, quelles sont-elles ?

Vd. De nouveaux collègues se sont-ils associés à votre projet ? Ou, êtes-vous moins nombreux qu'auparavant ?

Ve. Avez-vous récement ressenti le besoin de mettre à jour votre culture informatique ? Si oui, sous quelle forme ? Sinon, pourquoi ?

Vf. Comment évaluez-vous aujourd'hui les coût et bénéfices de votre investissement ? (sur les plans symboliques, financiers, intellectuels...)

Vg. Comment jugez-vous aujourd'hui le contenu du Web dans votre discipline ?

Vh. Que pensent aujourd'hui vos collègues de votre projet initial et de son évolution ?

Vi. Pensez-vous que leurs représentations d'Internet se sont modifiées depuis votre initiative ?

3.4.6 Bilan et projets futurs

VIa. Avec le recul, quel bilan feriez-vous de cette période comprise entre votre investissement premier et aujourd'hui ?

VIb. Quels sont vos projets actuels ?

VIc. À votre avis, dans quelle mesure Internet a-t-il modifié les logiques et les thèmes de recherche ? (recherche documentaire, diffusion d'outils et de méthodes informatiques, pratiques éditoriales, travail collaboratif et fédération d'équipes, internationalisation de la recherche, etc.)

3.4.7 Compléments

Tous commentaires sur votre expérience personnelle, votre culture initiale en informatique, sur l'évolution de votre milieu (collègues, étudiants, décideurs, ministère, etc.), sur la comparaison entre celui-ci et d'autres (français, étrangers) sont bienvenus, anecdotes comprises.

3.4.8 Informations personnelles

votre âge :
votre sexe :
votre statut :
votre institution :
votre ville :

3.4.9 Elargissement de l'enquête

Si vous pensez que les témoignages de collègues ou d'amis me seront précieux pour comprendre le plus finement possible les transformations des pratiques et des enjeux au sein des sciences humaines, n'hésitez pas à me transmettre leurs coordonnées. Je les contacterai avec plaisir.

D'avance, encore merci

Éric Guichard, etc.

Chapitre 4

Requêtes de Goosta

4.1 Les mille premières requêtes

Entre parenthèses, sont rappelés le rang et le nombre total d'occurrences. Procédure **Nett2**.

sexe (1, 284462), immobilier (2, 93755), sex (3, 91962), chat (4, 84861), mp3 (5, 79200), meteo (6, 77851), emploi (7, 74660), caramail (8, 73966), horoscope (9, 73071), anpe (10, 70122), sexe gratuit (11, 69967), sncf (12, 67406), porno (13, 63447), webcam (14, 60406), annuaire (15, 56386), voissa (16, 56268), yahoo (17, 55198), fond ecran (18, 51067), humour (19, 50307), pages jaunes (20, 49712), annuaire inverse (21, 44681), pokemon (22, 43856), gay (23, 43769), cinema (24, 43312), zoophilie (25, 42210), bourse (26, 41724), ecran veille (27, 41625), jeux (28, 41471), lycos (29, 40938), traduction (30, 39452), rencontre (31, 38918), sms (32, 37986), boursorama (33, 37981), impots (34, 37689), genealogie (35, 37353), charme (36, 37332), manga (37, 37185), alizee (38, 36943), mariage (39, 35515), voyage (40, 34342), cul (41, 33905), air france (42, 33163), fnac (43, 32818), musique (44, 32770), warez (45, 31887), tf1 (46, 31344), itineraire (47, 30832), hotmail (48, 30309), argus (49, 29913), photo (50, 29865), loto (51, 28851), moteur recherche (52, 28490), sfr (53, 28188), lingerie (54, 28176), napster (55, 28067), divx (56, 27922), crack (57, 27814), apec (58, 26754), britney spears (59, 26498), dictionnaire (60, 26113), moto (61, 26084), ratp (62, 26047), rencontres (63, 25908), degriftour (64, 25642), location (65, 25513), nouvelles frontieres (66, 25508), paris (67, 25496), voyages (68, 25382), wanadoo (69, 24369), gros seins (70, 24058), gites france (71, 23523), blagues (72, 23511), credit agricole (73, 23392), agence voyage (74, 23215), sms gratuit (75, 23171), minitel (76, 23105), radio (77, 22956), eminem (78, 22862), fonds ecran (79, 22853), corse (80, 22682), euro (81, 22662), maroc (82, 22552), lettre motivation (83, 22349), dvd (84, 22227), photos (85, 21939), echangisme (86, 21769), caisse epargne (87, 21194), sims (88, 21056), nokia (89, 20973), buffy (90, 20643), football (91, 20627), monde (92, 20516), automobile (93, 20457), journal officiel (94, 20230), stage (95, 20163), m6 (96, 20049), skyrock (97, 19963), sonnerie portable (98, 19931), gratuit (99, 19811), impot (100, 19755), hentai (101, 19605), carte france (102, 19519), video (103, 19491), peugeot (104, 19423), pamela anderson (105, 19413), msn search (106, 19399), france telecom (107, 19284), fellation (108, 19192), jeux video (109, 19035), mylene farmer (110, 18799), renault (111, 18590), michelin (112, 18514), astrologie (113, 18481), sex gratuit (114, 18257), voiture (115, 18177), pornographie (116, 18168), sodomie (117, 17679), location appartement (118, 17622), petites annonces (119, 17468), loterie (120, 17419), vacances (121, 17417), cuisine (122, 17297), test (123, 17130), paroles chansons (124, 17100), canal (125, 17021), banque (126, 16959), google (127, 16906), emulateur (128, 16687), bebe (129, 16547), cartes voeux (130, 16540), location voiture (131, 16517), telecharger (132, 16485), playstation (133, 16468), cannabis (134, 16347), erotisme (135, 16106), concours (136, 16045), laure sinclair (137, 16023), digimon (138, 16004), philosophie (139, 15861), tuning (140, 15836), kamasutra (141, 15739), adsl (142, 15723), clipart (143, 15660), cv (144, 15432), wallpaper (145, 15274), carte (146, 15208), pages blanches (147, 15135), societe generale (148, 15085), sexe amateur (149, 15084), xxx (150, 14993), poste (151, 14956), lolita (152, 14863), erotique (153, 14756), job (154, 14756), assurance (155, 14740), tarot (156, 14727), jennifer lopez (157, 14709), elections municipales (158, 14620), winzip (159, 14525), recettes cuisine (160, 14487), romeo juliette (161, 14289), grossesse (162, 14256), martinique (163, 14243), parisien (164, 14222), lyon (165, 14188), insee (166, 14148), dawson (167, 14057), espagne (168, 14051), harry potter (169, 13939), location vacances (170, 13899), images (171, 13895), aol (172, 13800), pmu (173, 13754), cartes (174, 13742), hotel (175, 13738), guadeloupe (176, 13695), mangas (177, 13670), encyclopedie (178, 13619), itineris (179, 13499), sonnerie (180, 13429), sport (181, 13428), france (182, 13380), plan paris (183, 13353), credit lyonnais (184, 13350), prenom (185, 13344), amour (186, 13308), carte voeux (187, 13307), marseille (188, 13284), tunisie (189, 13160), voyeur (190, 13132), ski (191, 13096), onisep (192, 13081), particulier particulier (193, 13081), javascript (194, 13054), fun radio (195, 13045), carrefour (196, 13013), race (197, 12955), gif (198, 12948), camping (199, 12925), escort girl (200, 12907), meteo france (201, 12900), seins (202, 12878), cartes virtuelles (203, 12774), toulouse (204, 12754), cracks (205, 12716), rugby (206, 12714), copernic (207, 12600), cheval (208, 12475), salope (209, 12430), publicite (210, 12397), club med (211, 12381), altavista (212, 12298), amateur (213, 12266), histoires erotiques (214, 12236), egypte (215, 12208), bondage (216, 12146), kiwee (217, 12104), garou (218, 12088), francaise jeux (219, 12045), credit (220, 12009), education nationale (221, 12007), diddl (222, 11980), ibazar (223, 11922), plan (224, 11919), masturbation (225, 11908), naturisme (226, 11905), voyance (227, 11819), palm (228, 11785), roms (229, 11771), recettes (230, 11743), universite (231, 11659), traducteur (232, 11594), code route (233, 11577), vendee globe (234, 11552), playboy (235, 11538), economiseur ecran (236, 11536), foot (237, 11449), prenoms (238, 11362), television (239, 11361), telechargement (240, 11355), sexy (241, 11333), bretagne (242, 11330), billet avion (243, 11322), algerie (244, 11302), stars nues (245, 11252), chien (246, 11213), clitoris (247, 11206), formation (248, 11164), interim (249, 11153), convention collective (250, 11143), x (251, 11133), web cam (252, 10990), europe (253, 10989), calendrier (254, 10899), encheres (255, 10873), bonjour (256, 10845), blague (257, 10841), logement (258, 10841), netclub (259, 10835), telephone (260, 10822), annonces (261, 10812), psg (262, 10755), centrale (263, 10681), italie (264, 10654), tchatche (265, 10611), charmed (266, 10568), france2 (267, 10528), presse (268, 10497), bmw (269, 10478), compagnie aerienne (270, 10451), gifs (271, 10397), stages (272, 10380), fessee (273, 10377), mappy (274, 10336), iti (275, 10334), reunion (276, 10286), voiture occasion (277, 10246), portable (278, 10234), carte anniversaire (279, 10233), elections (280, 10208), bricolage (281, 10200), peche (282, 10179), nrj (283, 10138), ma-donna (284, 10131), drivers (285, 10123), anniversaire (286, 10098), gifs animes (287, 10098), saint valentin (288, 10097), fakes (289, 10085), caf (290, 10030), redoute (291, 10019), screensaver (292, 10004), enneigement (293, 9968), coiffure (294, 9956), image (295, 9950), simpsons (296, 9949), freeware (297, 9917), sonneries (298, 9871), liberation (299, 9865), tatouage (300, 9854), msn (301, 9843), free (302, 9812), rap (303, 9765), roswell (304, 9747), femme (305, 9717), nantes (306, 9698), legifrance (307, 9695), recette cuisine (308, 9669), ouest france (309, 9662), figaro (310, 9604), louvre (311, 9604), jenna jameson (312, 9569), carnaval (313, 9544), sony (314, 9474), air liberte (315, 9466), porno gratuit (316, 9461), south park (317, 9430), partitions (318, 9403), chatte (319, 9356), lesbienne (320, 9341), recherche emploi (321, 9335), nue (322, 9334), om (323, 9310), telephone portable (324, 9309), crous (325, 9286), histoire (326, 9258), darty (327, 9257), recrutement (328, 9245), bordeaux (329, 9243), madagascar (330, 9215), echos (331, 9178), informatique (332, 9177), driver (333, 9159), ferrari (334, 9159), canabis (335, 9151), sante (336, 9147), lesbiennes (337, 9135), pochette cd (338, 9101), equipe (339, 9095), agence immobiliere (340, 9083), casting (341, 9082), dess (342, 9071), libertysurf (343, 9040), theatre (344, 8996), modelisme (345, 8947), edf (346, 8941), sida (347, 8939), formule (348, 8938), page jaune (349, 8933), ministere finances (350, 8930), femmes nues (351, 8913), prefecture (352, 8874), animaux (353, 8842), moretv (354, 8838), tabatha cash (355, 8836), disney (356, 8818), bnp (357, 8791), cabinet recrutement (358, 8784), femmes mures (359, 8761), citroen (360, 8753), plan ville (361, 8751), bd (362, 8693), tomb raider (363, 8686), magie (364, 8680), celebrites nues (365, 8642), iup (366, 8637), venise (367, 8635), portugal (368, 8626), montpellier (369, 8612), icones (370, 8607), moliere (371, 8604), bite (372, 8597), cartes postales (373, 8553), friends (374, 8545), pariscope (375, 8527), offre emploi (376, 8508), simpson (377, 8504), asterix (378, 8502), gamez (379, 8468), microsoft (380, 8448), travesti (381, 8413), tennis (382, 8409), iut (383, 8393), partition (384, 8383), grece (385, 8372), film (386, 8342), droit travail (387, 8313), atlas (388, 8306), hard (389, 8288), echangiste (390, 8266), hacking (391, 8264), fnaim (392, 8253), pap (393, 8246), assurances (394, 8217), handball (395, 8195), irlande (396, 8190), canada (397, 8187), golf (398, 8127), hotels (399, 8119), oreka

(400, 8065), vache folle (401, 8064), offres emploi (402, 8018), assedic (403, 8009), stars (404, 7986), locations vacances (405, 7985), bouygues (406, 7965), securite sociale (407, 7955), iufm (408, 7901), appz (409, 7878), bananalotto (410, 7869), annonces immobilieres (411, 7849), office tourisme (412, 7838), multimania (413, 7837), londres (414, 7800), islam (415, 7786), droit (416, 7772), ile reunion (417, 7762), lolitas (418, 7762), fievre aphteuse (419, 7750), transexuel (420, 7716), tourisme (421, 7677), tablatures (422, 7641), cpam (423, 7583), avion (424, 7572), winamp (425, 7569), fleurs (426, 7567), gnutella (427, 7556), helene segara (428, 7530), ikea (429, 7522), penis (430, 7512), paroles (431, 7507), decoration (432, 7493), mode (433, 7476), ugc (434, 7448), zoophile (435, 7448), linux (436, 7418), gites (437, 7417), alyssa milano (438, 7414), salopes (439, 7404), allocine (440, 7400), eau (441, 7393), yamaha (442, 7391), histoire erotique (443, 7388), rennes (444, 7386), hp (445, 7349), vin (446, 7326), ile maurice (447, 7282), counter strike (448, 7257), peine mort (449, 7250), agences voyages (450, 7238), gite france (451, 7186), hotel paris (452, 7186), celine dion (453, 7177), auto (454, 7161), restaurant (455, 7160), mutuelle (456, 7155), dreamcast (457, 7148), gsm (458, 7144), audi (459, 7125), cuba (460, 7122), rome (461, 7112), jeu (462, 7106), journaux (463, 7094), wallpapers (464, 7088), regime (465, 7075), assurance auto (466, 7068), argus automobile (467, 7057), vtt (468, 7044), point croix (469, 7031), hacker (470, 7008), mairie paris (471, 7006), ministere (472, 6995), casta (473, 6951), recette (474, 6940), emulation (475, 6911), internet (476, 6892), sonneries portables (477, 6882), call girl (478, 6878), java (479, 6871), france3 (480, 6865), australie (481, 6859), flash (482, 6850), tahiti (483, 6848), escort (484, 6823), jardinage (485, 6798), provence (486, 6778), electromenager (487, 6777), matrix (488, 6773), carte routiere (489, 6765), basket (490, 6756), pamela (491, 6753), fonds ecrans (492, 6745), thalassotherapie (493, 6742), vagin (494, 6734), femmes (495, 6723), annuaire sexe (496, 6721), mail (497, 6711), qi (498, 6708), dessin anime (499, 6703), thalasso (500, 6697), sms gratuits (501, 6685), bernouilli (502, 6667), jaquette (503, 6667), lille (504, 6653), dolly golden (505, 6642), vvf (506, 6636), etajv (507, 6634), ntm (508, 6634), particulier (509, 6618), 35 heures (510, 6609), ejaculation (511, 6607), astalavista (512, 6591), futuroscope (513, 6582), bingopoly (514, 6577), collants

(515, 6574), ministere interieur (516, 6572), nice (517, 6568), pacs (518, 6564), banque france (519, 6555), ucpa (520, 6551), forum (521, 6540), bob marley (522, 6536), celebrites (523, 6525), electronique (524, 6525), karaoke (525, 6524), meubles (526, 6524), lara croft (527, 6523), senegal (528, 6516), municipales (529, 6511), strasbourg (530, 6510), final fantasy (531, 6504), mercedes (532, 6468), coloriage (533, 6467), sexyloo (534, 6466), auchan (535, 6444), giclouz (536, 6443), magazine (537, 6440), voila (538, 6435), grenoble (539, 6432), monster (540, 6431), craig david (541, 6426), medecine (542, 6426), faire part (543, 6421), bouygues telecom (544, 6418), alsace (545, 6382), decathlon (546, 6382), chine (547, 6370), julia channel (548, 6368), pacte loups (549, 6356), maison (550, 6349), sperme (551, 6341), fesses (552, 6338), nike (553, 6333), coquine (554, 6329), photographie (555, 6320), guide routard (556, 6311), tabac (557, 6303), gouvernement (558, 6294), tunning (559, 6286), itineraires (560, 6265), ecrans veille (561, 6260), kama sutra (562, 6244), journal (563, 6234), musculation (564, 6233), afp (565, 6231), job ete (566, 6207), culotte (567, 6204), poesie (568, 6195), alcatel (569, 6175), visual basic (570, 6174), surf (571, 6163), tele (572, 6151), final fantasy 9 (573, 6138), pochettes cd (574, 6126), aquarium (575, 6125), cned (576, 6120), dragon ball (577, 6114), string (578, 6112), icq (579, 6096), inde (580, 6092), nomade (581, 6092), u2 (582, 6091), rochelle (583, 6083), association (584, 6078), moteurs recherche (585, 6066), canal plus (586, 6060), agence voyages (587, 6039), skate (588, 6030), drogue (589, 6029), carte postale (590, 6028), referencement (591, 6019), urssaf (592, 5999), pipe (593, 5987), racisme (594, 5986), conseil general (595, 5976), fff (596, 5976), chiens (597, 5971), siemens (598, 5961), tintin (599, 5958), dessins animes (600, 5955), nu (601, 5950), sunset beach (602, 5949), sonneries portable (603, 5947), tablature (604, 5921), equitation (605, 5916), baldur s gate (606, 5906), ogm (607, 5904), hotel formule (608, 5902), banques (609, 5894), carte virtuelle (610, 5882), 3 suisses (611, 5880), cliparts (612, 5874), sous jupes (613, 5861), pissing (614, 5855), astronomie (615, 5853), bts (616, 5841), hack (617, 5840), programme tele (618, 5828), dhea (619, 5822), music (620, 5821), xxl (621, 5808), laetitia casta (622, 5787), midtown madness (623, 5777), jaquette cd (624, 5766), sagem (625, 5751), theme bureau (626, 5747), pierre vacances (627, 5742), assemblee nationale

(628, 5733), montagne (629, 5733), pompier (630, 5732), ile re (631, 5730), chocolat (632, 5727), guitare (633, 5722), psychologie (634, 5719), sakura (635, 5719), parole chanson (636, 5718), voyages degriffes (637, 5685), japon (638, 5670), sorcellerie (639, 5663), nom domaine (640, 5657), se loger (641, 5651), retraite (642, 5647), scato (643, 5644), ovidie (644, 5642), dragon ball z (645, 5625), scooter (646, 5625), gif anime (647, 5620), assurance moto (648, 5619), picasso (649, 5616), credit mutuel (650, 5606), netmeeting (651, 5592), julia chanel (652, 5586), billets avion (653, 5580), repondeur (654, 5580), anal (655, 5575), rire (656, 5570), afpa (657, 5565), conforama (658, 5564), anastacia (659, 5551), gays (660, 5550), f1 (661, 5536), kompass (662, 5524), sexualite (663, 5514), rtl (664, 5508), html (665, 5502), parfum (666, 5486), brigitte lahaie (667, 5485), screen saver (668, 5477), election (669, 5476), police (670, 5476), broderie (671, 5468), citations (672, 5468), rigoler (673, 5461), emplois (674, 5459), new york (675, 5451), chasse (676, 5434), voeux (677, 5424), carmen electra (678, 5409), test qi (679, 5409), sonnerie nokia (680, 5403), salaire (681, 5402), sylvia saint (682, 5402), dessin (683, 5392), lettre demission (684, 5383), chambre commerce (685, 5360), mature (686, 5359), turquie (687, 5351), telegramme (688, 5346), fram (689, 5343), zelda (690, 5342), allemagne (691, 5327), 4x4 (692, 5317), poeme (693, 5317), vietnam (694, 5308), casino (695, 5304), lingerie sexy (696, 5297), message repondeur (697, 5294), tv (698, 5293), annecy (699, 5289), greta (700, 5288), code travail (701, 5285), programme tv (702, 5284), meteo neiges (703, 5280), decodeur canal (704, 5269), fr3 (705, 5269), cci (706, 5268), etudiant (707, 5263), euridile (708, 5256), taux change (709, 5249), vanessa paradis (710, 5228), neige (711, 5227), election municipale (712, 5222), notaire (713, 5220), taxi (714, 5214), calcul impots (715, 5213), argus auto (716, 5212), annuaire international (717, 5208), moniteur (718, 5207), php (719, 5205), revebebe (720, 5194), divorce (721, 5181), angers (722, 5158), sarah michelle gellar (723, 5157), asp (724, 5156), location voitures (725, 5142), jeux videos (726, 5141), livres (727, 5136), femme nue (728, 5135), cadremploi (729, 5129), teens (730, 5127), voila mail (731, 5117), danse (732, 5116), magie blanche (733, 5113), coca cola (734, 5110), pollution (735, 5108), roller (736, 5107), barcelone (737, 5101), citegay (738, 5100), c (739, 5094), aufeminin (740, 5091), fonction publique

(741, 5074), massage (742, 5073), honda (743, 5072), themes bureau (744, 5070), casio (745, 5068), teletravail (746, 5067), photo sexe (747, 5066), nouvelle frontiere (748, 5061), beaute (749, 5030), free sex (750, 5030), fetichisme (751, 5024), france 3 (752, 5022), aubade (753, 5019), logitelnet (754, 5018), banque populaire (755, 5017), olympique marseille (756, 5012), logo (757, 5007), partouze (758, 5004), upskirt (759, 4997), voitures (760, 4995), alize (761, 4991), appartement (762, 4991), credit immobilier (763, 4991), geographie (764, 4991), judo (765, 4984), education (766, 4980), maitresse (767, 4977), calcul impot (768, 4976), porsche (769, 4974), st valentin (770, 4973), piercing (771, 4967), violence (772, 4964), mobile (773, 4961), chaussures (774, 4960), creation entreprise (775, 4951), leclerc (776, 4944), libertin (777, 4941), annuaire telephonique (778, 4940), mexique (779, 4934), velo (780, 4920), cnam (781, 4910), quebec (782, 4899), adoption (783, 4892), express (784, 4892), moyen age (785, 4890), surcouf (786, 4887), itineraire routier (787, 4874), camping car (788, 4872), betisier (789, 4867), aeroport (790, 4866), agriculture (791, 4863), maupassant (792, 4855), dauphin (793, 4842), annonce (794, 4841), plongee (795, 4839), star (796, 4837), locations (797, 4828), musee (798, 4823), peugeot 307 (799, 4816), plagne (800, 4808), castorama (801, 4801), piscine (802, 4801), hewlett packard (803, 4796), ophelie winter (804, 4796), orientation (805, 4789), offspring (806, 4784), bac francais (807, 4783), france inter (808, 4775), chats (809, 4753), volkswagen (810, 4753), bateau (811, 4751), lemel (812, 4751), virus (813, 4751), techno (814, 4742), voix nord (815, 4737), amatrice (816, 4735), phototheque (817, 4734), auto plus (818, 4729), animation (819, 4726), concours administratifs (820, 4720), location maison (821, 4719), diablo (822, 4703), sonnerie telephone (823, 4701), webcams (824, 4696), effet serre (825, 4694), marketing (826, 4692), mairie (827, 4687), dietetique (828, 4683), internet gratuit (829, 4677), midi (830, 4675), autoplus (831, 4667), salon (832, 4657), volcan (833, 4654), angleterre (834, 4650), faire-part (835, 4650), universite paris (836, 4646), cartographie (837, 4640), gites ruraux (838, 4634), real player (839, 4626), pute (840, 4615), feminin (841, 4606), concert (842, 4601), nues (843, 4598), karen lancaume (844, 4595), generateur mobicarte (845, 4591), robe mariee (846, 4589), belgique (847, 4588), mpeg (848, 4586), gendarmerie (849, 4574), comptabilite (850, 4573), location saisonniere (851, 4567), rouen (852, 4567), travail domicile (853, 4565), ericsson (854, 4553), aom (855, 4549), architecture (856, 4549), location immobiliere (857, 4541), telephone mobile (858, 4530), victor hugo (859, 4515), yahoo.fr (860, 4514), jeux concours (861, 4510), logiciel (862, 4504), nancy (863, 4504), home cinema (864, 4502), chevaux (865, 4498), versailles (866, 4490), thailande (867, 4483), centrale particuliers (868, 4482), newlook (869, 4477), paris match (870, 4475), discotheque (871, 4474), budweiser (872, 4469), filles nues (873, 4468), pub (874, 4467), karting (875, 4461), compagnies aeriennes (876, 4460), philips (877, 4455), nero (878, 4454), goa (879, 4450), cyclisme (880, 4441), laure sainclair (881, 4440), environnement (882, 4437), curriculum vitae (883, 4426), guide michelin (884, 4422), hot (885, 4410), amsterdam (886, 4402), enchere (887, 4394), ssii (888, 4393), rallye (889, 4391), mp3 gratuit (890, 4388), roms snes (891, 4380), rotten (892, 4379), electricite (893, 4365), mir (894, 4365), bac (895, 4360), tryo (896, 4358), tabata cash (897, 4354), abcoeur (898, 4345), annu (899, 4328), union europeenne (900, 4322), cartes anniversaire (901, 4320), celebrite (902, 4317), infirmiere (903, 4314), nasa (904, 4311), mariah carey (905, 4309), peinture (906, 4309), arbre genealogique (907, 4302), brad pitt (908, 4302), bigard (909, 4301), loterie gratuite (910, 4299), nudisme (911, 4298), voyeurisme (912, 4295), bresil (913, 4293), icone (914, 4293), location immobilier (915, 4293), papier peint (916, 4293), wap (917, 4292), seisme (918, 4285), telerama (919, 4280), clonecd (920, 4279), formule1 (921, 4272), zola (922, 4270), ford (923, 4264), stade france (924, 4264), playmate (925, 4260), warhammer (926, 4258), sonneries nokia (927, 4252), paroles chanson (928, 4247), chamonix (929, 4246), snowboard (930, 4243), bijoux (931, 4242), art (932, 4239), spartateur (933, 4228), vache (934, 4227), champagne (935, 4224), dix commandements (936, 4223), verite si je mens (937, 4223), camif (938, 4221), ardeche (939, 4218), tierce (940, 4216), turf (941, 4207), penetration (942, 4201), satellite (943, 4191), nudite (944, 4189), filles (945, 4179), telephonie (946, 4178), epson (947, 4174), maeva (948, 4174), baise (949, 4169), inspection travail (950, 4156), republique dominicaine (951, 4154), sexe amateur gratuit (952, 4153), promovacances (953, 4152), lyrics (954, 4148), beatles (955, 4146), serie tv (956, 4141), melodie portable (957, 4136), cirque (958, 4135), films (959, 4133), tennis table (960, 4132), communication (961, 4125), aquariophilie (962, 4118), hommes nus (963, 4118), enfants (964, 4114), station ski (965, 4114), armee (966, 4113), hebergement (967, 4113), evangelion (968, 4112), jaquettes (969, 4112), location appartement paris (970, 4108), dido (971, 4106), enfoires (972, 4099), annuaires (973, 4097), dopage (974, 4097), suisse (975, 4095), mathematiques (976, 4089), meteo neige (977, 4089), transport (978, 4083), bmx (979, 4078), fake (980, 4078), napoleon (981, 4073), securite routiere (982, 4073), more tv (983, 4071), clemence arnaud (984, 4067), antivirus (985, 4066), carte geographique (986, 4066), videos (987, 4066), cnfpt (988, 4064), voitures occasion (989, 4063), emulateurs (990, 4062), cours anglais (991, 4061), restaurants (992, 4059), mariage figaro (993, 4055), poemes (994, 4055), liberty surf (995, 4054), franchise (996, 4053), bleem (997, 4051), gaumont (998, 4051), bdsm (999, 4043), dijon (1000, 4042).

4.2 Les mille premiers mots

Entre parenthèses, sont rappelés le rang et le nombre total d'occurrences. Procédure **Nett2**.

sexe (1, 688278), france (2, 576184), paris (3, 529294), gratuit (4, 487042), location (5, 419503), photo (6, 393489), photos (7, 369153), carte (8, 330797), jeux (9, 316871), video (10, 306657), emploi (11, 277851), immobilier (12, 269857), sex (13, 260077), hotel (14, 238580), mp3 (15, 236779), annuaire (16, 217983), vente (17, 215247), ecran (18, 202996), musique (19, 196277), nue (20, 193140), ecole (21, 190641), saint (22, 183092), cours (23, 182182), informatique (24, 181966), site (25, 175418), plan (26, 174622), formation (27, 172863), internet (28, 172215), histoire (29, 167693), gay (30, 163923), lyon (31, 162932), porno (32, 162539), image (33, 162532), telecharger (34, 161895), voiture (35, 157773), voyage (36, 154363), portable (37, 153957), crack (38, 153685), vacances (39, 153364), club (40, 149946), dans (41, 148828), agence (42, 148476), maison (43, 148145), chat (44, 147349), francais (45, 146719), meteo (46, 144207), cinema (47, 142497), cartes (48, 140755), tourisme (49, 140612), cd (50, 139673), femme (51, 138283), femmes (52, 138168), annonces (53, 137015), recherche (54, 136476), logiciel (55, 134390), moto (56, 130615), code (57, 127558), amateur (58, 123424), 2001 (59, 123398), automobile (60, 123383), dvd (61, 122864), webcam (62, 122592), free (63, 122196), credit (64, 118998), concours (65, 117719), telephone (66, 116904), occasion (67, 116528), images (68, 115158), universite (69, 114114), droit (70, 113994), auto (71, 113971), travail (72, 112678), mariage (73, 111323), sonnerie (74, 110280), 2000 (75, 108572), entreprise (76, 107447), francaise (77, 106289), sport (78, 106137), journal (79, 105952), nues (80, 105302), nokia (81, 104261), air (82, 103953), rencontre (83, 102115), toulouse (84, 102097), radio (85, 100313), film (86, 99397), centre (87, 99380), association (88, 99271), sms (89, 98653), cuisine (90, 98290), jeu (91, 98240), monde (92, 97500), moteur (93, 97284), s (94, 95773), fond (95, 94848), web (96, 94278), ile (97, 94172), achat (98, 93861), voyages (99, 93532), une (100, 92056), pages (101, 91627), ski (102, 90568), lycee (103, 89655), marseille (104, 89091), societe (105, 88863), jean (106, 88279), pokemon (107, 87338), humour (108, 86878), ville (109, 86427), horoscope (110, 86146), presse (111, 85997), sncf (112, 85628), stage (113, 85201), enfants (114, 85074), pc (115, 84671), anpe (116, 83937), prix (117, 83923), commerce (118, 83843), art (119, 83580), camping (120, 83401), conseil (121, 83182), corse (122, 82903), tv (123, 82546), paroles (124, 82141), lettre (125, 81945), assurance (126, 81832), magazine (127, 81721), st (128, 81569), bourse (129, 81101), eau (130, 81011), caramail (131, 80993), gestion (132, 80025), banque (133, 79428), windows (134, 79427), elections (135, 79345), football (136, 78983), ministere (137, 78866), manga (138, 78722), appartement (139, 78427), impots (140, 77248), sonneries (141, 76702), canal (142, 76554), enfant (143, 76344), bts (144, 76076), test (145, 75981), recette (146, 75801), traduction (147, 75772), warez (148, 75377), cul (149, 75334), bordeaux (150, 75151), 3 (151, 74406), sexy (152, 74034), bac (153, 73830), espagne (154, 73535), telechargement (155, 73451), and (156, 73042), maroc (157, 72191), veille (158, 70885), seins (159, 70567), salon (160, 70399), restaurant (161, 70043), electronique (162, 69931), ligne (163, 69713), loi (164, 69345), dictionnaire (165, 69329), telecom (166, 68658), particulier (167, 68557), bretagne (168, 67993), avion (169, 67428), roms (170, 67376), locations (171, 67224), office (172, 67017), vie (173, 66880), securite (174, 66595), charme (175, 66498), aux (176, 66451), amour (177, 65985), driver (178, 65970), guide (179, 65865), municipales (180, 65837), hotels (181, 65570), annonce (182, 65215), erotique (183, 65154), dessin (184, 65002), bois (185, 64983), rencontres (186, 64448), anglais (187, 64427), pierre (188, 64168), education (189, 64003), lingerie (190, 63700), nantes (191, 63566), agricole (192, 63288), materiel (193, 63211), publicite (194, 63074), genealogie (195, 62564), homme (196, 62561), gros (197, 62521), perso (198, 62112), gratuites (199, 61934), son (200, 61615), c (201, 61600), videos (202, 61297), page (203, 61162), officiel (204, 61026), peugeot (205, 60493), theatre (206, 60437), marche (207, 60413), feature (208, 60390), gratuite (209, 60339), immobiliere (210, 60295), argus (211, 60187), yahoo (212, 59969), recettes (213, 59841), europe (214, 59476), gites (215, 59370), divx (216, 59313), renault (217, 58930), nationale (218, 58861), sante (219, 58737), playstation (220, 58524), cv (221, 58442), chambre (222, 58354), sites (223, 57636), creation (224, 57386), montpellier (225, 57283), voissa (226, 56923), logement (227, 56741), jeunes (228, 56425), chansons (229, 56056), euro (230, 55957), jaunes (231, 55866), emulateur (232, 55771), groupe (233, 55451), download (234, 55381), communication (235, 55200), technique (236, 55079), lille (237, 54611), mairie (238, 54266), voitures (239, 54205), fonds (240, 54146), nord (241, 54104), liste (242, 53715), aide (243, 53476), programme (244, 53326), stars (245, 53167), mobile (246, 53124), http (247, 53075), zoophilie (248, 52781), impot (249, 52583), reseau (250, 51970), calcul (251, 51831), sous (252, 51824), hard (253, 51503), television (254, 51443), chanson (255, 51387), tour (256, 51360), inverse (257, 51166), midi (258, 51150), mer (259, 51036), mac (260, 50689), plus (261, 50566), xxx (262, 50503), petites (263, 50354), livre (264, 50315), dess (265, 50173), alizee (266, 50087), chateau (267, 50077), itineraire (268, 50044), gsm (269, 49852), musee (270, 49826), nice (271, 49826), avec (272, 49825), filles (273, 49687), nu (274, 49373), star (275, 49093), voeux (276, 48997), etude (277, 48974), caisse (278, 48969), gratuits (279, 48623), recrutement (280, 48449), bebe (281, 48448), partition (282, 48297), britney (283, 48218), pro (284, 47837), transport (285, 47827), international (286, 47817), spears (287, 47794), comment (288, 47736), provence (289, 47465), foot (290, 47377), animaux (291, 47148), guerre (292, 47142), rennes (293, 46866), calendrier (294, 46835), magasin (295, 46823), chien (296, 46814), reunion (297, 46806), job (298, 46644), peche (299, 46538), linux (300, 46338), service (301, 46268), italie (302, 46146), serie (303, 46073), entreprises (304, 46069), anniversaire (305, 46057), grenoble (306, 46057), tele (307, 45884), cote (308, 45643), economie (309, 45391), belgique (310, 45340), sfr (311, 45296), codes (312, 45093), drivers (313, 44867), final (314, 44581), faire (315, 44385), suisse (316, 44344), mode (317, 44132), offre (318, 43728), bureau (319, 43711), forum (320, 43666), peinture (321, 43335), alpes (322, 43219), title (323, 43217), election (324, 43124), fille (325, 43116), loire (326, 43086), biographie (327, 43035), college (328, 42710), nouvelles (329, 42670), convention (330, 42658), gif (331, 42578), java (332, 42565), lycos (333, 42465), pas (334, 42417), pays (335, 42312), motivation (336, 42252), est (337, 42228), loto (338, 41904), strasbourg (339, 41900), philosophie (340, 41835), modele (341, 41732), vin (342, 41694), jeune (343, 41492), pub (344, 41196), wanadoo (345, 41166), or (346, 40963), epargne (347, 40935), hebergement (348, 40755), marketing (349, 40736), grand (350, 40695), histoires (351, 40608), black (352, 40309), ecoles (353, 40212), agences (354, 40132), construction (355, 40033), age (356, 39879), contrat (357, 39740), girl (358, 39733), culture (359, 39509), dreamcast (360, 39479), terre (361, 39191), agriculture (362, 39189), films (363, 39135), iut (364, 39028), pamela (365, 38997), sud (366, 38863), michel (367, 38808), sociale (368, 38619), tuning (369, 38575), canada (370, 38510), dragon (371, 38485), ball (372, 38398), fantasy (373, 38227), net (374, 38122), boursorama (375, 38109), historique (376, 37917), rom (377, 37742), anderson (378, 37740), 3d (379, 37709), police (380, 37595), tf1 (381, 37498), alsace (382, 37478), nouvelle (383, 37469), fnac (384, 37401), sony (385, 37314), astuces (386, 37126), offres (387, 37062), cheval (388, 36910), partitions (389, 36902), famille (390, 36887), mail (391, 36856), bateau (392, 36819), siecle (393, 36805), aeroport (394, 36763), tunisie (395, 36728), marie (396, 36722), general (397, 36694), escort (398, 36555), enseignement (399, 36552), federation (400, 36510), 4 (401, 36450), tennis (402, 36440), haute (403, 36363), sims (404, 36346), guadeloupe (405, 36323), systeme (406, 36313), rugby (407, 36281), revue (408, 36183), adsl (409, 36122), sagem (410, 36121), numerique (411, 36013), soluce (412, 36006), animes (413, 35829), bd (414, 35826), david (415, 35765), palm (416, 35726), music (417, 35668), resultats (418, 35541), flash (419, 35528), golf (420, 35302), erotiques (421, 35244), gite (422, 35143), temps (423, 35142), dessins (424, 35141), ordinateur (425, 35038), station (426, 34976), centrale (427, 34966), emplois (428, 34878), laure (429, 34706), blagues (430, 34660), martinique (431, 34650), nom (432, 34532), portables (433, 34510), guitare (434, 34505), buffy (435, 34461), nus (436, 34436), solution (437, 34414), route (438, 34289), audio (439, 34240), electrique (440, 34194), maisons (441, 34175), vendee (442, 34157), val (443, 34120), danse (444, 34036), adresse (445, 34030), national (446, 34028), architecture (447, 34003), petit (448, 33929), medecine (449, 33921), etudes (450, 33835), 5 (451, 33818), louis (452, 33789), psx (453, 33729), parc (454, 33719), institut (455, 33651), napster (456, 33512), logo (457, 33479), hentai (458, 33436), afrique (459, 33434), immobilieres (460, 33201), ouest (461, 32985), hommes (462, 32954), direct (463, 32856), publique (464, 32830), aol (465, 32785), vetements (466, 32653), acces (467, 32560), minitel (468, 32559), hot (469, 32455), environnement (470, 32424), collective (471, 32314), politique (472, 32308), produits (473, 32260), satellite (474, 32252), maladie (475, 32215), sans (476, 32175), iup (477, 32125), eminem (478, 32103), etudiant (479, 32020), farmer (480, 32005), industrie (481, 31991), fabrication (482, 31946), couple (483, 31859), studio (484, 31854), anime (485, 31741), cable (486, 31723), hotmail (487, 31707), alcatel (488, 31658), gifs (489, 31641), meubles (490, 31541), michelin (491, 31497), top (492, 31431), game (493, 31393), montagne (494, 31387), region (495, 31366), romeo (496, 31316), action (497, 31293), information (498, 31221), poste (499, 31140), magie (500, 31098), msn (501, 31004), fonction (502, 30984), public (503, 30978), stages (504, 30960), comparatif (505, 30847), new (506, 30773), production (507, 30771), siemens (508, 30755), frontieres (509, 30746), bmw (510, 30740), social (511, 30650), salle (512, 30646), rap (513, 30625), seine (514, 30482), compagnie (515, 30469), voila (516, 30351), ratp (517, 30303), metiers (518, 30276), patch (519, 30222), tours (520, 30156), bas (521, 30154), domicile (522, 30151), resume (523, 30128), egypte (524, 30065), carnaval (525, 30052), in (526, 30043), algerie (527, 30004), decoration (528, 29946), point (529, 29896), qualite (530, 29888), echangisme (531, 29871), rouge (532, 29795), definition (533, 29789), prefecture (534, 29759), memoire (535, 29737), coiffure (536, 29693), syndicat (537, 29673), formule (538, 29670), generale (539, 29663), lopez (540, 29620), grossesse (541, 29607), festival (542, 29304), jacques (543, 29267), animation (544, 29183), ete (545, 29179), sciences (546, 29136), analyse (547, 29128), taux (548, 29123), fleurs (549, 29031), union (550, 29010), neige (551, 29006), rhone (552, 28972), parole (553, 28957), hopital (554, 28921), fichier (555, 28872), livres (556, 28825), table (557, 28768), mecanique (558, 28749), projet (559, 28726), cannabis (560, 28708), nancy (561, 28703), jardin (562, 28645), arts (563, 28605), pieces (564, 28587), jennifer (565, 28529), jaquette (566, 28510), paul (567, 28459), logiciels (568, 28352), info (569, 28323), etat (570, 28310), on (571, 28270), salaire (572, 28268), concert (573, 28253), billet (574, 28248), repondeur (575, 28196), mylene (576, 28128), retraite (577, 28090), super (578, 28090), marne (579, 28010), londres (580, 28008), cam (581, 27964), celebrites (582, 27959), m6 (583, 27923), apec (584, 27908), quebec (585, 27893), mort (586, 27861), machine (587, 27844), chimie (588, 27667), republique (589, 27615), yamaha (590, 27614), interim (591, 27583), mangas (592, 27573), allemagne (593, 27508), surf (594, 27506), modem (595, 27502), ingenieur (596, 27499), usa (597, 27347), petite (598, 27301), valentin (599, 27222), etranger (600, 27208), raider (601, 27155), ambassade (602, 27138), clipart (603, 27094), fellation (604, 27085), loterie (605, 27066), tout (606, 27039), javascript (607, 26976), espace (608, 26975), mc (609, 26945), noir (610, 26941), encheres (611, 26882), edition (612, 26828), citroen (613, 26821), microsoft (614, 26787), grande (615, 26770), prenom (616, 26762), papier (617, 26720), sodomie (618, 26714), serveur (619, 26711), argent (620, 26701), lecteur (621, 26689), spectacle (622, 26683), academie (623, 26682), normandie (624, 26658), boy (625, 26656), assurances (626, 26578), message (627, 26497), freeware (628, 26429), decodeur (629, 26409), particuliers (630, 26407), fabricant (631, 26361), port (632, 26327), savoie (633, 26318), pyrenees (634, 26306), meuble (635, 26304), melodie (636, 26298), bibliotheque (637, 26293), traducteur (638, 26261), geographie (639, 26242), physique (640, 26202), maurice (641, 26155), salope (642, 26065), cracks (643, 26056), cabinet (644, 26049), philips (645, 26047), chasse (646, 26016), degriftour (647, 25976), premiere (648, 25919), 7 (649, 25889), astrologie (650, 25885), routiere (651, 25825), developpement (652, 25819), violence (653, 25791), vetement (654, 25775), editions (655, 25740), protection (656, 25695), vehicule (657, 25645), disney (658, 25642), naissance (659, 25540), rome (660, 25528), ecrans (661, 25498), grece (662, 25445), pret (663, 25376), virus (664, 25376), texte (665, 25341), angleterre (666, 25283), theme (667, 25267), telephonique (668, 25214), tomb (669, 25181), disque (670, 25177), encyclopedie (671, 25163), scolaire (672, 25135), visual (673, 25128), rouen (674, 25125), evolution (675, 25089), lolita (676, 25054), martin (677, 25047), carrefour (678, 24991), mpeg (679, 24955), park (680, 24943), traitement (681, 24939), municipale (682, 24936), laboratoire (683, 24910), europeenne (684, 24853), portugal (685, 24844), sophie (686, 24806), croix (687, 24805), ans (688, 24802), hp (689, 24771), droits (690, 24751), lettres (691, 24709), francois (692, 24685), piscine (693, 24671), blanc (694, 24633), domaine (695, 24632), basket (696, 24616), juliette (697, 24549), venise (698, 24532), patrick (699, 24421), aix (700, 24370), village (701, 24290), pornographie (702, 24272), fiche (703, 24240), galerie (704, 24199), qui (705, 24124), marine (706, 24118), e (707, 24116), cap (708, 24106), cafe (709, 24085), anti (710, 24084), generateur (711, 24078), voyeur (712, 24058), amateurs (713, 23966), collection (714, 23934), modelisme (715, 23906), prenoms (716, 23882), programmation (717, 23842), html (718, 23801), pollution (719, 23710), contre (720, 23709), b (721, 23644), charles (722, 23628), sa (723, 23588), zola (724, 23464), declaration (725, 23450), ventes (726, 23450), fete (727, 23435), vins (728, 23411), var (729, 23396), fun (730, 23348), wallpaper (731, 23348), regime (732, 23346), part (733, 23283), echangiste (734, 23280), vendre (735, 23279), rock (736, 23219), management (737, 23193), automobiles (738, 23170), arabe (739, 23166), pochette (740, 23165), hack (741, 23155), economique (742, 23119), vache (743, 23109), acheter (744, 23073), vtt (745, 23072), mutuelle (746, 23053), je (747, 23021), chaussures (748, 23013), honda (749, 22968), soleil (750, 22942), gays (751, 22938), massage (752, 22912), psychologie (753, 22892), nude (754, 22882), multimedia (755, 22866), search (756, 22828), classe (757, 22790), metier (758, 22762), resultat (759, 22762), moyen (760, 22757), associations (761, 22755), express (762, 22748), porte (763, 22716), micro (764, 22714), schema (765, 22700), bande (766, 22666), figaro (767, 22615), se (768, 22597), chez (769, 22551), armee (770, 22522), equipe (771, 22498), mathematiques (772, 22484), car (773, 22455), comptabilite (774, 22402), beaute (775, 22362), occasions (776, 22354), plans (777, 22328), base (778, 22304), boutique (779, 22297), mont (780, 22295), commercial (781, 22290), journaux (782, 22286), script (783, 22274), versailles (784, 22229), culotte (785, 22227), velo (786, 22222), exemple (787, 22204), fournisseur (788, 22196), salopes (789, 22184), ford (790, 22152), helene (791, 22151), counter (792, 22127), statistiques (793, 22125), appz (794, 22107), inde (795, 22101), blanche (796, 22099), liberte (797, 22082), batiment (798, 22058), reseaux (799, 22010), basic (800, 22007), voile (801, 21958), erotisme (802, 21948), sinclair (803, 21948), chatte (804, 21914), angers (805, 21888), player (806, 21858), catalogue (807, 21855), ftp (808, 21820), travaux (809, 21819), correspondance (810, 21800), design (811, 21787), strike (812, 21759), professionnel (813, 21712), casta (814, 21705), photographie (815, 21691), magic (816, 21680), iso (817, 21676), tablatures (818, 21615), tarot (819, 21608), poeme (820, 21587), plongee (821, 21560), 9 (822, 21552), lyonnais (823, 21530), creer (824, 21523), skyrock (825, 21513), francaises (826, 21503), shop (827, 21458), vacance (828, 21455), itineris (829, 21445), finances (830, 21414), portail (831, 21409), atlas (832, 21371), dj (833, 21350), documentation (834, 21249), restaurants (835, 21247), optique (836, 21186), alimentaire (837, 21146), chinois (838, 21145), bouygues (839, 21111), parisien (840, 21109), organisation (841, 21104), alternance (842, 21103), libre (843, 21102), masturbation (844, 21055), japon (845, 21047), nuit (846, 21039), distribution (847, 21014), mars (848, 20996), alimentation (849, 20944), ericsson (850, 20902), astuce (851, 20900), bar (852, 20899), bite (853, 20899), clip (854, 20897), bricolage (855, 20825), entre (856, 20818), rue (857, 20803), mobilier (858, 20792), claude (859, 20787), bernard (860, 20770), penis (861, 20753), medical (862, 20750), coupe (863, 20746), chiens (864, 20736), casting (865, 20599), sports (866, 20570), poesie (867, 20508), interieur (868, 20506), cash (869, 20501), villa (870, 20447), philippe (871, 20432), sejour (872, 20427), anne (873, 20379), lara (874, 20372), rapport (875, 20344), virtuelles (876, 20319), methode (877, 20318), ejaculation (878, 20307), chocolat (879, 20272), pilote (880, 20264), live (881, 20255), litterature (882, 20248), numero (883, 20226), jeunesse (884, 20211), editeur (885, 20127), science (886, 20125), mere (887, 20115), noire (888, 20113), chine (889, 20102), elevage (890, 20082), irlande (891, 20081), langue (892, 20025), espagnol (893, 20015), apprendre (894, 20011), train

(895, 19996), winzip (896, 19980), peintre (897, 19970), robert (898, 19957), fakes (899, 19956), chaine (900, 19934), geneve (901, 19902), services (902, 19891), home (903, 19887), lecture (904, 19886), m (905, 19879), max (906, 19878), grosses (907, 19861), 6 (908, 19840), audi (909, 19827), donnees (910, 19787), aerienne (911, 19732), comite (912, 19644), postales (913, 19640), reims (914, 19639), edf (915, 19599), j (916, 19591), vanessa (917, 19584), civil (918, 19582), insee (919, 19548), heures (920, 19520), sons (921, 19472), 35 (922, 19462), ip (923, 19460), tablature (924, 19455), madagascar (925, 19432), restauration (926, 19387), coquine (927, 19376), excel (928, 19371), oise (929, 19367), maritime (930, 19366), dijon (931, 19335), digimon (932, 19275), vol (933, 19267), course (934, 19253), media (935, 19251), defense (936, 19244), bondage (937, 19237), crous (938, 19200), lesbiennes (939, 19192), i (940, 19190), finance (941, 19175), voyance (942, 19165), origine (943, 19146), justice (944, 19138), coeur (945, 19113), mures (946, 19108), naturisme (947, 19092), annee (948, 19091), tarif (949, 19083), poitiers (950, 19078), brest (951, 19067), cher (952, 19047), 4x4 (953, 19016), race (954, 19009), york (955, 18997), blanches (956, 18944), laurent (957, 18927), real (958, 18914), loisirs (959, 18911), stade (960, 18899), que (961, 18886), isere (962, 18874), georges (963, 18871), email (964, 18866), controle (965, 18853), graphique (966, 18848), ferrari (967, 18778), iufm (968, 18773), nt (969, 18710), genie (970, 18708), energie (971, 18699), fichiers (972, 18695), etats (973, 18660), pascal (974, 18651), discount (975, 18650), belle (976, 18638), plage (977, 18618), maupassant (978, 18614), professionnelle (979, 18610), economiseur (980, 18565), electricite (981, 18520), jour (982, 18477), handball (983, 18463), rallye (984, 18441), solaire (985, 18438), unis (986, 18435), orleans (987, 18430), affaires (988, 18427), caen (989, 18389), grosse (990, 18380), julia (991, 18346), sida (992, 18341), moliere (993, 18325), orientation (994, 18291), chauffage (995, 18290), feminin (996, 18272), harry (997, 18266), installation (998, 18246), laser (999, 18208), dawson (1000, 18198).

4.3 Variation de l'importance des requêtes

TAB. 4.1 *Les requêtes dont l'importance baisse grandement chez les GR000 (leur rang initial se retrouve multiplié par plus de 5 avec le panel GR000). Ne sont conservées ici que les requêtes dont la fréquence dans GR000 dépasse 10. Le type F repère les requêtes dont la fréquence initiale dépasse 500.*

rang GR000	requête	rg initial	freq. GR000	freq. init.	type
833	impot	100	33	19755	F
1000	race	197	31	12955	F
1045	parisien	164	30	14222	F
1068	loto	51	30	28851	F
1222	xxx	150	28	14993	F
1368	diddl	222	27	11980	F
1373	code route	233	26	11577	F
1389	ratp	62	26	26047	F
1591	ibazar	223	25	11922	F
1663	iti	275	24	10334	F
1691	sexy	241	24	11333	F
1720	libertysurf	343	24	9040	F
1814	citroen	360	24	8753	F
1893	madagascar	330	23	9215	F
2028	loterie	120	23	17419	F
2337	bouygues	406	22	7965	F
2462	hard	389	21	8288	F
2770	erotique	153	20	14756	F
2970	ucpa	520	20	6551	F
3075	baldur s gate	606	20	5906	F
3118	hotels	399	20	8119	F
3185	calendrier	254	20	10899	F
3335	urssaf	592	20	5999	F
3396	icq	579	19	6096	F
3426	html	665	19	5502	F
3520	palm	228	19	11785	F
3523	dreamcast	457	19	7148	F
3527	moretv	354	19	8838	F

suite sur la page suivante

suite de la page précédente					
rang GR000	requête	rg initial	freq. GR000	freq. init.	type
3631	oreka	400	19	8065	F
3649	basket	490	19	6756	F
3676	culotte	567	19	6204	F
3687	mercedes	532	18	6468	F
3774	tabac	557	18	6303	F
3776	monster	540	18	6431	F
3818	angers	722	18	5158	F
3824	transexuel	420	18	7716	F
3834	femmes mures	359	18	8761	F
3844	chine	547	18	6370	F
3903	encheres	255	18	10873	F
3941	gamez	379	18	8468	F
3971	australie	481	18	6859	F
3993	gays	660	18	5550	F
4073	mobile	773	17	4961	F
4133	tintin	599	17	5958	F
4176	pochettes cd	574	17	6126	F
4181	string	578	17	6112	F
4184	vvf	506	17	6636	F
4249	casino	695	17	5304	F
4277	cartes voeux	130	17	16540	F
4290	upskirt	759	17	4997	F
4315	logiciel	862	17	4504	F
4404	xxl	621	17	5808	F
4406	chats	809	17	4753	F
4430	hot	885	17	4410	F
4484	telephone	260	16	10822	F
4501	libertin	777	16	4941	F
4548	carte voeux	187	16	13307	F
4690	senegal	528	16	6516	F
4751	guitare	633	16	5722	F
4764	ardeche	939	16	4218	F
4853	voix nord	815	16	4737	F
4898	voitures	760	16	4995	F
4912	chasse	676	16	5434	F
					suite sur la page suivante

suite de la page précédente

rang GR000	requête	rg initial	freq. GR000	freq. init.	type
4922	adoption	783	16	4892	F
4924	pmu	173	16	13754	F
4971	tv	698	16	5293	F
5235	karting	875	15	4461	F
5280	casta	473	15	6951	F
5294	pollution	735	15	5108	F
5301	enfoires	972	15	4099	F
5323	anal	655	15	5575	F
5392	filles nues	873	15	4468	F
5423	patrick bruel	1043	15	3877	F
5425	opel	1057	15	3817	F
5522	bresil	913	15	4293	F
5606	station ski	965	14	4114	F
5622	camif	938	14	4221	F
5692	cnfpt	988	14	4064	F
5735	bas nylon	1117	14	3603	F
5779	poitiers	1084	14	3722	F
5819	sonneries nokia	927	14	4252	F
5880	sncm	1080	14	3741	F
5891	kournikova	1110	14	3618	F
5928	mathematiques	976	14	4089	F
5972	teens	730	14	5127	F
6042	dragon	1054	14	3822	F
6235	plans	1196	14	3406	F
6295	infirmiere	903	13	4314	F
6325	ogm	607	13	5904	F
6343	netmeeting	651	13	5592	F
6391	karaoke	525	13	6524	F
6399	autocad	1243	13	3262	F
6456	livres	727	13	5136	F
6467	comite entreprise	1161	13	3497	F
6515	lolitas	418	13	7762	F
6522	tchatche	265	13	10611	F
6565	location appartement paris	970	13	4108	F
6637	amatrices	1222	13	3315	F

suite sur la page suivante

suite de la page précédente					
rang GR000	requête	rg initial	freq. GR000	freq. init.	type
6648	livre	1008	13	4007	F
6719	avion	424	13	7572	F
6726	porsche	769	13	4974	F
6733	cirque	958	13	4135	F
6735	gendarmerie	849	13	4574	F
6739	altalavista	1299	13	3164	F
6743	election	669	13	5476	F
6770	aviation	1096	13	3657	F
6778	cd	1041	13	3883	F
6822	voile	1190	13	3424	F
6824	danone	1027	13	3929	F
6832	delphi	1350	13	3074	F
6841	andorre	1312	13	3134	F
6858	vanessa demouy	1204	13	3378	F
6910	angleterre	834	13	4650	F
6939	travelprice	1020	13	3961	F
6959	trucalacon	1171	13	3461	F
7003	wallpapers	464	13	7088	F
7005	yahoo.fr	860	13	4514	F
7016	giclouz	536	13	6443	F
7032	texto	1071	13	3766	F
7056	annuaires	973	13	4097	F
7099	asse	1150	13	3516	F
7100	repondeur	654	13	5580	F
7125	pret immobilier	1253	13	3236	F
7127	chanson	1062	13	3800	F
7160	sex amateur	1422	12	2932	F
7170	tribune	1288	12	3181	F
7255	autoradio	1195	12	3412	F
7278	mail	497	12	6711	F
7294	coquines	1341	12	3085	F
7298	tps	1182	12	3436	F
7415	crete	1218	12	3334	F
7423	liberty surf	995	12	4054	F
7577	titeuf	1211	12	3349	F
					suite sur la page suivante

suite de la page précédente					
rang GR000	requête	rg initial	freq. GR000	freq. init.	type
7712	laure sainclair	881	12	4440	F
7756	gymnastique	1385	12	2997	F
7758	fiat	1199	12	3396	F
7776	effet serre	825	12	4694	F
7789	cd cover	1373	12	3019	F
7841	honda	743	12	5072	F
8014	snes	1491	12	2835	F
8015	vente correspondance	1047	12	3866	F
8041	velo	780	12	4920	F
8107	cours	1514	12	2804	F
8110	voyages degriffes	637	12	5685	F
8150	dna	1175	12	3448	F
8170	tfou	1203	12	3378	F
8243	java	479	11	6871	F
8285	page blanche	1339	11	3086	F
8342	horloge	1616	11	2673	F
8390	promovacances	953	11	4152	F
8469	mythologie	1233	11	3291	F
8472	annuaire sexe gratuit	1612	11	2677	F
8528	dopage	974	11	4097	F
8545	patch	1670	11	2603	F
8550	cadeaux	1605	11	2685	F
8556	appartement	762	11	4991	F
8577	coquine	554	11	6329	F
8617	marilyn manson	1064	11	3791	F
8685	syntec	1382	11	3001	F
8751	michael page	1184	11	3434	F
8768	video sexe	1417	11	2938	F
8777	thalasso	500	11	6697	F
8785	bafa	1474	11	2858	F
8817	playmates	1012	11	3995	F
8827	cdrwin	1486	11	2842	F
8893	madrid	1181	11	3437	F
8931	dj	1678	11	2592	F
8942	besancon	1235	11	3286	F
					suite sur la page suivante

suite de la page précédente					
rang GR000	requête	rg initial	freq. GR000	freq. init.	type
8946	demenagement	1038	11	3892	F
8957	larousse	1401	11	2971	F
9025	excel	1311	11	3136	F
9053	troyes	1352	11	3069	F
9069	vivendi	1338	11	3089	F
9126	train	1562	11	2739	F
9172	cameleon	1488	11	2840	F
9184	scream	1821	11	2409	F
9268	black	1042	11	3881	F
9278	caramail.com	1265	11	3223	F
9360	salons	1689	11	2569	F
9385	minijupe	1685	11	2577	F
9416	lsa	1518	11	2792	F
9417	discount	1647	11	2630	F
9455	logo nokia	1790	11	2447	F
9457	france info	1628	11	2657	F
9468	vins	1132	11	3567	F
9474	occasion	1197	11	3399	F
9520	infographie	1708	11	2543	F
9555	amitie	1477	11	2853	F
9569	mc solaar	1309	11	3144	F
9571	cliparts	612	11	5874	F
9577	shakespeare	1874	11	2350	F
9579	reims	1010	11	3998	F
9582	bete gevaudan	1635	11	2644	F
9591	annuaire portable	1835	10	2389	F
9608	neige	711	10	5227	F
9647	macif	1493	10	2832	F
9654	mirc	1655	10	2622	F
9659	discussion	1920	10	2303	F
9706	france television	1852	10	2373	F
9776	charmes	1155	10	3513	F
9793	sm	1124	10	3591	F
9817	collants	515	10	6574	F
9826	location logement	1807	10	2425	F
					suite sur la page suivante

suite de la page précédente					
rang GR000	requête	rg initial	freq. GR000	freq. init.	type
9827	astuces	1452	10	2894	F
9892	photographe	1438	10	2912	F
9899	partouze	758	10	5004	F
9922	airfrance	1135	10	3560	F
9928	appz	409	10	7878	F
9943	celebrites	523	10	6525	F
10123	auto journal	1576	10	2723	F
10171	femmes poilues	1944	10	2285	F
10175	notaires	1419	10	2935	F
10201	couples	1601	10	2692	F
10233	polynesie	1418	10	2937	F
10256	finances	1627	10	2657	F
10374	expose	2074	10	2160	F
10393	spartateur	933	10	4228	F
10422	jeuxvideo	1879	10	2345	F
10425	bijouterie	1978	10	2249	F
10499	enceinte	1814	10	2419	F
10522	paques	1296	10	3168	F
10604	restaurants	992	10	4059	F
10623	ti89	1696	10	2556	F
10625	telecharger mp3	2114	10	2126	F
10667	script	2013	10	2216	F
10694	trenet	1818	10	2413	F
10719	angel	2038	10	2197	F
10727	dialogue	1425	10	2929	F
10731	prague	1644	10	2631	F
10775	winrar	2090	10	2149	F
10840	exhibition	1028	10	3922	F
10842	alfa romeo	1543	10	2766	F
10865	voitures occasions	1588	10	2710	F
10881	couple echangiste	1643	10	2631	F
10887	barcelone	737	10	5101	F
10888	hertz	2075	10	2160	F
10893	dreamweaver	1516	10	2797	F
10926	stepstone	1500	10	2822	F
					suite sur la page suivante

suite de la page précédente					
rang GR000	requête	rg initial	freq. GR000	freq. init.	type
10977	revebebe	720	10	5194	F
10989	netclub	259	10	10835	F
11007	avoriaz	1157	10	3506	F
11052	anyway	1364	10	3040	F
11055	maths	2070	10	2164	F
11065	mobiles	1923	10	2300	F
11102	fax	1654	10	2622	F
11138	polices	2212	10	2046	F
11144	albator	1956	10	2275	F
11207	texas instrument	2129	10	2117	F
11231	golf	398	10	8127	F
11250	bois	1266	10	3220	F
11269	revolution francaise	2249	10	2027	F
11276	planete	2221	10	2042	F
11297	gif anime	647	10	5620	F
11322	usa	1851	10	2377	F

TAB. 4.2 *Les requêtes qui prennent une importance particulière avec GR000 (leur rang initial se retrouve divisé par plus de 5). Ne sont conservées ici que les requêtes dont la fréquence dans GR000 dépasse 25. Le type F repère les requêtes dont la fréquence initiale dépasse 500.*

rang GR000	requête	rg initial	freq. GR000	freq. init.	type
3	fond ecran	18	306	51067	F
8	moteur recherche	52	275	28490	F
11	fievre aphteuse	419	234	7750	F
32	sarah michelle gellar	723	171	5157	F
35	stars nues	245	164	11252	F
74	final fantasy 9	573	114	6138	F
75	call girl	478	113	6878	F
76	hommes nus	963	112	4118	F
92	journaux	463	100	7094	F
103	zoophilie gratuit	2273	94	2011	F
108	kama sutra	562	92	6244	F
115	c	739	90	5094	F
124	geographie	764	86	4991	F
149	seine maritime	12667	78	420	
157	select	14126	77	376	
161	joconde	2366	76	1936	F
163	arbre genealogique	907	76	4302	F
164	location espagne	1074	75	3759	F
165	jessica biel	1846	75	2383	F
166	gladiator	1462	75	2879	F
172	laure sinclaire	1361	75	3056	F
194	sexe grosse salope	36149	72	148	
201	flavie flament	2426	71	1887	F
212	musique film	2209	70	2049	F
219	rodez	4133	68	1192	F
229	cap petite enfance	11361	67	466	
246	webcam direct	10992	64	482	
256	walt disney	1354	63	3068	F
270	moteur recherche sexe	3211	62	1487	F
281	fete internet	2772	61	1686	F

suite sur la page suivante

suite de la page précédente					
rang GR000	requête	rg initial	freq. GR000	freq. init.	type
283	final fantasy ix	2979	61	1585	F
294	fbi	1531	60	2781	F
307	louis 16	13775	58	385	
320	etats unis	1921	57	2302	F
330	lolo ferrari	1867	56	2356	F
347	location particulier	3206	54	1489	F
350	insolite	1752	54	2495	F
365	iut informatique	4977	52	1008	F
373	baladeur cd mp3	39308	52	135	
375	lingerie fine	3109	51	1527	F
376	gestion patrimoine	4533	51	1095	F
383	instruments musique	2841	51	1652	F
386	punk	2770	50	1687	F
389	magazine sante	27478	50	195	
401	capital risque	5472	49	924	F
407	emulateur super nintendo	10749	49	492	
408	alfred hitchcock	28694	49	186	
425	sleepy hollow	12631	47	421	
433	ubu roi	19749	47	271	
438	liaisons dangereuses	16170	46	330	
440	rachat credit	2321	46	1975	F
455	machine coudre	4827	45	1034	F
458	marketing olfactif	16432	45	325	
461	hamlet	12299	45	431	
462	licence professionnelle	3046	45	1557	F
463	diplodocus	16672	45	320	
469	lutte chomage	3023	44	1568	F
498	moteur electrique	4592	43	1080	F
506	verhaeren	44862	43	119	
511	american history	4427	42	1118	F
513	bouledogue	10907	42	485	
515	jeunes filles nues	4073	42	1211	F
517	jupe cuir	39143	42	136	
518	credit ligne	6409	42	802	F
519	location villa espagne	5460	42	926	F
					suite sur la page suivante

suite de la page précédente

rang GR000	requête	rg initial	freq. GR000	freq. init.	type
521	bombe	4086	42	1205	F
523	fond ecran gratuit	16212	42	329	
530	arts plastiques	5295	42	950	F
533	bts informatique gestion	5165	41	973	F
535	confessions erotiques	5774	41	884	F
536	camping var	7111	41	724	F
544	kabylie	3373	41	1422	F
547	materiel agricole	3152	41	1507	F
549	groupe sanguin	20925	40	256	
553	asptt	6470	40	795	F
554	as400	3246	40	1475	F
556	livraison domicile	2992	40	1580	F
561	parodies chansons	20947	40	256	
567	pulp fiction	9268	40	565	F
573	syndicat professionnel	20535	40	261	
575	achat ligne	3865	40	1264	F
577	mc cain	29263	40	183	
580	fusions acquisitions	14112	39	376	
582	base donnees	3117	39	1523	F
588	restauration collective	5998	39	853	F
589	liane foly	3660	39	1326	F
592	scat	7532	39	687	F
593	egalite homme femme	34871	39	153	
596	camping herault	11244	39	471	
597	ecole danse	10087	39	521	F
611	boulogne	5406	38	932	F
616	betterave	18092	38	295	
619	chevalier	3723	38	1307	F
630	kelly preston	19135	38	279	
634	escorte girl	5398	38	934	F
636	aprilia	4783	38	1042	F
659	mondes engloutis	12181	37	435	
661	jardinerie	3407	37	1409	F
670	immobilier yvelines	13605	37	390	
671	viking	5054	37	995	F

suite sur la page suivante

suite de la page précédente					
rang GR000	requête	rg initial	freq. GR000	freq. init.	type
679	declaration droits homme citoyen	20403	37	262	
682	location voilier corse	44324	37	120	
688	natasha henstridge	14378	36	370	
695	corriges bac	7230	36	713	F
700	ufolep	4342	36	1139	F
703	bruit	3946	36	1245	F
717	shining	28819	36	186	
718	pin up	4329	36	1142	F
721	banque ligne	4647	36	1068	F
725	gateau chocolat	4456	35	1111	F
733	1942	50474	35	105	
740	anti pokemon	4081	35	1208	F
744	photos films	26666	35	201	
746	pessac	12906	35	411	
748	sexe nylon	73843	35	72	
750	janet jackson	6304	35	816	F
753	pokemon or argent	6139	35	835	F
760	mot passe caramail	27630	35	194	
767	roi soleil	26959	35	199	
768	fleurs bleues queneau	26480	35	202	
770	ados nus	18036	34	296	
780	materiel cuisine	26648	34	201	
784	ecole superieure commerce	13514	34	393	
797	comptines	6824	34	756	F
798	cea	4605	34	1076	F
801	semi marathon paris	17084	34	313	
818	fournisseurs logiciels comptabilite	67046	34	79	
819	orl	10029	34	524	F
824	produits bio	12398	33	428	
829	enseignement prive	5402	33	933	F
830	wass up	18959	33	282	
831	ma sorciere bien aimee	51247	33	104	
834	remorque voiture	58554	33	91	
843	messages gratuits	19664	33	272	
847	concours cpe	10368	33	508	F
					suite sur la page suivante

suite de la page précédente

rang GR000	requête	rg initial	freq. GR000	freq. init.	type
848	equipement cuisine	42791	33	124	
850	hautes pyrenees	14659	33	363	
871	maurice trintignant	18115	33	295	
877	elections communales	45389	32	117	
881	orcad	11797	32	449	
887	scuio	13560	32	392	
891	camping cap agde	28417	32	188	
893	pepin bref	66505	32	80	
898	clubs libertins	36467	32	146	
901	pokemon gold silver	31291	32	171	
904	pornic	6250	32	821	F
905	supply chain management	15678	32	341	
911	eclairage spectacle	63064	32	84	
917	landerneau	15229	32	350	
925	sophie thalmann nue	52891	32	101	
927	oloron	17529	32	305	
928	kiosque	5651	32	899	F
932	autorun	6800	32	758	F
946	camping espagne	5037	32	997	F
951	oreille	8527	31	611	F
953	jupes plissees	27049	31	198	
958	infographie lyon	56594	31	94	
962	montage video	4900	31	1022	F
966	cours espagnol	6555	31	786	F
972	nicolas anelka	7831	31	664	F
981	terrains batir	7107	31	725	F
985	soundfont	23771	31	226	
988	charleville mezieres	23277	31	230	
990	index	10866	31	487	
994	serveur application	45594	31	117	
996	code ape	7781	31	667	F
998	chauffeur	10690	31	494	
1010	sexe gratuis	10290	31	512	F
1011	vente automobile	5835	31	875	F
1018	catherine zeta jones	5861	30	871	F

suite sur la page suivante

suite de la page précédente					
rang GR000	requête	rg initial	freq. GR000	freq. init.	type
1022	dragee	12677	30	419	
1024	particularisme corse	54444	30	98	
1026	camions occasion	29342	30	182	
1028	adobe premiere	5201	30	967	F
1031	immobilier cannes	27894	30	192	
1039	michele gasc	34695	30	154	
1047	adjudication immobiliere	54615	30	97	
1053	mont marsan	7200	30	716	F
1065	code barres	17936	30	298	
1075	fete grands meres	28913	30	185	
1081	jean luc lahaye	47075	30	113	
1083	jeux ps2	36078	30	148	
1085	abortion	48557	30	109	
1094	soirees parisiennes	29600	29	181	
1096	presse masculine	12802	29	415	
1098	tangram	9727	29	539	F
1104	instrument musique	5823	29	876	F
1105	charisma carpenter nue	38182	29	139	
1106	scratch	9133	29	572	F
1107	bouquet fleurs	18760	29	284	
1110	image minotaure	89371	29	59	
1114	mask	10937	29	484	
1117	recettes cuisine marocaine	59840	29	89	
1128	moi nue	53291	29	100	
1137	jessica biel nue	16320	29	327	
1142	immobilier pays basque	10165	29	517	F
1145	cote moto	11984	29	442	
1148	corinne hermes	13331	29	398	
1149	regles grammaire	58552	29	91	
1155	skateboarding	27950	29	192	
1164	gites ruraux bretagne	66268	29	80	
1173	torse nu	48078	29	111	
1181	forum emploi	20021	28	267	
1192	beachcomber	19324	28	276	
1194	armenager	20897	28	256	
					suite sur la page suivante

suite de la page précédente

rang GR000	requête	rg initial	freq. GR000	freq. init.	type
1196	dessin humoristique	6382	28	805	F
1200	sexe femme mure	41324	28	129	
1202	collectionneur	6642	28	777	F
1210	telecharger logiciel	13932	28	381	
1218	regression logistique	31115	28	172	
1226	cholula	36931	28	144	
1241	locations saisonnieres corse	33598	28	159	
1242	magazine cinema	15181	28	351	
1262	restauration rapide	6985	27	738	F
1263	bruay buissiere	64442	27	82	
1269	ch tons	83342	27	63	
1270	stage mecanique	63519	27	84	
1275	concept	24469	27	219	
1278	diester	27720	27	193	
1285	highlander 4	60950	27	87	
1290	bergerac	6501	27	792	F
1293	imprimante olivetti	95883	27	55	
1295	locations ile re	56039	27	95	
1297	cours auditeur ecole louvre	76397	27	69	
1298	cannet	20474	27	261	
1303	salon recrutement	31864	27	168	
1306	msa	7515	27	689	F
1310	paris sud encheres	80350	27	66	
1312	kelly trump	10730	27	493	
1326	roms dragon ball z	59206	27	90	
1335	cosmetologie	10328	27	510	F
1337	rouge noir stendhal	24207	27	222	
1340	fillettes nues	16034	27	333	
1341	explosifs	9867	27	532	F
1343	fable	15951	27	335	
1345	galfione	7178	27	718	F
1347	environnement magazine	36961	27	144	
1351	musulman	13130	27	405	
1356	hairy pussy free	95804	27	55	
1359	affaire hakkar	43046	27	123	

suite sur la page suivante

suite de la page précédente					
rang GR000	requête	rg initial	freq. GR000	freq. init.	type
1364	photos acteurs	21472	27	250	
1370	provence alpes cote azur	30750	26	174	
1371	secretaires sexy	17149	26	312	
1372	corriges bts	26723	26	200	
1381	segura	23654	26	227	
1383	ska p	7866	26	661	F
1385	morilles	37281	26	143	
1400	logiciel transaction immobiliere	97734	26	54	
1403	immobilier landes	12479	26	425	
1405	localisation gsm	55461	26	96	
1410	sculture	24628	26	218	
1417	ecran veille matrix	19041	26	280	
1422	mobilier france	10286	26	512	F
1428	provence cote azur	60480	26	88	
1430	web cam gratuit	40457	26	132	
1434	paul morand	30054	26	178	
1435	plantu	8251	26	632	F
1438	table ronde	9271	26	565	F
1446	interpol	10752	26	492	
1448	societe assurance	82942	26	64	
1455	grenade	7663	26	675	F
1459	1945	52968	26	100	
1461	immobilier haute savoie	19266	26	277	
1463	college vieux colombier	77450	26	68	
1464	gillain	13187	26	403	
1465	extraterrestre	8800	26	592	F
1466	culture americaine	52104	26	102	
1468	telecharger dreamcast	94913	26	56	
1471	christopher pike	61511	26	86	
1472	fleurs artificielles	43906	26	121	
1473	castes	42740	26	124	
1481	armentieres	20555	26	260	
1483	saint emilion	12879	26	412	
1494	jenny mc carthy	13842	26	383	
1495	solution resident evil	53442	26	100	
					suite sur la page suivante

suite de la page précédente

rang GR000	requête	rg initial	freq. GR000	freq. init.	type
1496	holly marie combs	8545	26	609	F
1501	baie mont saint michel	56362	26	94	
1505	beauce	27282	25	196	
1508	fromages france	60237	25	88	
1509	ours peluche	11444	25	463	
1513	petites annonces rencontres	14933	25	357	
1515	echauffement	29833	25	179	
1520	trucs astuces pc	43733	25	122	
1524	obesite enfant	64840	25	82	
1531	vente distance	24437	25	220	
1533	hitchcock	10129	25	519	F
1535	bac stl	39601	25	134	
1537	95b	22777	25	235	
1543	depannage televiseur	63038	25	84	
1545	amina	24125	25	222	
1547	patrick swayze	22369	25	240	
1548	mariage paris	27906	25	192	
1551	position	8308	25	628	F
1560	amateurs nues	17599	25	303	
1561	voyage venise	10963	25	483	
1564	prenoms russes	91351	25	58	
1566	harem	26897	25	199	
1567	ulm pendulaire	84367	25	63	
1573	generateur nomad	8038	25	647	F
1575	e procurement	43150	25	123	
1583	duplex	18098	25	295	
1587	amanda lear nue	46048	25	115	
1592	overclock amd	56117	25	95	
1593	camping calvi	75193	25	70	
1610	pays cathare	59137	25	90	
1611	syndrome	13499	25	394	
1612	ecole esthetique	10675	25	495	
1619	lettre emploi	52736	25	101	
1639	coloriage enfant	11156	25	475	
1640	monte cristo	45880	25	116	

suite sur la page suivante

suite de la page précédente					
rang GR000	requête	rg initial	freq. GR000	freq. init.	type
1641	cederoms	76311	25	69	
1642	phlebologie	54448	25	98	

Bibliographie

[AL91] Jacques Anis et Jean-Louis Lebrave (dir.). *Texte et ordinateur. Les mutations du Lire-Écrire.* Éditions de l'espace européen, La Garenne-Colombes, 1991. Actes du colloque interdisciplinaire tenu à l'Université Paris X Nanterre, 6–7–8 juin 1990.

[AM85] Jean-Loup Amselle et Elikia M'Bokolo (dir.). *Au cœur de l'ethnie.* La Découverte, Paris, 1985.

[AM95] Albert et Loy Morehead (dir.). *The new American Webster handy college dictionnary.* Penguin Books, New York, 1995. Troisième édition.

[And91] Jacques André. Nouveaux caractères pour une nouvelle typographie. In Anis et Lebrave [AL91], pages 43–52.

[And95] Olivier Andrieu. *Internet, guide de connexion.* Eyrolles, Paris, 1995.

[Ani98] Jacques Anis. *Texte et ordinateur. L'écriture réinventée?* De Boeck Université, Paris et Bruxelles, 1998.

[BA02] Valérie Beaudouin et Houssem Assadi. Usages des moteurs de recherche: une approche centrée utilisateurs. In *Sixièmes Journées internationales d'Analyse Statistiques des Données Textuelles (JADT)*, 2002.

[Bal99] Francis Balle. *Médias et Sociétés.* Montchrestien, Paris, 1999. Neuvième édition.

[BC01] Stefana Broadbent et Francesco Cara. Évolution des usages de l'internet. In Guichard [Gui01], pages 156–164.

[BCP83] Pierre Bourdieu, Jean-Claude Chamboredon, et Jean-Claude Passeron. *Le métier de sociologue*. Mouton, La Haye, 1983. Quatrième édition.

[BCT01] Francis Balle et Laurent Cohen-Tanugi (dir.). *Dictionnaire du web*. Dalloz, Paris, 2001.

[BJ96] Marc Baratin et Christian Jacob (dir.). *Le pouvoir des bibliothèques*. Albin Michel, Paris, 1996.

[Bou84] Pierre Bourdieu. *Homo academicus*. Éditions de Minuit, Paris, 1984.

[Bou01] Pierre Bourdieu. *Science de la science et réflexivité*. Raisons d'agir, Paris, 2001.

[Bra96] Hervé Le Bras. *Le peuplement de l'Europe*. La documentation française, Paris, 1996.

[Bra00] Hervé Le Bras. *Essai de Géométrie Sociale*. Odile Jacob, Paris, 2000.

[Bre00] Philippe Breton. *Le culte de l'Internet*. La Découverte, Paris, 2000.

[BV01] Valérie Beaudouin et Julia Velkovska. Un univers relationnel sur l'internet : forum et pages personnelles. In Guichard [Gui01], pages 217–224.

[Cas98] Manuel Castells. *La société en réseau*. Fayard, Paris, 1998.

[Cat88] Nina Catach (dir.). *Pour une théorie de la langue écrite*. Éditions du CNRS, Paris, 1988.

[CEF98] CEFRIO. Internet : accès et utilisation au Québec, 1998.

[Coh97] Daniel Cohen. *Richesse du monde, pauvreté des nations*. Flammarion, Paris, 1997.

[CT99] Laurent Cohen-Tanugi. *Le nouvel ordre numérique*. Odile Jacob, Paris, 1999.

[DCN98] Roberto Di Cosmo et Dominique Nora. *Le hold-up planétaire*. Calmann-Lévy, Paris, 1998.

[Des93] Alain Desrosières. *La politique des grands nombres.* La Découverte, Paris, 1993.

[d'I99] Alain d'Iribarne. Les SHS et les usages des TIC: disciplines, approches, acteurs. In *Ingénieurs et chercheurs en sciences humaines: quelle contribution à la problématique du RNRT ?*, École nationale supérieure des télécommunications, Paris, 1999.

[Djo01] Mahieddine Djoudi. Conception d'assistants à la navigation. In Guichard [Gui01], pages 165–173.

[Duc98] Christine Ducourtieux. L'usage de l'internet en lettres et en sciences humaines à l'École normale supérieure, 1998. *in* «Internet et les chercheurs», rapport final délivré au ministère de l'Enseignement Supérieur et de la Recherche, juillet 1998, pages 45–55.

[Dup02] Gabriel Dupuy. *Internet ; géographie d'un réseau.* Ellipses, Paris, 2002.

[Edg98] David Edgerton. De l'innovation aux usages. dix thèses éclectiques sur l'histoire des techniques. *Annales Histoire, Sciences Sociales*, 4–5:815–837, 1998. Histoire des techniques.

[eSP96] Philippe Breton et Serge Proulx. *L'explosion de la communication.* La Découverte, Paris, 1996. Seconde édition.

[Feb68] Lucien Febvre. *Le problème de l'incroyance au* XVI^e^ *siècle, la religion de Rabelais.* Albin Michel, 1968. Seconde édition (première ed. 1942).

[Gal92] John K. Galbraith. *Brève histoire de l'euphorie financière.* Seuil, Paris, 1992.

[Gan99] Jean-Gabriel Ganascia. *2001, l'odyssée de l'esprit.* Flammarion, Paris, 1999.

[Gen87] Gérard Genette. *Seuils.* éditions du Seuil, Paris, 1987.

[GN97] Éric Guichard et Gérard Noiriel (dir.). *Construction des nationalités et immigration dans la France contemporaine.* Presses de l'ENS, Paris, 1997.

[Goo97] Jack P. Goody. *Entre l'oralité et l'écriture*. Presses Universitaires de France, Paris, 1997.

[Goo00] Jack P. Goody. *The Power of the Written Tradition*. Smithsonian Institution Press, Washington et London, 2000.

[Gri98] Maurizio Gribaudi (dir.). *Espaces, temporalités, stratifications. Exercices sur les réseaux sociaux*. Éditions de l'École des hautes études en sciences sociales, Paris, 1998.

[Gui97] Éric Guichard. Un coupable idéal ? *Terminal*, pages 181–190, 1997. Numéro 71–72.

[Gui01] Éric Guichard (dir.). *Comprendre les usages de l'Internet*. Éditions Rue d'Ulm, Paris, 2001.

[Hal90] David Hall. La bible et ses lecteurs. In Universalis [Uni90], pages 270–271.

[Hal97] Serge Halimi. *Les nouveaux chiens de garde*. Liber-Raisons d'agir, Paris, 1997.

[Her99] Clarisse Herrenschmidt. Les origines de l'écriture. *Le débat*, 106:37–65, 1999.

[HR01] François-Xavier Hussherr et Julien Rosanvallon. *e-communication*. Dunod, Paris, 2001.

[Hui95] Christian Huitema. *Et Dieu créa l'Internet*. Eyrolles, Paris, 1995. Seconde édition.

[Jac96] Christian Jacob. Lire pour écrire : navigations alexandrines. In Baratin et Jacob [BJ96], pages 47–83.

[Jan00] Bernard J. Jansen. The effect of query complexity on web searching results. *Information Research*, 6(1), October 2000. URL : http://informationr.net/ir/6-1/paper87.html.

[Jea00] Yves Jeanneret. *Y a-t-il (vraiment) des Technologies de l'Information ?* Septentrion, Villeneuve d'Asq, 2000.

[JP00] Bernard J. Jansen et Udo Pooch. Web user studies : A review of web searching studies et a framework for future research. *Journal of* *the American Society for Information Science and technology*, 2000. URL : http://dois.mimas.ac.uk/DoIS/data/Articles/jullqmcmjy:-2001:v:52:i:3:p:235-246.html.

[Kuh90] Thomas S. Kuhn. *La tension essentielle*. Gallimard, Paris, 1990.

[Lac97] Guy Lacroix. *Le mirage internet. Enjeux économiques et sociaux*. Vigot, Paris, 1997.

[Laj98] Jacques Lajoie. Les moteurs de recherche du réseau internet comme indicateurs des besoins intimes. *Revue Québécoise de Psychologie*, 19(2), 1998.

[Laj01] Jacques Lajoie. Richesse et diversité des requêtes. In Guichard [Gui01], pages 187–195.

[Lat89] Bruno Latour. *La science en action*. collection Folio. Gallimard, Paris, 1989.

[Lat96] Bruno Latour. Ces réseaux que la raison ignore : laboratoires, bibliothèques, collections. In Baratin et Jacob [BJ96], pages 23–46.

[Les98] Lawrence Lessig. Jefferson's nature, 1998. ULR : http://cyber.law.harvard.edu/works/lessig/NatureD3.pdf.

[Mar90] Henri-Jean Martin. L'apparition du livre imprimé en Occident. In Universalis [Uni90], pages 158–159.

[Mar96] Henri-Jean Martin. *Histoire et pouvoirs de l'écrit*. Albin Michel, Paris, 1996. Seconde édition. Avec la collaboration de Bruno Delmas.

[Mat01] Paul Mathias. Anatomie d'une abondance. In Guichard [Gui01], pages 130–138.

[MB00] R. McArthur et P. D. Bruza. The ranking of query refinements in interactive web-based retrieval. In *Proceedings of the Information Doors Workshop*, 2000. URL : http://www.guidebeam.com/infodoors.pdf.

[Mor80] Louis-Marie Morfaux. *Vocabulaire de la philosophie et des sciences humaines*. Armand Colin, Paris, 1980.

[Nau93] Daniel Naulleau. Avec Socrate, tout est possible à la SNCF. *Terminal*, pages 13–24, 1993. numéro 61.

[Nor99] Donald A. Norman. *The invisible computer*. MIT Press, Cambridge, MA, 1999.

[OP96] Jean-Marc Offner et Denise Pumain (dir.). *Réseaux et territoires*. l'Aube, Paris, 1996.

[Pal96] Gilles Palsky. *Des chiffres et des cartes*. Comité des travaux historiques et scientifiques, Paris, 1996.

[Per89] Jacques Perriault. *Logique de l'usage. Essai sur les machines à communiquer*. Flammarion, Paris, 1989.

[Pet90] Armando Petrucci. La lecture des clercs. In Universalis [Uni90], pages 266–267.

[Rob85] Paul Robert. *Le Petit Robert*. Le Robert, Paris et Montréal, 1985. Dictionnaire alphabétique et analogique de la langue francaise ; rédaction dirigée par A. Rey et J. Rey-Debove. Seconde édition.

[S+98] Craig Silverstein et al. Analysis of a very large altavista query log. Note #1998-014 ; 26 octobre 1998, Digital Equipment Corporation, Systems Research Center, 1998. URL : http://gatekeeper.dec.com/pub/DEC/SRC/technical-notes/abstracts/src-tn-1998-014.html.

[Sal01] Jean-Michel Salaün. Documents numériques et universités françaises. In Guichard [Gui01], pages 112–117.

[sdb98] Conseil supérieur des bibliothèques (dir.). *Rapport pour les années 1996–1997*. Association du Conseil supérieur des bibliothèques, Paris, 1998. URL : http://www.enssib.fr/autres-sites/csb/rapport97/csb-rapp97integral.html.

[Ser01] Thomas Serval. Nature de l'échange par l'internet. In Guichard [Gui01], pages 23–32.

[Sim89] Gilbert Simondon. *Du mode d'existence des objets techniques*. Aubier, Paris, 1989. Troisième édition.

[Sol01] Sergei Soloviev. Bibliothèque universelle et culture de l'éphémère. In Guichard [Gui01], pages 59–63.

[Sve90] Jesper Svendro. La lecture dans l'antiquité. In Universalis [Uni90], pages 262–263.

[Tou99] François-Michel Le Tourneau. Le 'rézo' de l'ens: une expérience réussie. *Le Courrier de l'ENS*, 1999. Numéro 46.

[Uni90] Encyclopædia Universalis (dir.). *Grand Atlas des Littératures*. Encyclopædia Universalis France, Paris, 1990.

[VB+02] Christian Licoppe Valérie Beaudouin et al. Parcours sur internet : analyse des traces d'usage, mai 2002.

[Ver82] Jean-Pierre Vernant. *Mythe et pensée chez les Grecs*, volume 1. François Maspero, Paris, 1982.

[Vit90] Jean-Pierre Vittu. La contrefaçon. In Universalis [Uni90], pages 320–321.

[Wol99] Dominique Wolton. *Internet, et après ?* Flammarion, Paris, 1999.

Index des URL cités

Cet index complète la bibliographie précédente, notamment pour les articles et rapports qui ne sont disponibles que sur le web.

clinton4.nara.gov/WH/New/html/20000121_3.html, 20
interactif.lemonde.fr/article/0,5611,3324–223491-0,FF.html, 262
perso.club-internet.fr/octavia0/classiques.htm, 145
ww.macfergus.com/niels/dmca/cia.html, 85
www.blinkenlights.com/classiccmp/apple2history.html, 50
www.bmlisieux.com/maj.htm, 86
www.europarl.eu.int/tempcom/echelon/mandate_fr.htm, 262
www.info.unicaen.fr/bnum/jelec/Solaris/d04/4charpin.html, 18
www.iris.sgdg.org/actions/lsi/evol/art14.html, 262
www.let.leidenuniv.nl/history/ivh/frame_theorie.html, 130
www.ling.udel.edu/colin/tools/BBEdit/grep_search.html, 68
www.multimania.com/uzine/OURS/afp2.htm, 398
www.netreporter.fr/netframe/chronique/html/airont.htm, 398
www.olf.gouv.qc.ca/ressources/internet/fiches/8388786.htm, 33
www.recherche.gouv.fr/appel/2002/societeinfo.htm, 22
www.renater.fr/GIPRenater/historique.htm, 114
www.senat.fr/Vforum/13/forum.html, 260
www.unine.ch/antic/ressources.html, 145
www.W3.org/History.html, 130

– **Rapports**

europa.eu.int/ISPO/infosoc/backg/bangeman.html#chap1, 21
mission-dti.inria.fr, 35
ww.pner.org/html/activduprog/doc2b1.asp, 184
www.education.gouv.fr/rapport/silberman, 59, 213
www.senat.fr/rap/r97-331-t1/r97-331-t10.html, 179

– **Sites**

barthes.ens.fr, 143, 144
barthes.ens.fr/atelier, 144
barthes.ens.fr/carto, 146
barthes.ens.fr/clio, 146

Table des figures

Liste des tableaux

Table des matières

L'internet : mesures des appropriations d'une technique intellectuelle

Résumé

Nous montrons tout d'abord comment l'informatique et l'internet transforment l'écriture : le code et son support visuel sont maintenant distants, la notion de texte s'étend aux programmes, puis aux sons et aux images. La production de listes, graphiques et hypertextes en est facilitée, mais la maîtrise des outils de manipulation de l'écrit requiert de nouveaux savoir-faire, qui élargissent potentiellement nos perspectives intellectuelles.

Ces interactions entre support matériel, système de signes et psyché — comme la production d'institutions autour de l'écriture tendant à nier son statut technique — n'ont rien de révolutionnaire. En nous inspirant des travaux de Jack Goody et de Pierre Bourdieu, et à partir de l'internet, nous détaillons les relations réciproques entre l'écriture et la production scientifique.

Au fil de deux enquêtes, des chercheurs en sciences humaines témoignent de la façon dont l'édition savante électronique transforme les conditions de l'argumentation scientifique et de la diffusion de la connaissance. Le coût d'un tel engagement est élevé, intellectuellement comme socialement ; mais la compréhension des relations entre l'outillage mental, les modalités de sa transmission, et l'économie symbolique du monde universitaire, s'en trouve accrue.

Enfin, nous donnons une preuve *interne* de la réflexivité de l'écriture : à partir de listes produites par des centaines de milliers de personnes, et archivées par des machines, nous montrons que ces textes électroniques ne peuvent s'appréhender sans outils spécifiques, mais qu'alors, les uns et les autres donnent à voir une réalité linguistique méconnue. Nous prouvons que les études sociologiques à partir de telles archives sont non seulement réalisables, mais aussi scientifiquement pertinentes. Et les résultats obtenus invitent à questionner la dépendance des concepts des sciences sociales face aux outils de mesure et d'écriture qui, implicitement, les organisent.