

Cahiers **GUT**enberg

☞ UNE POLICE POUR LA SCIENCE. DE
L'IMPACT DU TRAITEMENT DE TEXTE SUR
L'ACTIVITÉ SCIENTIFIQUE

☞ Emmanuel SAINT-JAMES

Cahiers GUTenberg, n° 13 (1992), p. 39-54.

http://cahiers.gutenberg.eu.org/fitem?id=CG_1992__13_39_0

© Association GUTenberg, 1992, tous droits réservés.

L'accès aux articles des *Cahiers GUTenberg*

(<http://cahiers.gutenberg.eu.org/>),

implique l'accord avec les conditions générales

d'utilisation (<http://cahiers.gutenberg.eu.org/legal.html>).

Toute utilisation commerciale ou impression systématique

est constitutive d'une infraction pénale. Toute copie ou impression

de ce fichier doit contenir la présente mention de copyright.

Une police pour la science. De l'impact du traitement de texte sur l'activité scientifique

Emmanuel SAINT-JAMES

Bull, direction de la recherche
rue Jean Jaurès, B.P. 53
78340 Les Clayes-Sous-Bois

Ils se révélaient fortement attachés à leur planche à dessin, aimant leur métier et faisant preuve, dans leur métier, d'une activité remarquable ; mais si on leur avait proposé d'appliquer à eux-mêmes, et non plus à leurs machines, la hardiesse de leurs idées, ils eussent réagi comme si on leur eût demandé de faire d'un marteau l'arme d'un meurtre. (...) On peut encore comprendre qu'un ingénieur soit absorbé par sa spécialité, au lieu de déboucher dans la vastitude et de la liberté du monde de la pensée, quoique ses machines soient livrées jusqu'au confins de la terre ; car on ne lui demande pas plus d'être capable de faire bénéficier son âme privée de l'audace de l'âme technique, qu'à une machine de pouvoir s'appliquer à elle-même les équations infinitésimales qui sont à la base de sa conception. Mais on ne peut en dire autant de l'informatique.

Robert Musil

L'homme sans qualité (page 44).

Introduction

L'utilisation du traitement de texte pour rédiger des résultats scientifiques a aujourd'hui largement débordé la communauté des informaticiens. Il devient un outil obligé, détrônant la machine à écrire et la linotype, comme le livre supplanta le manuscrit il y a quelques siècles. Nous voudrions réfléchir ici aux conséquences de ce changement, précisément en le comparant avec le précédent de l'imprimerie.

Certaines des idées avancées choqueront peut-être ; elles résultent pourtant d'un vécu : la rédaction, à l'aide d'outils informatiques, d'un *Traité de programmation* de quatre cents pages, dont la suite de cet article constitue la conclusion¹. Nous ne parlerons pas de son contenu sémantique, qui est de peu d'importance ici, mais nous mentionnerons souvent ses caractéristiques syntaxiques. Nous espérons que le lecteur nous pardonnera ce manque de modestie, car c'est ce qui nous empêche de tomber dans une techno-fiction chimérique, dont l'intelligence artificielle, après bien d'autres, s'est fait la spécialité.

Nous articulerons notre exposé en posant la question de l'automatisation des cinq actions nécessaires à la rédaction d'un ouvrage de ce genre :

- sélection des sujets ;
- élaboration du plan ;
- formulation des phrases ;
- choix des mots ;
- dessin des caractères.

Les cinq sections qui suivent examinent ces problèmes par difficulté croissante, c'est-à-dire à l'envers. Toutefois, elles concerneront moins la possibilité de cette automatisation, ce qui nécessiterait un nouvel ouvrage, que ses conséquences sur l'activité scientifique elle-même. En effet, que certaines innovations techniques aient permis des expérimentations irréalisables auparavant, cela n'est plus à démontrer. Mais ce qui nous intéresse se situe, pourrait-on dire, à un *méta-niveau*. Il s'agit de percevoir l'impact d'une technique sur l'*imagination scientifique*.

On connaît les thèses désormais classiques [2] sur l'art du XIX^e siècle : l'impressionnisme est issu des nouvelles sensations provoquées par la vitesse des locomotives et la foule désordonnée des grandes villes. On a également établi [8] que la nouvelle organisation du travail imposée par la révolution industrielle était responsable du développement des sciences humaines. À notre connaissance, on manque de semblables analyses des sciences exactes. Aussi, nous relayerons l'histoire des sciences par celles des idées, de l'art et des sociétés pour trouver le fil conducteur dans les bouleversements imposés par l'informatique.

¹ On en trouvera l'introduction dans : Emmanuel Saint-James, « Introduction du *Traité de Programmation Applicative (de Lisp à l'Assembleur en passant par le λ -calcul)* », *BIGRE* 73, juin 1991 (numéro spécial *Les langages applicatifs dans l'enseignement de l'informatique*), 86-96. {Ndlr}.

1. Les caractères

Cet ouvrage a réduit volontairement son ensemble de caractères à ceux d'un clavier usuel, si l'on excepte les chapitres de mathématiques pures, où l'on a respecté l'usage séculaire de signes typographiques d'ailleurs peu nombreux. Trois polices seulement ont été utilisées : deux pour différencier le texte français des textes informatique et mathématique, et une pour signaler à la fois de la syntaxe abstraite dans ceux-ci et la première occurrence d'un terme technique dans le texte français. On s'est également refusé les soulignés et les mots en majuscules.

Cette attitude s'oppose à l'exploitation systématique des innovations en matière d'impression. On ne voit pas, en effet, de quel droit une découverte exigerait un nouvel apprentissage pour être communiquée, discutée, améliorée et surtout réfutée, ni pourquoi elle se priverait des idées de chercheurs aux ressources matérielles limitées.

Cette préoccupation sociale se double d'une argumentation épistémologique. L'ordinateur permet le contrôle expérimental du raisonnement ; qui s'en prive risque des incorrections qu'une analyse syntaxique corrigerait, des truismes qu'un optimiseur éliminerait, des incohérences qu'une inférence de types dénoncerait. Aujourd'hui, écrire des formules sans les vérifier électroniquement, c'est ignorer le progrès. Et ce n'est pas seulement à l'époque de la naissance de l'informatique que l'on régresserait, mais jusqu'à la préhistoire du calcul, car les mathématiques ont peu changé leur méthode depuis. Paradoxalement, la première des sciences est la dernière à bénéficier des produits industriels ; lui interdire relève de l'habitude.

Néanmoins, transcrire tous les énoncés mathématiques dans le jeu de caractères du premier terminal venu n'observerait que la lettre de ce nouvel état d'esprit. Cette matérialité soudaine d'activités intellectuelles est encore balbutiante ; la considérer avec intérêt n'est pas l'adopter sans réserve. L'écriture bi-dimensionnelle des mathématiques permet une concision hors de portée des langages de programmation actuels. Indices, exposants et barre de fractions reposent sur la signification de l'emplacement d'un signe par rapport à ses voisins, procédé que les analyseurs lexicaux amputent d'une dimension, trahissant une conception contemporaine du ruban perforé. Attendre que l'informatique comble ce handicap est justifié. Mais définir une syntaxe platement linéaire sur un alphabet ostensiblement ésotérique est désormais suspect : la moindre fantaisie graphique demande un travail considérable qui aurait plus profité à l'expressivité. À notre époque, une

syntaxe inédite devrait s'accompagner de vérificateurs au moins aussi puissants que le plus sophistiqué des compilateurs.

Voir dans cette rigueur un dirigisme maniaque serait méconnaître un précédent que la bibliophilie [7] nous remémore : l'émergence de la notion mathématique de *variable* grâce à l'imprimerie. En effet, l'assurance d'une reproduction à l'identique de toutes les occurrences d'une lettre autorise sa substitution aux périphrases manuscrites, dont la longueur prévenait les confusions. À l'heure de l'expansion automatique d'abréviations par les éditeurs de texte, écrire *x* ou *la première quantité* ou encore *la distance entre les deux centres* semble indifférent : la réduction du nombre de caractères n'est plus qu'un cas particulier de sa modulation. Mais relisons les algébristes de l'époque. Ils vantèrent *une science brève et claire* permise par la fiabilité du système d'impression. La concision est la condition de l'intelligibilité d'un discours ; et cette règle vaut pour tous les modes d'expression : *seul le mémorisable est compréhensible* remarque un compositeur [13] qui s'appuya sur des invariants musicaux insoupçonnés pour produire des chefs-d'œuvre d'athématisme et d'atonalité.

Mais revenons à cette influence du matériel sur l'intellectuel, apparemment si contraire à la plus immatérielle des sciences. On a peine à croire que l'abstraction algébrique repose sur la standardisation d'une mécanique. Pourtant le paradoxe est explicable. Un texte rédigé sur un support déjà connu, par exemple un alphabet séculaire, profite des acquis antérieurs comme d'une seconde nature. L'effort de déchiffrement devient inconscient. Poursuivre la standardisation de la communication scientifique, voilà ce que suggère l'imprimerie mariée aux mathématiques, sans regret pour la grâce des écritures cursives.

2. Les mots

Cet ouvrage définit plus de cent termes. Un programme a vérifié l'absence de références en avant dans ces définitions, qui totalisent plus de deux mille mots. À l'inverse, il a confirmé l'utilité de cette centaine de mots en assurant que chacun d'eux avait au moins une occurrence dans un texte qui en contient mille fois plus.

De telles vérifications sont humainement irréalisables. L'idée même de les effectuer n'a peut-être jamais effleuré les esprits avant l'apparition de l'informatique. Mais elles répondent à un besoin qui lui est antérieur. Le

discours scientifique, en effet, a toujours recherché un vocabulaire rigoureux. La détection automatisée des références en avant commence enfin à satisfaire cette vieille exigence. Et des contrôles semi-automatiques ont permis d'aller plus loin.

La première occurrence d'un mot est toujours significative de son importance. La *fonction*, héroïne de cet ouvrage, y apparaît avant même la première définition, et dans le texte de la deuxième. Autrement dit, ce terme fondamental n'est pas défini. Y remédier serait repousser le problème aux mots dont on se servirait. C'est un choix, justifié par la longévité de ce concept dans l'histoire des sciences. Plus généralement, un programme a exhibé tous les mots indéfinis utilisés dans les définitions. Ils indiquent les connaissances indispensables à sa lecture et leur mise en évidence a permis de tempérer un jargon trop laxiste.

On a proscrit les emprunts au vocabulaire de l'électronique, puisque l'informatique en est indépendante, ainsi que les adjectifs relatifs au comportement du matériel. Les noms trop vagues, les sigles, les américanimes, les néologismes n'ont même pas été admis entre guillemets, artifice syntaxique que nous nous sommes interdit : tous ces mots ont été bannis, sans égard pour leur popularité. De ce fait, cet ouvrage se caractérise autant par les mots qu'il définit que par ceux qu'il rejette.

L'histoire des sciences nous invite à cette témérité. On se rappelle [4] que pendant des décennies, la physique admit la notion d'*éther* dont il ne reste plus rien aujourd'hui. Une détection des indéfinis dans les textes d'alors aurait certainement soulevé la question de son utilité. En tout cas, qu'une science déjà solidement constituée se fourvoie si longtemps dans un tel pont-aux-ânes suggère de traiter avec circonspection de trop grands succès lexicaux. L'unanimité provient moins souvent de l'objectivité que d'une équivocité dont chacun s'accommode.

Précisément, traquer l'ambiguïté fut un autre souci. L'analyse automatique en a révélé quelques unes, invitant à distinguer des termes souvent confondus comme appel et application, fermeture et clôture, recyclage et récupération. Ces dédoublements sont nés des contradictions ou des truismes trouvés dans les explications initiales. Si appel et application étaient synonymes, *une macro écrasante modifie son appel à sa première application* serait un charabia incompréhensible. Montrer qu'une liaison ne conditionne pas la portée exige de différencier fermeture et clôture. Expliquer les performances contrastées de deux exécutions itératives est impossible si l'on

désigne pareillement un recyclage et une récupération. Prédire l'échec industriel des machines abstraites ne vaut que pour une frontière bien établie entre langages évolués et langage d'assemblage.

On le voit : dénoncer les abus de langage et l'inflation du vocabulaire scientifique n'est pas un purisme ; c'est la condition même de sa validité. Il faut opposer à la dégradation de notre environnement sémantique, une véritable *écologie lexicale*. Et c'est justement parce que les machines ne comprennent pas, qu'elles peuvent infailliblement détecter l'usage inédit d'un mot, risquant d'en altérer le sens. Figurer le vocabulaire, comme on figea naguère les caractères, c'est préluder à un bouleversement aussi déterminant que le symbolisme algébrique.

Symétriquement à la polysémie, on a surveillé la synonymie. Déjà, les sciences naturelles durent identifier les espèces vivantes sous des latitudes différentes, décrites par des auteurs différents, dans des langues différentes. À l'heure des télécommunications, on ne devrait plus perdre du temps avec ces duplications, involontaires ou non. Les synonymes sont les parasites de la science : ils en freinent le développement tant ils en pompent la sève. Programmons leur éradication.

On peut imaginer l'avenir de la communication scientifique sur la base d'une analyse systématique des mots employés dans un article. Les termes techniques seraient automatiquement repérés, d'où une fiabilité infiniment supérieure à nos actuelles listes de mots-clés, empiriquement établies. À l'inverse, chaque terme nouvellement introduit devra s'accompagner d'une preuve de non-synonymie. Au bout de cette procédure, apparaît la possibilité de breveter un concept scientifique.

Cette perspective peut choquer, tant elle risque d'entraver la libre circulation des idées. Mais elle ne surprend pas de la part d'une époque qui dépose des brevets pour un être vivant obtenu par manipulation génétique. On se souvient de la rente exceptionnelle que les inventeurs de la photographie obtinrent de leur gouvernement. Le député rapporteur du projet expliqua [1], avec une compétence qui laisse rêveur un contemporain de la découverte du virus du sida, pourquoi le résultat des longues années de recherche dans la *chambre obscure* ne pouvait être breveté : savoir que le bromure d'argent s'impressionne définitivement permettait à chacun de profiter immédiatement de ces patientes recherches. Or, en informatique, de tels cas sont quotidiens. C'est faute de légiférer sur ces questions que les équilibres économiques entre pays producteurs d'informatique ont

été rompus et que la recherche fondamentale, privée de ses retours sur investissements, est délaissée.

Néanmoins, cette informatisation du lexique scientifique pose de graves questions. On sait que le fichage informatisé des noms propres, même à un codage près, nécessite une autorisation légale. Un sociologue [10] a pu avancer le paradoxe que cette informatisation protège peut-être plus les libertés que les méthodes manuelles, dont les abus sont moins rapidement décelables. Mais on aurait tort de conclure que le problème du fichage est aujourd'hui résolu ; car celui des noms communs n'est pas aussi anodin qu'il n'y paraît.

On peut déjà en voir le danger dans les traductions. La locution anglaise *artificial intelligence* ne peut se traduire littéralement : le sens anglais de *intelligence* est plutôt *savoir*, sens qui ne s'est conservé en français que dans l'expression désuète *intelligence avec l'ennemi*, qui signifie une communication de renseignements, non une aide à la compréhension de, par exemple, ces problèmes. Si on l'avait traduit par le modeste *savoir artificiel*, ou même par *raisonnement artificiel*, qui nous paraît plus juste, il est probable que le crédit accordé à cette activité en eût été différent.

Et ce cas n'est rien, face au dynamisme imprévisible d'une langue. Dans cet ouvrage, nous avons dénoncé les métaphores et les métonymies comme autant d'obstacles pédagogiques, que seule une étude historique pouvait surmonter. Mais elles ne sont pas fortuites. Avant qu'un concept soit bien dégagé, il est fatal de le désigner métaphoriquement par un mot commun, avant qu'un glissement de sens ne le consacre entièrement. Et à l'inverse, le phénomène peut rejoindre le sens premier : s'il devenait possible de breveter un concept scientifique, disparaîtrait le sens métaphorique de la formule *le marché de la science*, par laquelle la sociologie désigne *la circulation des valeurs symboliques entre des producteurs qui, à la limite, n'ont pour clients que leurs concurrents* [3]. Priver un mot de ses sens figurés favoriserait l'acquisition de la science constituée, mais risquerait de gêner l'élaboration de connaissances nouvelles.

3. Les phrases

Cet ouvrage contient quatre mille phrases, censées expliquer six cents sessions ou fonctions. Quiconque comprend ces programmes trouve ces explications redondantes. Un souhait en découle : déduire ces paraphrases automatiquement à partir du texte informatique.

Cet ouvrage s'y est préparé, en s'imposant des contraintes non seulement typographiques et lexicales, mais aussi syntaxiques. Les notes en bas de page ont été confinées à des remarques terminologiques, liées en particulier aux questions de traduction. On conserve ainsi la séquentialité d'un résultat informatique, ces problèmes devant être éliminés par la standardisation du lexique. Citons également l'absence de formulations interrogatives ou exclamatives, des points de suspension et de la locution *et cætera* : ils s'opposent trop à l'assurance, la sérénité et la précision de l'exposé scientifique. À chaque fois que nous n'avons pas respecté ces contraintes, les phrases écrites étaient effectivement contestables, partisanses ou incomplètes. Pour autant qu'il nous soit permis d'en juger, leur remplacement n'a pas amené cet ouvrage en dessous de la qualité littéraire moyenne des textes scientifiques contemporains.

Ainsi, on s'est rapproché du laconisme des analyseurs de programmes. L'inférence de type, en effet, est déjà la production d'éclaircissements sur le but d'une fonction. Exécuter itérativement une récursivité, anticiper une recopie inutile, c'est maîtriser, au moins en partie, le coût d'un algorithme. Si l'on pouvait pousser ces techniques jusqu'à une rédaction automatique en français, le discours scientifique atteindrait une objectivité insurpassable. On évacuerait ces sempiternels jugements de valeur que commande une époque obsédée par la transformation d'un résultat scientifique en avantages technologiques.

Ce rapprochement des langages formels et des langues humaines n'est pas un postulat, mais le résultat d'une observation. Comparons les principes de l'informatique et de la littérature. Passons rapidement sur l'utilisation obligatoire d'un lexique : l'équivalence entre une variable non déclarée et un barbarisme est incontestable. De même, on trouve une commune obéissance à des règles de syntaxe, même si ces règles sont moins rigides ici que là : un lexème incongru est un solécisme, que le langage soit formalisé ou non. À la frontière de la syntaxe et de la sémantique, on peut identifier amphibologie et programme non typable : un membre de phrase référant un mot trop éloigné est comme une fonction s'appliquant sur un argument inapproprié, surtout par suite d'un parenthésage erroné. De même, un pléonasme n'est pas sans rapport avec une redondance de calcul.

Mais venons-en au style. Dans les deux domaines, les répétitions sont mal vues. Plus généralement, la concision est recherchée. Bien écrire, dit-on, c'est trouver une formulation dont on ne peut rien retirer. On atteint le maximum d'expression en remplaçant une périphrase par le mot juste et on

aide encore plus la mémorisation par des procédés comme la versification, les rimes, les allitérations, l'inversion de génitifs ou le détournement de formules connues. Or, le programmeur n'a pas d'autres buts : éliminer les branches de programme inutiles, choisir l'instruction la plus courte ou la plus rapide, réutiliser une fonction déjà écrite, réduire la taille des composants du programme pour qu'ils tiennent chacun sur un segment de la mémoire. Gagner un octet, obtenir l'octosyllabe, la différence n'est pas si grande.

On peut aussi bien tirer la conclusion que le programmeur fait de la prose sans le savoir, ou que tout écrivain scrupuleux est un informaticien qui s'ignore. Mais on ne se formalisera pas de cette parenté entre littérature et programmation : que les analyseurs de programmes nous aient permis de traquer systématiquement les fautes de français et les maladroites de style dont nous sommes coutumiers plaide en sa faveur. Les algébristes des débuts de l'imprimerie et les écrivains soucieux de leur style avaient une doctrine commune : une pensée libre son fond en contraignant sa forme. L'opposition entre lettres et sciences n'a pas deux siècles ; elle n'en aura pas plus.

Le grec et le latin servaient autrefois de critère de sélection pour les classes scientifiques. Le thème, en tant qu'exercice scolaire sur une langue morte, a cédé la place à la maîtrise des langages formels : on préfère aujourd'hui une belle expression Lisp à un latin de cuisine. Un parallèle avec l'histoire de l'architecture se dessine. À la fin du XIX^e siècle, les ingénieurs formés à la physique ont rivalisé avec les architectes pétris d'histoire de l'art. On sait de quels anathèmes furent frappés ces nouveaux venus, osant délaïsser la pierre et le bois pour employer des matériaux considérés comme froids : métal, verre ou béton. Ils construisirent pourtant des tours radieuses, avant que la standardisation ne produise des cités infernales.

L'histoire, sans doute, se répétera. Et plus particulièrement avec le français, dont la précision l'a déjà désigné comme langue du droit par excellence. Après quelques chefs-d'œuvre où l'écriture informatisée n'étouffera pas le style, s'installera une production littéraire comparable à trop d'habitations modernes : irréprochable en théorie, insupportable en pratique.

· En définitive, l'informatisation de la syntaxe n'est pas moins exempte de dangers que celle du lexique. Nous avons parlé de comportement schizophrénique d'un système, suite à une perturbation dans l'internement de ses symboles. Cette insistance sur la collision des vocabulaires psychiatrique et informatique n'était pas fortuite. Une société qui

interne systématiquement ses schizophrènes, même les plus inoffensifs, s'interrogera sur les programmes qui n'internent pas leurs symboles, ou, plus généralement, qui énoncent des phrases en contradiction avec leur comportement. Rien de tel qu'une erreur dans un message d'erreur pour perdre son temps et presque sa raison. La loi punit déjà les outils menaçant notre intégrité physique ; l'informatique pose aujourd'hui la question de la préservation de notre intégrité psychique. Dès lors, on peut se demander si l'on ne devrait pas réglementer la production automatique de phrases mettant en cause l'utilisateur du programme.

À l'inverse, réduire le discours scientifique à des phrases reconnaissables par des grammaires formelles serait dangereux. Que la partie du français qui leur échappe soit minime ou non, peu importe ; car à mesure que le discours formalisable sera automatisé, notre intérêt pour cet inexplicable s'accroîtra.

Il ne s'agit pas seulement ici de défendre une activité poétique que d'aucuns trouveront dépassée. Comme le lexique, la syntaxe d'un domaine scientifique nouveau est floue, tant que la causalité que l'on cherche à établir est mal assurée. Contraindre la forme de son discours n'est productif que si les règles sont déduites des propriétés de ce dont on veut rendre compte, non pas imposées autoritairement de l'extérieur. Ce serait une censure proche de l'appauvrissement des dictionnaires, condamnant le discours scientifique, et plus généralement la liberté de pensée, à des allusions. Inutile d'en donner des exemples historiques ou littéraires : il en vient mille neuf cent quatre-vingt-quatre à l'esprit.

4. Le plan

Cet ouvrage se décompose en parties, chapitres, sections, paragraphes et phrases. En partant de l'ensemble des concepts à expliquer, on a d'abord repéré des degrés de complexité, que l'on a ordonnés pour éliminer les références en avant. On a ainsi obtenu les quatre parties. Sur chacune d'entre elles, on a recommencé l'opération afin d'aboutir aux chapitres. On a agi pareillement sur chacun d'eux pour trouver les différentes sections, puis les paragraphes et enfin les phrases, chaque phrase exprimant une et une seule idée. En bref, cet ouvrage consacré à la récursivité a été conçu récursivement.

Bien entendu, plusieurs ajustements ont dû être effectués en cours de rédaction. On a parfois introduit un découpage non prévu initialement, un concept sous-jacent n'étant pas explicité. De même qu'un *menu déroulant*

n'est lisible que s'il contient moins de sept entrées, aucune subdivision de cet ouvrage ne contient plus de sept unités de niveau inférieur ; c'est aussi le cas de cet article, même si l'on n'a pas nommé les sous-sections, cette question nous semblant secondaire. Enfin, une analyse statistique du nombre de mots nouveaux par chapitre a permis de répartir plus uniformément les points d'achoppements. Les chapitres encore trop riches lexicalement, en particulier ceux où figurent beaucoup de synonymes, concernent, et ce n'est pas un hasard, des domaines toujours en gestation.

De la sorte, cet ouvrage est dépourvu de redites, ou tout au moins les signale explicitement pour assurer la compréhension, exactement comme on ne tolère des redondances dans un programme qu'à des fins de sécurité. La densité résultante d'informations peut sembler inadéquate à la pédagogie. Mais elle constitue un *méta-apprentissage* de la lecture d'un programme qu'on cherche à comprendre. De même que l'électronique vise à réduire l'encombrement des mémoires, l'informatique poursuit une plus grande concentration des idées sur un même nombre de caractères. Il faudrait que les produits de l'homme, comme ceux de la nature, ait une densité inversement proportionnelle à leur volume. On cesserait enfin de rapporter l'intérêt d'un ouvrage à sa taille, mais plutôt à celle des textes qu'il dispense de lire.

D'autre part, le graphe des références entre concepts n'établissant qu'un ordre partiel, nous avons expérimenté plusieurs extensions à un ordre total, afin de réduire les références en arrière trop éloignées. Il n'est que trop tentant de présenter chronologiquement des découvertes scientifiques ; mais il n'en sortira jamais un exposé rationnel et pédagogique. De même qu'un choix de 26 caractères, aussi beaux soient-ils, ne donne pas nécessairement une belle typographie, une suite d'exemples, aussi représentatifs soient-ils, ne suffit pas à produire un discours intelligible : c'est leur succession que le lecteur doit ressentir comme allant de soi. L'intérêt d'un traité réside moins dans les techniques décrites, qui sont le plus souvent connues de longue date, que dans leur articulation et leur compatibilité.

Cette compatibilité est ce qui reste à automatiser. On vise le *zéro défaut* dans des produits matériels ; une procédure de qualification des produits intellectuels pourrait bien nous amener le *zéro contradiction* dans les ouvrages scientifiques. Mais ici, de nouveau, certaines interrogations apparaissent.

Elles sont assez bien illustrées par le premier article consacré à Lisp [11]. C'est là que figure le magnifique programme de dérivation algébrique,

étudié dans le chapitre consacré aux fonctionnelles récursives. Rappelons qu'il repose sur une combinaison de **maplist** et de **eq** appliqué aux listes. Mais au début de l'article, il est dit que **eq** est indéfini sur les listes. Étant donné l'importance de Lisp dans l'histoire de l'informatique, on est pris d'un sentiment de malaise à l'idée qu'un système expert en communications scientifiques aurait rejeté un texte aussi capital. La compétence a parfois des effets pervers.

Cet exemple n'est encore pas très grave, car il s'agissait d'une faute d'inattention, à laquelle les auteurs auraient pu remédier si on la leur avait signalée. Mais l'histoire des sciences est pleine d'ouvrages où l'auteur feint de condamner des idées qu'il a en fait grand plaisir à exposer. C'est par crainte de la censure qu'il est obligé de conclure des centaines de pages en faveur d'une théorie, par le regret de son incompatibilité avec le dogme en vigueur. Mais l'essentiel est passé : on a installé le doute quant aux explications courantes, même si on leur substitue d'autres qui ne sont pas toujours plus scientifiques. Et il ne faudrait pas croire que ces stratagèmes appartiennent à des époques passées, dominées par l'obscurantisme. Que l'on songe au temps mis par la biologie moléculaire [12] ou la tectonique des plaques [9] pour s'imposer au cours de ce siècle. L'on a déjà remarqué [6] que toute société est fondée sur un certain état des connaissances ; le remettre en cause, c'est contester ceux qui en profitent et ont les moyens de se défendre.

Le style n'est que l'ordre et le mouvement qu'on met dans ses pensées a dit un savant écrivain [5]. L'ordinateur, comme son nom l'indique, est idéal pour ce premier travail. Mais c'est le deuxième qui donne son intérêt à un texte, voire en livre la clé quand il paraît contradictoire ou incohérent. Dans les lignes, il y a l'ordre, entre les lignes, il y a le mouvement. C'est, si l'on veut, ce qui détermine le *méta-plan*.

5. La sélection

Cet ouvrage explique des concepts et les illustre par des exemples ; ce sont les données de départ d'un programme qui le rédigerait plus ou moins automatiquement. Elles proviennent de la littérature scientifique, à une reformulation près nécessaire à l'unité du discours. On les a sélectionnés sur la base de leurs rapports : les phénomènes singuliers, les théories dont les théorèmes n'éclairent qu'elles-mêmes ont été rejetés, sans égard pour le nombre parfois élevé d'articles qui leur sont consacrés. De nouveau, cette

attitude s'objective par le graphe des relations entre les termes. Et se profile à l'horizon la production automatique des bibliographies.

À première vue, elle se fonderait sur les lectures de l'auteur. Mais cette réduction de la bibliographie à une tranche d'autobiographie recèle un grave arbitraire. Il est rare qu'on prenne connaissance d'une découverte par le premier texte qui la décrit. Celui-ci, en effet, est souvent confus, dans son fond comme dans sa forme. De ce fait, les résultats essentiels sont immanquablement repris dans d'autres textes, éventuellement par d'autres auteurs, et une présentation plus avantageuse leur vaudra une diffusion plus large. Se produit alors un phénomène de boule de neige : cette disponibilité plus grande amènera la substitution de la copie à l'original dans les bibliographies, et cela non seulement parce que statistiquement elle sera plus souvent lue, mais aussi parce que, pour minimiser le travail des rapporteurs, on citera de préférence ce qui leur est le plus accessible. En dernière analyse, c'est par prévenance, voire par désintéressement, qu'un générateur de bibliographies puisera prioritairement dans les œuvres des membres du comité de lecture.

Le problème peut sembler mineur. Mais un précédent laisse penser le contraire. La découverte de l'imprimerie a provoqué la diffusion massive des connaissances dues à des civilisations à transmission manuscrite voire orale. Il s'en suivit une occultation de cet héritage qui affecta même le lexique. Des étymologies aussi ethnocentristes que fantaisistes parachevèrent cette évolution. Depuis peu, l'histoire des sciences a rétabli tellement de filiations, qu'on en vient à se dire que la locution *pays en voie de développement*, qui se voudrait bienveillante, est en définitive une insulte à un passé infiniment plus glorieux que celui de bien des pays dits développés. Mais entretemps, que d'injustices. Pour les valeurs symboliques aussi, on ne prête qu'aux riches.

Plus on arrive à examiner l'activité scientifique d'un point de vue historique impartial, et plus naît l'impression que l'innovation n'est pas localisable : c'est un phénomène continu d'amélioration et d'adaptation infinitésimales d'une technique déjà connue. Une idée semble toujours avoir été déjà émise, éventuellement sous une autre forme, pour une autre application, avec d'autres métaphores. Nous n'avons d'ailleurs aucune illusion sur les quelques concepts ou fonctions élaborés exprès pour cet ouvrage : comme les autres, ils ont certainement déjà été publiés, y compris l'idée qu'exprime cette phrase.

À ce risque d'une bibliographie partielle et partiale, s'ajoute une autre remarque. Il est surprenant que l'activité scientifique se soit accommodée d'un terme aussi lourdement connoté que celui de *bibliographie* : la science fonctionne par démonstrations, s'opposant consciemment aux arguments d'autorité d'écritures plus ou moins sacrées. Laisser la référence bibliographique prendre la place des résultats d'expériences est contraire à l'esprit scientifique. On peut admettre qu'un article y oblige, mais ce ne serait pas acceptable dans un traité, qui est précisément l'occasion de confronter rationnellement les connaissances, délaissant la question de leur paternité. Les techniques nécessaires à, par exemple, l'appel par nécessité, les exécutions itératives, les modes de liaison et la récupération ont été souvent exposées ; leur interaction l'a été beaucoup moins. Rédiger un traité est une affaire de scientifique ; établir une bibliographie est de la responsabilité d'un historien.

L'on s'achemine vers une rédaction scientifique exclusivement électronique : plus personne ne voudra se rendre dans une lointaine bibliothèque à la complétude illusoire, lorsqu'il suffira d'un *clic* de souris pour voir apparaître sur son écran ou son imprimante un texte rédigé à l'autre bout de la planète et à une autre époque. Aujourd'hui, ceux qui ne pourront s'équiper des moyens de reproduction informatisée sont menacés d'exclusion de la mémoire collective. Un retournement de situation s'opère, que l'on pourrait trouver bien mérité si ses conséquences n'étaient pas aussi graves : une légende tenace situe la découverte de l'imprimerie bien loin du lieu de son invention réelle ; un effet de boomerang est en train d'installer non loin de là la naissance de l'ordinateur.

Conclusion

Par une certaine ironie de l'histoire, l'homme n'a surmonté la complexité de la nature qu'en enfant démesurément ses propres productions. Une vie ne suffit pas à explorer une bibliographie. On est mieux équipé aujourd'hui pour affronter une forêt tropicale qu'une jungle éditoriale. À la crainte superstitieuse, a succédé l'intimidation technocratique. L'esprit scientifique se doit de combattre une dérive aussi flagrante de son but initial.

Cette question concerne tous les scientifiques, mais plus particulièrement l'informaticien. Ce qui freine la science aujourd'hui, c'est moins quelques rares énigmes encore irrésolues, que la maîtrise des relations entre les connaissances scientifiques déjà connues. Nous en avons donné des exemples

en informatique, mais c'est encore plus net d'une discipline à l'autre. Un physicien sait calculer le gain de productivité d'une nouvelle technique, et un sociologue connaît le coût social d'un changement de mode vie imposée par une innovation ; mais personne ne domine cette réaction en chaîne. Or c'est ici que l'informatique a sa place : elle cherche moins à créer des connaissances, qu'à exhiber leur relation.

On s'occupe aujourd'hui de cartographier le génome humain, malgré les atteintes aux libertés individuelles qui risquent d'en découler. Ce que l'informatique propose aujourd'hui, c'est de cartographier les connaissances. Il s'agit de donner à chacun la possibilité de retrouver la source exacte d'une information, de mesurer sa fiabilité, de tester sa compatibilité avec une autre. La mise en œuvre d'un tel projet, ses retombées sociales et économiques sont floues. Il nous faut, une fois encore, expérimenter. Nous attendons d'un physicien qu'il vive près de la centrale nucléaire qu'il a conçue, d'un chimiste qu'il s'administre le médicament qu'il a fabriqué. C'est à nous, informaticiens, d'appliquer nos méthodes à notre activité quotidienne, si nous voulons prouver, contrairement à une idée reçue, que l'informatique peut être un facteur de liberté.

Références

Cet article épistémologique s'autorise quelques références que nous refusons déjà à un traité scientifique ; mais il va de soi que nous déplorons cette concession.

- [1] ARAGO *Rapport sur le daguerréotype (Musée d'Orsay)*
- [2] WALTER BENJAMIN *Sur quelques thèmes baudelairiens*
- [3] BOURDIEU *Homo Academicus*
- [4] BRILLOUIN *La science et la théorie de l'information*
- [5] BUFFON *Discours de réception à l'Académie française*
- [6] CANGUILHEM *Galilée : la signification de l'œuvre et la leçon de l'homme*
- [7] CIFOLETTI *Les algébristes français du XVII^e siècle à la Bibliothèque nationale*
- [8] FOUCAULT *Les mots et les choses*
L'archéologie du savoir

- [9] GOHAU *Une histoire de la géologie*
- [10] HEILMAN « Le policier, l'ordinateur et le citoyen », *Culture Technique*
numéro 21
- [11] MC CARTHY ET ALII *Recursive functions of Symbolic Expressions &
their computations by machines*
- [12] ROSTAND *Histoire de la génétique*
- [13] SCHOENBERG *Le style et l'idée*