INTRODUCTION
QUELQUES REPÈRES HISTORIQUES

Il est inutile de rappeler que nous vivons dans une civilisation transformée par l'informatique qui a ouvert des possibilités gigantesques en gain de temps et en facilité de traitement, a créé de nouvelles fonctionnalités, a permis l'accès aux informations et aux archives. Mais en même temps de nouvelles questions sont apparues : où sont stockées les données ? Qui y a accès ? Sont-elles en sécurité ? Sommes-nous dépendants de machines, de logiciels, de sociétés commerciales ? Faute de posséder les bases d'une véritable culture numérique, qui ne se limite pas à l'utilisation d'un traitement de texte ou à la participation à un réseau social, le citoyen du XXIe siècle en est bien souvent réduit à être de plus en plus un consommateur et non un utilisateur averti et libre de ces technologies numériques. Par exemple, tout un chacun est-il sûr de pouvoir nommer son OS, saisir du texte dans une autre langue que la sienne sur un ordinateur, savoir quand utiliser un éditeur de texte plutôt qu'un traitement de texte, définir son identité numérique, dire ce qu'est un cookie ou encore pourquoi les résultats dans Google traduction sont parfois inutilisables ?

C'est pour démythifier l'informatique d'aujourd'hui qu'il est si important d'étudier celle d'hier, car tout objet numérique résulte certes de croisements entre les techniques mais révèle aussi les intérêts économiques et les visions politiques d'une société. Avoir des repères historiques précis permet de comprendre son environnement informatique et d'être en mesure de le reprendre en main.

Les premiers ordinateurs

Avant le XXe siècle

Sans s'attarder, il est quand même peut-être nécessaire de rappeler que tout n'est pas né au XXe siècle : le premier calculateur analogique date du…

«Vade-mecum informatique pour les Lettres et Sciences humaines», Estelle Debouy
ISBN 979-10-413-0002-0 Presses universitaires de Rennes, 2025, www.pur-editions.fr

IIe siècle avant J.-C. Il s'agit du mécanisme d'Anticythère qui a modélisé la course des astres à l'aide d'un système d'engrenages si complexe qu'il faut attendre plus d'un millénaire pour voir apparaître des systèmes comparables dans les horloges du Moyen Âge.

À partir de la Renaissance, deux mouvements suscitent l'accroissement de la demande en matière de calcul et de traitement de l'information : la révolution scientifique et industrielle et la formation des États modernes. Ainsi Pascal conçoit-il dans les années 1640 sa Pascaline, considérée comme la première machine à calculer (additions et soustractions), pour aider son père, receveur des impôts, à effectuer ses fastidieux calculs fiscaux. C'est à la même époque qu'est inventée la règle à calcul qui restera le moyen matériel de calcul le plus couramment employé par les scientifiques, les ingénieurs et les étudiants jusqu'à l'apparition des calculettes électroniques vers 1970. À la fin du siècle, en 1694, ayant pris connaissance des travaux de Pascal, Leibniz met au point la première machine permettant la multiplication et la division.

Enfin, citons l'invention, datée de 1838, qui constitua une révolution dans les communications et qui repose sur un système de signes permettant d'utiliser l'énergie électrique pour transmettre des informations à distance : le morse (qui doit son nom à son inventeur). Quelques années plus tard, en 1876, est inventé le système permettant de transmettre de la voix à distance à l'aide de l'électricité : le téléphone.

Programmes, langages, terminologie : tout s'invente au XXe siècle

La première moitié du XXe siècle, avec la seconde révolution industrielle et la production de masse, se caractérise par l'apparition de nouveaux besoins de calcul et de traitement d'information. Les guerres ne font qu'accentuer ces tendances, comme le montrent notamment les problèmes de calcul et de correction automatique des nouveaux systèmes d'armes ou encore la mécanisation du travail après guerre en raison d'une pénurie de main-d'œuvre. En 1945, on voit apparaître de puissants calculateurs qui soulèvent cependant des questions essentielles : quel type de mémoire utiliser, comment rendre ces machines fiables, comment communiquer avec les machines – on parlera bientôt de programmation. La première réponse est apportée par Von Neumann qui invente un concept absolument inédit : le programme enregistré.

«Vade-mecum informatique pour les Lettres et Sciences humaines», Estelle Debouy
ISBN 979-10-413-0002-0 Presses universitaires de Rennes, 2025, www.pur-editions.fr

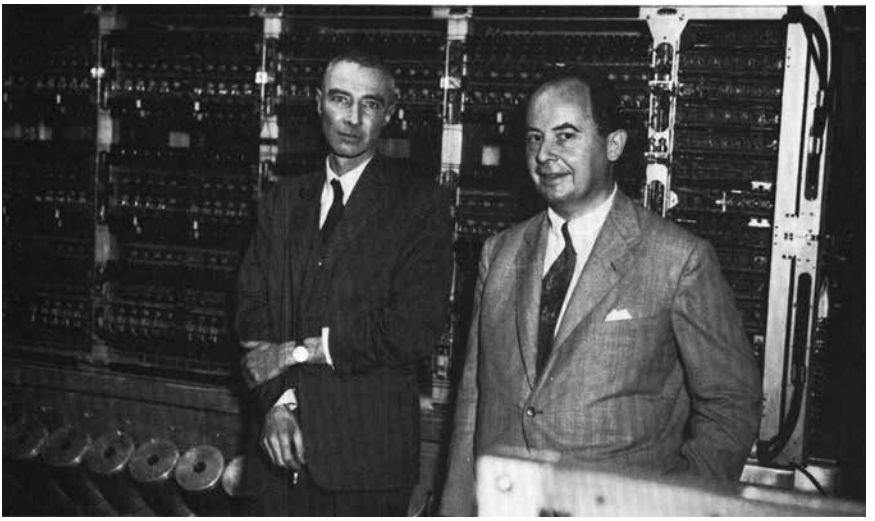

Fig. 1. – John Von Neumann (à droite) et J. Robert Oppenheimer en 1952 devant le premier ordinateur.
Source : [https://commons.wikimedia.org/wiki/File:Oppenheimer_%26_Neumann.jpg].

C'est à cette époque qu'est mis au point le premier langage évolué encore mondialement utilisé : le FORTRAN (FORmula TRANslator). Quelques années plus tard apparaît le COBOL (Common Buisness Oriented Language) qui, réagissant à la prolifération des langages, jette les bases d'une standardisation, puis le BASIC, langage de programmation simplifié à l'usage des étudiants.

Des premières machines à l'empire d'IBM

L'informatique naissante participe au changement des équilibres mondiaux : les États-Unis, sortis renforcés de la seconde guerre mondiale, sont décidés à devenir leaders sur la scène internationale et deviennent, de fait, le berceau de la troisième révolution industrielle, grâce à la synergie entre recherche universitaire et industrie, soutenue par d'énormes investissements et la foi dans l'innovation.

À partir des années cinquante, les entreprises entrevoient le potentiel des ordinateurs développés par les universitaires et prennent le risque d'en réaliser des versions commerciales. C'est le début de la domination d'IBM sur le marché, qui recrute d'ailleurs Von Neumann comme consultant. Pour commercialiser en France son *electronic data processing machine*, IBM-France souhaite le désigner par un nom moins indigeste[1]. Sollicité par la direction de l'usine de Corbeil-Essonnes, François Girard, alors responsable du service promotion générale publicité, décida de consulter un de ses anciens maîtres, Jacques Perret, professeur de philologie latine à la Sorbonne. À cet effet il

1. Cf. la figure 2 qui présente la publicité diffusée aux États-Unis en 1955.

écrit une lettre soumise à la signature de Christian de Waldner, Président d'IBM France. Le 16 avril 1955, le professeur Perret lui envoie une réponse dont voici un extrait[2] :

> « En Sorbonne, le 16 IV 1955
> Cher Monsieur,
> Que diriez-vous d'ordinateur ? C'est un mot correctement formé, qui se trouve même dans le *Littré* comme adjectif désignant Dieu qui met de l'ordre dans le monde. […] Systémateur serait un néologisme, mais qui ne me paraît pas offensant ; il permet systématisé ; – mais système ne me semble guère utilisable – combinateur a l'inconvénient du sens péjoratif de combine ; combiner est usuel donc peu capable de devenir technique ; combination ne me paraît guère viable à cause de la proximité de combinaison. […] Congesteur, digesteur évoquent trop congestion et digestion. Synthétiseur ne me paraît pas un mot assez neuf pour désigner un objet spécifique, déterminé comme votre machine. En relisant les brochures que vous m'avez données, je vois que plusieurs de vos appareils sont désignés par des noms d'agent féminins (trieuse, tabulatrice). Ordinatrice serait parfaitement possible et aurait même l'avantage de séparer plus encore votre machine du vocabulaire de la théologie. [...]
> Vôtre, Jacques Perret. »

L'« ordinateur » IBM 650 peut alors commencer sa carrière. Protégé pendant quelques mois par IBM-France, le mot fut rapidement adopté par un public de spécialistes, de chefs d'entreprise et par l'administration. IBM-France décida de le laisser dans le domaine public.

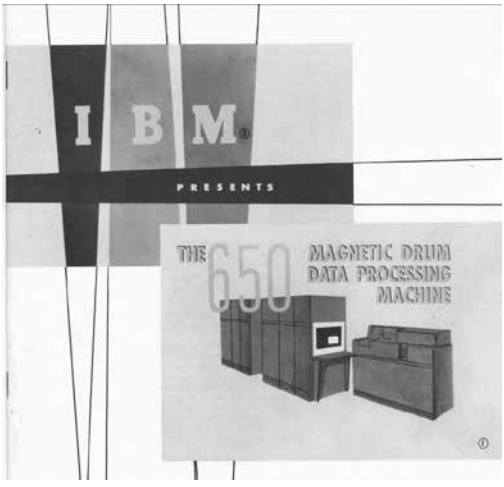

Fig. 2. – Publicité pour l'IBM 650.
Source : [https://www.computerhistory.org/brochures/doc-4372956d2be71/].

2. Loïc Depecker reproduit la version numérisée de la lettre et l'analyse : Depecker Loïc, « Que diriez-vous d'"ordinateur"? », *Bibnum*, vol. « Calcul et informatique », 2017, [http ://journals.openedition.org/bibnum/534], consulté le 2 février 2024.

D'ailleurs depuis la fin des années cinquante, plusieurs spécialistes imaginaient des termes pour désigner les activités liées à l'ordinateur : « informatique » est un mot-valise inventé en 1962 en fusionnant les termes « information » et « automatique »[3].

C'est en 1957 qu'est fondé le département d'Intelligence artificielle (IA) au Massachusetts Institute of Technology. À quand remonte l'IA ? Durant la Seconde Guerre mondiale, Alan Turing aida les Anglais à déchiffrer les codes secrets de l'armée allemande[4]. C'est lui qui, le premier, posa cette question fascinante : les machines peuvent-elles penser ? Pour y répondre, il imagina un test qui porte aujourd'hui son nom. Il s'agit pour un jury de poser à l'aveugle des questions à un ordinateur et à un être humain : on considère que le test est réussi si le jury est incapable de déterminer quelles sont les réponses apportées par l'ordinateur. A. Turing a même postulé que dans les années 2000, l'ordinateur serait capable de tromper le jury pendant cinq minutes pour 70 % des cas. L'autoapprentissage est au cœur de cette problématique car l'ordinateur apprend à partir de ses expériences en capitalisant sur la somme de données qu'il a stockées et analysées. C'est ainsi que fonctionnent les antispams par exemple. C'est pourquoi, au lieu de jeter à la poubelle ses spams, il est beaucoup plus judicieux de les enregistrer dans sa boîte d'indésirables afin de permettre aux algorithmes antispam d'améliorer leurs performances.

L'ordinateur pour saisir du texte

C'est toujours dans les années soixante qu'apparaît la nécessité de représenter chaque caractère en code traitable par l'ordinateur. Tout commence par une constatation très simple : les premiers informaticiens parlaient anglais. Et l'anglais s'écrit avec peu de choses : deux fois 26 lettres (majuscules et minuscules), 10 chiffres, une trentaine de signes de ponctuation, de signes mathématiques, sans oublier le symbole dollar. Au total, 95 caractères, auxquels il faut ajouter 33 caractères de « contrôle » comme le retour à la ligne par exemple. On a alors attribué des numéros à ces 128 valeurs à l'intérieur de 7 bits, c'est-à-dire 7 chiffres qui indiquent 0 ou 1, le huitième bit étant utilisé à des fins de vérification. Chaque caractère correspond donc à une séquence de 0 et de 1 à 7 chiffres : l'ASCII était né.

3. Puisqu'on évoque l'invention du vocabulaire de l'informatique, rappelons d'où vient le mot *bug* : en 1947, un gros calculateur d'Harvard tomba en panne ; après des recherches, on découvrit un insecte (*bug* en anglais) coincé dans un relais…
4. C'est l'objet du film *The Imitation Game* dirigé par Morten Tyldum et écrit par Graham Moore (2014).

Binaire	Hexadécimal	Décimal	Caractères ASCII
0100011	23	35	#
0100100	24	36	$
0100101	25	37	%
1000100	44	68	D
1000101	45	69	E
1000110	46	70	F

Tab. 1. – Extrait de la table ASCII.

Comme son nom l'indique, l'ASCII permet de saisir des documents en anglais mais non dans des langues comprenant des caractères accentués par exemple. Dans ces conditions, comment saisir du français ou de l'allemand ? En étendant le code ASCII à 8 bits, il a été possible d'ajouter les caractères accentués et divers autres symboles utilisés par les langues d'Europe de l'Ouest. Le huitième bit est affecté différemment d'un programme à l'autre. Dans la plupart des cas, le poste supplémentaire est utilisé pour répondre aux besoins nationaux. Cependant, les 128 premiers caractères sont toujours conservés dans leur forme originale et donc lisibles sur tous les ordinateurs. C'est pourquoi les adresses mails, les URL des sites Web et le langage HTML sont écrits en ASCII de base, c'est-à-dire sans aucun caractère accentué[5].

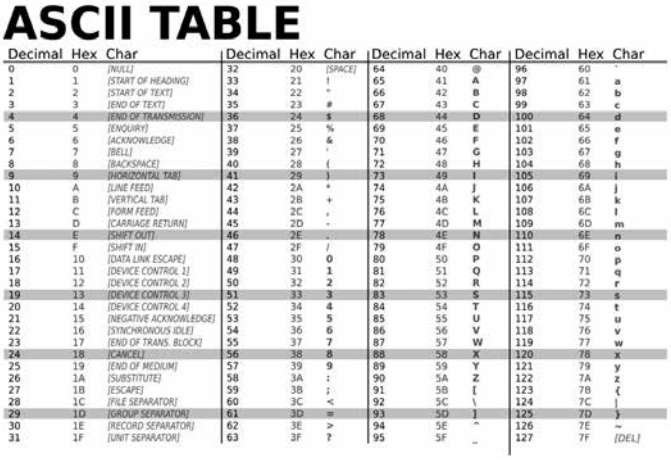

Fig. 3. – Table ASCII.
Source : [https://commons.wikimedia.org/wiki/File:ASCII-Table.svg].

Mais, les polices asiatiques, les alphabets cyrillique, grec, hébreu, etc. étaient toujours absents de l'ASCII. Pour résoudre durablement tous ces

5. En 2003 cependant le standard d'Internet Internationalizing Domain Names in Applications (IDNA) a été créé : il est donc possible aujourd'hui d'avoir un nom de domaine comportant un « c » cédille ou un « i » tréma par exemple.

problèmes de langues, au début des années 2000, s'est formé un consortium regroupant des grands noms de l'informatique et de la linguistique : le consortium Unicode. Sa tâche était de recenser et numéroter tous les caractères existant dans toutes les langues du monde. Est donc né un jeu universel de caractères : l'unicode. En 2007, le standard publié comportait environ 60 000 caractères et aujourd'hui plus de 130 000. Cela a représenté un énorme travail de normalisation de tous les systèmes d'écriture, incluant notamment le sens d'écriture, les superpositions de signes, les ligatures, etc. L'un des encodages les plus utilisés est l'UTF-8 car il présente l'avantage d'être compatible avec l'ASCII, de sorte que les parties écrites avec l'alphabet latin de base d'un texte codé en UTF-8 seront à peu près lisibles même avec un logiciel qui ne comprend pas cet encodage.

L'informatique personnelle

Revenons aux années soixante : IBM domine le marché et, plus généralement, les États-Unis ont la mainmise sur l'informatique, ce dont témoigne la prise de contrôle de Bull et d'Olivetti par General Electric en 1964. Les ordinateurs, qui se miniaturisent, évoluent rapidement, ce qui fait prédire à Gordon Moore, directeur de recherche, un doublement annuel de la puissance des ordinateurs : on parle de la « loi de Moore », qui fut observée jusqu'en 2016. La miniaturisation fait entrer massivement l'ordinateur dans les entreprises et les administrations : la petite machine d'Olivetti qu'on voit sur la figure ci-dessous fut un *best-seller* dans les bureaux d'études.

Fig. 4. – L'Olivetti Programma 101.
Source : [https://commons.wikimedia.org/wiki/File:Olivettiunderwood_programma101.jpg].

C'est à cette époque qu'est fondée la science informatique, à l'instigation de Donald Knuth qui apporte les bases théoriques à ce qui relevait jusque-là du bricolage algorithmique. C'est en préparant la publication de son étude qu'il conçut un logiciel complet de mise en page de documents scientifiques, TeX, qui est aujourd'hui un outil incontournable dans le domaine de la publication.

On commence aussi à étudier l'interface homme-machine. En ce qui concerne l'interface logicielle, c'est en 1968 que sont inventés la plupart des éléments maintenant standards en informatique personnelle tels que le système de fenêtrage, le bureau, etc. ; en ce qui concerne l'interface matérielle, c'est l'époque où apparaissent la souris et les supports de stockage externes qui deviennent de plus en plus petits alors que les capacités de stockage ne cessent d'augmenter : on passe des disquettes de 8 pouces d'une capacité de 80 Ko à des disquettes de 5 pouces 1/4 puis de 3 pouces 1/2 d'une capacité de 1,4 Mo, accessibles grâce aux deux lecteurs nommés `A :` et `B :`. Initialement dépourvus de disques durs, les micro-ordinateurs des années quatre-vingt utilisaient en effet la disquette comme support de stockage de masse avant qu'elle ne soit supplantée dans les années quatre-vingt-dix par des supports de plus grande capacité. Quand on développa des ordinateurs avec une mémoire propre interne, comme A et B étaient pris, on a naturellement choisi C pour cette nouvelle source de mémoire[6].

À la même époque, un réseau est conçu par des scientifiques, avec un soutien de la Défense, pour connecter les ordinateurs entre eux : c'est la naissance d'Arpanet, l'ancêtre d'Internet. En pleine guerre froide, il a d'emblée un usage de renseignement militaire et permet, par exemple, de détecter des essais nucléaires soviétiques en Norvège. Ce réseau s'étend progressivement et permet l'envoi de message entre deux machines : c'est le premier email, qui date de 1971. Ray Tomlinson, son inventeur, décide alors d'utiliser l'arobase pour séparer les deux parties d'une adresse électronique, soit le nom du correspondant (sophie.fonsec) du nom de domaine qui désigne le serveur où est hébergé le courrier (univ-poitiers.fr)[7]. Aujourd'hui, pour éviter le spam, l'arobase est parfois remplacée par d'autres caractères (ex. : [arobase] en toutes lettres) afin d'empêcher les robots qui scrutent les pages Web de capturer les adresses mails dans le but d'envoyer des courriers indésirables.

Unix et Linux

C'est encore dans les années soixante-dix qu'apparaît un petit système d'exploitation multi-tâches et multi-utilisateurs : Unix. Il connaît une grande descendance *via* les nombreuses versions développées à partir d'un code source librement diffusable, l'Open Source. C'est ainsi notamment qu'en 1984

6. C'est ce qui explique que, sous Windows, les documents de l'utilisateur soient accessibles depuis le disque C. Ex. : `C :\ MesDocuments\article.pdf` (cf. chap. 1).
7. On pourra lire avec profit l'article « Histoire de l'arobase », [https://www.arobase.org/culture/arobase-histoire.htm], consulté le 2 février 2024.

«Vade-mecum informatique pour les Lettres et Sciences humaines», Estelle Debouy
ISBN 979-10-413-0002-0 Presses universitaires de Rennes, 2025, www.pur-editions.fr

Richard Stallman, chercheur en intelligence artificielle au MIT, propose une alternative gratuite d'Unix : GNU, un système d'exploitation non seulement gratuit mais aussi libre. Quelle est la différence ? Un programme libre est un programme dont on peut avoir le code source, c'est-à-dire la « recette de fabrication », et qu'on peut donc copier, modifier, redistribuer. Au contraire, Windows est un système propriétaire dont le code source est conservé par Microsoft. L'idée qu'il faut « ouvrir le code » (*open source*) est très importante car c'est ce partage des connaissances qui permet à l'informatique d'évoluer plus vite. Quelques années plus tard, en 1987, apparaît un clone libre d'Unix, Minix, créé par le professeur Andrew Tanenbaum à des fins pédagogiques, volontairement réduit afin qu'il puisse être compris entièrement par ses étudiants en un semestre. C'est ce système qui servit de source d'inspiration à Linus Torvalds pour créer son noyau Linux (contraction de Linus et d'Unix). Ce projet était complémentaire avec celui de Richard Stallman : ce dernier créait les programmes de base (programme de copie de fichier, de suppression de fichier, éditeur de texte) et Linus Torvalds s'occupait du « cœur » du système d'exploitation, d'où l'appellation GNU/Linux. C'est aujourd'hui l'un des systèmes d'exploitation les plus répandus au monde, preuve de la force du modèle du développement *open source*[8].

À quoi ressemble cet OS si différent des autres ? On pourrait dire qu'il est léger, rapide, souple, complet, stable, fiable, et gratuit par essence. Sinon, il est difficile de répondre, car ce système est très personnalisable. En effet, pour répondre aux besoins de tous, différentes distributions ont été créées. De quoi s'agit-il ? Il n'existe rien de semblable sous Windows[9]. Il s'agit d'un ensemble de logiciels assemblés autour du noyau Linux pour former un système cohérent. D'une distribution à l'autre, les différences tiennent dans la convivialité, le nombre de logiciels distribués, la fréquence de leur mise à jour, l'ampleur de la communauté qui suit la distribution. On compte aujourd'hui plus de trois cents distributions Linux. Voici quelques conseils pour choisir la sienne :

- pour ceux qui débutent : Ubuntu ou Mint ;
- pour ceux qui souhaitent un système réputé : Debian, OpenSuse ou Fedora ;
- pour ceux qui veulent toujours la dernière version (on parle de *rolling release*) : Arch-Linux ou Void ;
- pour ceux qui personnalisent leurs applications dans les moindres détails : Slackware.

Terminons en mentionnant un système dérivé d'Unix : Berkeley Software Distribution (BSD). Contrairement à la licence des systèmes Linux, BSD

8. Il ne faut pas s'y tromper : bien que Linux soit quasiment inexistant auprès du grand public (autour de 4,5 % du marché des systèmes d'exploitation de bureau en 2024), un grand nombre d'appareils électroniques ou encore d'infrastructures scientifiques et de serveurs sont dominés par Linux… même les serveurs de Microsoft.
9. Si on veut, malgré tout, recourir à une comparaison, c'est un peu comme la différence entre Windows Famille et Windows Professionnel.

«Vade-mecum informatique pour les Lettres et Sciences humaines», Estelle Debouy
ISBN 979-10-413-0002-0 Presses universitaires de Rennes, 2025, www.pur-editions.fr

autorise qu'on réutilise les sources, même pour les inclure dans un produit commercial non *open source*. Résultat : MacOS X est un dérivé de FreeBSD…

Windows et MacOS : l'ère de l'ordinateur personnel

À partir des années quatre-vingt, l'ordinateur, qui avait conquis les entreprises et les administrations une décennie plus tôt, entre maintenant chez les particuliers. En 1975, la revue *Popular Electronics* annonce qu'avec la sortie de l'Altair 8800, « *The era of personal computing in every home* [...] *has arrived.* » Ainsi, pour un prix de 400 $ (soit environ 1 500 € aujourd'hui), l'Altair met à la disposition de milliers de passionnés ce qui était jusque-là réservé aux scientifiques. Le génie de Bill Gates est d'avoir compris que c'était là le début de l'ère de l'ordinateur personnel : il lance Microsoft qui n'est au départ qu'une petite entreprise de logiciel jusqu'à ce qu'IBM la choisisse pour réaliser le système d'exploitation de son IBM PC[10]. À peu près au même moment, Steve Jobs fonde Apple et lance le Macintosh, le premier micro-ordinateur convivial, à l'interface graphique révolutionnaire, qui connaîtra un succès mondial. C'est notamment le premier ordinateur à avoir un affichage graphique en couleurs !

La question des libertés informatiques

Il ne faudrait pas imaginer qu'il faut attendre 1984 – date pourtant symbolique depuis le texte d'Orwell – pour voir émerger une réflexion sur les libertés informatiques. En effet, dix ans plus tôt, en 1974, une affaire défraya la chronique : le ministère de l'Intérieur autorisa le croisement des fichiers informatiques administratifs en utilisant le numéro de sécurité sociale, créant une base de données centralisée de toute la population : c'est le projet SAFARI, qui fut révélé par *Le Monde* (cf. la figure ci-dessous).

• • • LE MONDE — 21 mars 1974 — Page 9

JUSTICE

Tandis que le ministère de l'intérieur développe la centralisation de ses renseignements

Une division de l'informatique est créée à la chancellerie

« Safari » ou la chasse aux Français

De vastes ambitions

« A la hussarde »

Fig. 5. – Extrait du journal *Le Monde* du 21 mars 1974.
Source : [https://podcasts.lemonde.fr/lheure-du-monde/202403220300-projet-safari-lhistoire-du-scandale-lorigine-de-la-creation].

10. Si l'acronyme PC est aujourd'hui employé banalement, on le doit à IBM qui inventa le PC.

Comme le projet suscita une vive opposition, il fut retiré et fut créée la CNIL afin de réfléchir au devenir des libertés individuelles et publiques dans la quête permanente d'information[11].

C'est à la même époque que les dirigeants français prennent conscience de la nécessité de développer des services performants de transmission de données. La France se dote d'un réseau télématique utilisant des terminaux simples et peu onéreux : le Minitel, qui permet de se connecter, *via* le réseau téléphonique, à des services en ligne. Il est aujourd'hui considéré comme l'une des expériences de services en ligne antérieures au Web les plus réussies au monde. Il a notamment permis à des millions de Français de se familiariser avec l'utilisation des réseaux numériques et à des milliers de développeurs de créer des services qui ont pu ensuite basculer sur Internet.

Fig. 6. – Un Minitel.
Source : [https://commons.wikimedia.org/wiki/File:Minitel_terminal.jpg].

L'ère des réseaux

Ce qui marque un tournant dans l'histoire de l'informatique, à partir des années quatre-vingt-dix, c'est bien évidemment l'émergence d'Internet, résultat de l'utopie d'un partage du savoir universel déjà cher aux Encyclopédistes des Lumières et de l'utopie informaticienne d'augmenter les moyens mis au service de l'intelligence humaine. Internet est essentiellement un réseau physique, c'est-à-dire une infrastructure réseau formée de moyens de transmission et de moyens de commutation, alors que le Web désigne l'ensemble des documents HTML accessibles *via* le protocole HTTP, comme on l'étudiera.

11. Cf. [https://www.cnil.fr/fr/maitriser-mes-donnees], consulté le 2 février 2024.

Premiers usages d'Internet

Au début, l'utilisation d'Internet était limitée par la difficulté à y chercher de l'information. La solution est inventée par deux chercheurs, Tim Berners-Lee et Robert Cailliau, qui développent un ensemble de protocoles et de logiciels qu'ils nomment *World Wide Web*. Ce système combine un logiciel de serveur, un navigateur et un éditeur de documents permettant de créer des pages Web en ligne et d'y insérer des liens hypertextes reliés à d'autres documents situés sur n'importe quel autre site Internet. Le succès du Web est tel qu'il sera confondu avec Internet.

L'utilisation du réseau ne s'est pas limitée à la consultation de pages Web. Il a été utilisé pour s'envoyer des messages électroniques. Longtemps, la consultation du courrier électronique a nécessité l'utilisation d'un logiciel de messagerie car l'email d'un côté et le Web de l'autre étaient deux composantes bien distinctes de l'Internet. C'est en 1994 qu'apparurent les premiers Webmails, mais ce n'est qu'en 1996 que ce type de service permettant de consulter son courriel depuis un site Web se popularisa, avec le lancement de Hotmail (racheté par Microsoft en 1997). Pendant dix ans, les Webmails, aux fonctionnalités limitées, furent considérés comme une roue de secours permettant aux internautes nomades de consulter leurs messages en déplacement, ou comme des services ciblant les internautes néophytes. Mais en 2004 Google lança Gmail, une messagerie adoptant pleinement les standards du courrier électronique. Aiguillonnée par Gmail, la concurrence se mit alors elle aussi à proposer des Webmails plus évolués.

On constate donc qu'à partir des années quatre-vingt-dix, tout est allé très vite : en 1990, on compte environ cent mille ordinateurs connectés et, trois ans plus tard, plus de deux millions. Afin d'encadrer l'évolution du chantier du Web, Tim Berners-Lee propose en 1994 de monter un consortium ouvert, le W3C dont la tâche est d'assurer l'interopérabilité et la standardisation des technologies Web. Mais ce n'est pas le seul défi qu'il faut surmonter : il convient d'être en mesure de se repérer dans cette gigantesque jungle qu'est devenu le Web.

Avant que Google ne s'impose comme le moteur de recherche le plus utilisé au monde, il eut des précurseurs : en 1990, un étudiant de l'université McGill à Montréal met au point Archie, un logiciel qui permettait de repérer des fichiers disséminés sur des serveurs FTP connectés à Internet (cf. chap. v). L'année suivante, des scientifiques de l'université du Minnesota développent Gopher, un dispositif de publication qui permet de donner accès à des documents en ligne *via* des menus de navigation organisés sous forme hiérarchique. C'est en 1995 que des chercheurs de Digital Equipment développent le premier moteur capable d'indexer rapidement la plupart des pages Web et de lancer une recherche d'images, de fichiers audio et vidéo : Altavista.

Mécontents des moteurs de recherche existants, Larry Page et Sergei Brin créent Google en 1997. Si la société est connue pour son moteur de recherche, elle s'est largement diversifiée (Google Earth, Google Maps, l'OS Android, Youtube, etc.) au point de devenir l'une des plus grandes entreprises mondiales dont la position dominante suscite de plus en plus d'inquiétudes. Google s'est donné pour mission « d'organiser l'information à l'échelle mondiale et de la rendre universellement accessible et utile ». Le terme « google » est un jeu de mots sur « googol » qui désigne le chiffre 1 suivi de 100 zéros : il suggère ainsi que la mission de Google est d'organiser l'immense quantité d'informations disponibles sur le Web.

Du Web 1.0 au Web 4.0

On le voit à travers le cas de Facebook notamment[12], à partir des années 2000, le Web devient collaboratif : alors que dans les années 1990, le Web (rétroactivement nommé *Web 1.0*) est statique et linéaire, laissant uniquement à l'internaute la possibilité de consommer de l'information – c'est l'époque des hyperliens avec l'apparition du protocole HTTP –, il évolue progressivement vers une plate-forme donnant davantage de possibilités de diffusion et de partage de contenus par les internautes. L'utilisateur n'est plus un simple consommateur passif mais prend part à la production d'informations et à l'évaluation de leur valeur. En 2005, Tim O'Reilly baptise cette mutation progressive du Web statique vers un Web participatif ou social *Web 2.0*. Les utilisateurs s'emparent du Web pour y écrire autant que pour y lire. Ces écrits, ainsi que toutes leurs interactions, deviennent des *big data* convoitées par un marché qui y voit une opportunité économique lucrative.

Mais ce Web est dépassé depuis longtemps : avant que n'apparaisse dans les années 2020 le concept de *Web 4.0* permettant de connecter individus et objets dans un environnement Web omniprésent, on a vu apparaître au début des années 2010 l'expression *Web 3.0* pour désigner le Web sémantique[13]. L'idée était de parvenir à un Web intelligent, où les machines comprendraient le langage naturel et la signification de l'information sur le Web. Les pages Web étant actuellement lisibles uniquement par l'Homme, le Web sémantique a pour objectif de les rendre lisibles par les ordinateurs. Pour cela,

12. Notons que, depuis octobre 2021, la maison-mère se nomme Meta, et non plus Facebook Inc., en référence au métavers (« au-delà de l'univers ») considéré comme le graal des interactions sociales par Marc Zuckerberg, où la frontière entre le réel et le virtuel se brouille jusqu'à disparaître complètement.
13. Sur le Web sémantique comme nouvelle étape dans l'évolution du Web, cf. Bertails Alexandre, Herman Ivan et Hawke Sandro, « Le Web sémantique », *Annales des Mines – Réalités industrielles,* novembre 2010/4, p. 84-89, [https://www.cairn.info/revue-realites-industrielles1-2010-4-page-84.htm], consulté le 2 février 2024.

les informations doivent faire l'objet d'un langage structuré décrivant ces données[14] : c'est ce qu'on appelle « métadonnées », c'est-à-dire des données sur les données. Elles existent depuis longtemps : citons seulement les notices bibliographiques d'un livre ou encore le format, la date, la géolocalisation d'une photo par exemple. Mais il s'agit, avec le Web sémantique, de recourir à des traitements automatiques de données lisibles par les machines pour tout système produisant des données, ce qui n'est pas sans poser le problème de la frontière entre données publiques et usages privés.

L'intelligence artificielle ou l'enjeu du siècle[15]

Comment l'ordinateur devient-il plus intelligent ? Comment apprend-il ? On en revient à la notion d'IA née il y a plusieurs décennies. Il s'agit pour un système informatique d'imiter l'intelligence humaine. Ainsi, en 2011, IBM lance Watson contre les meilleurs candidats du jeu télévisé américain Jeopardy : en plus de sa capacité d'analyse d'un nombre considérable de connaissances, Watson a été capable de déjouer les pièges tendus par le présentateur et même de comprendre certains jeux de mots. Plus récemment encore, c'est vers l'Internet des objets que s'est tournée cette technologie, avec les voitures autonomes par exemple. Comment cela fonctionne-t-il ?

More data, better data

Pour qu'une machine apprenne, elle a besoin de données à analyser et sur lesquelles s'entraîner, les *big data*. Pour avoir une idée de la masse de données dont il est question, prenons une comparaison : la Bibliothèque nationale de France (BnF) qui, pendant longtemps, constitua l'horizon ultime du savoir, compte quatorze millions de volumes, soit 14 téraoctets (14 To) de données ; or le poids des données échangées chaque jour sur Facebook est d'environ 500 téraoctets (500 To)… Le *machine learning* permet d'extraire de la valeur en provenance de sources de données massives et variées sans avoir besoin d'un humain. Les données sont l'instrument qui permet à l'IA de comprendre et d'apprendre la façon dont les humains pensent. Plus un système reçoit de données, plus il apprend et plus il devient précis, ce que résume bien l'adage : « *more data, better data* ». Il s'agira donc d'entraîner la machine afin qu'elle apprenne à accomplir les tâches généralement assurées par les animaux et les hommes : percevoir, raisonner et agir. Comment ? Grâce aux réseaux

14. Différents langages de description des données permettent d'organiser et de partager des informations dans l'environnement du Web, dont l'un des plus reconnus est certainement le RDF. Ainsi par exemple, les données de data.bnf.fr sont restructurées, regroupées, enrichies par des traitements automatiques et publiées selon ce langage de description. Cf. chap. VII.
15. Je reprends ici en partie le titre de l'étude de Sadin Éric, *L'intelligence artificielle ou l'enjeu du siècle. Anatomie d'un antihumanisme radical*, Paris, L'Échappée, 2018.

«Vade-mecum informatique pour les Lettres et Sciences humaines», Estelle Debouy
ISBN 979-10-413-0002-0 Presses universitaires de Rennes, 2025, www.pur-editions.fr

de neurones artificiels organisés en plusieurs couches à la façon d'un millefeuille : ce qu'on va appeler *deep learning* ou apprentissage profond. Comme son nom l'indique, Google DeepMind utilise cette technologie pour créer des machines intelligentes capables d'apprendre et d'acquérir de nouvelles connaissances à la manière des êtres humains.

C'est cette technologie qui est utilisée pour la reconnaissance de texte et la traduction[16]. Prenons trois exemples, le début de l'*Énéide* de Virgile (latin), du *Don Quichotte* de Cervantes (espagnol) et d'*Oliver Twist* de Dickens (anglais). Ces extraits ont été soumis à Google traduction[17] :

> « Les armes et l'homme que je chante, qui sont venus d'abord des côtes de Troy en Italie, où il s'était enfui et que Lavinian il est venu au rivage, et il était sur terre et sur la mer par beaucoup de Juno en compte de la colère de la violence féroce des dieux ;
>
> Dans un endroit de la Manche, dont je ne veux pas me rappeler le nom, vivait il n'y a pas longtemps un gentilhomme de ce genre avec une lance dans un chantier naval, un vieux bouclier, un cerf maigre et un lévrier courant. Une marmite de quelque chose de plus de vache que de mouton, du salpicón la plupart des soirs, des duels et des quebrantos le samedi, des lantejas le vendredi, un peu de palomino en plus le dimanche, consommaient les trois parties de son hacienda.
>
> Parmi les autres édifices publics d'une certaine ville, qu'il sera prudent de ne pas mentionner pour plusieurs raisons, et auxquels je n'attribuerai aucun nom fictif, il en est un qui est autrefois commun à la plupart des villes, grandes ou petites, c'est-à-dire un hospice ; et c'est dans cet hospice qu'est né ; à un jour et à une date que je n'ai pas besoin de me donner la peine de répéter, puisqu'ils ne peuvent avoir aucune conséquence possible pour le lecteur, dans ce stade de l'affaire en tout cas ; l'élément de la mortalité dont le nom est préfixé à l'entête de ce chapitre. »

Que constate-t-on ? Qu'il ne faut pas faire confiance aveuglément à l'intelligence artificielle ! On voit en effet que le texte latin traduit en français est à peu près illisible, qu'il en est pour ainsi dire de même avec le texte de Cervantes, alors que la traduction du texte anglais est bien plus correcte. Pourquoi ? Parce que le moteur de traduction automatique repose sur l'analyse des millions de livres numérisés. On comprend bien que le corpus latin est beaucoup moins

16. Pour une réflexion sur la traduction automatique, cf. Deneufbourg Antoine, « Traduction automatique : la dangereuse "sagesse des foules" », *The Conversation*, octobre 2021, [https://theconversation.com/traduction-automatique-la-dangereuse-sagesse-des-foules-169376], consulté le 2 février 2024.

17. Selon des tests en aveugle réalisés avec des traducteurs humains professionnels, les traductions de l'outil DeepL sont considérées comme meilleures et plus naturelles que celles de Google traduction. Les experts ont ainsi noté que le système offre des traductions plus précises, mais aussi plus nuancées que ses concurrents. Malheureusement DeepL ne reconnaît pas le latin, raison pour laquelle nous n'avons pas pu l'utiliser ici…

important que les autres, et que la langue de Cervantes est moins représentée que la langue espagnole moderne, d'où des résultats médiocres pour les traductions de Virgile et Cervantes. Comme un algorithme apprend tout seul, plus il a de données, plus il se perfectionne… Ces modèles évoluent très vite : à l'heure où je relis ces lignes avant publication, de nouvelles IA ont fait leur apparition et proposent des traductions parfaites. C'est le cas, par exemple, de Claude dont la traduction du début de l'*Énéide* est impeccable [https://claude.talkai.info/fr/].

«Vade-mecum informatique pour les Lettres et Sciences humaines», Estelle Debouy
ISBN 979-10-413-0002-0 Presses universitaires de Rennes, 2025, www.pur-editions.fr