FRANÇOIS-BERNARD HUYGHE

L'art de l'incertitude : le code paradoxal

De haut en bas : **Champollion,** déchiffrage de hiéroglyphes *(Cours d'égyptologie)* ; **Machiavel,** emplacement des troupes codé par des lettres de l'alphabet *(L'Art de la guerre*, 1546) ; **Léonard de Vinci,** écriture en miroir *(Étude des divers effets de lumière provenant d'une unique source unique).* D.R.

Pourquoi la carte est-elle moins grande que le territoire ? Le nombre des mots inférieur à celui des choses ou événements qu'ils décrivent ? Comment des combinaisons de lettres peuvent-elles déclencher une guerre, inspirer l'amour, insuffler la foi, ou faire rire des millions de gens ? Réponse : par le code.
Par définition, il sert à produire plus d'information, plus de mémoire, plus de compréhension avec moins de symboles.

Il instaure un ordre intellectuel entre éléments physiques, sons, couleurs, figures dont il permet l'intellection. Il réutilise les mêmes signes suivant les mêmes règles et rend pourtant une multiplicité de sens. Cet invariant qui produit des variations suppose un principe d'économie : un répertoire fini d'éléments plus une combinatoire qui en détermine les rapports. La difficulté commence quand nous cherchons à définir ce qu'est un bon code, donc quel est ce « plus » qu'il est supposé créer. Plus de précision : davantage de monosémie, une meilleure adéquation entre l'objet représenté et sa représentation ? Plus de signification, plus de richesse expressive offerte à l'interprétation ? Plus de communication ? Le meilleur code serait alors le plus facile à comprendre, à partager, voire à apprendre.

Si le critère du bon code était de produire plus d'information au sens de la théorie du même nom ? La réponse est connue : un bon code est très entropique, c'est-à-dire qu'il utilise les symboles à sa disposition de manière équiprobable si bien qu'ils apparaissent avec la même fréquence statistique. C'est un code non redondant qui ne répète pas inutilement les mêmes signes ; il est efficace car il véhicule un maximum de bits (des choix binaires entre deux possibilités) pour chaque symbole qu'il utilise. Mais une telle définition intéresse l'ingénieur des Telecom ou le spécialiste de la compression numérique et ne concerne guère la conversation ou la correspondance usuelle.

Toute langue naturelle suppose une dépense inutile. Non seulement, nous multiplions mots et expressions qui ne servent qu'à prolonger le ronron de la communication, mais la langue est notoirement redondante. Cela importe assez peu dans la vie quotidienne où il nous est plus facile de deviner un message où un élément fait défaut ou est altéré. L'économie d'effort ou la facilité de la relation compense ce qui se perd en contenu.

Pareilles notions donnent lieu à une formulation mathématique. Si un alphabet de 26 lettres était employé de manière absolument équiprobable, en lisant un texte, il y aurait une chance sur 26 de deviner la lettre suivante, comme à pile ou face il y a une chance sur deux de deviner le prochain coup, même si le joueur croit distinguer des séries signifiantes. L'information (et l'entropie qui se confondent ici parfaitement) de chaque lettre serait égale au logarithme de 26 soit 4,7 bits. Dans la réalité, une fois encore, il en va tout autrement. Une langue comme l'anglais est redondante, quasiment aux trois quarts, et une lettre n'y apporte qu'un bit d'information en moyenne. C'est ce que démontra Shanon, un des pères de la théorie de la communication lorsqu'il soumit des textes à des groupes de cobaye et mesura combien ils étaient capables d'anticiper la prochaine lettre. La langue parfaite-

ment économe (qui véhiculerait la plus grande quantité d'information avec le moins de redondance) est sans doute aussi utopique que la langue parfaite (qui produirait la qualité la plus précise de signification).

La quête de l'aléa

Ce rapport entre moins et plus prend une importance toute particulière si un impératif de confidentialité s'ajoute aux contraintes du code linguistique. Il est un cas où il est facile de définir un « bon » code : le code secret. Sa valeur se mesure à sa résistance. Plus il est difficile à deviner, moins son usage en révèle sur sa nature, meilleur il est. Il doit tout dire sur tout sans rien dire sur soi. Représenter sans indiquer. Le code secret (ici nous emploierons le terme générique de code secret, même si les puristes préfèrent parler de chiffre pour désigner un système de substitution des lettres) est régi par un ordre. Des règles préalables, impératives, explicites établissent les correspondances les plus arbitraires entre le chiffré et le clair mais aussi les plus strictes afin d'éliminer erreur ou ambiguïté. Le code doit feindre le désordre, imiter le hasard, décourager la quête du sens par une forme molle sur une structure rigide. En tant qu'il est code, et permet d'émettre des messages, il réduit l'incertitude, en tant qu'il est secret, il produit de l'incertitude. Tout en n'ayant qu'une bonne interprétation, le message codé suggère toutes les interprétations. Idéalement, le violeur de code serait confronté à la situation du visiteur de la bibliothèque de Babel décrite par Borges : entre toutes les combinaisons possibles de toutes les lettres ou signes qu'il peut supposer, rien ne lui indique quelle est celle qu'il recherche.

Pour le cryptologue, la question est : comment faire parvenir un message à A que lui seul comprenne, étant entendu que a) le message inscrit sur un support peut tomber entre des mains adverses b) l'interprétation ne doit pas requérir la présence de l'émetteur c) le message doit voyager séparément du code. Idéalement la liste des conventions doit se transmettre sans laisser de traces et se retenir de mémoire.

Il est possible de camoufler les signaux du message, de telle manière que seul le véritable destinataire puisse les lire. Il saura alors où trouver ou comment discerner le texte écrit mais invisible. L'auteur du message occulte alors la chose et non le sens : il recourt à des cachettes sur le corps humain, à des encres sympathiques, à des supports truqués, à des microfilms, etc. Mais pareils procédés dont le nom savant est stéganographie s'apparentent plus à

la ruse qu'au code.

Pour coder vraiment deux grandes voies : substituer ou déplacer, jouer le répertoire ou la combinatoire. Dans le premier cas, il faut remplacer chaque élément clair par un élément codé, lettre par lettre, mot par mot, ou phrase par phrase. Dans le second, il faut réorganiser l'ordre des éléments signifiants : changer la première lettre pour la dernière, la seconde pour l'avant-dernière ou encore inventer une sorte de verlan. Plus toutes les complications que l'on peut imaginer (combinaison des deux procédés, adjonction de leurres qui ne signifient rien, etc.).

A priori, le cryptologue devrait l'emporter à tous les coups face au cryptanalyste qui tente de deviner. Quoi de plus simple que de compliquer ? Or l'expérience nous enseigne tout le contraire. Les performances du cryptanalyste ne cessent de s'améliorer et le rêve du code inviolable de s'éloigner.

Examinons le point de vue du cryptanalyste. Il n'a que deux méthodes à sa disposition, qu'il utilise ensemble ou concurremment. La première vise à reconstituer le cheminement mental du cryptologue et, par déduction ou intuition, à deviner la règle qu'il suit. L'autre méthode consiste à procéder par essais et erreurs : tester telle combinaison d'éléments en clair qui pourrait correspondre à telle combinaison d'éléments codés. Le cryptanalyste se trouverait devant le problème de la bibliothèque de Babel évoqué plus haut, s'il n'existait une analogie minimale entre le codé et le clair, elle-même reflet des régularités du texte clair et si cet ordre ne transparaissait pas indirectement.

Dès le IXe siècle, les Arabes remarquent que des lettres apparaissent plus souvent qu'à leur tour ; de la fréquence d'un signe dans le texte codé ils présument de son identité dans le clair et dressent des tables de fréquence. À la Renaissance, les spécialistes que s'arrachent les cours d'Europe cassent des chiffres par des recettes empiriques. Ils savent que telle lettre a de fortes chances d'être associée à telle autre, comme en français un « q » appelle un « u », et un « h » suit souvent un « c », que tel mot bref devant un mot plus long pourrait bien être un article, etc. Ces remarques sur la prédictibilité de la langue se retrouvent dans des textes comme « le scarabée d'or » d'Edgar Poe, lui-même remarquable cryptanalyste. Sherlock Holmes résout de la même manière le mystère des « hommes qui dansent » : un message chiffré dont les lettres sont remplacées par des silhouettes en mouvement. Le principe de la théorie de l'information est là : mesurer la prolifération du signifiant au regard de la rareté d'apparition du signifié. À langue prédictible, code déchiffrable.

Le problème du cryptologue devient d'inventer un code de transposition qui ne reflète pas mais annule les « défauts » de la langue naturelle : ne pas produire assez de plus avec du moins, ne pas exploiter suffisamment son répertoire. Un bon code secret ne devrait jamais rendre le même par le même. D'où plus de règles. D'où des combinatoires de combinatoires, des clefs changeantes déterminant l'emploi de tel répertoire à tel point du texte.

À l'extrême, le chiffreur idéal changerait de système à chaque lettre, puisant dans un répertoire différent, modifiant le principe de transposition. Quand bien même le cryptanalyste disposerait d'un temps infini et explorerait tous les sens possibles de x lettres ou éléments, il aurait le choix entre tous les messages possibles composés de x lettres ou éléments. Mais un tel code idéal semblerait fonctionner à rebours de son principe premier : il serait plus vaste et plus complexe que l'ensemble des messages. La carte camouflerait le territoire.

Impossible hasard

Dans la réalité, nous sommes incapables de mémoriser tant de règles. Il faut soit que le codeur limite la complication à des procédés transmissibles de vive voix, relativement simples, soit qu'il recoure à un support de mémoire, tel un livre de code. Il tombe de Charybde en Scylla : il doit maintenant faire connaître les conventions au destinataire et un tel texte en clair peut être dérobé, falsifié, copié, perdu… Tout repose à présent sur le secret du livre et non plus sur le mystère du message, ce qui ne se prête guère à un usage répétitif. Une bureaucratie du secret, service d'espionnage, diplomatie, armées, ne peut pas s'offrir le luxe de multiplier de tels procédés, pas plus que nous ne pouvons employer un dictionnaire en dix volumes pour mener une conversation ordinaire. Plus de règles, plus de temps, plus de mémoires, c'est trop.

S'il y a davantage de conventions à respecter, il faut en confier la garde soit à une mémoire humaine, d'où risque de perte, soit à une mémoire physique, d'où risque de vol. La seule solution alternative serait de recourir à un code sans trace. Ainsi, pendant la guerre du Pacifique, l'armée américaine confia ses transmissions radio à des Indiens navajos. Ils parlaient une langue rare, non écrite, et aucun linguiste nippon n'avait séjourné sous le tipi : c'était utiliser des hiéroglyphes sans pierre de Rosette. Mais c'est là l'exception.

Une autre réponse rationnelle est le recours à la machine sur qui l'on décharge la tâche de compliquer la mémoire. Elle effectue un grand nombre d'opérations plus vite et plus sûrement qu'un homme. Si un protocole en modifie les réglages chaque jour, le code employé lundi devient sans intérêt s'il est découvert mardi. D'où l'introduction d'un facteur temps : un code vaut ce que vaut son temps de résistance à un nombre élevé de tentatives de viol par essais et erreurs. C'est la solution que choisissent les Allemands pendant la seconde guerre mondiale. La fameuse machine Enigma comporte un jeu de rotors mobiles dont les positions sont convenues entre les correspondants. Complexe par le nombre de combinaisons, et partant les manières de chiffrer le même texte, qu'elle permet, Enigma est pourtant d'un usage enfantin : l'opérateur tape son texte clair sur un clavier et les lettres chiffrées apparaissent à l'instant. Le réglage quotidien assure une sécurité supplémentaire : le temps qu'un unique message soit décrypté, les positions des rotors d'Enigma auront changé et tout sera à refaire.

En riposte, les Alliés jouent le medium contre le code et inventent les machines qui pensent pour vaincre les machines qui brouillent. Les premiers cryptanalystes qui s'attaquent à Enigma, une équipe polonaise, imaginent des groupes de machines travaillant de façon coordonnée pour tester bien plus de combinaisons par essais et erreurs qu'aucun groupe de chercheurs dans le même temps. Ces machines, baptisées « bombes » en raison de leur tic-tac bruyant, ne sont pas encore des ordinateurs programmables, mais c'est déjà un pas gigantesque.

Des équipes anglaises qui recueilleront l'héritage des « bombes » casseront finalement le code d'Enigma, et contribueront à la victoire. Et cela, grâce à la fameuse machine de Turing, inventée par le génial mathématicien pour vaincre Enigma. L'idée est de créer des machines à « computer », programmables qui divisent des opérations complexes en une série d'opérations binaires, et exécutent des tâches en fonction du résultat de l'opération précédente. Même si la machine de Turing n'existe jamais que sur le papier, elle est la base de l'informatique moderne. Si bien qu'à travers la théorie de l'information de Shanon et les travaux de Turing, la cryptologie a déterminé deux des grands événements du Xx siècle : l'issue de la seconde guerre mondiale et la révolution informatique.

La guerre du plus

Désormais la cryptologie envahit la vie quotidienne, et ce pour des raisons de divers ordres :
- techniques : le nombre des messages, non seulement écrits, mais via Internet, par téléphone cellulaire, par satellite etc. qui sont susceptibles d'être interceptés, la possibilité dans l'espace d'agir de loin sur des mémoires, la faculté dans le temps de retracer l'histoire de tout message ou de toute connexion.
- économiques : l'économie dite de l'immatériel multiplie à la fois les transactions à distance qu'il faut sécuriser, et les données qu'il faut protéger contre l'intrusion, le sabotage, le piratage.
- symboliques pour ne pas dire idéologiques : la peur de Big Brother, le souci de confidentialité de millions d'internautes, l'action militante, ludique ou délictueuse de « hackers » dont le but est de s'introduire dans des sites ou bases de données.

La prolifération des secrets accompagne la banalisation du code qui se fait aussi code d'identification. Tel est le cas d'un numéro de carte bancaire ou du mot de passe d'un ordinateur. Il donne l'accès à un compte en banque, à une mémoire, etc. Ce code probant complète le code de transmission qui garantit la confidentialité d'une correspondance. Suivant la formule de Deleuze, dans notre société de contrôle, le mot d'ordre est remplacé par le mot de passe. Il importe avant tout de savoir qui a accès à quelle connaissance.

En effet, le numérique, ce code de codes, qui réduit à des séries de 0 et de 1 signes, sons et images, ce simplificateur absolu, requiert comme par compensation des moyens sophistiqués de compliquer. Du fait de la numérisation, l'information est devenue soit désirable (bases de données, images satellites, monnaie électronique, messages cryptés, mais aussi œuvres, propriété intellectuelle…), soit vulnérable (des logiciels, des mémoires, des sites, des réseaux qui peuvent être attaqués) soit redoutable (des virus informatiques, mais aussi des « rumeurs électroniques », de la désinformation). D'où tous ces besoins contradictoires : prouver qui l'on est pour ne pas être dérobé ou imité, ne pas laisser de traces de ce que l'on a fait pour ne pas être fiché, identifier ce que l'on a produit pour qu'il ne soit pas reproduit, cacher ce que l'on a appris pour que sa trace ne soit pas altérée… Et, plus que par des lois ou des verrous, ces besoins peuvent être résolus par des logiciels.

Entre-temps, en effet, la guerre cryptologique du plus s'est faite informatique. Le codage est délégué à des algorithmes : les éléments du texte clair deviennent des sériées de 0 et de 1 qui, elles-mêmes, sont comme « brassées » suivant un ordre. Ainsi, la notion de « plus » (plus d'opérations, plus de complexité, plus de règles) se mesure en bits. Un des systèmes les plus populaires le D.E.S., Data Encryption Standard d'IBM transpose le texte en séries de 64 bits, puis, au cours de 16 étapes successives, échange et transpose ces blocs suivant un rythme déterminé par sa clé secrète à 56 bits. Le destinataire procède à l'inverse, dans ce système dit à clé symétrique. Il existe aussi des systèmes à clef publique où chacun communique un algorithme à son éventuel correspondant, qui permet d'en recevoir des messages cryptés, mais se réserve la clef privée qui seule permet le déchiffrement : chiffrement et déchiffrement sont non réversibles, asymétriques. D'autres procédés expérimentaux reposent sur des principes d'incertitude de la mécanique quantique et sur l'envoi de messages sous forme d'électrons polarisés qui ne peuvent être interceptés sans altérer le message. D'autres expériences encore visent à permettre l'identification d'un interlocuteur à distance par l'image de sa rétine par exemple (si bien qu'ici, c'est son code génétique qui sert de code secret). Bref, il devient évident que le code est le domaine des technologies de pointe et que l'invention de procédés de dissimulation et de complication devient un enjeu scientifique crucial. Sans que pour autant la victoire définitive du cryptologue soit encore certaine.

Secrets et algorithmes

En principe, plus longue la clé, plus complexes les opérations, meilleure la sécurité. On a coutume de dire que tel algorithme résisterait à tant d'ordinateurs, de telle puissance travaillant pendant tant de siècles. Ce peut être présomptueux, comme l'a démontré le récent « craquage » de D.E.S. : la montée en puissance des ordinateurs, qui de surcroît travaillent en chaîne à briser les codes, rend de tels calculs très éphémères. Reste pourtant la notion de nombre d'essais et erreurs qui se mesure en puissance mathématico-informatique : la cryptanalyse n'a plus rien à voir avec l'ingéniosité d'un Edgar Poe. Un service secret (tel la National Security Agency américaine, premier employeur de mathématiciens au monde et responsable du projet Echelon qui permet l'interception des communications dans le monde entier) peut ou ne peut pas casser une clef de tant de bits, dans un délai de

tant d'heures ou de jours. Il en va de même pour un groupe de pirates informatiques travaillant en pool : leur puissance de calcul est leur puissance de frappe. La décision du gouvernement français de relever de 40 à 128 bits la longueur des clés librement disponibles, et donc de ne plus les classer comme matériel stratégique, permet au citoyen de se doter des armes de la cryptologie qui étaient autrefois celles du stratège et de l'espion. La défense de l'intimité et de la vie privée, devenue une activité conflictuelle et quasi martiale, suscite cette curieuse revendication : le droit pour tous de dissimuler, la démocratisation du secret.

Le cryptologue producteur d'une œuvre fermée est un anti-poète : s'il impose toujours plus de contraintes formelles au message, c'est pour le rendre impénétrable : le code doit qu'il invente doit produire du plus et non du moins. Mais, du coup, il est contraint à une curieuse fuite en avant qui dure quelques siècles et traverse plusieurs médiasphères. D'abord, plus de règles d'où plus de traces, puis des mécaniques pour exécuter, puis des mécanismes pour calculer et virtualiser : par étapes, les opérations sont déléguées de la pensée aux données, de l'interprète à l'inscription puis à l'instrument. Et pour finir le code se fait medium.