

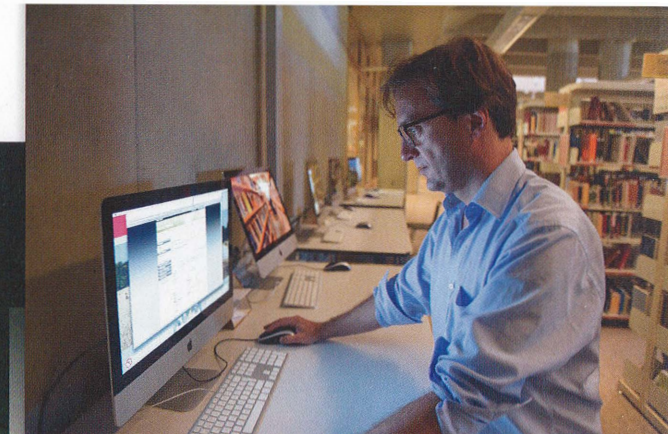
Les manuscrits à l'ère du Big Data

Les progrès des algorithmes et l'essor de la numérisation font entrer les manuscrits constitutifs de l'Ancien Testament dans une nouvelle ère. Et promettent de nombreuses avancées sur les sources, la compréhension et la recomposition de ces textes sacrés.

RONEN ZVULUN - REUTERS/PATRICK CHAPUIS

La numérisation des manuscrits de la mer Morte offre aux scientifiques des images multispectrales de haute résolution.

Le chercheur David Hamidovic développe une base de données rassemblant les images de chaque lettre des manuscrits de Qumrân.



Après avoir révolutionné de nombreux domaines, l'ordinateur est en passe de faire accomplir à la science biblique un pas de géant. « Une nouvelle génération de chercheurs voit le jour : les digital native nés avec l'ordinateur et les jeux vidéos, qui ont atteint la quarantaine et disposent de moyens de recherche et de laboratoires », analyse David Hamidovic, professeur à l'Institut romand des sciences bibliques de l'université de Lausanne. La Bible entre ainsi dans l'ère des Digital Humanities (humanités numériques). Ce domaine de recherche pluridisciplinaire, basé sur l'utilisation d'outils numériques au service de la recherche en littérature, en sciences humaines et en sciences sociales, est en plein essor. Avec des retombées potentielles dans la plupart des disciplines qui s'attachent à l'étude des textes anciens. Les études portent principalement sur les manuscrits de Qumrân, une collection majoritairement composée de parchemins et de papyrus découverts entre 1947 et 1956 à l'intérieur de jarres cachées dans onze grottes de Cisjordanie, sur la rive nord-ouest de la mer Morte. Cette bibliothèque d'une valeur historique inestimable regroupe officiellement 930 manuscrits, dont un tiers concerne l'Ancien Testament : du Pentateuque (Genèse, Exode, Lévitique, Nombres, Deutéronome), aux livres des Prophètes (Isaïe, Jérémie...), en passant par les livres poétiques ou encore les Cinq Rouleaux. Bref, un véritable trésor comprenant les plus anciens textes bibliques écrits en hébreu jamais découverts à ce jour, à savoir un fragment d'un rouleau des livres de Samuel datant du III^e siècle avant notre ère. À l'exception de quelques manuscrits retrouvés complets, comme le rouleau d'Isaïe, la plupart des textes ont été découverts éparpillés par petits bouts, façon puzzle, formant un ensemble d'environ 30 000 fragments. « Le problème avec les manuscrits de Qumrân, c'est que la plupart sont des frag-

ments comportant peu de phrases entières : à peine quelques mots, et encore, quand on a de la chance ! », détaille Daniel Stoekl, directeur d'études à l'École pratique des hautes études. Depuis presque soixante-dix ans, les chercheurs essaient donc de reconstituer le puzzle. Un vrai travail de bénédictin, exigeant de disposer d'une solide culture historique et d'une fine connaissance des textes et des formes d'écritures anciennes.

SUR LA TRACE DES SCRIBES

Depuis quelques années, l'ordinateur leur apporte une aide précieuse pour valider (ou pas) certaines hypothèses de restitution, l'exercice revenant à reconstituer le texte abîmé ou manquant. « Des logiciels de dessins assistés par ordinateur permettent d'analyser les photos des manuscrits et de créer une base de données des lettres dessinées par le scribe, explique David Hamidovic. Il suffit ensuite de vérifier si la restitution proposée est plausible, en écrivant le texte tel que le scribe l'aurait fait, avec la bonne taille de lettres et les bons espaces, et voir si le résultat est compatible avec les fragments. Certes, la technique n'est pas infaillible – un scribe peut parfois faire varier son écriture –, mais c'est plus précis que lorsque l'on fait cette vérification à la main. Cette technique a permis d'écarter des relectures autrefois jugées admises de certains manuscrits de la mer Morte. » Il est toutefois plus facile pour un ordinateur d'écarter une restitution erronée que de proposer une hypothèse. « Pour un ordinateur, cela revient à résoudre une seule équation, avec de nombreuses inconnues », explique Michaël Langlois, chercheur et maître de conférences à l'université de Strasbourg. L'ordinateur est également un allié de choix pour aider les biblistes à trouver des éléments de réponse à une question qui les taraude depuis toujours : qui a écrit et recopié ces textes ? « Quand on parle du livre de Samuel, par exemple, le grand public pense souvent que l'on parle d'un livre écrit par Samuel. Or ce dernier meurt après le premier quart du récit et ne peut donc pas l'avoir écrit lui-même, explique Michaël Langlois. En fait, il s'agit d'un livre dont Samuel est le personnage principal et non l'auteur. Il en va de même pour de nombreux textes, comme le Pentateuque, attribué à Moïse : il ne peut avoir écrit seul ces cinq livres regroupant des styles et des langages différents. » L'origine des cinq premiers livres de la Bible a fait l'objet de nombreuses théories. Jusque dans les années 1970, les chercheurs leur attribuaient quatre sources, dites jahviste, élohiste, sacerdotale et deutéronomiste, conformément à l'hypothèse « documentaire » prédominante jusqu'à cette époque. « Depuis, la majorité des biblistes s'accorde sur le fait que les choses ne sont pas aussi simples », admet Michaël Langlois. Ce que semblent confirmer les algorithmes. En 2011, une équipe de l'université de Tel Aviv pilotée par Nashum Dershowitz a ainsi annoncé avoir mis au point un logiciel à même d'identifier les différents coauteurs d'un texte. Le principe : l'étude de la préférence d'un auteur pour tel mot plutôt qu'un autre. « Des auteurs différents n'utilisent pas les mots avec la même fréquence, résume Nashum Dershowitz. Pour s'assurer que ces différences sont bien dues au style, et pas au sujet traité, il faut s'intéresser aux mots fonctionnels [articles, pronoms, adverbes...] ou aux synonymes. L'algorithme parvient alors à trouver des corrélations

UN LOGICIEL A ÉTÉ MIS AU POINT POUR IDENTIFIER LE STYLE DE CHAQUE RÉDACTEUR



Les manuscrits de Qumrân forment, dans leur majorité, un puzzle de 30 000 fragments. Le plus ancien remonte au III^e s. av. J.-C. (Extrait d'un manuscrit.)

DEBBY HILL - UPI - VISUAL PRESS AGENCY

RONEN ZVULUN - REUTERS



La restitution d'un texte abîmé ou manquant se fait à l'aide de logiciels de dessins assistés par ordinateur. Les manuscrits, une fois photographiés en haute résolution, sont mis en ligne.

Origine des parchemins : la mémoire de l'eau

Le numérique n'est pas la seule innovation au service des manuscrits de la mer Morte. Une nouvelle méthode pourrait bien aider les biblistes à savoir si tous les manuscrits de Qumrân ont été confectionnés sur place, en se basant sur les traces laissées par l'eau lors de la fabrication des encres et des parchemins. « L'eau de la mer Morte et des

sources alentour présente une caractéristique unique : un ratio chlore/bromine très bas, dû à une forte teneur en bromine », explique Ira Rabin, chercheuse au Federal Institute for Materials and Testing de Berlin. D'où l'idée d'utiliser la fluorescence de rayons X (une technique non destructive) pour mesurer ce ratio résiduel dans les parchemins, ce qui

permet de déterminer si leur tannage (qui nécessite de l'eau) a été réalisé sur place ou non. « Certains des 150 fragments testés ont révélé que les parchemins n'avaient pas tous été fabriqués localement. Si on pouvait tester de la sorte tous les fragments, on pourrait répondre à de nombreuses questions que l'analyse textuelle ne parvient pas à trancher. » La chercheuse tra-

vaille déjà sur une évolution de cette méthode, la fluorescence X haute résolution en mode confocal. Contrairement à la microfluorescence, cette technique ne pénètre pas toute l'épaisseur du parchemin, mais juste la surface. Ce qui permettra de tester l'encre seule, et de savoir plus précisément si elle aussi a été fabriquée près de la mer Morte ou pas. P.-Y. B.

tions entre les différents choix faits par chaque auteur. » Pour démontrer son efficacité, ils ont réuni aléatoirement chapitres et versets des livres des prophètes Jérémie et Ezéchiel pour les fusionner dans un seul texte. Résultat : le logiciel est parvenu à identifier et à séparer les extraits ainsi mélangés avec un taux de réussite supérieur à 95 %. « Le taux atteint 99 % quand on mélange des chapitres entiers ; il est légèrement inférieur pour les versets », détaille Nashum Dershowitz. Le chercheur a, depuis, utilisé cette même méthode sur la Genèse, et les Nombres. Résultat : le logiciel scinde le texte en deux parties correspondant assez fidèlement à la division du Pentateuque en deux sources : sacerdotale et non sacerdotale. « La séparation faite par notre algorithme correspond à la vision scientifique d'une division sacerdotale/non sacerdotale pour 90 % des versets », précise cette étude publiée en 2015 dans le *Journal of Biblical Literature*.

DU STYLE À LA FORME DES LETTRES

Outre le style orthographique, les logiciels de reconnaissance d'image permettent d'identifier des formes d'écritures (taille des lettres, espacement...) propres à chaque scribe. « Je suis en train de développer une base de données des mains de tous les scribes ayant travaillé sur les manuscrits de Qumrân, révèle David Hamidovic. L'idée est de comparer toutes les formes de lettres trouvées pour voir s'il y a des mains communes. On a déjà réussi à rapprocher des manuscrits et ainsi identifié des scribes ayant participé à dix manuscrits différents. Et découvert des similitudes sur les sujets abordés. L'un d'entre eux parle ainsi du Messie dans plusieurs passages, ce qui peut laisser penser que les scribes avaient peut-être des sujets de prédilection, voire des projets éditoriaux ».

Le numérique permet aussi de mettre ces précieux documents, fragiles et autrefois difficiles d'accès, à la disposition de tous. Il en est ainsi de logiciels comme BibleWorks, un outil d'exégèse regroupant des textes originaux et plus de 200 traductions de la Bible dans 40 langues, dont la première version date de 1992. Accordance, regroupe, lui, plusieurs éditions de la Septante (la traduction de la Bible hébraïque en langue grecque), enrichies d'un moteur de recherche en langues anciennes (hébreu biblique, araméen...). Ces outils numériques donnent un sérieux coup de vieux aux volumineuses collections de « Concordance », qui, totalisant plus de 1 500 pages, listent tous les mots utilisés dans la Bible, leur origine (araméenne, hébraïque ou grecque), et les versets dans lesquels ils figurent. « L'informatique et les bases de données facilitent grandement la recherche de co-occurrences : l'apparition de plusieurs mots dans un même texte. En quelques mois, on arrive à faire ce qui auparavant aurait exigé des années de travail », détaille Michaël Langlois.

Le numérique ne se contente pas de démocratiser les contenus, il s'étend aussi aux supports. La numérisation des manuscrits de la mer Morte entreprise par Google depuis 2010, à la demande de l'Autorité des antiquités israéliennes, fait entrer l'Ancien Testament de plain-pied dans l'ère du Big Data : les manuscrits ont été photographiés en haute résolution et mis en ligne sur une bibliothèque

1 – The Digital Dead Sea Scrolls : <http://dss.collections.imj.org.il>
www.deadseascrolls.org.il



Le rouleau d'Isaïe (120 av. J.C.), le plus complet des 930 manuscrits de Qumrân, est consultable en ligne.



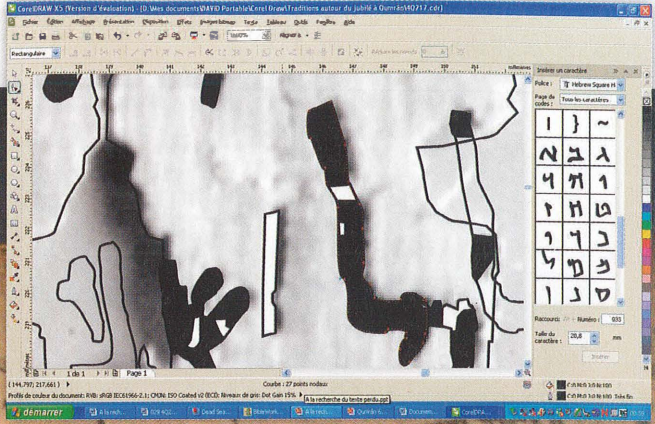
Ce microscope digital permet d'avoir une image de très haute définition et de faire des images infrarouges. On peut ainsi y déceler des détails invisibles à l'œil nu.

PATRICK CHAPIUS

BAZ RATNER-REUTERS / D. HAMIDOVIC

DES ANALYSES EXIGENT DU MATÉRIEL AUSSI PUISSANT QUE CELUI DES ASTRONOMES

À partir d'un répertoire de lettres dessinées, on peut restituer une lettre abîmée en glissant le bon dessin trouvé sur les traces visibles sur le fragment.



gratuite accessible à tous⁽¹⁾. Celle-ci permet de zoomer pour observer la forme des lettres. On arrive même à y distinguer les pores de la peau des parchemins. De plus, les clichés ont été réalisés sous différentes longueurs d'onde, de l'infrarouge à l'ultraviolet, permettant de déceler des détails invisibles à l'œil nu, tels que des traces d'écritures effacées ou noircies par le temps. « Grâce à ce nouveau type de photo, on a pu déchiffrer des mots jusqu'ici invisibles, sur les bords d'un fragment d'un manuscrit du livre des Psaumes », s'enthousiasme Michaël Langlois. « Ce n'est pas parfait : ces photos ne sont qu'une représentation en 2D de documents en 3D, voire en 4D si on prend en compte leur évolution dans le temps », tempère Daniel Stoeckl. Mais cela a changé notre vie : auparavant, on devait se contenter de photos imprimées qu'il fallait demander à l'Autorité des antiquités israéliennes. »

L'EXPÉRIENCE DE L'ŒIL HUMAIN

Le numérique va-t-il permettre pour autant de lever tous les mystères que l'Ancien Testament nous réserve encore ? Pas sûr, car l'exercice souffre encore de certaines limites. Techniques tout d'abord : si le *text mining* (la fouille de texte) n'est pas très exigeant en ressources pour les textes bibliques, les bases de données et les logiciels basés sur la reconnaissance et le traitement d'images nécessitent du matériel informatique puissant. « Certaines analyses exigent du matériel aussi puissant que celui utilisé par les astronomes, ou pour l'étude du génome, détaille Daniel Stoeckl. L'ordinateur est doué pour jongler rapidement avec un grand nombre de données. Mais l'humain doit toujours être présent pour vérifier ses résultats. » « Il faut entraîner l'ordinateur et le nourrir avec des données : des textes dont on connaît l'auteur. Ce qui suppose que ces données sources soient fiables. Le problème, c'est que le corpus est souvent restreint, et que l'on ne dispose que de peu de données fiables. Il faut rester attentif aux hypothèses faites en amont sur lesquelles se basent les algorithmes », prévient Michaël Langlois. « La machine peut apporter de réels services pour les concordances : elle est rapide, concède Émile Puech, orientaliste et professeur à l'École biblique et archéologique française de Jérusalem, considéré comme un des experts mondiaux du déchiffrement. En revanche, je ne crois guère à l'utilisation de l'ordinateur pour la restitution : les scribes n'étaient pas des machines, et leur écriture pouvait varier. À mon sens, rien ne peut remplacer l'expérience, l'œil et l'esprit humain dans ce domaine. » Un avis en partie partagé par Daniel Stoeckl : « Pour le déchiffrement et l'analyse contextuelle de données très fragmentaires, l'ordinateur reste moins performant qu'un humain expérimenté. » Du moins pour l'instant.

Pierre-Yves Bocquet