

Euratom et la genèse des humanités numériques en Europe (1957-1970)

Edgar LEJEUNE

RÉSUMÉ

En 1957, les États membres de la CECA signent le traité EURATOM dans le but de construire une politique européenne de l'énergie. Un Centre commun de recherche (CCR) est créé et s'installe à Ispra, en Italie. À la marge d'initiatives visant à développer de nouveaux types de réacteurs nucléaires, une équipe du CCR, le CETIS, s'engage dans le développement d'une machine documentaire capable d'aider les techniciens à se documenter plus efficacement sur les travaux publiés sur le nucléaire civil dans le reste du monde. Dans ce cadre, le CETIS collabore avec une autre équipe située à quelques dizaines de kilomètres, dirigée par Roberto Busa, collaborateur de IBM et véritable pionnier des humanités numériques dans le monde. Le CETIS finance également des projets d'humanités numériques poursuivant des objectifs comparables aux siens, comme celui de l'archéologue Jean-Claude Gardin, faisant du laboratoire d'Ispra l'une des infrastructures majeures du développement des humanités numériques en Europe dans les années 1960.



Les opératrices du Centro per l'Automatizzazione dell'Analisi Letteraria produisent les quelques milliers de cartes perforées de l'*Index Thomisticus* dans la salle de codage de Gallarate en juin 1967. Illustration reproduite par Steven E. Jones dans *Roberto Busa, S. J., and the Emergence of Humanities Computing : The Priest and the Punched Cards*, New York, Routledge, 2016, p. 123. Source : Busa Archive, Università Cattolica del Sacro Cuore, Milan.

À la fin des années 1950, alors que les premiers projets impliquant l'utilisation des ordinateurs en sciences humaines et sociales sont envisagés en Europe, les lieux dans lesquels les spécialistes des SHS et informaticiens peuvent collaborer pour mener ce type de recherches sont extrêmement rares. Le Centre commun de recherche, créé en 1958 dans le cadre du traité EURATOM, constitue à ce titre l'une des infrastructures clés pour le développement des humanités numériques en Europe.

Ispra, berceau méconnu des humanités numériques ?

En 1957, les États membres de la Communauté européenne du charbon et de l'acier signent à Rome deux textes fondateurs : le traité instituant la Communauté économique européenne (CEE) et le traité EURATOM, instituant la Communauté européenne de l'énergie atomique. L'organe exécutif d'EURATOM est doté d'un budget de 215 millions de dollars pour construire une Europe de l'énergie.

Le principal chantier consiste en l'accompagnement des États membres dans l'installation et le développement de nouveaux types de réacteurs nucléaires (projets Orgel, Dragon, SUSPOP, etc.). Au cœur du dispositif, un Centre commun de recherche (CCR) est mis en place sur quatre sites : à Mol (Belgique), à Karlsruhe (RDA), à Petten (Pays-Bas) et à Ispra (Italie). C'est sur ce dernier site, où travaillent les deux tiers du personnel d'Euratom (1 500 personnes), que se développent dans les années 1960 certains projets pionniers en humanités numériques.

L'un des problèmes auxquels font face les personnels du CCR touche à la circulation des connaissances portant sur le nucléaire civil. Dans un texte publié en 1960, le président de la commission Euratom, Étienne Hirsch (1901-1994), explique :

« La quantité de documents et d'articles publiés dans le monde au sujet des utilisations pacifiques de l'énergie atomique ne cesse d'augmenter et atteint des proportions considérables. De ce fait, les techniciens perdent un temps précieux à chercher dans les bibliothèques les documents ou références dont ils ont besoin. D'autre part, par ignorance des travaux publiés en langues slaves ou orientales, ils procèdent parfois à des études dont les résultats sont rendus publics depuis longtemps. »

En réponse à ce défi informationnel, Euratom se dote d'un centre de documentation comprenant une bibliothèque, une section d'information sur les traductions de documents techniques rédigés en langues slaves et orientales, mais aussi un groupe de recherche sur l'information scientifique, le Centre européen de traitement de l'information scientifique (CETIS).

Ce groupe de recherche, dirigé par l'informaticien Paul Braffort (1923-2018), se fixe pour objectif de mettre au point une machine documentaire capable d'enregistrer, de classer et de restituer à la demande une quantité importante d'informations scientifiques. L'idée générale de ce travail est d'exprimer la « substance » d'un texte scientifique *via* la construction d'un langage documentaire, afin de permettre des recherches rapides par mots-clés dans de grands ensembles de documents et des analyses automatiques de textes. Pour réaliser ce projet, le CETIS collabore avec d'autres groupes de recherche européens engagés sur des voies similaires.

Relations avec un voisin d'exception : le centre de Roberto Busa à Gallarate

Au début des années 1960, le CETIS collabore notamment avec l'équipe du père jésuite Roberto Busa, qui mène, en partenariat avec la firme américaine IBM, un projet d'édition sur cartes perforées des œuvres de Saint Thomas d'Aquin. Lancé en 1947, l'*Index Thomisticus* doit permettre la génération automatique de concordances. Il est aujourd'hui considéré comme le premier projet relevant des humanités numériques par de nombreux spécialistes.

En 1956, deux ans avant la création de l'antenne italienne du CCR d'Euratom, Busa crée à Gallarate, à moins de trente kilomètres d'Ispra, un centre dédié à l'application des calculateurs électroniques à l'étude du langage naturel : le Centro per l'Automatizzazione dell'Analisi Letteraria (CAAL). Busa y forme de jeunes étudiants - surtout des jeunes femmes - à la manipulation et la production de cartes perforées, pour enregistrer en machine les quelque 11 millions de mots de l'œuvre de Thomas d'Aquin. Dans les années 1960, le CETIS puisera dans cette main-d'œuvre locale pour poursuivre ses propres recherches sur la documentation scientifique.

Mais la relation entre le CETIS et Busa va au-delà. C'est grâce à l'entremise de ce dernier que le centre d'Euratom reçoit, au début des années 1960, le système de traduction automatique « IBM-Georgetown », qui permet de réaliser des traductions automatiques de textes scientifiques depuis le russe. Busa et son équipe fournissent d'ailleurs eux-mêmes un corpus de résumés d'articles enregistrés sur cartes perforées pour tester l'efficacité du système. En contrepartie, Busa profite des échanges avec les informaticiens d'Euratom pour faire avancer son propre projet, et aurait même bénéficié, selon Steven E. Jones, de financements de la part d'Euratom.

Jean-Claude Gardin à Ispra : la naissance du langage documentaire SYNTOL

En février 1960, le CETIS lance par ailleurs une série de séminaires dans le but de constituer un groupe de nouveaux collaborateurs. Près de 80 auditeurs y assistent. Dans les mois qui suivent, un appel à contrats est lancé et 45 projets portés par des linguistes, des philologues, des logiciens ou des archéologues sont financés. C'est par exemple le cas du travail d'établissement d'un langage documentaire universel pour la description des articles en sciences sociales, le SYNTOL (Syntagmatic Organization Language), dirigé par l'archéologue et spécialiste de l'automatique documentaire français Jean-Claude Gardin (1925-2013).

Mené entre 1960 et 1962 en collaboration avec l'association Marc Bloch, ce projet profite de ce financement pour payer les heures de calculs nécessaires à la réalisation d'analyses-test, réalisées notamment sur le calculateur électronique du CETIS à Ispra, mais également pour définir et programmer des méthodes d'indexation documentaire originales avec les informaticiens du CETIS. Conçu comme un « ensemble de règles et de procédures concernant l'enregistrement et la recherche des informations scientifiques », SYNTOL partage les mêmes objectifs que les chercheurs du CETIS : traduire un texte écrit dans un langage naturel en un langage normalisé, et permettre la recherche automatique de documents.

Dès 1963, plusieurs chercheurs en sciences sociales, parmi lesquels Maurice Coyaud (1934-2015), Françoise Héritier (1933-2017) ou Gardin lui-même expérimentent ce système

documentaire sur des corpus de psychosociologie, d'anthropologie et d'archéologie. Les méthodes de l'indexation documentaire sont dès lors considérées comme essentielles non seulement pour donner accès à de grands corpus de documents, mais également pour les étudier à l'aide de méthodes mathématiques. Elles se diffusent ensuite vers d'autres disciplines comme l'histoire médiévale ou la linguistique.

Les deux exemples présentés rapidement ici illustrent le rôle qu'a joué le laboratoire d'Ispra dans les premiers moments de l'histoire des humanités numériques. Tout d'abord, cette antenne du CCR constitue dans les années 1960 l'une des très rares zones d'échange entre informaticiens, linguistes, mathématiciens et spécialistes des SHS en Europe puisque, lorsqu'ils existent, les centres de calculs universitaires sont dédiés aux travaux des sciences « exactes ». Ensuite, Euratom est l'une des rares institutions à financer des recherches en sciences humaines, contribuant ainsi de manière décisive au développement de méthodes originales d'indexation documentaire, qui ont durablement transformé l'accès aux ressources et aux informations scientifiques jusqu'à nos jours.

BIBLIOGRAPHIE

JONES, Steven E., Roberto Busa, S. J., and the Emergence of Humanities Computing : The Priest and the Punched Cards, New York, Routledge, 2016.

LÉON, Jacqueline, « [Documentation automatique et analyse automatique de discours. Spécificité des réceptions de Harris en France](#) », *Histoire de l'automatisation des sciences du langage*, Lyon, ENS Éditions, 2015.

PLUTNIAK, Sébastien, « Assyrian merchants meet nuclear physicists: history of the early contributions from social sciences to computer science. The case of automatic pattern detection in graphs (1950s-1970s) », *Interdisciplinary Science Reviews*, Maney Publishing, 46 (4), 2021, p. 547-568.

Source URL:

<https://ehne.fr/encyclopedie/thematiques/civilisation-materielle/le-numerique/euratom-et-la-genese-des-humanites-numeriques-en-europe-1957-1970>