



**HAL**  
open science

# Mathématisations et automatisations des sciences du langage : des tournants conceptuels ou technologiques ?

Emanuel Bertrand

► **To cite this version:**

Emanuel Bertrand. Mathématisations et automatisations des sciences du langage : des tournants conceptuels ou technologiques ?. *Revue d'histoire des sciences humaines*, 2018, *Penser par Écoles*, 32, pp.288-292. halshs-02174452

**HAL Id: halshs-02174452**

**<https://shs.hal.science/halshs-02174452>**

Submitted on 5 Jul 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Mathématisations et automatisations des sciences du langage : des tournants conceptuels ou technologiques ?

Jacqueline Léon, *Histoire de l'automatisation des sciences du langage*, Lyon, ENS Éditions, 2015, 218 pages

Disons-le d'emblée, l'ouvrage de Jacqueline Léon, *Histoire de l'automatisation des sciences du langage*, est une véritable somme, qui analyse et synthétise l'histoire récente d'une discipline – la linguistique – et qui impressionne par l'étendue des connaissances que l'auteure mobilise, voire par son érudition. Écrit par une authentique spécialiste de l'histoire des sciences du langage, le livre ne se contente pas d'être une remarquable synthèse de vingt années de recherches de son auteure, mais il explore aussi des sources nouvelles et propose des mises en perspective originales. Travaillant principalement sur une période allant de 1948 à aujourd'hui, J. Léon ne s'interdit pas des incursions dans les étapes antérieures du développement de la linguistique. Elle propose aussi une périodisation précise et argumentée de cette histoire, qui se décline, en fonction des perspectives adoptées, en plusieurs variantes. Surtout, l'ambition de l'ouvrage est de faire une véritable histoire transnationale des sciences du langage, qui ne se contente jamais d'une simple approche comparatiste, même si elle s'appuie avant tout sur trois, voire quatre, terrains privilégiés : les États-Unis, la Grande-Bretagne, la France, et, dans une moindre mesure, l'URSS. L'auteure varie à l'envi les échelles d'analyse. Elle alterne ainsi les enquêtes précises à propos de contextes nationaux et l'étude des circulations de concepts et de pratiques, aussi bien entre les quatre pays concernés qu'au sein d'instances nationales et internationales. Pour observer ces circulations, elle se focalise sur les travaux de divers comités, ainsi que sur des rapports, des colloques, des sociétés savantes et des financements. L'auteure met ainsi en évidence de façon très convaincante l'enchevêtrement complexe entre les enjeux institutionnels et politiques, les cultures scientifiques locales, les mobilisations de disciplines externes aux sciences du langage, et le travail de certains acteurs pivots, institutionnels ou scientifiques, ces derniers n'étant d'ailleurs pas tous issus de la communauté des sciences du langage.

Du double point de vue théorique et méthodologique, le concept d'« horizon de rétrospection », développé en 1987 par l'historien et philosophe du langage Sylvain Auroux (Auroux, 1987), occupe une place centrale dans l'ensemble de l'ouvrage. Pour résumer la notion en quelques mots, l'horizon de rétrospection d'un champ disciplinaire désigne l'ensemble des travaux passés qui font référence pour les acteurs actuels de ce champ. L'approche historiographique choisie consiste à identifier les grands changements d'orientation qui jalonnent l'histoire récente des sciences du

langage – « tournant de l'automatisation » et « tournant des grands corpus » – et à déterminer si ces changements s'accompagnent (ou pas) de modifications de l'horizon de rétrospection de la linguistique, et, le cas échéant, selon quelles modalités intellectuelles et sociales s'effectuent ces modifications d'horizon. Évidemment, ces questions n'appellent jamais de réponse simple et univoque, et leur appréhension dépend fortement des contextes nationaux. Néanmoins, cette focalisation sur les changements d'horizon de rétrospection permet de ne jamais séparer artificiellement travaux théoriques, pratiques expérimentales et discours institutionnels, ou encore, pour le dire autrement, de toujours mobiliser histoire des idées et histoire institutionnelle en parallèle. En particulier, cette façon de faire l'histoire des sciences du langage permet de mettre à distance les (fausses) évidences souvent partagées par les linguistes de l'époque étudiée à propos de l'histoire de leur propre champ, « dans un souci constant de légitimation » (p. 26).

En ce qui concerne l'objet d'enquête précis, J. Léon s'intéresse à l'automatisation des sciences du langage. Ce concept d'automatisation est à distinguer de deux autres, avec lesquels il possède des zones de recouvrement : la formalisation et la mathématisation. La périodisation générale proposée par l'auteure permet de saisir la signification précise de ce qu'elle désigne par « automatisation ». Cette périodisation se développe sur trois échelles temporelles distinctes, la troisième étant celle à laquelle l'ouvrage est plus directement consacré. La première périodisation, à peine ébauchée, correspond à une mise en perspective du langage sur les temps les plus longs, et est ponctuée par trois « révolutions technologiques » (p. 9, 181) : l'écriture, la grammatisation des langues vernaculaires (Auroux, 1994), et, enfin, l'automatisation des sciences du langage (p. 181-183). La deuxième échelle de périodisation utilisée est celle de la « mathématisation » du langage, qui se développe, selon l'auteure, en deux phases distinctes : la « première mathématisation du langage » correspond à la formalisation proposée par l'École de Vienne dans les années 1930, et dans le cadre de laquelle les mathématiques sont instituées comme un langage parmi d'autres, mais qui permet, sur leur modèle, le développement de nouveaux langages formels (p. 6-7) ; la « seconde mathématisation », que J. Léon préfère appeler « automatisation-mathématisation », constitue précisément l'objet central de son livre, et est indissociable de l'émergence de la programmation sur ordinateur (p. 7). Enfin, la troisième périodisation, à plus courte échelle de temps, concerne justement cette automatisation-mathématisation des sciences du langage, et se déploie en deux étapes : celle de la traduction automatique et du traitement automatique des langues, de 1948 à 1966 (p. 13), et celle du traitement informatique des grands corpus textuels, ou *corpus turn*, dans les années 1990 (p. 7, 157-176). Cette dernière périodisation est d'ailleurs déclinée selon différentes scansion, en fonction des différents contextes nationaux étudiés. En outre, cette périodisation de l'automatisation des sciences du langage entre en dialogue avec une périodisation plus générale de la linguistique et de ses

différents courants académiques, dans un jeu complexe d'influences réciproques et de recherches de légitimation internes à la discipline.

Lorsqu'elle considère la première étape de l'automatisation-mathématisation du langage, l'auteure distingue deux phases, très différentes sur le plan épistémique, mais marquées toutes les deux par une forte intervention d'institutions publiques. La première phase, qualifiée d'« événement constitutif du tournant de l'automatisation » (p. 9), concerne le développement, aux États-Unis, de la traduction automatique comme « technologie de guerre », dans le cadre de la guerre froide, à partir de 1948. La linguistique ne fait pas partie des « sciences de la guerre » (Dahan, Pestre, 2004), et la traduction automatique (TA) se développe donc, paradoxalement, sans les linguistes, et avec pour objectif essentiel de « produire des traductions en série de textes scientifiques du russe vers l'anglais » (p. 15). Dans une seconde phase, les linguistes vont devoir s'approprier ce domaine connexe au leur, mais à la création duquel ils n'ont pas participé. Cela va constituer pour eux un changement d'horizon de rétrospection, et le vecteur de ce changement sera la linguistique computationnelle, dont la légitimité comme « nouvelle linguistique » est décrétée par un rapport de 1966 d'un comité du National Research Council (p. 13-17). Il s'agit donc là d'une histoire institutionnelle complexe, à la fois états-unienne et internationale, dont les principaux acteurs collectifs sont publics et dont J. Léon identifie deux acteurs individuels essentiels : Warren Weaver<sup>1</sup>, dont le rapport de 1949<sup>2</sup> entraîne la création de grands centres universitaires de traduction automatique, dont un au Massachusetts Institute of Technology (MIT), haut lieu des sciences de la guerre ; Yehoshua Bar-Hillel, philosophe du langage et linguiste, embauché au MIT en 1951 pour travailler sur la traduction automatique, et qui élabore une « syntaxe opérationnelle » pour la TA (p. 30). En 1960, à la demande de la National Science Foundation, et après deux ans d'enquête, Bar-Hillel publie un rapport-bilan sur les résultats obtenus par la vingtaine d'équipes qui travaillent sur la TA dans le monde (Bar-Hillel, 1960). Très négatif, ce rapport provoquera l'arrêt, à partir de 1964, de la plupart des financements états-unien pour la TA. Bar-Hillel est un acteur clé de l'histoire de l'automatisation des sciences du langage, car il est un des très rares linguistes parmi les chercheurs en traduction automatique. Il joue ainsi un rôle central pour intégrer les apports de la TA aux sciences du langage, en instituant la linguistique computationnelle comme nouvel horizon de rétrospection pour les linguistes. Notre auteure étudie en détail comment s'interpénètrent alors les enjeux de politique de guerre froide, de politique scientifique états-unienne, et les enjeux plus directement disciplinaires. Elle montre

<sup>1</sup> Mathématicien et ingénieur, acteur central de la culture scientifique de guerre aux États-Unis, Warren Weaver se considère lui-même comme « un facilitateur, un vulgarisateur plus qu'un véritable chercheur » (p. 20). Ce sont ces talents de vulgarisateur qui font de lui le co-auteur du fameux ouvrage de 1949 de présentation de la théorie de l'information par Claude E. Shannon (Weaver, Shannon, 1975).

<sup>2</sup> Ce rapport intitulé « Translation » sera publié sous forme académique en 1955 (Weaver, 1955).

notamment comment ces mouvements institutionnels autour de la TA aboutissent à une réécriture interne de l'histoire de la linguistique états-unienne : cette histoire interne insiste alors sur une prétendue rupture nette entre la linguistique structurale américaine et la linguistique chomskienne (p. 41), là où J. Léon se prononce en faveur d'une « approche continuiste de l'histoire de la linguistique américaine » (p. 94, 181 note 5)<sup>3</sup>.

La modalité selon laquelle s'effectue, en France, cette première étape de l'automatisation-mathématisation du langage est complètement différente de celle mise en évidence pour les États-Unis (et la Grande-Bretagne). Cette différence tient à plusieurs spécificités françaises : les sciences de la guerre n'y ont jamais eu d'existence institutionnelle ; l'équipement informatique y est bien moins développé qu'aux États-Unis ; la traduction automatique n'y apparaît qu'en 1959, dans le cadre de la politique gaullienne de recherche scientifique (p. 121). C'est pourquoi l'auteure parle, à propos de la France, de réception externe de l'automatisation-mathématisation, et pas d'intégration d'un nouvel horizon de rétrospection (p. 119-133).

Enfin, à propos de la seconde étape de l'automatisation-mathématisation des sciences du langage – le tournant des grands corpus du début des années 1990 – J. Léon montre bien qu'elle ne correspond aucunement à l'émergence d'une « nouvelle linguistique » (p. 158). En effet, ce tournant ne représente pas une rupture conceptuelle : les corpus étaient disponibles depuis les années 1960, mais c'est la technologie des ordinateurs des années 1990 qui rend leur exploitation possible (p. 175). Il s'agit donc bien plus d'un tournant technologique que d'un tournant conceptuel. Cela amène même l'auteure, en conclusion de son ouvrage (p. 181-183), à se demander, en revenant sur la périodisation du langage sur le temps long, si l'automatisation du langage est bel et bien une nouvelle (la troisième) révolution technologique des sciences du langage, et pas, plus trivialement, une simple conséquence de la révolution de l'informatique sur les sciences du langage.

Pour conclure, le livre de Jacqueline Léon repose sur une bibliographie extensive, voire exhaustive, sur de nombreuses archives institutionnelles, françaises notamment, ainsi que sur un outillage méthodologique et historiographique rigoureux, et sur une épistémologie explicite de l'histoire des disciplines scientifiques. Néanmoins, on pourra regretter qu'il adopte une forme de récit qui, faisant l'impasse sur la définition des concepts et notions clefs de la linguistique contemporaine – linguistique computationnelle, structuralisme néo-bloomfieldien, grammaire générative et transformationnelle – semble parfois s'adresser uniquement aux spécialistes des sciences du langage ou de leur histoire, et demande un investissement substantiel au lecteur profane<sup>4</sup>.

<sup>3</sup> L'auteure insiste notamment (p. 78-95) sur les liens étroits entre les approches du linguiste structuraliste américain Zellig Harris (Harris, 1951) et la grammaire générative de Noam Chomsky (Chomsky, 1957), même si, à partir des années 1960, ce dernier s'efforce de nier toute influence de Harris sur ses propres travaux (p. 95).

<sup>4</sup> Le lecteur potentiel pourra se référer, entre autres, à l'ouvrage introductif de Georges Mounin (1971).

Cela est d'autant plus dommage qu'il s'agit là de recherches qui dépassent largement le cadre des sciences du langage, et qui approfondissent avec acuité des questions centrales de l'histoire des sciences et des disciplines scientifiques.

Emanuel Bertrand

Centre Alexandre-Koyré, UMR 8560 (CNRS/EHESS/MNHN) ; Institut de recherche interdisciplinaire en sciences sociales (IRISSO, UMR 7170, CNRS/université Paris-Dauphine) ; École supérieure de physique et de chimie industrielles de la ville de Paris (ESPCI Paris) ; Université de recherche PSL

## Bibliographie

- Auroux, S.**, 1987, « Histoire des sciences et entropie des systèmes scientifiques. Les horizons de rétrospection », dans Schmitter, P. (éd.), *Zur Theorie und Methode der Geschichtsschreibung der Linguistik*, Tübingen, Narr, p. 20-42.
- Auroux, S.**, 1994, *La révolution technologique de la grammatisation*, Liège, Mardaga.
- Bar-Hillel, Y.**, 1960, « The present status of automatic translation of languages », dans Alt, F. L. (éd.), *Advances in Computers*, vol. 1, New York/Londres, Academic Press, p. 91-163.
- Chomsky, N.**, 1957, *Syntactic Structures*, Londres, Mouton.
- Dahan, A. et Pestre, D.** (éd.), 2004, *Les sciences pour la guerre (1940-1960)*, Paris, Éditions de l'EHESS.
- Harris, Z. S.**, 1951, *Methods in Structural Linguistics*, Chicago, University of Chicago Press.
- Mounin, G.**, 1971, *Clefs pour la linguistique*, Paris, Seghers.
- Weaver, W. et Shannon, C. E.**, 1975 [1949], *Théorie mathématique de la communication*, Paris, Retz/CEPL.
- Weaver, W.**, 1955, « Translation », dans Locke, W. N. et Booth, A. D. (éd.), *Machine Translation of Languages, 14 essays* Cambridge/New York, MIT Press/John Wiley, p. 15-23.