HAL
open science

Recension : Jacqueline Léon (2015), Histoire de l'automatisation des sciences du langage

Naomi Truan

▶ **To cite this version:**

Naomi Truan. Recension : Jacqueline Léon (2015), Histoire de l'automatisation des sciences du langage. Lectures, 2015. halshs-01734546

HAL Id: halshs-01734546

https://shs.hal.science/halshs-01734546v1

Submitted on 14 Mar 2018

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.

Lectures
Les comptes rendus

Jacqueline Léon, *Histoire de l'automatisation des sciences du langage*

Naomi Truan

Éditeur
Centre Max Weber

Édition électronique
URL : http://lectures.revues.org/17935
ISSN : 2116-5289

Ce document vous est offert par
Bibliothèque Diderot de Lyon

Référence électronique
Naomi Truan, « Jacqueline Léon, *Histoire de l'automatisation des sciences du langage* », *Lectures* [En ligne], Les comptes rendus, 2015, mis en ligne le 05 mai 2015, consulté le 14 juin 2017. URL : http://lectures.revues.org/17935

Ce document a été généré automatiquement le 14 juin 2017.

© Lectures - Toute reproduction interdite sans autorisation explicite de la rédaction / Any replication is submitted to the authorization of the editors

Jacqueline Léon, Histoire de l'automatisation des sciences du langage

Naomi Truan

1 Jacqueline Léon propose ici une synthèse des recherches menées entre 1997 et 2010 et ayant abouti en 2010 à un mémoire d'habilitation à diriger des recherches. L'ouvrage, situé en histoire et épistémologie des sciences du langage, propose une synthèse de l'histoire de l'automatisation et de la mathématisation (deux termes que l'auteure s'attache à distinguer) des sciences du langage depuis les années 1950 dans trois grandes aires culturelles : les États-Unis et la Grande-Bretagne, l'ex-URSS et la France. Alors que la traduction automatique impose avec brusquerie un « tournant de l'automatisation » (p. 5) dans un contexte de rivalité accrue entre les États-Unis et l'ex-URSS aux débuts de la guerre froide, l'ouvrage entend étudier non seulement le contexte d'apparition de ces nouvelles techniques et pratiques, mais aussi, empruntant à Auroux (1987, 2007[1]) les concepts d'« horizon de rétrospection », entendu comme l'ensemble des connaissances[2] précédentes possédées par un sujet lors de l'activité cognitive productrice d'un (nouvel) acte de savoir et « horizon de projection », compris comme la capacité à se projeter vers l'avenir[3], leur portée à court et moyen terme pour les sciences du langage. Ces dernières connaissent en outre un second « tournant » au début des années 1990, plus connu des linguistes, puisque celui-ci a débouché sur une nouvelle manière de concevoir la discipline, voire sur une « nouvelle linguistique » (Leech 1992[4]), bien que l'auteure demeure sceptique quant à cette appellation, estimant que le « renouveau de l'empiricité en linguistique », tout en étant un véritable « bouleversement », « ne justifie pas qu'on puisse en appeler à l'émergence d'une nouvelle discipline » (p. 158). Il s'agit de la linguistique de corpus[5].

2 L'ouvrage se décompose en neuf chapitres de longueur égale. Tandis que les six premiers sont consacrés à l'apparition d'une linguistique automatisée dans les années 1950-1960, les chapitres 7 et 8 se concentrent sur la particularité d'une réception française tardive et

sur les applications de l'automatisation des sciences du langage en France, notamment dans les domaines de la documentation automatique puis de l'analyse automatique du discours. Enfin, le dernier chapitre s'attache au *corpus turn* des années 1990, notamment autour des sources britanniques ayant contribué à l'émergence de l'analyse automatisée de grands corpus (voir note 5).

3 Comme l'auteure le souligne à juste titre en conclusion, l'apport de l'ouvrage réside principalement dans la différenciation effectuée entre temps long et temps court dans l'histoire des sciences. Ainsi, l'apparition de la traduction automatique (TA) dans le contexte de guerre froide relève d'un « événement » (p. 177) « sans ancrage disciplinaire spécifique dans des centres universitaires » (p. 14). Il s'agit en effet de (faire) traduire automatiquement des textes du russe vers l'anglais, sans pour autant inscrire cette entreprise dans une réflexion linguistique. Dotée de moyens financiers et institutionnels importants, la TA connaît un franc succès, malgré des traductions souvent « approximative[s] et grossière[s] » (p. 24) et l'absence d'horizon de rétrospection, ce qui conduit l'auteure à parler d'une technologie de guerre « coupé[e] de son passé, voire sans passé du tout » (p. 27).

4 Le « nouveau tournant de la mathématisation de la linguistique » (p. 41), symbolisé par l'émergence de la linguistique computationnelle (ou linguistique informatique) autour de la figure de Bar-Hillel décrit au chapitre 2, se construit par contre dans l'horizon de rétrospection de la TA (p. 56). Il vise à « automatiser le traitement des langues, par la mise en algorithme de l'analyse syntaxique » (p. 41). Ce tournant impose de prendre des distances avec la TA, conçue par et pour l'industrie, afin de faire émerger la linguistique computationnelle en tant que nouveau domaine et nouvelle discipline.

5 La révolution dans l'enseignement des langues initiée par le Mouvement de la Réforme (chapitre 3), qui prête une attention accrue à l'apprentissage d'une langue en contexte et à l'oral, s'effectue dans la continuité de la linguistique appliquée qui se développe dans les années 1960. Des linguistes tels que Bloomfeld, Bloch, Trager ou Joos sont cette fois-ci davantage impliqués « dans l'effort de guerre » (p. 46), bien que le projet engendre des réticences chez les enseignant·e·s de langue, notamment parce que « l'enseignement des langues est encore associé à l'apprentissage de la langue écrite, à l'aide de la traduction » (p. 53).

6 Le chapitre 4, plus technique, expose les ambiguïtés de la théorie de l'information, qui donne lieu à des travaux très hétérogènes et prête dès ses débuts à confusion, notamment en raison des deux acceptions d'« information » entre « information-contenu » et « information-signal » (p. 57)[6]. Comme la TA auparavant, et contrairement à la linguistique computationnelle, la théorie de l'information « est importée de l'extérieur, en pièces détachées et à la demande » (p. 56).

7 La réception du tournant de l'automatisation pour les néo-bloomfeldiens, « plus un réseau qu'un groupe, auquel Bloomfeld lui-même ne se serait pas identifié » (p. 79), est l'objet du chapitre 5. Automatisation et rationalisation sont distinguées, notamment autour de l'exemple de Voguelin et Harris, qui proposent une nouvelle approche de la traduction où est fait appel à la compétence de linguistes, contrairement à la TA qui s'était faite en-dehors de la linguistique.

8 Le chapitre 6 présente les traditions britannique et russe côte à côte, permettant une confrontation fructueuse de deux « acteurs majeurs de la guerre froide » (p. 98), ainsi que les travaux de lexicologie française, qui ne relèvent pas d'« une véritable

mathématisation du langage d'ordre logico-mathématique ou statistique » (p. 111). L'auteure distingue ainsi, à la suite de Gardin (Gardin et Jaulin 1968[7]), « formalisation » (qui ne suppose pas calcul) et « automatisation », montrant que les travaux britannique, russe et français sont comparables, malgré leur diversité, par l'importance accordée à la sémantique pour développer de nouvelles méthodes en TA, quand la tradition américaine autour du modèle computationnel de Bar-Hillel et Chomsky (chapitre 2) se consacre prioritairement à « l'analyse syntaxique logico-mathématique » (p. 97).

9 Le retard de l'informatique en France, mais aussi la faible réception des travaux américains parmi les linguistes français dans les années 1960, expliquent l'apparition tardive de la seconde mathématisation en France (chapitre 7). Les « opérateurs de passage » (p. 123) prennent ici la forme de figures clefs (Schützenberger ou Gross) ou de sociétés savantes, dont certaines ne font pas partie du milieu académique. Malgré ce décalage dans l'appropriation de la TA, la France est précurseur dans le domaine de la documentation automatique et l'analyse automatique de discours, présentées au chapitre 8. C'est dans ce cadre que le concept d'analyse en chaîne de Harris est découvert en France, bien qu'il soit vivement critiqué (Coyaud), voire quasiment passé sous silence (Pêcheux). L'auteure montre en quoi la référence au texte *Discourse Analysis* de Harris (1952) comme texte fondateur de l'Analyse du Discours (AD) à la française relève d'une reconstruction, rappelant les débats actuels sur la pertinence et l'héritage de Harris pour la linguistique.

10 Le dernier chapitre, consacré à la linguistique de corpus, est le seul à faire état d'enjeux méthodologiques et théoriques suscitant encore des discussions passionnées parmi la communauté des linguistes en 2015. Contrairement à la TA des années 1960, qui s'est faite en rupture avec son horizon de rétrospection et à l'écart des linguistes, « l'automatisation des données [...] introduit un changement radical dans la pratique des linguistes » (p. 157), donnant lieu à deux grandes tendances : *corpus-driven* et *corpus-based*[8].

11 Cet ouvrage bien écrit présente un intérêt certain pour tout·e linguiste désireux ou désireuse d'en apprendre davantage sur la constitution de sa discipline et le succès de la linguistique de corpus aujourd'hui, qui explique qu'« il devient rare d'ailleurs qu'un linguiste, de quelque "obédience" qu'il soit, se passe des possibilités offertes par les grands corpus » (p. 157), mais aussi de façon plus générale pour un·e historien·ne des sciences. Soulignant qu'intégration n'est pas synonyme de transfert, l'auteure différencie ainsi « intégration d'un nouvel horizon de rétrospection » (dans le cas de la TA, p. 177), « intégration par adaptation » (théorie de l'information, p. 58), « intégration par confrontation » (« nouvelle linguistique » britannique, p. 178) et « réception externe » (linguistes français, p. 119), fournissant des concepts féconds pour tout·e épistémologue.

12 Malgré un effort d'explicitation non négligeable et des illustrations détaillées, l'ouvrage peut demeurer d'un abord difficile pour quiconque ne disposerait pas de connaissances préalables en linguistique (en sus des index des noms et des notions, un glossaire aurait par exemple été le bienvenu). On pourra par exemple être désarçonné·e par le fait que les néo-bloomfeldiens, mentionnés dès la page 10 puis plusieurs fois ensuite, ne soient réellement décrits qu'en page 78, ou que l'ouvrage soit conduit à faire quelques retours en arrière dans la chronologie, malheureusement inévitables dès lors que plusieurs traditions scientifiques sont décrites simultanément. Ces remarques ne sauraient toutefois entacher l'intérêt de l'ouvrage, qui présente de manière dynamique les intrications entre acteurs, rapports de force et enjeux de pouvoirs dans le renouveau des sciences du langage. Ainsi, les éléments biographiques autour de « personnalités clefs »

(p. 179), mais aussi la description des mécanismes institutionnels et financiers contribuant à la création d'une revue ou d'un pôle de recherche (au sein des universités ou non, et, parmi ces dernières, avec l'appui de linguistes ou non) contribuent à expliquer de façon nuancée les rôles de passeurs de certains chercheurs et la réception de Harris ou Chomsky en fonction des aires culturelles.

NOTES

1. Auroux Sylvain, « Histoire des sciences et entropie des systèmes scientifiques. Les horizons de rétrospection », in Schmitter, Peter (dir.), *Geschichte der Sprachtheorie 1: Zur Theorie und Methode der Geschichtsschreibung der Linguistik*, Tübingen, Narr, 1987, p. 20-42 [1ère parution en 1986 dans *Archives et Documents de la SHESL*, 7, p. 1-26] et Auroux Sylvain, *La question de l'origine des langues* suivi de *L'historicité des sciences*, Paris, PUF, 2007. Voir aussi Auroux Sylvain, « Les modes d'historicisation », *Histoire Épistémologie Langage*, Tome 28, Fascicule 1, *Histoire des idées linguistiques et horizons de rétrospection*, 2006, p. 105-116.

2. Ou « ensemble de références » (Auroux 1987, p. 29, cité ici par Jacqueline Léon page 6).

3. « Parce qu'il est limité, l'acte de savoir possède par définition une épaisseur temporelle, un horizon de rétrospection, aussi bien qu'un horizon de projection » (Auroux 1995, p. 49, cité ici par Jacqueline Léon page 6).

4. Leech, « Corpora and theories of linguistic performance », in Jan Svartvik (dir.), *Directions in Corpus Linguistics. Proceedings of Nobel Symposium, 4-8 August 1991*, Berlin/New-York, Mouton de Gruyter, 1992, p. 105-122.

5. Comme le rappelle Teubert : « La linguistique de corpus n'est pas entrée en scène avant les années soixante du siècle dernier. Le corpus est une collection, réglée par des principes, de données du langage empirique, de textes (ou de fragments de textes), qui sont des échantillons d'un discours donné, dotés en conséquence d'une valeur représentative. » (Teubert Wolfgang, « La linguistique de corpus : une alternative [version abrégée] », *Semen* [online], 27, 2009, p. 185-211).

6. Ainsi, « information » réfère, pour les ingénieurs, physiciens et mathématiciens, à « la grandeur qui caractérise différents modes de communication » liée à la « capacité de stockage ou de traitement des ordinateurs » (p. 57). Cette acception est, selon l'auteure, au plus près du sens étymologique d'information comme « mise en forme d'un signal » (p. 57). Nonobstant, « information » peut aussi faire référence au « traitement des messages en tant que porteurs de sens » (p. 57), redonnant une dimension sémantique au terme, qui glisse vers « information-contenu ».

7. Gardin Jean-Claude et Jaulin Bernard (dir.), *Calcul et formalisation dans les sciences de l'homme, Actes des Journées Internationales d'études sur les méthodes de calcul dans les sciences de l'homme, Rome 4-8 juillet 1966, Unesco et MSH Paris*, Paris, Éditions du CNRS, 1968.

8. Selon Tognini-Bonelli, toutefois non citée par Jacqueline Léon, les études *corpus-based* utilisent les données du corpus pour explorer une théorie une hypothèse et ainsi la valider, la réfuter ou la redéfinir, tandis que les études *corpus-driven* estiment que c'est le corpus lui-même qui devrait être la source d'hypothèses linguistiques (Tognini-Bonelli Elena, *Corpus linguistics at work*, Amsterdam/Philadelphia, John Benjamins, 2001).

AUTHOR

NAOMI TRUAN

Agrégée d'allemand, en préparation de thèse en linguistique comparée (anglais/allemand) et analyse du discours politique à Paris IV (Centre de Linguistique en Sorbonne CeLiSo - EA 7332) et à la Freie Universität Berlin (Institut für Englische Philologie)https://cmb.hu-berlin.de/fr/lequipe/profil/naomi-truan/