

**LA CRITIQUE LITTÉRAIRE  
ET L'ORDINATEUR**

\*

\*

\*

**LITERARY CRITICISM  
AND THE COMPUTER**

---

**BERNARD DERVAL — MICHEL LENOBLE**

**Éditeurs**

**LA CRITIQUE LITTÉRAIRE  
ET L'ORDINATEUR**

\*

\*

\*

**LITERARY CRITICISM  
AND THE COMPUTER**

---

**BERNARD DERVAL — MICHEL LENOBLE**

**Éditeurs**

Dépot légal — 3<sup>e</sup> trimestre 1985  
Bibliothèque nationale du Québec

Copyright © 1985  
B. Derval et M. Lenoble  
ISBN 2-9800398-0-2

## TABLE DES MATIERES — TABLE OF CONTENTS

PREFACE .....	vii
INTRODUCTION .....	xiii
MICHEL H. LENOBLE	
Traitement informatique du texte littéraire présentation et réflexions .....	1
PAUL A. FORTIER	
Using the Computer for Literary Criticism Theoretical Underpinnings and Cost Factors	37
SUSAN S. HOCKEY	
Literature and the Computer at Oxford University .....	53
JOHN B. SMITH	
Arras and Literary Criticism .....	79
ROSANNE G. POTTER	
Computer-assisted Research on Literary Texts the Problem of "Messy Data Sets" .....	95
ETIENNE BRUNET	
La phrase de Zola .....	111
LES AUTEURS — CONTRIBUTORS .....	159

## PREFACE

A conversation on literary methods and research  
 Begat a conversation on computers  
 Which begat a conference

Literary Criticism and the Computer  
 University of Victoria, January 1984

Because we needed to know what it was possible to do in our field of French literature, we decided to organize a two-day presentation of the state of the art and invited a bevey of experts to make known the fruits of their labours to the interested and the uninformed.

That conference begat this book which contains those papers which are most directly useful to literary critics. It is a volume of information for us and for our colleagues.

Desire begat energy  
 Energy begat funding

Our thanks to SSHRCC and the University of Victoria for the generosity which made the conference possible; also to the Dean of Engineering, University of Victoria and the Dean of Arts, University of Western Ontario for their support.

Funding begat participants  
Participants begat papers  
Papers begat interest  
Interest attracted two editors

Our thanks to Bernard Derval and Michel  
Lenoble for their work on this volume.

Possibility of publication begat more funding

Our thanks to Concordia University without  
whose aid this book would not exist.

We conceived of the project  
The labour is over  
In your hands is the brainchild of multiple  
parentage  
We hope it will serve you well.

Louise Forsyth (Univ. of Western Ontario)

Mair Verthuy (Concordia Univ.)

Jennifer Waelti-Walters (Univ. of Victoria)

## PREFACE

D'une conversation sur les méthodes de recherche  
en critique littéraire

Naquit une conversation sur les ordinateurs  
D'où naquit un colloque

La critique littéraire et l'ordinateur  
Université de Victoria, janvier 1984

Parce que nous avons envie de savoir ce qu'il était possible de faire dans notre "domaine réservé" de la littérature française et québécoise, nous entreprîmes d'organiser une rencontre de deux jours qui devait nous permettre de faire le tour de la question; à la-dite rencontre nous invitâmes une nuée d'expert(e)s qui purent ainsi présenter les fruits de leurs recherches aux néophytes intéressé(e)s.

Du colloque naquit ce livre qui contient les communications les plus susceptibles d'intéresser les critiques littéraires professionnels ou amateurs. Pour nos collègues et pour nous, c'est une mine de renseignements précieux.

De la volonté naquit l'énergie  
De l'énergie naquit une subvention

Nous tenons ici à exprimer toute notre reconnaissance au CRSH et à l'Université de Victoria dont la générosité rendit possible cette rencontre; notre gratitude s'adresse également au Doyen de la Faculté de Génie de l'Université de Victoria et au Doyen de la Faculté des Arts de l'Université de Western Ontario dont l'appui fut très apprécié.

Des subsides naquirent les intervenant(e)s  
Des intervenant(e)s naquirent les communications  
Des communications naquit un projet  
Ce projet attira deux éditeurs

Tous nos remerciements vont à Bernard Derval et Michel Lenoble pour le travail immense qu'ils ont effectué lors de la préparation de ce volume.

De l'idée de publier naquirent d'autres subventions

Tous nos remerciements au Doyen des Sciences Humaines de l'Université Concordia sans qui ce livre n'existerait pas.



Nous conçûmes ce projet  
Il vient de voir le jour  
Ce rejeton aux parents multiples,  
Qu'il réponde à vos désirs.

Louise Forsyth (Univ. de Western Ontario)

Mair Verthuy (Univ. Concordia)

Jennifer Waelti-Walters (Univ. de Victoria)

## INTRODUCTION

En janvier 1984 se tenait à l'Université de Victoria en Colombie Britannique un colloque sur le thème "L'ordinateur et la critique littéraire". L'entreprise était assez innovatrice car elle se situait en marge d'une tradition bien établie dans le "monde du traitement de textes". En effet, outre le fait qu'il ne soit pas organisé par une des grandes associations ou un des grands centres réputés dans le domaine mais bien par des néophytes désireux de faire le point sur la question et de voir ce que le traitement de textes pouvait leur offrir comme facilités dans le quotidien de leur recherche, il est remarquable que l'ère des grands colloques, rassemblant tous les chercheurs utilisant l'ordinateur tire à sa fin et qu'est venu le temps des réunions organisées moins en fonction de l'outil ou de la méthode qu'en fonction d'un domaine ou d'un objet d'étude.

En fait, c'est un signe de santé évident du traitement de textes, non plus discipline isolée voire ignorée mais bien méthodologie

heuristique suffisamment élaborée et établie pour être adoptée en tant que telle par une communauté spécifique de chercheurs.

Dans la courte histoire des chercheurs qui ont fait du traitement automatique du texte le centre de leurs préoccupations, la première génération des fondateurs et la deuxième des bâtisseurs de grands laboratoires font maintenant place à la génération toute récente des "utilisateurs" pour qui le traitement automatique est chose acquise et, en fait, le nouveau paradigme heuristique.

Comme toute nouvelle discipline, le traitement de textes a connu une phase de tâtonnement et de développement avant d'inaugurer une phase de réflexion et d'évaluation critique des résultats obtenus. C'est exactement dans cette perspective et dans cet esprit que le colloque s'est déroulé : en effet, il était encourageant de constater qu'une très large majorité des intervenants ne se complaisaient plus à faire état des mille et un problèmes qu'ils avaient dû habilement contourner, mais au contraire, venaient présenter les résultats concrets de leur recherche et amorçaient une réflexion théorique et méthodologique très lucide des avantages et des inconvénients de leur démarche.

Loin d'être la marque d'un reflux ou d'une incertitude, ce glissement d'intérêt témoigne plutôt des interrogations soulevées par la troisième génération (les auteurs ici regroupés sont à cet égard très représentatifs); les "utilisateurs", ayant cédé le fardeau de la programmation aux spécialistes, se consacrent plutôt à mesurer les difficultés d'utilisation, les limites voire la validité des modèles et des outils informatiques.

Depuis que le traitement automatique du texte littéraire existe, il a connu des applications tellement nombreuses et tellement variées que les nouveaux adeptes de la discipline en distinguent difficilement les grands enjeux, les buts concrets et la portée véritable voire la pertinence de bien des résultats. Michel Lenoble dresse une double typologie des programmes et des unités traitables avant d'esquisser quelques questions fondamentales auxquelles la discipline se doit de répondre si elle prétend asseoir définitivement sa crédibilité.

Le traitement automatique du texte littéraire entraîne des dépenses considérables et nécessite une gestion de projet beaucoup plus serrée que par le passé. Outre cet aspect, Paul Fortier analyse également les diverses conceptions du texte qui sous-tendent la plupart des recherches littéraires automatisées.

Les applications de l'ordinateur ne se limitent pas strictement à l'analyse de contenu mais s'intéressent aussi à toutes les opérations jalonnant le processus informatisé de la recherche littéraire. Susan Hockey dirige à l'Université d'Oxford un centre national de services à la communauté universitaire qui offre des facilités pour la préparation, l'entrée, le traitement et la sortie du texte ainsi que, mis à part les programmes d'analyse de contenu, des progiciels d'étude de filiation, de datation et d'attribution de manuscrits.

Au cours des années et au fil des projets, les centres de traitement de texte ont constitué des bibliothèques de magnétoscrits assez considérables. Oxford s'enorgueillit d'une belle collection multilingue de textes sur support informatique. En France, le Trésor de la langue française (TLF) accroît sans cesse son corpus de textes de 1789 à nos jours. En collaboration avec l'Université de Chicago, le TLF a connu un regain de croissance et de vigueur. Cette énorme base de données textuelles est maintenant accessible à tous par bien des réseaux de communication; l'exploitation des textes est réalisée à l'aide du progiciel ARRAS dont John Smith nous présente les principaux attraits.

L'analyse de contenu connaît deux approches très différentes : la première est purement quantitative et statistique : la stylométrie; cette dernière est actuellement dans une phase de révision de sa démarche. Plutôt que de "torturer" les textes et de renoncer à ses propres principes de recherche en soumettant l'oeuvre à des modèles statistiques trop rigides et souvent inopérants, les chercheurs en littérature préfèrent ne plus sacrifier les données et leur approche mais bien revoir, adapter voire créer de toutes pièces des outils statistiques appropriés. Rosanne Potter montre comment et pourquoi elle a été amenée à opérer ce renversement de perspective.

La deuxième approche de l'analyse de contenu est plus qualitative et se sert des outils devenus traditionnels : listes de fréquences, dictionnaires et concordances. Ces types de recherches sont basés sur le mot. Etienne Brunet aborde l'étude de la phrase de Zola par le biais de traits linguistiques infiniment discrets et souvent négligés : les signes de ponctuation, dont il exploite les relevés avec l'unique intention de venir enrichir, d'un point de vue original, la compréhension des "universaux" de l'oeuvre de Zola et du discours romanesque.

Ce recueil est donc le reflet d'un moment de réflexion, d'une pause marquée dans l'incessante fébrilité de nos recherches quotidiennes en constante mutation.

Bernard Derval  
Michel Lenoble

## INTRODUCTION

In January 1984, a conference on Literary Criticism and the Computer took place at the University of Victoria in British Columbia. It was rather innovative since it broke a long-standing tradition in the world of text processing. Indeed, it is striking that this conference was neither organised by one of the major associations nor by one of the outstanding centers renowned in the field, but rather by neophytes eager to discover the state of the art and to see in which way their everyday research could be facilitated with the help of text processing. It is of some interest to note that the era of plenary conferences which gathered researchers who utilize the computer, is coming to an end and that now is the time for seminars focusing not so much on the tool or the method as on the particular field or object of study.

In fact, this is an evident sign of good health for text processing, no longer an iso-



lated or even unrecognized discipline but rather a heuristic methodology sufficiently elaborated and established to be adopted as such by any specific group of scholars.

During the relatively short history of those who have devoted the best of their abilities to text processing, the first generation of founders and the second, builders of large laboratories, are now giving way to a new generation of "users" who take text processing for granted and, in fact, consider it as the new heuristic paradigm.

Just as all other developing fields, text processing has gone through periods of germination and expansion before coming to a rather reflective stage, critically evaluating the results. It is in this very perspective and with that aim in mind that this conference took place : indeed, it was rather encouraging to see that the vast majority of the speakers no longer limited themselves to reporting the numerous difficulties they had overcome, but on the contrary, they came to present their concrete results and also to put forth a lucid theoretical and methodological reflexion on the advantages and disadvantages of their research procedures.

It would be mistaken to see this shift of interest as a sign of withdrawal or lack of confidence. It is much more a question of problems encountered by the third generation and in this respect the authors found here are significant; problems which arise at a time when revolutionary research instruments are no longer being produced; we now have rather to evaluate the difficulties and perhaps the limitations if not the reality of such research.

From its beginnings, the text processing has been applied in the literary domain in such a great variety of ways that newcomers to the discipline have difficulty identifying the issues, the concrete aims and the veritable scope or even the relevance of most of the results. In his chapter, Michel Lenoble draws a double typology of programs and computable units before outlining some fundamental questions that the discipline must answer in order to definitively establish its credibility.

Text processing requires considerable expenditures and necessitates much tighter project management than in the past. In addition to this aspect, Paul Fortier analyses the diverse conceptions of the text which underlie the majority of automated literary research.

Computer applications are not strictly limited to content analysis but also embrace any operation performed by literary scholars. Susan Hockey heads Oxford University's national center offering computer facilities to the entire academic community : textual preparation, input, processing and output as well as program packages for filiation, dating and authorship studies apart from content analysis programs.

In the course of time and from each project, research centers have acquired a considerable number of magnetic versions of texts. Oxford takes pride in its substantial library of magnetoscripts in several languages. In France, Le trésor de la langue française (TLF) is constantly adding to its corpus of texts from 1789 to the present. In its collaboration with the University of Chicago, the TLF is experiencing a renewal of vitality and growth. This huge data base is now accessible to all through many communication networks; the text processing is achieved through the ARRAS package whose main advantages are presented here by John Smith.

Content analysis has two very different approaches, the first, stylometrics, being purely quantitative and statistical. This approach is currently being reevaluated. Rather than deforming textual data and renouncing one's

own research principles by subjecting the work to statistical models which were too inflexible and often inoperative, literary researchers now choose not to abandon their approach or data but instead attempt to revise, adapt or even create appropriate statistical tools from scratch. Rosanne Potter shows how and why she was led to this change of perspective.

The second approach to content analysis is more qualitative and makes use of tools which have now become traditional : frequency lists, indexes and concordances. These listings are word-based. Etienne Brunet studies the sentence in Zola by means of minute linguistic features which are often neglected : punctuation marks whose distribution he interprets with the sole aim of enriching, from an original point of view, our understanding of the universals in Zola's work as well as in the romanesque discourse.

These articles are a reflective pause in the ever-ending evolution of our day-to-day research.

Bernard Derval  
Michel Lenoble