

De quoi parle la littérature ?

Gérald PURNELLE

BERNARD (Michel) : 1994, *De quoi parle ce livre? Élaboration d'un thésaurus pour l'indexation thématique d'œuvres littéraires* (Paris : Champion, Coll. «Travaux de linguistique quantitative», 54), 365 pp.

ISBN 2-85203-386-0.

L'auteur est membre de l'équipe de la Banque de Données d'Histoire Littéraire (BDHL), développée à l'Université de la Sorbonne-Nouvelle (Paris III). L'ouvrage qu'il publie est tiré de sa thèse de doctorat. Il a pour objet de décrire, de manière chronologique, l'élaboration d'un thésaurus de descripteurs utilisables pour l'indexation d'œuvres littéraires et d'évaluer les résultats de cette opération. Il s'est agi, en fait, d'appliquer à un corpus de références littéraires la technique du thésaurus.

Le titre dit bien, dans sa brièveté, quel est le propos de l'ouvrage : permettre de répondre à diverses questions relatives à la thématique de ces œuvres, non seulement « de quoi parle [tel] livre ? », mais aussi « quels sont les livres qui parlent de ... ? ».

L'entreprise ici décrite répondait, selon l'auteur, à une triple demande et, corollairement, à une triple problématique. Face à la masse grandissante des textes imprimés, plusieurs groupes marquent le souhait de pouvoir accéder, par voie informatique, à la thématique des œuvres. Les problèmes sont à la fois informatiques, littéraires et documentaires; les demandeurs sont la BDHL elle-même, la littérature comparée et la corporation des documentalistes.

La BDHL est une banque de données bibliographiques et factuelles, qui regroupe, pour 500 auteurs et 1 600 œuvres de langue française, des données biographiques, des titres, des dates, des éditeurs. L'équipe de la BDHL a ressenti le besoin d'ajouter à ces données des informations relatives au contenu des œuvres : il fallait arriver à décrire les thèmes de chacune d'entre elles.

Des autres demandeurs, les uns, historiens de la littérature, pourront utilement recourir à cette banque de données pour leurs études thématiques; les autres, documentalistes et bibliothécaires, attendent de disposer d'un instrument qui leur permette d'indexer aussi facilement la littérature que les autres livres.

✉ Université de Liège; L.A.S.L.A.; place du 20-Août, 32; B-4000 Liège (Belgique).

Fax : +32 41 23 25 45

E-mail : gpurnell@vm1.ulg.ac.be

La question fondamentale qui se posait à l'auteur (et qu'il examine dès la p. 23) était donc : qu'est-ce qu'un thème littéraire ? Après avoir passé en revue les définitions du terme proposées par ses prédécesseurs (essentiellement des comparatistes et des théoriciens), il paraît en arriver à une conception qui a le mérite d'être directement orientée vers ses objectifs : la comparaison des œuvres entre elles, leur catalogage en bibliothèque ou en banques de données : le thème est ce qui permet de rapprocher des œuvres diverses, les thèmes d'une œuvre pouvant se combiner; le thème est le nom que nous donnons à un point commun entre plusieurs œuvres.

L'auteur évoque ensuite la spécificité de l'indexation d'œuvres littéraires, et les problèmes qu'elle pose, notamment aux documentalistes. Il a raison d'opposer ces œuvres aux autres ouvrages : comme il l'indique, la littérature ne peut actuellement faire l'objet d'une indexation thématique dans les bibliothèques au même titre que les ouvrages techniques, et *Germinal* ne peut être mis sur le même pied qu'une histoire de la mine; consécutivement, les thésaurus utilisés jusqu'à présent ne peuvent servir pour la littérature.

Suivent une description de la technique du thésaurus et des règles qui président au choix de descripteurs qui le constituent, et une évocation de diverses tentatives précédentes, automatiques ou non, et portant sur la littérature. L'intérêt du thésaurus, dans ce cas, repose dans le contrôle de la terminologie employée.

Revenant à la définition du thème littéraire, l'auteur se contente finalement d'une définition strictement pragmatique, le thème littéraire étant « ce que les spécialistes de littérature appellent thème », dans le sens où l'on peut parler « d'universaux thématiques indiscutables ».

Dans les chapitres suivants l'auteur décrit la confection de son thésaurus puis sa mise en œuvre. Par diverses comparaisons (avec la technique de l'indexation automatique, du résumé, etc.), il évalue la pertinence et l'efficacité de sa méthode. Enfin, il examine les perspectives qui s'offrent à son thésaurus et à la BDHL ainsi complétée : la statistique thématique, l'accès intelligent à des bases de données textuelles telles que FRANTEXT ou même la lecture automatique de textes. En annexe figure notamment le thésaurus complet, ainsi que la liste des œuvres indexées, chacune étant accompagnée de la liste des descripteurs qui lui ont été attribués, c'est-à-dire des thèmes que les indexeurs y ont trouvés.

L'entreprise de M. Bernard fut vaste. Elle a donné naissance à un instrument riche et puissant. Assurément, il est à même de remplir les missions que son auteur lui assignait. Un point, toutefois, reste à mon sens épineux. Il est lié à la question, somme toute non résolue, de la définition du thème littéraire; il est en outre tributaire d'une relative divergence entre les deux finalités du thésaurus : indexation documentaire ou études thématiques.

Les descripteurs retenus sont quasi exclusivement abstraits, ou conceptuels : amour, mort, femme, poète, rêve, nature, souvenir, etc. Leur liste est en outre le produit d'une réduction drastique, opérée par généralisation ou substitution de synonymes partiels. Le but était ici d'éviter les synonymies, les redondances, les ambiguïtés, l'éparpillement du signifié. Enfin les descripteurs attribués à une œuvre ne sont pas

hiérarchisés (en principaux et secondaires, par exemple). Cette méthode me paraît présenter des conséquences embarrassantes. L'auteur formule d'ailleurs lui-même, très honnêtement, les objections qui viennent à l'esprit de son lecteur (pp. 179-180).

D'une part, l'indexation paraît pécher par abondance. Chaque ouvrage étant en moyenne affecté d'un nombre important de descripteurs, un grand nombre de ceux-ci se retrouvent en de nombreux endroits (ainsi une œuvre sur trois parle de l'amour), ce qui, paradoxalement, réduit leur utilité.

Les réflexions statistiques auxquelles l'auteur se livre pâtissent de cette extrême abstraction de la thématique telle qu'elle est contenue dans la banque. Compter toutes les œuvres qui « parlent » d'amour, ou de mystère, ou de poète, a-t-il un sens, dès lors que la façon dont elles en parlent et ce qu'elles en disent se trouvent oblitérés par des termes aussi vagues ?

Le fait d'attribuer conjointement deux ou plusieurs thèmes à une même œuvre est-il toujours innocent ? Je songe notamment au cas des recueils de poésie. Peut-on dire que tout *Le Cornet à dés* de Max Jacob parle de « mort, guerre, cheval, féerie, humour, mythologie, dieu, beaux-arts, édifice, lumière, artiste, religions constituées, messie » ? Et que dire de l'indexation des *Essais* de Montaigne ? Dans le cas de *Madame Bovary*, l'association, par simple juxtaposition dans une liste, des thèmes « amour », « adultère », « ennui » et « bourgeoisie » est pertinente ; mais que dire de la cooccurrence, par exemple, de « vin » et « oiseau » pour *Alcools* ?

On trouve en outre, disséminés parmi les descripteurs, des termes qui renvoient à l'ouvrage lui-même et non à ses thèmes ; ainsi d'« humour » pour Jacob ci-dessus, ou de « surréalisme » pour toute œuvre surréaliste. Il ne s'agit pas de ce dont parlent ces ouvrages ; la nature des descripteurs est donc ambiguë.

Pour les études d'histoire littéraire comparée, le thésaurus décrit me semble présenter un intérêt certain, mais variable, comme le montrent, à mon sens, les quelques esquisses auxquelles se livre l'auteur. Les unes sont riches d'enseignements (ainsi la fluctuation des places occupées par les thèmes de l'amour et de la mort au cours des siècles), mais d'autres pâtissent de la trop grande généralisation de l'indexation (notamment la comparaison de deux œuvres sur base statistique).

En ce qui concerne la documentation, c'est-à-dire le recours à ce thésaurus pour rechercher, en bibliothèque, toutes les œuvres littéraires parlant d'un thème — démarche qui, selon l'auteur, devrait notamment intéresser les enseignants — les mêmes inconvénients viennent tempérer les mêmes qualités.

En conclusion, reconnaissons que le but que s'était imposé l'auteur n'était pas facile. Les exigences de la rigueur technique lui imposaient une démarche dont il est le premier à reconnaître qu'elle se heurte, dans le cas de la littérature, à la spécificité de ce domaine. Il n'est pas difficile de dire de quoi parle un manuel de menuiserie, un précis de métaphysique ou un traité de chimie : ils parlent précisément de menuiserie, de métaphysique et de chimie. Mais de quoi parlent *La Recherche*, *Les Fleurs du Mal* ou *Phèdre* ? Comme le suggère Michel Bernard lui-même, aucune réponse ne pourra satisfaire le moindre lecteur attentif de ces œuvres. Néanmoins, l'aventure valait la peine d'être tentée.