Conception d'une base de données géomatique pour l'étude de l'espace narratif de *Nedjma* de Kateb Yacine.

Juliette Morel

Université Rennes 2, CELLAM, France.

Résumé : La démarche développée dans cet article s'intègre dans le cadre des humanités numériques à double titre, puisqu'elle s'inspire des modèles de la géomatique pour proposer un nouvel outil numérique à l'analyse littéraire. Il s'agit de construire une base de données narratologique capable de rendre compte de la non-linéarité d'une structure narrative comme celle des œuvres de Kateb Yacine et d'explorer l'espace littéraire ainsi construit.

Mots clefs: humanités numériques, littérature, géomatique, espace littéraire, base de données, narratologie, Kateb Yacine.

La démarche que je développe dans ma thèse de doctorat se situe au carrefour de trois grandes tendances des humanités numériques : l'explosion des initiatives dédiées à la rencontre des outils informatiques et de l'analyse littéraire, notamment à travers les outils de textométrie, la géomatique, et l'utilisation de la visualisation comme outil de communication et méthode de traitement. Mon objectif est d'emprunter les méthodes de modélisation de la géomatique pour explorer graphiquement la structure narrative non-linéaire et fragmentée du roman *Nedjma* de Kateb Yacine [1] et plus généralement pour cartographier l'espace littéraire ouvert au sein du « Cycle de Nedjma »[1], ensemble d'œuvres où s'intègre ce roman. Pour cela, je propose de construire une base de données sur le modèle des bases de données géographiques, afin de référencer l'œuvre et de l'aborder de manière géomatique.

Appréhender de manière nouvelle l'œuvre de Kateb Yacine

Kateb Yacine est un auteur algérien né en 1929 et mort en 1989. Il a écrit en Français, entre 1956 et 1970, un ensemble d'œuvres rassemblées sous le terme de « Cycle de Nedjma », dont fait partie *Nedjma*, son roman majeur. Le déroulement narratif de ce roman ne correspond pas à l'enchaînement chronologique des événements de la diégèse (histoire factuelle racontée dans le livre) : les ellipses, les analepses et les ruptures spatio-temporelles sont courantes au point qu'il est impossible de déterminer clairement une trame principale. Les fragments narratifs sont en moyenne très courts. Certains événements diégétiques sont répétés mot pour

1 Ensemble cohérent d'œuvres de Kateb Yacine écrites en français dans les années 1950 et 1960, partageant une figure centrale, Nedjma, la femme aimée, un même système de personnages, un même contexte historique et géopolitique, etc.. Le Cycle de Nedjma comprend *Nedjma* (1956), *Le Cercle des représailles* (1959), *Le Polygone étoilé* (1966) et les nombreux fragments partageant les caractéristiques citées, en partie réunis dans *L'Œuvre en fragments* (1986).

mot ou réécrits à plusieurs reprises. Ces reprises ou réécritures se font non seulement à l'intérieur du roman mais également dans les autres œuvres composant le Cycle. Dès lors émerge un réseau dense de correspondances immanentes. Les nombreuses répétitions permettent de temps à autres de se repérer au sein de cet ensemble qui bouleverse les codes et la linéarité classiques de la narration romanesque. Mais parfois, de légers décalages dans les récits - de lieux, de temps ou de personnages - égarent à nouveau, voire davantage, le lecteur. Finalement, un mouvement cyclique s'impose : le début est repris tel quel à la fin du roman ; mais encore l'inversion in extremis de deux épisodes perturbe…

Si l'organisation de *Nedjma* n'est pas linéaire d'un point de vue géographique ou temporel, c'est que son principe est à chercher du côté de l'espace littéraire construit par le récit. Autrement dit, la structure de *Nedjma* n'est pas imposée par la diégèse mais structurée selon les modalités de la narration qui décrivent une figure spatiale. Kateb Yacine définit lui-même son processus d'écriture en termes spatiaux :

« *Quand j'ai écrit Nedjma, on a beaucoup parlé de technique, etc., mais au fond* ***le mouvement****, si on veut tout réduire à un seul mouvement, c'est tout bêtement l'écrivain qui commence à créer* ***en partant d'un point****. À la fin, il arrive à un autre point, il se relit et il s'aperçoit qu'il n'a pas tout dit. Quand il se remet à écrire, ce qui sort n'est pas tout ce qu'il avait en lui. Alors on* ***repart****,* ***on revient à son point de départ*** *et on essaie une autre foulée, revenir au même point mais par* ***un autre chemin****. Et entre temps, on défriche.* » [2]

C'est au sein de ce mouvement, sur les lignes et dans les formes qui se tracent, que réside la cohérence narrative de *Nedjma*. Celle-ci suit précisément une logique géométrique qui peut être schématisée par une figure récurrente dans l'œuvre de Kateb Yacine : le polygone étoilé.

Le recours aux technologies numériques, et à la géomatique en particulier, permet d'appréhender la problématique de la structure géométrique de l'œuvre en s'affranchissant de la linéarité du discours critique et en concrétisant enfin – c'est-à-dire en éprouvant – l'hypothèse de la spatialité de l'œuvre littéraire. En effet, l'analyse littéraire classique, dans un premier temps au moins, a tendance à « mettre de/dans l'ordre », c'est-à-dire à rétablir la chronologie et la contiguïté géographique, afin de comprendre et d'analyser le récit tel qu'on en a l'habitude, selon le principe de progression du raisonnement cartésien. Au contraire, questionner les données littéraires issues de ce roman à travers une base de données construite à cette fin permet de remplacer la progression et l'étalement temporels du discours critique par une vision en coupe, simultanée et potentiellement globale de la matière littéraire. Il s'agit en fait simplement d'adopter une nouvelle perspective, puisque la construction d'une base de données oblige à adopter une logique différente, de l'ordre de l'algorithmie et du modèle relationnel. Cette démarche permettra de naviguer dans l'œuvre grâce à des requêtes ciblées, d'obtenir des résultats exhaustifs et de découvrir probablement d'autres dynamiques narratives ; tout du moins d'affiner la connaissance de celles dont nous avons déjà l'intuition. En un mot, l'objectif est d'étudier la structure particulière de l'œuvre sans avoir à la déstructurer ni la restructurer, c'est-à-dire sans la dénaturer.

Une base de données géomatico-narratologique

Parmi les outils les plus répandus actuellement en analyse littéraire numérique figure l'encodage en xml. Cet outil a l'avantage d'impliquer une implémentation bien moins lourde que la construction d'une base de données ainsi

qu'une interopérabilité et une généricité plus importantes (notamment grâce à sa normalisation par le Text Encoding Initiative (TEI)). L'encodage du texte de Kateb Yacine et l'exploration des potentialités offertes par cette technique se sont donc imposés comme une première étape nécessaire à la définition d'une méthode d'analyse numérique optimale selon l'objectif exposé plus haut. Cet encodage a consisté à identifier des groupes de mots selon leur fonction (marqueurs de temps et de lieu, noms de personnage, thèmes, etc.) et à définir des unités narratives relativement grandes sur un principe d'unités spatio-temporelles (changement d'unité, c'est-à-dire de balise, quand il y a rupture spatiale et/ou temporelle). L'encodage en xml ne s'est pas révélé pleinement satisfaisant parce que le modèle de l'arbre hiérarchique et dichotomique sur lequel il est fondé correspond à un schéma de pensée trop contraignant et construit un modèle de données trop strict (notamment parce qu'il interdit le chevauchement des balises). Un tel schéma n'est pas adapté à l'objet littéraire en question, marqué par l'hétérogénéité, l'irrégularité et le foisonnement. Une fameuse métaphore botanique développée par Gilles Deleuze et Felix Guattari permet de comprendre cette inadaptation : le xml encode en arbre alors que l'objet littéraire étudié est davantage de l'ordre du rhizome, caractérisé par la connexion (entre ces différentes parties), la fragmentation en tant qu'elle est rupture non-signifiante, la multiplicité et l'hétérogénéité [3]. Le modèle de la base de données relationnelle, qui met par définition en relation des éléments hétérogènes, parait plus appropriée pour traiter et rendre compte de cette construction rhizomatique.

La finalité concrète de mes recherches est donc de construire une base de données relationnelle référençant un ensemble narratif – *Nedjma* dans un premier temps, mais à terme tout le Cycle de Nedjma – sur le modèle des bases de données conçues pour référencer, traiter et interroger des données géographiques. Selon l'idée qu'il est possible de prendre au mot l'hypothèse de la spatialité de l'œuvre littéraire (au moins pour son expérimentation), il s'agit de considérer les unités narratives découpant le texte à différentes échelles comme des unités spatiales, comme des espaces ponctuels, comme des lieux par lesquels passe la ligne du texte. Ces unités narratives (métaphoriquement) spatiales seraient donc localisables les unes par rapport aux autres, et au sein d'ensembles spatiaux hiérarchiquement supérieurs, comme les sous-chapitres, les chapitres, l'ouvrage et finalement l'ensemble d'ouvrages, le Cycle. Il est donc question de construire un(des) système(s) de coordonnées propre(s) au livre afin de donner à chaque fragment narratif des coordonnées qui le situent dans l'œuvre, à la fois dans la matérialité du livre, mais aussi dans la progression de la diégèse. C'est sans doute à la rencontre des deux (en définissant par exemple un x de l'ordre de la matérialité de l'œuvre et un y de l'ordre de la progression de la diégèse) que l'espace littéraire – dans le sens de volume, étendue, profondeur ouverts entre les faits racontés et la manière de les raconter[2] – pourra être réellement approché et représenté. Il s'agira donc d'expérimenter différents systèmes de coordonnées, utilisant les différentes mesures possibles de la matérialité de l'œuvre littéraire[3] – le numéro de page, le numéro de ligne, le numéro du sous-chapitre ou du chapitre, le rang chronologique du fragment narratif, etc. – et ainsi d'explorer différentes modélisations, pour éprouver leur pertinence et la pertinence *in fine* du fait de

2 Selon une conception typiquement genettienne du récit [4].

3 Dans un premier temps, l'idée est de transposer cet espace littéraire dans un espace euclidien, pour pouvoir l'appréhender dans des SIG, mais évidemment une réflexion sur le type d'espace mis en jeu est menée parallèlement. L'espace littéraire est-il un espace non-euclidien ? Y-a-t'il des espaces non-euclidiens qui seraient plus à même d'accueillir un espace littéraire comme celui construit par Kateb Yacine ? Auquel cas, la question à se poser est : comment utiliser les SIG pour traiter et représenter cet espace non-euclidien ? Comment mesurer et rendre compte cartographiquement de cet espace ?

considérer l'œuvre dans sa spatialité. Cette localisation systématique permettra enfin d'étudier précisément les rapports des unités narratives entre elles, par exemple à travers des notions de proximité ou d'éloignement au sein du système de coordonnées choisi.

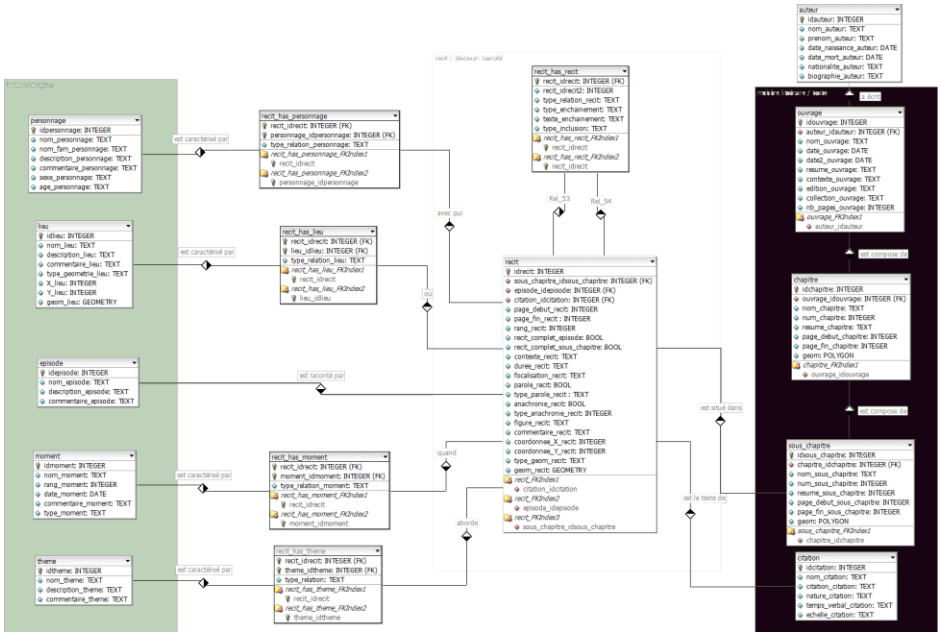

Fig. 1. Schéma UML de la base de données construite pour étudier la structure narrative du Cycle de Nedjma de Kateb Yacine

Pour affiner la structure de la base de données et l'adapter parfaitement à son objet, certaines notions de narratologie notamment développées par Gérard Genette dans *Figure III* au début des années 1970[4], sont venues compléter le modèle géomatique lors de sa conception. C'est ainsi que la base de données créée différentie d'un côté les éléments de la « diégèse » (avec une table référençant les personnages, une table lieux, une table moments de l'action, une table thèmes et une table épisodes diégétiques, sur fond gris clair sur le schéma UML) et de l'autre les éléments textuels (avec une table référençant les ouvrages, une table chapitres, une table sous-chapitres et une table citations, sur fond gris foncé). Au centre se trouvent les informations concernant les « récits » (table récit), que j'ai définis comme les unités narratives cohérentes de base, c'est-à-dire les citations situées dans le texte (par un attribut page et leur liaison avec la table des sous-chapitres) et dans la diégèse (par leur liaison avec les tables référençant les éléments diégétiques cités plus haut). Cette dichotomie permet de définir le récit comme le véritable objet de cette base de données, et non plus seulement le texte ou le lexique comme c'est le cas en général dans les études littéraires utilisant des outils numériques.

Cette base de données permettra par exemple de mettre en évidence les réécritures – qui correspondent à des relations entre un seul et même objet dans la table épisode et plusieurs objets dans la table récit – ; les répétitions à la lettre – qui correspondent à des relations entre un seul et même objet dans la table citation et plusieurs objets dans la table récit – ; les échos thématiques – qui correspondent à des relations entre un seul et même objet dans la table thème et plusieurs objets dans la table récit. Grâce à l'étude de ces relations, on peut imaginer quantifier précisément le phénomène de la répétition chez Kateb Yacine et ainsi juger de son importance : en

créant par exemple un indice qui comparerait les épisodes et les passages répétés par rapport à ceux qui ne le sont pas. Un autre exemple d'application à partir de cette base de données peut être la mesure du bouleversement chronologique à l'œuvre dans *Nedjma*, à travers la comparaison des rangs chronologiques rétablis dans la table des moments de l'action et des rangs effectifs dans la progression des pages de l'œuvre, contenus dans la table des récits. Ce phénomène pourra également être visualisé grâce à la transposition du rapport entre la distance temporelle entre deux épisodes au sein de la diégèse et la distance physique des deux bouts de texte correspondant au sein de l'objet livre. Il s'incarnera par une distance géométrique plus ou moins grande, reportée dans un repère orthonormé. Enfin, cette base de données permettra d'étudier de manière exhaustive la façon dont on passe d'un lieu à un autre dans le texte (comment passe-t-on d'un lieu page 1 à un autre lieu page 2 ?). La comparaison visuelle entre la proximité des lieux dans l'espace géographique (sur une carte géographique du monde) et la contiguïté narrative de ces lieux au sein de l'espace littéraire (figurée cartographiquement par des liens tracés entre ces lieux) sera intéressante pour étudier l'éclatement, ou au contraire la linéarité, et l'étendue géographique du récit. Cette carte se fondera sur la relation que tisse la base de données entre les lieux (table lieux) et le texte (table citation) à travers le récit (table récit), cette dernière contenant l'information de l'ordre narratif.

Le réel potentiel de cette base de données se révèlera lors de son exploration. Il reste beaucoup à faire dans le domaine de la transposition d'outils numériques à l'analyse littéraire et l'essentiel est de valoriser l'exploration et la démarche expérimentale, mais aussi la complémentarité entre les méthodes numériques. L'encodage en xml est utile à l'étape de la collecte, la description et la prise en main des données ; les outils de textométrie permettent d'avoir une vision globale et immédiate de l'œuvre et d'établir des hypothèses pertinentes à approfondir autrement ; enfin le modèle de la base de données parait approprié pour l'analyse narratologique d'une structure littéraire complexe et rhizomatique telle que celle du Cycle de Nedjma de Kateb Yacine. Le tout est de ne jamais perdre de vue la spécificité de son objet, question centrale des humanités numériques.

Références

1. Kateb Yacine : *Nedjma*. Edition du Seuil, « Point », Paris, France (1956).
2. Kateb Yacine : *Le Poète comme un boxeur : entretiens 1958-1989*. Textes réunis et présentés par Gilles Carpentier avec le concours de l'IMEC. Éditions du Seuil, Paris, France (1994).
5. Gille Deleuze et Felix Guattari : *Mille Plateaux, Capitalisme et schizophrénie 2*. Les Editions de Minuit, Paris (1980).
4. Gérard Genette : *Discours du récit*. Editions du Seuil, « Point essais », Paris, France ([1972], [1983], 2007).