Analyse littéraire et traitement automatique de corpus de littérature orale
Une approche expérimentale

Ursula BAUMGARDT
Plidam-Inalco, Paris

Résumé : Je voudrais présenter un exemple d'utilisation de base de données textuelles dans l'analyse d'un corpus de littérature orale. L'approche proposée est expérimentale, en ce sens que je ne me réfère pas à une école ni à une méthode vérifiée et confirmée que j'appliquerais à l'objet dont il est question ici. Je retracerai simplement le cheminement de mes interrogations qui, partant de la littérature africaine francophone, ont intégré la littérature orale en peul. L'une des questions concerne les représentations de l'espace dans des contes oraux. La réflexion porte sur les relations pouvant être éventuellement établies entre le traitement automatique de corpus et l'analyse littéraire.

Mots-clés : littérature orale, corpus, littérature africaine francophone, contes oraux

Je voudrais présenter un exemple d'utilisation de base de données textuelles dans l'analyse d'un corpus de littérature orale. L'approche proposée est expérimentale, en ce sens que je ne me réfère pas à une école ni à une méthode vérifiée et confirmée que j'appliquerais à l'objet dont il est question ici. Je retracerai simplement le cheminement de mes interrogations qui, partant de la littérature africaine francophone, ont intégré la littérature orale en peul. L'une des questions concerne les représentations de l'espace dans des contes oraux. La réflexion porte sur les relations pouvant être éventuellement établies entre le traitement automatique de corpus et l'analyse littéraire. Le premier peut ouvrir à la seconde des perspectives nouvelles et inattendues. Mohand Akli Salhi (2015) a, par exemple, développé une approche de la métrique kabyle ; dans le même sens, Patrice Pognan et Mohand Akli Salhi (2016) ont analysé la rime dans la même langue[1].

[1] Mohand Akli Salhi et Patrice Pognan (2016) ont analysé des textes de poésie pour en extraire les mots en position de rime, l'application pouvant être un dictionnaire des rimes kabyles. Auparavant, Mohand

1 Un premier objet d'observation : littérature africaine francophone

Le point de départ est pour moi la littérature africaine francophone[2] ; d'où un rapide retour dans ce domaine.

L'interrogation portait d'abord sur la dimension politique de l'écriture romanesque des auteurs africains francophones dans les années 1970, et mon approche était alors sociologique. Cependant, plusieurs romans ont soulevé des questions relatives au fonctionnement des textes. J'ai essayé d'y répondre en me référant à la sémiotique textuelle, mais sans l'appliquer selon un schéma préétabli[3].

Dans cette première étape, mon corpus était *Le Cercle des Tropiques* (1972) d'Alioum Fantouré. Il s'agit du premier roman de l'auteur, qui est guinéen ; il y développe, à propos d'un pays imaginaire, une réflexion politique approfondie sur une société postcoloniale.

Dans un contexte où le régime politique en Guinée était très controversé, la question qui se pose ici est de savoir à quelle réalité le texte réfère-t-il ? Selon une lecture « biographique » associant un texte littéraire à la vie de son auteur, on pourrait être tenté de stipuler que le texte renvoie à la Guinée. Or, pour éviter une telle extrapolation, il s'agit d'identifier le référent[4] au niveau textuel, en analysant les références au réel hors-texte, notamment celles qui renvoient à l'espace[5]. En tant que genre narratif, le roman est propice à une telle analyse car il crée un espace dans lequel s'inscrit l'action[6]. Pour cette raison, on peut s'attendre à ce qu'un texte narratif long présente suffisamment de matériaux textuels pour que ce niveau puisse être appréhendé. L'analyse prend en compte les axes paradigmatique et syntagmatique tout en s'intéressant aux articulations entre eux, l'objectif premier étant dans le cas décrit, l'identification de l'espace de référence.

La méthode d'analyse textuelle que j'ai appliquée consiste à relever tout terme lié à l'espace, dans sa qualité d'indice spatial. J'utilise « indice », non pas « marqueur ». La différence entre les deux réside dans leur fonction : l'indice n'identifie pas priori-

Akli Salhi (2015) a analysé des textes de poésie du point de vue de la structuration métrique et des formes les plus fréquentes.

[2] Hermann Hofer, romaniste, professeur à l'université de Marburg, m'a encouragée à explorer ce domaine, à ce moment encore peu connu en Allemagne.

[3] À la différence de A. J. Greimas, défenseur entre autres du « schéma actantiel » et du « Carré sémiotique », Claude Abastado, professeur à l'Université de Nanterre et directeur du « Centre de sémiotique textuelle », avait une approche focalisée sur l'observation des textes eux-mêmes dans leur spécificité. Je m'y réfère dans ma thèse, 1985, « 'Le Cercle des Tropiques' d'Alioum Fantouré. Texte et contexte », sous la direction de Claude Abastado.

[4] Ce à quoi renvoie une unité linguistique.

[5] La démarche que j'adopte se réfère nécessairement à la définition la plus générale possible de l'« espace », comme une « étendue, concrète ou abstraite », car l'analyse a pour objectif de relever des qualifications, des structurations et des représentations éventuelles de l'espace concerné.

[6] Quatre niveaux peuvent être distingués au niveau du texte romanesque : (1) *L'espace narratif* est le cadre dans lequel se situent les actions des personnages ou actants. (2) En revanche, l'action n'amène pas les personnages dans l'*espace évoqué* ; si c'est le cas, il participe de la première catégorie. (3) *L'espace référentiel* (ou hors-texte) peut être déduit à partir d'un ensemble d'indices. (4) Concernant *l'espace de la narration [ou du discours],* l'analyse s'intéresse entre autres à la représentation de l'espace dans lequel se situe le narrateur, au type de narrateur et à son positionnement par rapport aux faits narrés. Je ne prends pas en compte cette dernière question ici. Le contexte d'oralité connaît une dimension supplémentaire (5) *l'espace de la performance,* celui où se rencontrent l'énonciateur et le public. Jean Derive (2010, p. 201 à 222) a analysé cette problématique dans son article « Typologie et fonctions de quelques genres oraux du Manding à l'aune du critère de la spatialité ».

Published in : JAMBOROVA LEMAY, Diana ; PEREIRA, Christophe ; ŠMILAUER, Ivan (dir.), *Des langues calculables à l'homme incalculable. Hommage à Patrice Pognan*, Editions des archives contemporaines, ISBN : 9782813004260, doi : 10.17184/eac.9782813004260

tairement l'espace, sa présence est plutôt liée à la narration, qui peut être une forme de « mouvement » parfois concrétisé à travers des déplacements de personnages. Dans ce cas, la fonction d'indice spatial est secondaire. Cependant, à travers leur valeur référentielle, ces indices, auxquels s'ajoutent d'autres, permettent de situer l'espace narratif géographiquement et d'identifier l'espace hors-texte.

J'ai relevé tous les indices de manière conventionnelle, *i.e.* manuellement, je les ai triés et j'ai retenu les indices spatiaux pertinents dans le texte[7].

Le résultat est précis et significatif. En comparant tous les indices et par déduction, l'identification de l'espace référentiel est possible, il s'agit effectivement de la Guinée. Cette analyse donne accès à la vison politique exprimée à travers le roman et d'en appréhender la portée : la référence au régime de Sékou Touré est claire, il s'agit d'une critique finement argumentée et sans concession. C'est une première dans le roman africain francophone et qui mérite d'être soulignée.

2 Un deuxième objet d'observation : littérature orale

La comparaison entre « l'écrit » et « l'oral » est fréquente. Elle est erronée si elle consiste à transposer une approche élaborée à propos d'une littérature donnée pour l'appliquer à une autre littérature, pourtant radicalement différente dans son fonctionnement. Dans le cas présent, il s'agirait de passer d'un texte d'un auteur africain relevant de l'écriture littéraire en français, à la littérature orale (qui est devenue mon domaine de spécialisation[8]). Un retour rapide sur quelques spécificités de cette modalité d'expression[9] est nécessaire. Je les évoque à propos du conte.

La littérature orale est produite en performance, *i. e.* en contexte de communication directe et non médiatisée, réunissant le producteur et le public. La performance et la langue définissent ainsi une situation d'énonciation spécifique, non reproductible à l'identique. Elle suppose une relative homogénéité culturelle entre l'énonciateur et le public. Il est probable que les deux pôles de la communication, partageant la même langue, partagent aussi une certaine connaissance du contexte. Ainsi, l'appartenance du texte à une culture ou à un univers culturel est censée être partiellement connue ou du moins identifiable. Par conséquent, la performance implique la possibilité de créer un certain degré d'intercompréhension, en s'appuyant sur l'implicite[10]. La référentialisation opère sur cette base et a recours à des procédés spécifiques.

En effet, l'identité et l'appartenance culturelles du texte sont fortement déterminées par la langue. Le texte étant ainsi inscrit dans son contexte culturel, l'énonciateur

[7] Cette démarche participe de la sémiotique textuelle.

[8] Claude Gouffé et Roger Labatut de l'INALCO étaient mes professeurs respectivement de haoussa et de peul. J'ai découvert la littérature orale à travers les enseignements de Geneviève Calame-Griaule et son groupe de recherche organisé autour de l'ethnolinguistique. Jean Derive m'a offert la possibilité de réfléchir avec lui dans une perspective méthodologique et théorique sur la littérature orale. Je me réfère à ces travaux qui ont trouvé une concrétisation dans le projet « Encyclopédie des littératures en langues africaines » (ELLAF).

[9] Je me réfère dans cet aperçu à l'ouvrage, U. Baumgardt et J. Derive (dir.), 2010.

[10] La littérature orale est souvent analysée d'un point de vue anthropologique, notamment dans la perspective d'interroger les textes sur la réalité sociale et culturelle ; les contes peuls ne sont pas une exception. Or, les textes oraux et précisément les contes, fournissent généralement peu d'explications sur le contexte, car en règle générale, ils se situent dans un contexte d'homogénéité culturelle entre l'énonciateur et le public.

Published in : JAMBOROVA LEMAY, Diana ; PEREIRA, Christophe ; ŠMILAUER, Ivan (dir.), *Des langues calculables à l'homme incalculable. Hommage à Patrice Pognan*, Editions des archives contemporaines, ISBN : 9782813004260, doi : 10.17184/eac.9782813004260

peut évoquer le « Nous » et l'« Autre » sans explications supplémentaires, car le public comprendra aisément ce dont il est question[11]. Une perspective comparative, nouvelle, s'ouvre ici, à savoir l'incidence du contexte de communication sur les textes. Prenant comme exemple l'espace, il est possible d'observer des différences entre les deux modalités d'expression littéraire, l'écriture et l'oralité. Cependant, ce n'est pas le sujet principal ici.

3 Analyse « conventionnelle » : contes peuls

L'inscription dans le contexte culturel de la performance donne à l'énonciateur la possibilité d'intégrer une référentialisation spatiale minimale. C'est d'ailleurs cette spécificité générique qui permet aux contes oraux une circulation aisée d'une langue à une autre. L'analyse porte ici sur les représentations de l'identité culturelle. La question n'est donc pas en premier lieu d'identifier l'espace référentiel, comme c'était le cas du roman.

Pour respecter la cohérence méthodologique par rapport au corpus long du roman présenté ci-dessus, je me réfère à des textes narratifs, en l'occurrence à des contes. Genre proportionnellement beaucoup plus bref, il ne peut de toute évidence pas être comparé isolément au genre long et écrit, le roman[12].

Le corpus auquel je me réfère est le répertoire de la conteuse peule Goggo Addi (U. Baumgardt, 2000). Il comprend soixante contes, recueillis lors de deux recherches de terrain, en 1985 et 1989[13]. Il peut être considéré comme une unité textuelle dont les éléments constitutifs sont les contes. Ce répertoire forme effectivement un texte étendu et cohérent qui offre suffisamment de données pour une analyse de l'espace. C'est uniquement grâce à ce caractère exceptionnel que cette recherche est possible.

La question posée est de savoir certes, quel est l'espace référentiel, mais surtout, quelles sont les représentations culturelles exprimées dans le corpus et dans quelle mesure les représentations de l'espace y participent éventuellement.

Comme dans le cas exposé ci-dessus concernant le roman, j'ai utilisé une méthode « conventionnelle » en relevant manuellement tous les indices se rapportant à l'espace. Leur classification mène aux résultats suivants, que j'expose succinctement[14].

Malgré sa longueur, le corpus présente proportionnellement peu d'indices spatiaux. L'une des raisons pouvant être évoquée pour l'expliquer relève du contexte de com-

[11] L'inscription du texte dans son contexte culturel défini par la performance est l'une des conditions pour la circulation des textes et notamment des contes en oralité. Ici se réalise une véritable dialectique : le public dans une performance et dans une langue données comprend le contexte culturel : dans une autre performance du même texte, ce sera le même fonctionnement.

[12] Deux autres contraintes doivent être prises en considération. Ainsi, une dizaine de contes ne présenterait certainement pas assez de données. De même, un corpus numériquement plus important mais hétérogène du point de vue de l'énonciateur et des performances, serait inapproprié. En effet, l'inscription de chaque texte dans la situation d'énonciation singulière dans laquelle il est dit, implique une variabilité équivalente au nombre de performances.

[13] Hermann Jungraithmayr, africaniste à l'Université de Marburg, m'a aidée dans l'organisation et le financement des premiers séjours de recherche au Cameroun de même que, un peu plus tard, Janos Riesz de l'Université de Bayreuth. Romaniste, spécialiste de la littérature africaine francophone, il a soutenu mes recherches en littérature orale.

[14] Pour plus de détails, voir U. Baumgardt (2000, p. 43-61).

Published in : JAMBOROVA LEMAY, Diana ; PEREIRA, Christophe ; ŠMILAUER, Ivan (dir.), *Des langues calculables à l'homme incalculable. Hommage à Patrice Pognan*, Editions des archives contemporaines, ISBN : 9782813004260, doi : 10.17184/eac.9782813004260

munication. En effet, comme je l'ai dit, dans une situation d'homogénéité culturelle, la dénomination d'un élément narratif seule situe un espace. Les indices se distinguent selon leur fréquence : nombreux sont ceux qui ne figurent qu'une seule fois dans le corpus alors que d'autres sont récurrents. Par ailleurs, les indices ont des fonctions différentes et peuvent être classés en indices indirects et indices explicites. Les premiers n'ont pas pour fonction principale d'identifier l'espace référentiel. Ils n'ont pas forcément une fonction narrative précise, ils sont parfois mentionnés au cours d'une action et ils relèvent du niveau figuratif. C'est le cas de la faune (lion, girafe), de la flore (baobab) et de certaines pratiques culturelles (transhumance), entre autres. Les indices explicites, en revanche, désignent l'espace référentiel, comme un nom de pays (Âdamâwa), de ville (Garoua) ou de village (Layndé).

D'un autre côté, en contexte d'oralité, un énonciateur peut être amené à référer à la deixis. Celle-ci englobe l'espace de la performance, de même que l'aire culturelle dont fait partie la littérature orale produite dans la performance en question. La conteuse Goggo Addi dit par exemple :

« La nouvelle a fait le tour de toute l'Âdamâwa », sous-entendu, « elle est venue jusqu'ici puisque je la connais et que je te la transmets ». De même, pour mesurer la distance parcourue par un personnage qui se rapproche de sa destination, elle utilise dans cinq contes (19, 21, 43, 45, 60) un toponyme dans l'expression « comme d'ici Layndé », un village non loin de Garoua, la ville où elle habite et où ont lieu les performances.

Les indices spatiaux indirects (faune et flore entre autres) ramenés à la seule désignation et l'explicitation de la deixis sont prégnantes dans la référentialisation spatiale en contexte d'oralité.

Le corpus distingue clairement l'espace habité et l'espace non habité, chacun d'eux comprenant plusieurs sous-catégories. Cette distinction récurrente est liée au nombre important de contes du répertoire. En effet, puisque les actions se déroulent dans plusieurs contes et dans un temps relativement long, les références spatiales sont fréquentes, les deux niveaux, l'espace et le temps, étant souvent indissociables[15]. Dans le cas du répertoire de Goggo Addi, la séparation entre l'espace habité et l'espace non habité est proportionnellement la plus fréquente, les espaces surnaturels étant rares et les espaces évoqués étant « inconnus » du public car l'action ne s'y déroule pas.

La distinction entre deux types d'espaces les organise en « paradigmes spatiaux », comprenant eux-mêmes éventuellement plusieurs sous-catégories. L'analyse des occurrences dans leurs contextes narratifs (les actions des personnages) donne accès aux qualifications et aux représentations culturelles qui sont associées aux espaces respectifs. Ainsi, par exemple, la « brousse » *ladde,* sert de cadre de vie de manière passagère à plusieurs personnages (Baumgardt, 2000, p. 49-51). Elle a deux fonctions. C'est un refuge qui offre réconfort et réparation. De l'autre côté, c'est là qu'est envisagée la transgression des interdits les plus forts : le père, qui est le roi, y fait conduire son fils adulte pour le piéger et le tuer ; des parents abandonnent leur enfant en brousse.

[15] Mon analyse du répertoire de Goggo Addi ne prend pas en compte les relations, souvent effectivement étroites, entre l'espace et le temps.

Published in : JAMBOROVA LEMAY, Diana ; PEREIRA, Christophe ; ŠMILAUER, Ivan (dir.), *Des langues calculables à l'homme incalculable. Hommage à Patrice Pognan*, Editions des archives contemporaines, ISBN : 9782813004260, doi : 10.17184/eac.9782813004260

Dans ces deux fonctions, la « brousse » est définie par opposition à l'espace habité, mais également en articulation avec « le fleuve » *maayo*[16], frontière entre ces deux univers. La brousse participe ainsi des représentations culturelles du monde socialisé et non socialisé.

4 Traitement automatique de corpus de littérature orale

La question sur l'articulation éventuelle du traitement automatique de corpus et l'analyse littéraire s'est posée pour moi en plusieurs étapes à travers le projet d'« Encyclopédie des littératures en langues africaines » (ELLAF). Patrice Pognan est un pilier central du projet, je lui renouvelle les plus sincères remerciements de toute l'équipe. Nous avons réfléchi ensemble, Patrice Pognan, Aliou Mohamadou et moi-même, sur la production de matériaux d'enseignement à partir de textes littéraires. Dans ce contexte se sont profilées pour moi une idée et une dimension nouvelles que ma formation disciplinaire seule ne m'aurait pas permis d'entrevoir : analyser un corpus de littérature orale en ayant recours au traitement automatique.

Une telle démarche n'a de pertinence méthodologique que si elle est comparative, si elle est appliquée à un corpus proche[17] et si elle concerne le même niveau que dans les analyses précédentes. En effet, pour appréhender concrètement la portée de cette approche nouvelle, elle doit viser la même problématique, en l'occurrence les représentations de l'espace dans les contes de Goggo Addi. Par ailleurs, ne connaissant pas le domaine, une aide au niveau du traitement du corpus et de son analyse m'est indispensable. Elle m'a été apportée par Aliou Mohamadou qui a construit une base de données textuelles à partir du répertoire de Goggo Addi ; la présentation de cette base de données ci-dessous a été faite avec son concours.

4.1 Le corpus et son traitement

La base de données analyse les 60 contes du répertoire. Chacun des 48 785 mots pertinents, sur 49 786 mots graphiques[18] est indexé à des identifiants de contexte permettant de situer le mot à 4 niveaux : a) le conte où il apparaît ; b) le bloc de texte dans le conte (paragraphe ou strophe dans lesquels apparaît le mot) ; c) l'unité de contexte large (phrase, ensemble de phrases dont participe le mot à l'intérieur du bloc) ; d) et l'unité de contexte restreint à l'intérieur de la phrase (proposition, groupes de mots réduits servant d'unité d'observation).

La base de données définit ensuite une table de concordance des occurrences avec les contextes larges et avec les contextes restreints.

En troisième lieu, les occurrences sont lemmatisées, l'opération consistant à ramener les différentes formes d'un même mot, à une forme conventionnelle de citation ; cette opération est essentielle pour l'exploitation du corpus en raison de l'alternance consonantique de l'initiale des verbes, noms et adjectifs qui caractérise le peul, et en raison

[16] Dans le cadre de la présente étude, je ne peux pas analyser cette relation.

[17] Comme dans le cas de la comparaison entre littérature écrite et orale, la proximité concerne au moins l'énonciateur, le genre littéraire, la longueur et le degré de complexité.

[18] Le mot graphique est défini par un séparateur espace ou un retour à la ligne. Ne sont pris en compte que les mots peuls ; les mots étrangers ou non identifiés ne sont pas retenus.

Published in : JAMBOROVA LEMAY, Diana ; PEREIRA, Christophe ; ŠMILAUER, Ivan (dir.), *Des langues calculables à l'homme incalculable. Hommage à Patrice Pognan*, Editions des archives contemporaines, ISBN : 9782813004260, doi : 10.17184/eac.9782813004260

des variantes libres, fréquentes chez la conteuse. La lemmatisation est accompagnée de l'étiquetage grammatical des occurrences.

D'autres opérations, plus complexes, sont effectuées sur le corpus, mais celles-ci n'intéressent pas directement l'analyse faite ici.

Le corpus ainsi traité a été examiné, à titre d'exemple et dans une perspective comparative, par rapport à un seul niveau, l'unité lexicale de *ladde* « brousse ».

4.2 Observation des données

L'observation est plus précise que dans l'analyse précédente du même corpus.

Le terme *ladde* « brousse » est fréquent dans le texte : il est mentionné 84 fois. Un autre aspect est mis en évidence, à savoir la répartition des occurrences : dans 25 contes, le terme n'est pas mentionné[19], l'action se déroule donc exclusivement dans un espace habité ; il est attesté dans les 35 autres contes, où l'action concerne les deux espaces. La distinction claire entre espace habité et non habité est confirmée et focalisée sur les relations entre les deux espaces. L'interrogation de la base de données fournit un résultat beaucoup plus détaillé, permettant d'identifier toutes les occurrences et la distribution du terme.

Mais l'apport le plus important est d'avoir la possibilité d'appréhender tout mot avec la totalité de ses cooccurrents.

Le terme *ladde* « brousse » apparaît souvent associé à *huunde* « chose » dans le syntagme complétif *huunde ladde* « chose de la brousse ; animal sauvage » ; l'expression est mentionnée 10 fois, dont 5 occurrences au pluriel *kuuje laɗɗe* avec, ce qui est rare, *ladde* au pluriel (*laɗɗe*). Le cooccurrent *huunde* « chose ; animal » est également attesté deux fois à l'augmentatif et une fois au diminutif. Dans le premier cas, il s'agit de *kuunga* (classe NGA) « animal hors norme ; monstre », où il est sujet de verbes de déplacement, partant de *ladde* « la brousse » vers le lieu habité ; dans le second cas, il s'agit du diminutif péjoratif, rare, *kuungum* (classe NGUM) « misérable petit animal », qui est situé « dans la brousse », soit : *kuungum nder ladde* « un sauvage perdu dans la brousse », l'expression étant employée par des conseillers du roi pour manifester leur mépris vis-à-vis de quelqu'un.

Les cooccurrents les plus fréquents de *ladde* sont des verbes. Ils permettent de mieux appréhender les représentations culturelles de la « brousse ».

Plusieurs de ces verbes, *yahgo* « aller », *dillugo* « partir » et *tawgo* « trouver » indiquent un simple déplacement, sans événement particulier. Mais dans un exemple[20], l'un des verbes, en fonction du statut social du personnage sujet, entraîne des événements particuliers : *mallum feere yehi haa ladde* « un marabout est allé en brousse » [1].

Un deuxième cas de figure concerne un déplacement intentionnel et inhabituel : *naatugo ladde hamaadaare* « entrer dans une brousse profonde » [10] ; *waancugo nder ladde*

[19] Ce sont les contes : 4 ; 5 ; 15 ; 16 ; 17 ; 19 ; 20 ; 22 ; 23 ; 25 ; 26 ; 28 ; 31 ; 32 ; 33 ; 35 ; 36 ; 38 ; 39 ; 40. Dans un travail ultérieur, il serait intéressant d'analyser s'il existe un lien entre l'absence ou la présence de *ladde* et la problématique abordée dans les contes concernés.

[20] Les numéros des contes dans le répertoire sont indiqués entre crochets.

Published in : JAMBOROVA LEMAY, Diana ; PEREIRA, Christophe ; ŠMILAUER, Ivan (dir.), *Des langues calculables à l'homme incalculable. Hommage à Patrice Pognan*, Editions des archives contemporaines, ISBN : 9782813004260, doi : 10.17184/eac.9782813004260

« se promener dans la brousse » [65] ; *yahgo ɗaɓɓita kusel haa ladde* « aller chercher de la viande en brousse » [03] (c'est une femme enceinte qui décide d'aller chercher du gibier).

Dans un troisième cas de figure, une action non commentée qui se déroule en brousse indique d'emblée qu'il s'agit d'une transgression : *mbarooga te'i mo haa nder ladde* « le lion l'épousa dans la brousse » [06] ; *O yehi o nyiɓani mo haa ladde, suudu njamndi* « Il alla lui construire une maison dans la brousse, une maison en dur » [18] ; il s'agit d'un roi qui, par cette action, soustrait sa fille au mariage ; *yahgo wudina mo nder ladde* « aller l'abandonner [son fils] dans la brousse » [12].

Un quatrième cas de figure, que je n'avais pas remarqué, concerne l'association d'un verbe de déplacement avec un autre qui en exprime le motif.

Plusieurs exemples sont relevés. Le premier est attesté deux fois dans le même conte : *o dogga o dilla ladde* [51 ; 2 fois] « Elle [l'épouse] fuguait et partait en brousse ». La « brousse » est qualifiée de « refuge », trait qui relève du même champs sémantique que « réparation de torts subis ». Les deux autres exemples expriment l'association de « sentiment », de « brousse » et de « distance » voire séparation tragique.

Ainsi, dans un conflit d'autorité entre la sœur et son frère cadet, orphelins, le frère refuse d'obéir et menace : *Mi tikkan mi naata ladde* [65 ; 4 fois] « Je me fâcherai et je m'en irai dans la brousse ».

Dans le troisième exemple, c'est un conflit entre fils et mère qui est illustré : la lionne appelle son fils pour l'allaiter, le lionceau refuse, *Nonnon o ɓerni, o dilli ladde* [11] « C'est ainsi qu'elle se fâcha [se mit en colère] et s'en alla en brousse ».

La convergence au même niveau de plusieurs éléments narratifs apporte des nuances importantes à la représentation de l'espace. C'est bien une confirmation de l'observation suivante : un conte seul ne fait pas sens, c'est au contraire dans la pluralité des textes que se construit ce dernier.

Pour clore ce point, je retiens plusieurs observations. Deux aspects étaient passés inaperçus dans l'analyse précédente : les formes linguistiques et la répartition entre contes présentant ou non *ladde.* Par ailleurs, le relevé des cooccurrences permet d'approfondir l'analyse précédente : la représentation de la brousse en tant qu'espace de transgression existe effectivement, mais elle se manifeste concrètement à travers les verbes associés à *ladde* ; ceux-ci annoncent que l'action envisagée ou réalisée relève de l'interdit. En revanche, il n'y a pas de cooccurrence de *ladde* qui indique explicitement que l'action qui se déroule est réparatrice de transgressions graves. Par exemple, un « figuier » *dunndeehi,* figure maternelle, offre protection à un enfant abandonné par ses parents dans la brousse [12] ; de même, des « oiseaux » *colli,* dans leur conversation, feront des révélations à un personnage pour le guérir [67].

Concernant les représentations de l'espace, plusieurs points ressortent de l'analyse. La mise en évidence du fonctionnement textuel souligne dans le répertoire de Goggo Addi la cohérence entre l'expression linguistique et la complexité des représentations culturelles qui fondent une authentique vision du monde. Je pense avoir rendu compte de cette dimension du répertoire dans mon analyse précédente (Baumgardt, 2000),

Published in : JAMBOROVA LEMAY, Diana ; PEREIRA, Christophe ; ŠMILAUER, Ivan (dir.), *Des langues calculables à l'homme incalculable. Hommage à Patrice Pognan*, Editions des archives contemporaines, ISBN : 9782813004260, doi : 10.17184/eac.9782813004260

mais sans avoir su l'étayer par une observation plus détaillée. Or, à travers l'approche expérimentale que j'ai essayé de présenter, il me semble que le résultat est plus fin et plus solide. Il va dans le sens d'une étude ethnolinguistique, celle-ci étant fondée sur une analyse très détaillée de la langue que permet de réaliser une base de données textuelles.

5 Conclusion

Ainsi, l'interrogeabilité illimitée d'une base de données textuelles obtenue par le traitement automatique de corpus – impossible à travers une méthode d'analyse conventionnelle – ouvre effectivement des perspectives nouvelles pour l'analyse littéraire.

Pour rester sur la problématique de l'espace, le relevé des indices pour en déterminer la nature et les fonctions laisse entrevoir la possibilité de définir une typologie des textes en fonction de la structuration de l'espace ; ainsi le croisement de critères initialement non identifiés permet d'affiner l'articulation entre l'espace non habité et l'espace habité, et leurs délimitations.

Pour aller plus loin, en littérature orale peule, je citerai un exemple ; il concerne la langue à un niveau qui n'a pas encore été exploré : le peul ne distinguant pas grammaticalement le masculin et le féminin, le pronom personnel singulier *o* signifie indifféremment « il » ou « elle » ; l'observation exhaustive des concordances du mot permettra d'établir deux paradigmes genrés fondés sur les cooccurrences ; celles-ci peuvent apporter des informations fondamentales sur les représentations culturelles du genre.

Une nouvelle approche devient possible à propos du conte, très fréquent en littérature orale. L'une de ses caractéristiques structurelles est la répétition, souvent « décriée », sans que le fonctionnement textuel et les fonctions de la répétition dans les contes n'aient été étudiés de manière approfondie. Pourtant, on peut penser qu'il existe plusieurs raisons à la récurrence de cette forme d'expression. Quel rôle joue-t-elle, par exemple, dans l'apprentissage de la langue, le développement de la mémoire et l'acquisition de structures cognitives ?

Bibliographie

BAUMGARDT Ursula (2000), *Une conteuse peule et son répertoire. Goggo Addi de Garoua, Cameroun,* Paris, Karthala, 548 p.

BAUMGARDT Ursula et DERIVE Jean (dir.) (2008), *Littératures orales africaines. Perspectives théoriques et méthodologiques,* Paris, Karthala, 439 p.

BAUMGARDT Ursula (2009), « L'espace en littérature orale africaine Quelques réflexions méthodologiques autour des indices spatiaux », *Cahiers de Littérature orale,* n° 65, p. 111-131.

DERIVE Jean (2009), « Des modalités de l'énonciation en littérature orale », *Cahiers de Littérature orale,* n° 65, p. 91-11.

DERIVE Jean (2010), « Typologie et fonctions de quelques genres oraux du Manding à l'aune du critère de la spatialité », *Journal des africanistes, L'expression de l'espace dans les langues africaines,* tome 579, fasc. 2, p. 201-222.

MOHAMADOU Aliou, *Concordanciel d'occurrences lexicales (COL). Base de données de textes peuls.* Inédit.

SALHI Mohand Akli (2015), *Poésie kabyle,* L'Odyssée Editions, Tizi-Ouzou, 187 p.

Published in : JAMBOROVA LEMAY, Diana ; PEREIRA, Christophe ; ŠMILAUER, Ivan (dir.), *Des langues calculables à l'homme incalculable. Hommage à Patrice Pognan*, Editions des archives contemporaines, ISBN : 9782813004260, doi : 10.17184/eac.9782813004260

POGNAN Patrice, SALHI Mohand Akli (2016), *9th Bayreuth-Frankfurt-Leidener Kolloquium zur Berberologie*, 20.-23. July 2016, Frankfurt am Main, main topic « Etudes sur la linguistique et la documentation écrite en berbère »

Encyclopédie des littératures en langues africaines, ELLAF, http://ellaf.huma-num.fr/

Published in : JAMBOROVA LEMAY, Diana ; PEREIRA, Christophe ; ŠMILAUER, Ivan (dir.), *Des langues calculables à l'homme incalculable. Hommage à Patrice Pognan*, Editions des archives contemporaines, ISBN : 9782813004260, doi : 10.17184/eac.9782813004260