

## **NOUVELLES PERSPECTIVES DE LECTURE D'UN TEXTE LITTÉRAIRE**

**Mirsini TZANAVARI**

myrsini@uom.gr

**Université de Macédoine, Thessalonique**

### *Résumé*

*Dans notre culture occidentale moderne, la littérature est si intrinsèquement liée aux livres qu'elle a longtemps été tenue à l'écart des développements de l'informatique. Les créateurs et les critiques ont même manifesté très tôt de profondes résistances à l'utilisation de l'ordinateur comme outil de lecture. Par ailleurs, l'apparition de chaque nouveau support d'inscription du texte a généralement provoqué une modification des modes de lecture. La numérisation des textes a pour conséquence la dématérialisation du livre, non seulement comme support de l'organisation textuelle, mais comme objet clos sur lui-même. C'est ainsi que le corpus remplace désormais le livre. Constitué des textes d'un même auteur, le corpus est la nouvelle unité de lecture sur support électronique. Qu'il s'agisse d'établir le vocabulaire d'un auteur pour le confronter à celui de ses contemporains ou de classer la liste des vocables d'une œuvre par ordre de fréquence décroissante, l'ordinateur ouvre la voie à de nouvelles lectures qui modifient singulièrement notre rapport aux textes.*

*A travers cet article, nous voulons mettre en lumière une série d'analyses statistiques selon le système de la cooccurrence généralisée (Viprey, 1997) menées en vue de détecter des groupements lexicaux thématiques au sein du corpus romanesque d'Hervé Bazin. Il s'agit de la méthode qui consiste à relever les réseaux d'association qu'entretiennent les vocables dans un environnement textuel déterminé. La démarche que nous allons suivre pour traiter statistiquement notre corpus est celle de l'environnement d'exploration hypertexte Astartex (Viprey, 1997).*

*Notre propos sera également d'examiner si ce champ d'investigation met à l'épreuve les hypothèses émises par la critique littéraire traditionnelle et permet d'apporter des informations que la lecture ordinaire ne livre pas ou qu'elle ne livrerait qu'au prix d'un travail fastidieux et sans commune mesure avec les bénéfices escomptés.*

*Mots clés. cooccurrence généralisée, Astartex, Bazin, exploration thématique.*

### **1. Introduction**

Les rapports qu'entretient la littérature avec l'univers de l'informatique sont complexes et l'enjeu de ce croisement va bien au-delà des problèmes posés par la simple transposition d'un support à l'autre. Ramener cet enjeu à des questions d'ergonomie de lecture serait méconnaître la nature particulière de la littérature et de son rapport aux supports matériels de

son inscription et de sa circulation. Il ne semble pas que tous ceux qui oeuvrent dans le champ littéraire aient pris la mesure des problèmes épistémologiques qui se posent à la littérature lorsque celle-ci se risque à quitter le support du livre qui a été pendant si longtemps le sien. La littérature est si intrinsèquement associée au livre que nous avons tendance à nous préoccuper de l'avenir de ce dernier plutôt qu'à interroger la nature du lien instauré il y a cinq siècles entre le texte et son support imprimé. En se libérant de ce lien, la littérature prend le risque de remettre en cause son identité, le risque de se perdre ou de se dissoudre dans la vaste mutation médiatique en cours.

La première interrogation doit porter sur la dématérialisation du texte et sur son caractère désormais virtuel. Par ailleurs, la digitalisation du texte le rend susceptible d'être soumis à un traitement informatique, voire même d'être généré par un programme. C'est ainsi que de la confrontation du texte avec l'image naissent de nouveaux modes de lecture de textes qui modifient singulièrement notre rapport aux textes. L'enjeu de l'article présent est de mettre en lumière une série d'analyses statistiques selon le système de la cooccurrence généralisée appliqué aux textes littéraires. Les outils fournis par la statistique textuelle ouvrent la voie à de nombreuses pistes de recherche dans ce domaine permettant de reconstruire objectivement la thématique majeure d'un corpus. Prolongeant la méthode de la cooccurrence généralisée, cette étude s'intéresse donc aux groupements lexicaux thématiques récurrents dans un corpus littéraire contemporain. Nous tenterons de tirer profit des dernières innovations de l'analyse des données textuelles en testant des outils d'analyse intégrés dans le logiciel d'exploration multidimensionnelle ASTARTEX pour mettre à jour la structure thématique des œuvres ; pour cette raison c'est la micro-distribution calculée sur les lemmes qui sera mise à l'œuvre pour tester la mise en série des textes du corpus par le recensement des cooccurrents et des séquences d'items ; le dépouillement du texte sous formes de lemmes permet le regroupement des formes qui se rattachent à la même entrée du dictionnaire en neutralisant les variations grammaticales. Par ailleurs, nous avons recours à une nouvelle projection cartographique de l'AFC, la technique de la projection « géodésique » pour saisir ces relations cooccurrentielles.

Nous utilisons pour cette recherche le corpus littéraire issu de l'œuvre d'un des écrivains les plus célèbres du XXème siècle : Hervé Bazin. La production d'Hervé Bazin est considérable, cependant le corpus numérisé employé dans cette étude comprend avant tout les seize romans de l'œuvre de l'écrivain de 1948 à 1994.

## 2. Le texte informatisé.

Dans notre culture occidentale moderne, la littérature est si intrinsèquement liée aux livres qu'elle a longtemps été tenue à l'écart des développements de l'informatique. Les créateurs et les critiques ont même manifesté très tôt de profondes résistances à l'utilisation de l'ordinateur comme outil de lecture. Paradoxalement, ce sont souvent des informaticiens ou des scientifiques d'horizons divers qui se sont intéressés les premiers aux possibilités nouvelles qu'offrait l'informatique à l'étude des œuvres littéraires.

Pour bien comprendre les enjeux de la digitalisation des textes littéraires, il faut tout d'abord revenir sur les techniques utilisées. Les ordinateurs permettent de coder des images, du son et du texte et de les traiter par programme. Dans ce processus, le texte occupe cependant une place singulière. A la différence de l'image et du son dont c'est la matérialité analogique qui est codée, le codage du texte peut se faire selon deux modalités fondamentalement différentes : le mode image et le mode texte.

## 3. Du livre au corpus

Le premier effet du codage informatique des textes est la dématérialisation du livre, non seulement comme support de l'organisation textuelle, mais comme objet clos sur lui-même. Les textes numérisés n'étant plus contraints par les limites du papier, ils peuvent être stockés sur des supports de très grande capacité, en ligne ou sur cédérom. Ce changement d'échelle modifie le rapport du lecteur au texte. Le corpus remplace le livre. Constitué des textes d'un même auteur ou de ceux d'un ensemble plus large, le corpus est la nouvelle unité de lecture sur support électronique.

Ce que l'utilisateur attend de l'examen d'une base de données textuelles établie sur un corpus, ce sont des informations que la lecture ordinaire ne livre pas ou qu'elle ne livrerait qu'au prix d'un travail fastidieux et sans commune mesure avec les bénéfices escomptés.

## 4. Cooccurrence

Une des préoccupations de l'analyse des données textuelles (ADT) est de passer de la statistique occurrentielle- étude de la distribution fréquentielle d'un terme dans un corpus partitionné-, à la statistique cooccurrentielle- étude du rapport fréquentiel entre deux termes co-

présents dans le corpus au sein de fenêtres co-textuelles délimitées (le paragraphe par exemple)-. On passe en effet d'une approche formelle ou positiviste du corpus à une approche contextualisante c'est à dire déjà sémantique. Avec la cooccurrence, la statistique textuelle met un pied dans une sémantique de corpus qui lui était jusqu'ici interdite et réaffirme par là sa vocation herméneutique. La recherche de cooccurrence renvoie à des notions de parcours interprétatifs, de mises en réseaux ou de mise en résonance<sup>1</sup> de textualité ou de texture. Elle s'inscrit en tout cas dans une « logique de contextualisation » qui est la condition de l'élaboration du sens et de l'interprétation.

#### 4.1. La cooccurrence comme forme minimale du contexte

Comme on le sait, le sens naît du/en contexte ; la contextualisation (et sa formalisation) est aussi la condition d'un traitement sémantique, interprétatif. A une extrémité, la « forme maximale » du contexte serait « tout le texte » (sauf à être, au de-là du texte, le corpus dans son ensemble).<sup>2</sup> A l'autre extrémité, nous posons ici que la « forme minimale » du contexte est la cooccurrence. Nous définissons en effet le phénomène de contextualisation d'un mot par un autre mot. (Charge à la statistique et au traitement informatique de repérer systématiquement toutes les attirances/répulsions lexicales cooccurentielles significatives et leur degré de significativité). Au sein du corpus, le contexte minimal d'un mot pôle n'est pas le syntagme ou la phrase. Ceux-ci sont trop nombreux et trop variables pour être synthétisables : il y aurait alors, en effet, autant de contextes (c'est à dire de sens) du mot que de syntagmes ou de phrases le contenant. Au contraire, le contexte minimal d'un mot est son cooccurrent attesté par la statistique, ou plutôt ses cooccurrents lexicaux attestés, systématiquement repérés. Par là, l'enjeu du traitement cooccurrentiel devient majeur, il s'agit d'objectiver par la statistique, le contexte minimal des mots nécessaire à leur compréhension/interprétation.

<sup>1</sup> Le traitement cooccurrentiel espère saisir l'instant où le vocabulaire d'un texte entre en résonance avec lui-même.

<sup>2</sup> Selon Rastier « Le contexte, c'est tout le texte ». Techniquement, la loi endogène du calcul de la cooccurrence- prend en considération l'ensemble du corpus comme la norme par rapport à laquelle s'individualisent des événements linguistiques.

## 4.2. La cooccurrence comme nœud du tissu textuel

L'étymologie de « texte » est souvent rappelée. Tisser, tissus, tissage : on convient généralement qu'un texte est un entrelacement de deux fils ou de deux axes, vertical et horizontal.

Le fil de chaîne horizontal représenterait la linéarité du texte ou l'axe syntagmatique ; il rendrait compte du déroulement et des combinaisons mises en place par l'auteur pour produire un texte.

Le fil de trame, lui, serait l'axe vertical ou l'axe paradigmaticque qui représenterait la dimension non-linéaire du texte. Il mettrait à jour, sous forme de nomenclature, la sélection (la sélection lexicale par exemple) opérée par l'auteur. Classiquement, nous avons alors à faire à l'inventaire des formes utilisées, au dictionnaire alphabétique du corpus, à l'index hiérarchique, à telle liste ou à tel tableau des spécificités, d'hapax etc. C'est dans ce cadre que la cooccurrence peut être considérée comme un nœud du tissu-texte ; l'endroit même où se noue le fil de trame et le fil de chaîne, l'axe paradigmaticque et l'axe syntagmatique d'une production textuelle, la tabularité et la linéarité du texte. La statistique met à jour une sélection lexicale : elle repère systématiquement par exemple les substantifs qu'un auteur aura sur-utilisés (sur-sélectionnés) pour les considérer comme les mots-pôles à traiter. Mais les mots de la sélection seront considérés dans le cadre d'une combinaison ou d'une fenêtre syntagmatique données. Le traitement de cooccurrence procède à la fois d'une décontextualisation lexicale (le mot extrait du corpus) et d'une recontextualisation (le mot remplacé dans son cotexte).

## 5. la cooccurrence généralisée

D'après J.-P. Massonnie, « C'est de la rencontre des mots, rencontre naturelle ou contre- nature, que naît le sens et, dans le cas des textes littéraires, que naît l'impression ».<sup>1</sup> Ce qui revient à dire que, dans un système comme la langue, le sens d'une unité lui vient, d'abord, des relations qu'elle entretient avec tous les autres mots du lexique. Ou encore, qu'un mot dans un texte prend tout son sens grâce à son environnement.

La méthode de la cooccurrence généralisée, telle qu'elle est définie par Jean-Marie Viprey, consiste à signaler un vocable sur représenté dans le contexte d'un autre vocable du même ensemble. Il est donc possible de mesurer l'organisation des cooccurents entre eux pour entrer plus

<sup>1</sup> Massonnie, J.-P., *Analyse informatisée des textes*, Les Belles Lettres, Besançon, 1990.

finement dans l'entrelacement lexical et déterminer des contextes minimaux qui font sens. Cette approche renvoie à la volonté, en ADT, d'offrir à l'analyste des parcours de lecture aux vertus herméneutiques afin de mieux interpréter les textes. Bref, elle constitue une contextualisation minimale nécessaire- elle balise le parcours interprétatif- mais, en dernier recours, seul le retour au texte permet l'interprétation.

## 6. Méthodologie : le logiciel ASTARTEX

Le traitement statistique des données textuelles par le logiciel ASTARTEX permet une approche des réseaux thématiques d'un corpus. Il autorise l'extraction des items les plus fréquents et permet la recherche de cooccurrence entre deux ou plusieurs items. Ce programme nous permet de calculer, dans un premier temps, la fréquence globale des occurrences, qui sera recueillie dans un vocabulaire fréquentiel, puis, de constituer une liste de vocables, candidats à l'analyse distributionnelle. A l'étape suivante, cet outil est à même de comptabiliser l'ensemble des cooccurrences des vocables retenus dans une matrice carrée. Puis, il classe ces vocables selon leurs parentés de profils cooccurentiels. En fait, au lieu de reposer sur une segmentation pré-établie d'un thésaurus préalablement constitué, ici les mots reconnaissent d'eux-mêmes leur parenté du seul fait de leur voisinage dans les mêmes contextes. Un autre outil, mis au point dans le cadre d'ASTARTEX, permet enfin de visualiser l'application de la micro-distribution des matrices collectées à l'aide de l'A.F.C. L'analyse de la micro-distribution est un procédé qui consiste à relever, par la ventilation des items sur les graphes d'analyses factorielles, les contextes d'occurrence d'un pivot individuel ou collectif dans les limites de la phrase dans le but de vérifier comment s'organise la résonance textuelle au niveau de la micro-distribution. Enfin, nous avons recours à la projection « géodésique » de l'A.F.C. pour livrer les sorties graphiques qui permet de mieux éclater la visualisation du nuage de points organisés sur un plan déterminé par les 3 premiers axes comme si l'on regardait de l'intérieur d'une sphère.

## 7. Corpus.

La production littéraire d'Hervé Bazin est vaste, s'étend sur plus de quarante ans et englobe plusieurs genres littéraires. Le corpus auquel nous avons eu recours recouvre la totalité de l'œuvre romanesque. Pour

ce travail, notre corpus a été lemmatisé au préalable avec le lemmatiseur Intex qui permet une lemmatisation automatique et efficace des données. L'objectif de cette étude étant l'exploration thématique du corpus Bazin, nous prenons en considération les calculs de l'A.F.C soumis aux substantifs lemmatisés. L'élimination des mots outils et des verbes à tout faire contribue également à privilégier les relations sémantiques ou thématiques plutôt que les rapports de dépendance syntaxique. Le choix des termes est automatique mais compte tenu de l'étendue du corpus, le programme de sélection retient 200 lemmes qui sont au - dessus d'un certain seuil d'effectif dans un environnement textuel déterminé ou empan cotextuel de dix vocables de part et d'autre du mot pivot.

## 8. Résultats et interprétation de la micro-distribution des substantifs dans le corpus entier

Nous nous donnons comme objectif d'observer, dans les graphes de la micro-distribution des substantifs de l'œuvre romanesque complète de Bazin, les réseaux d'attraction et de répulsion des couples de cooccurrences pour dégager les thèmes caractéristiques développés dans l'écriture bazinienne. (cf. figure 1) Nous relevons les premiers résultats à partir du graphe de l'AFC et nous tenons compte à la fois des résultats plus fins que produisent les graphes selon la projection géodésique pour mieux étayer nos arguments. (cf. figure 2)

Cinq groupements thématiques se dessinent sur le tableau de l'AFC. L'un au sud - est, a pour cooccurrents les plus saillants *année, mois, journée, semaine, jour, soir, heure, matin, nuit, minute, temps*, qui se rapportent à l'encadrement temporel et qui révèlent une structure temporelle fortement marquée et précise dans l'ensemble du corpus. La répétition infinie des rythmes temporels confère une réalité aux événements par les moyens d'une observation qui s'exaspèrent dans les petits détails. Les déclarations d'Hervé Bazin, à propos de son orientation de romancier réaliste dans son *Abécédaire*, sont révélatrices, il écrit : « Pour ma part ce qui m'intéresse au-delà du récit, c'est l'effet qu'il aura. Le seul miracle de l'écriture, c'est qu'elle porte l'évidence au carré ».<sup>1</sup>

Un autre groupe de termes se distingue au sud à l'ouest du graphe et il a trait au cadre de la vie quotidienne. On peut repérer, dans un premier temps, au sud et à l'intersection de la colonne verticale et horizontale, un ensemble de cooccurrents qui se réfèrent à l'espace extérieur avec *jardin, bureau, chemin, rue, voiture, route, place, train* ; une autre catégorie de

<sup>1</sup> Bazin, H., *Abécédaire*, Grasset, Paris, 1984, p.212.

termes à l'ouest qui concernent le foyer et la réalité quotidienne qui transparaissent dans les substantifs comme *fenêtre, cuisine, pièce, boîte, chaise, escalier, couloir, mur, chambre, lit, papier, téléphone, clef, drap, fauteuil*. Le quotidien pour Hervé Bazin se traduit par l'addition d'éléments mineurs de tous les jours qui, par leur accumulation, forment le tissu des habitudes. De tels éléments l'auteur prend un plaisir évident à dresser des listes copieuses, à faire de véritables inventaires représentatifs de notre époque. Il serait inopportun et pesant de s'en remettre à des mots comme *vérisme* ou *naturalisme* pour qualifier sa manière de peindre.

Il faut également souligner un petit groupe de substantifs qui se situe dans la même partie à l'ouest et est particulièrement concerné par les éléments de la nature : *arbre, ciel, lumière, ombre, eau, bruit, bois, vent, odeur, terre*. L'odeur de la rivière et celle de la terre humide se confondent pour l'écrivain avec l'émoi ressenti devant la vie. Il l'avoue lui-même à Pierre Moustiers lors d'une rencontre :

*« La terre mouillée, l'eau plate avec des nénuphars dessus, le ciel brouillé traversé par le vol irrégulier des bécassines, et le cri de la hulotte un peu plus tard dans le frémissement de ce qui s'endort, oui, la vie ».*<sup>1</sup>

Cet amour de la terre l'apparente à la famille des Rezeau qui, alors que ses membres confinent à la noblesse, sont des paysans et il s'est établi entre leur sol et leur race une longue intimité. C'est pourquoi, Hervé Bazin s'attache à la terre, non pas pour sa possession mais pour sa réalité concrète, pour la vie qu'elle dispense.

L'aire nord-ouest occupe une place considérable où se rejoint un ensemble de cooccurrents qui décrit les parties du corps humain tels que *jambe, nez, œil, doigt, paupière, menton, cou, cheveu, lèvres, bouche, langue, oreille, sourire, dent, cœur, corps, yeux, visage, bras, tête*. Nous pouvons interpréter le micro-univers du thème du corps comme l'indice que l'auteur souligne, au cours de son écriture, le témoignage des effusions, des émotions sans pour autant étouffer la tendresse et les doux sentiments. Il est vrai que notre auteur atteint parfois à la vraie audace qui consiste à tout dire sans intention néanmoins de nous scandaliser, encore moins à nous intéresser à des indiscretions d'alcôve. Son propos est de pourchasser la vérité, d'écrire comme on nettoie sa conscience en s'attardant aux recoins. Pierre Moustiers (1973 : 200) l'a très bien observé :

<sup>1</sup> Moustiers, P., Hervé Bazin ou le romancier en mouvement. Seuil, Paris, 1973, p.131.

« le réalisme sans concession, à force de précision maniaque, [...] anéantit tout sentiment sublime. .. mais inexplicablement, une tendresse infinie s'exhale en même temps qu'une odeur d'humanité sacrée ».<sup>1</sup>

Enfin, on observe que le nord-est du tableau est constitué d'un groupe de termes à très fort indice d'associativité qui est relatif aux membres de la cellule familiale : *oncle, parents, famille, cousin, mariage, amour, mari, frère, sœur, père, fils, garçon, fille, tante, mère, femme, enfant, homme*. L'emploi de ces cooccurrents renforce l'idée que ce romancier est l'historien de l'évolution des mœurs dans la seconde moitié du vingtième siècle. Le regard que porte l'auteur sur le modèle familial d'essence bourgeoise, la routine abrasive de la vie conjugale ainsi que l'attention des parents focalisée vers les enfants sont confirmés à la suite de l'A.F.C des réseaux d'association des cooccurrences.

L'analyse de la micro-distribution des substantifs dans le corpus entier a pu mettre en exergue les champs thématiques récurrents dans l'œuvre romanesque d'Hervé Bazin. Les thèmes auxquels l'écrivain témoigne de l'importance n'étonneront aucun lecteur d'Hervé Bazin. Nous trouvons, parmi les thèmes significatifs dans les ouvrages, celui qui se réfère au temps ; l'auteur s'est également attaché à l'abondance des informations sur la vie quotidienne. Nous n'aurions pas pu ne pas être en présence des éléments de la nature, auxquels l'œil et le cœur de l'écrivain sont si sensibles. Bazin aime aussi décrire les détails des relations intimes confirmés par la sur-représentation des cooccurrences relatives au corps.

Il s'avère enfin que le thème de prédilection que les critiques traditionnels associent à l'œuvre bazinienne, la famille, est validé par le tableau de la micro-distribution des substantifs dans le corpus entier.

L'analyse des réseaux d'association des cooccurrences nous a aidée à mettre en évidence des termes associés que l'on doit considérer comme représentant un aspect du style de l'œuvre romanesque de l'auteur, celui qui concerne les agencements lexicaux et thématiques fins. Il est aisé de constater que l'œil impartial de la statistique permet bien de confirmer un grand nombre d'études qui ont déjà été menées sur les thématiques dans l'œuvre de notre écrivain. Ce qui justifie notre démarche, c'est d'apporter une solution que l'on voudrait objective à un problème qui n'a longtemps reçu que des réponses subjectives. Il s'agit, en réalité, d'une confirmation croisée de la méthode ici appliquée et des intuitions de la tradition critique.

<sup>1</sup> Moustiers, P., *Hervé Bazin ou le romancier en mouvement*. Seuil, Paris, 1973, p.200.

## 9. Conclusion

La littérature aujourd'hui perçoit l'arrivée du numérique d'abord comme une menace et un risque. La numérisation du patrimoine littéraire qui est appelée à se poursuivre n'est pas seulement une entreprise de conservation des œuvres. Elle introduit de nouveaux modes de lecture qui peuvent faire craindre un abandon de la lecture intensive au profit de lectures extensives. De nouvelles pratiques se font jour, de nouvelles habitudes se créent. La méthode de la cooccurrence est l'une de ces pratiques. Celle-ci permet de repérer des couples de formes qui se rencontrent beaucoup plus souvent dans un environnement déterminé que ne le laissent prévoir des calculs fondés sur des modèles probabilistes. Cette statistique des rencontres fréquentes se révèle très efficace en ce qu'elle identifie les attractions lexicales en contexte autour d'un pôle. Pourtant nombre de questions n'ont pu être abordées.

La première d'entre elles est la question des unités linguistiques traitées ; les lemmes sont souvent utilisés dans la recherche des cooccurrences afin de limiter les entrées, mais l'on objectera que la lemmatisation consiste à projeter le sens dans le texte là où la recherche des cooccurrences se proposait de le chercher ; il y aurait là un vis de forme dans le procès de la démonstration.

La deuxième question porte sur la taille du cotexte, c'est à dire la fenêtre artificielle d'étude. Pourtant l'avantage du traitement statistique est qu'il peut s'affranchir d'unités qui n'ont de naturelles que le mot. La question rebondit sur l'orientation de la recherche des cooccurrences à l'intérieur de la fenêtre ou encore sur la prise en considération de l'empan entre les cooccurrents.

Enfin, la dernière question interroge la pertinence des calculs proposés : nous n'avons pas cherché à arbitrer les formules disponibles sur le marché scientifique et avons utilisé le calcul de la cooccurrence généralisée implémentée dans le logiciel ASTARTEX.

L'objectif de cette étude était ailleurs. Nous avons abordé la question par le biais de la cooccurrence en ADT qui doit offrir, dans le cadre d'une linguistique des textes, des parcours de lecture susceptibles de nourrir l'interprétation. Dans cet horizon, l'enjeu de l'ADT et de la cooccurrence est de ne pas se laisser enfermer dans une démarche purement lexicographique pour proposer des perspectives lexicologiques. Si l'objectif de la cooccurrence est de rendre compte du texte dans sa complexité, cela passe par la prise en compte d'unités textuelles. La cooccurrence est cette unité textuelle élémentaire, le contexte minimal, producteur de sens et matrice d'interprétation. Plus loin, si l'objectif de la

cooccurrence est de faire émerger des groupements thématiques par l'observation de l'organisation générale des vocables et non sur l'environnement de termes déterminés *a priori*, la méthode de la cooccurrence généralisée permet de formuler non seulement une analyse indépendamment des données empiriques, de centrer notre connaissance du vocabulaire d'un écrivain sur la structure du texte mais cette procédure offre également la possibilité au critique de suspendre l'interprétation. La conclusion que nous pouvons en tirer est que la latence interprétative donne la liberté au chercheur de livrer le plus tard possible le dialogue de deux instances : la première est le texte qui s'articule à travers ses parcours et la seconde est son intertexte qui est liée à la culture du lecteur et qu'elle apporte en fin de compte les résultats d'une lecture plus fine.

#### Bibliographie

- Bazin H., *Abécédaire*. Grasset, Paris, 1984.
- Clément J., Le e-book est-il le futur du livre ? *Les savoirs déroutés*. Presses de l'enssib, Paris, 2000.
- Lebart L. et Salem A., *Statistique textuelle*. Dunod, Paris, 1994.
- Martinez W., *Contribution à une méthodologie de l'analyse des cooccurrences lexicales multiples dans les corpus textuels*. Thèse de doctorat en Sciences du Langage, Université de la Sorbonne nouvelle-Paris3, sous la direction d'André Salem, 2003.
- Massonnie J.-P., *Analyse informatisée des textes*, Annales Littéraires de l'université de Besançon, Les Belles Lettres, Besançon, 1990.
- Moustiers P., *Hervé Bazin ou le romancier en mouvement*. Seuil, Paris, 1973.
- Rastier F., *Sémantique interprétative*. PUF, Paris, 1987.
- Rastier F., Passages. *Corpus*, 6, Paris, 2007, pp.25-54.
- Véronis J., Cartographie lexicale pour la recherche d'information. *Actes de TALN 2003*, Paris, 2003, pp.265-274.
- Viprey J.-M., *Dynamique du vocabulaire des Fleurs du mal*. Champion, Paris, 1997.
- Viprey J.-M., Analyse séquencée de la micro-distribution lexicale. *Actes des 7èmes JADT*, Louvain, 2004, pp.1165-1176.
- Viprey J.-M., Philologie numérique et herméneutique intégrative. *Sciences du texte et analyse de discours : enjeux d'une interdisciplinarité*, Slatkine, Paris, 2005, pp. 51-68.
- Viprey J.-M., Ergonomiser la visualisation AFC dans un environnement d'exploration textuelle : une projection « géodésique ». *Actes des 8èmes JADT*, Besançon, 2006a, pp.981-992.
- Viprey J.-M., Structure non-séquentielle des textes. *Langages*, Paris, 2006b, 163, pp.71-85.

#### ANNEXES

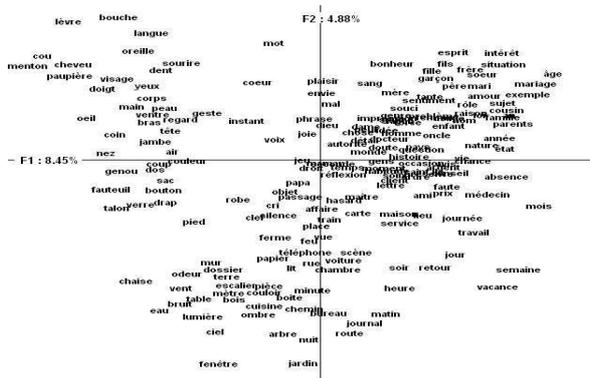


Figure 1. Micro-distribution des substantifs dans le corpus entier

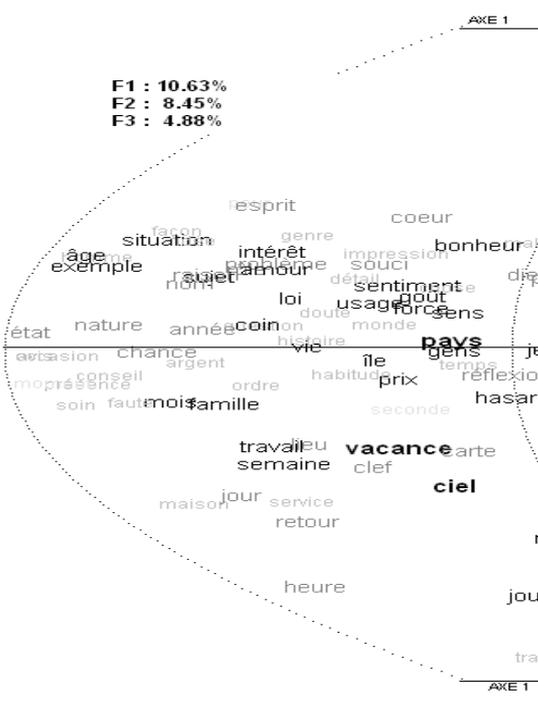


Figure 2. Micro-distribution-zoom des substantifs dans le corpus entier

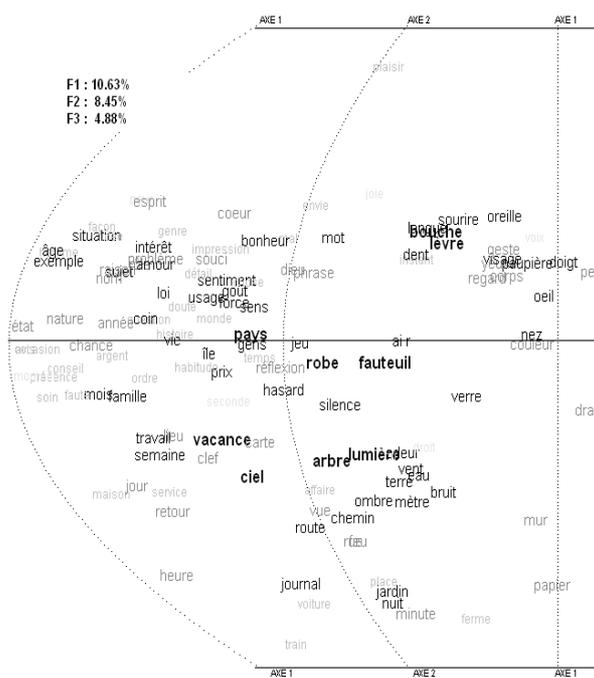


Figure 3. Micro-distribution-zoom des substantifs dans le corpus entier