

*Le fantastique : un genre reconnaissable
par l'outil informatique*

Séminaire d'Analyse du texte fantastique :
Littérature et informatique

Tania Collani
Università degli Studi di Bologna

Sommaire

Une déclaration d'intention

1. L'informatique appliquée aux sciences humaines : l'état actuel et un regard sur le futur

1.1. Logiciels d'analyse textuelle et théorie de la critique textuelle informatisée

2. Le fantastique : un genre

2.1. Le genre

2.2. Bref essai de caractérisation du fantastique

3. Est-ce qu'un logiciel d'analyse textuelle reconnaît le fantastique ?

3.1. Tropes

3.2. Définition d'un vocabulaire du fantastique : deux contes de Gautier

Conclusion

Bibliographie

Annexe

Une déclaration d'intention

Le propos de cette étude on se propose de donner un bref aperçu au sujet de l'informatique appliquée aux sciences humaines et de montrer comment, au cours des dernières décennies, cette matière a trouvé des applications plus spécialisées par rapport à une approche généraliste, que l'ampleur même de l'union de l'informatique et les disciplines humaines (littérature, lexicographie, philologie, histoire etc.) comporterait. Le but de ce travail étant d'appliquer un logiciel d'analyse textuelle (Tropes) à des textes fantastiques et de conceptualiser, à partir des résultats, un 'vocabulaire' du fantastique qui consente de reconnaître le genre d'une manière plus objective, on réservera une grande partie aux instruments qui permettent de mettre en acte une analyse littéraire informatisée : les bases des données textuelles accessibles sur Internet et les logiciels. Les outils techniques présentent cependant des limites et des manques ; en raison de cela, la figure du chercheur revêt un rôle primaire et à la base d'une analyse qui soit caractérisée par des critères de scientificité acceptable.

1. L'informatique appliquée aux sciences humaines : l'état actuel et un regard sur le futur

L'étude informatique appliquée aux sciences humaines, en italien *Informatica Umanistica* (IU) et connue dans l'aire anglo-saxonne avec le terme de *Humanities Computing* (HC), s'occupe de l'analyse, gestion et élaboration informatique de documents, données textuelles, images etc. pour les disciplines humaines (littérature, linguistique, philologie, philosophie, histoire, art etc.). Cela n'implique pas seulement l'acquisition, de la part des chercheurs qui opèrent dans ces domaines, des compétences techniques et pratiques, mais aussi, et surtout, la conscience d'un changement méthodologique comportant le traitement électronique de l'information, c'est-à-dire des contenus et des formes de leurs matières d'étude.

Cette discipline, bien que considérée, dans quelque cas, à l'instar d'un joli appellatif de fantaisie ne transmettant pas un vrai et propre savoir¹, constitue un champ

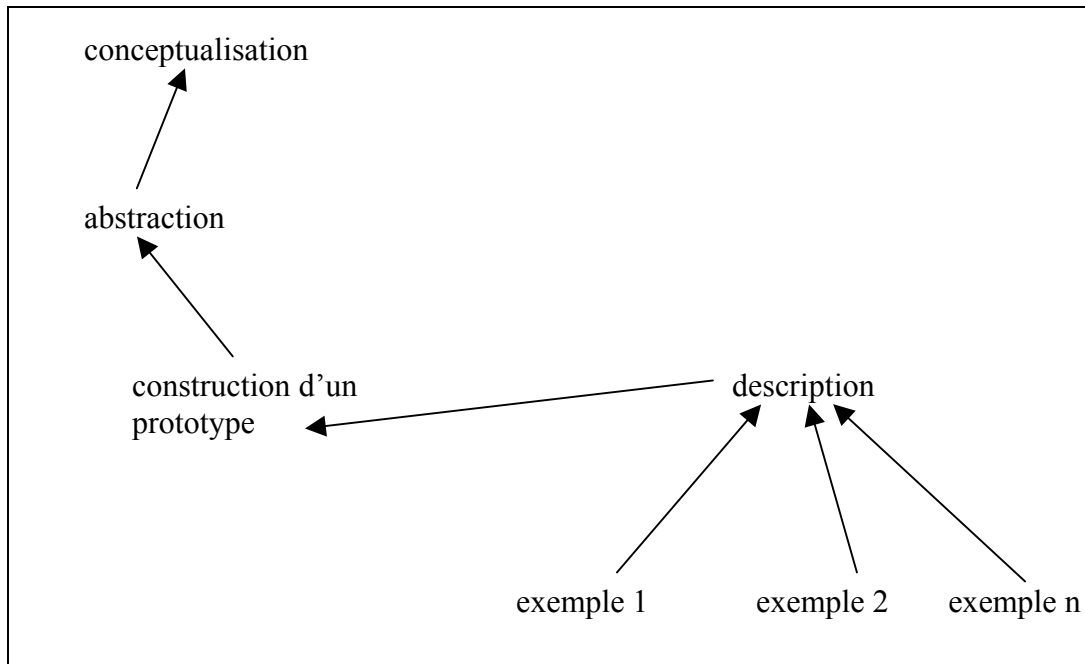
¹ Cf. BENEDETTI, G., « Troppe lauree brevi, 400 saranno cancellate », article in *Corriere della sera*, 6 juin 2002 : « ... molte facoltà hanno attivato una pletora di corsi di laurea che si distinguono l'uno

vaste qui attire nombreux chercheurs. L'analyse des textes informatisée, qui n'est qu'une branche de l'informatique des sciences humaines, représente un domaine riche en nouvelles possibilités pour la critique littéraire : l'automatisation de certaines tâches désagréables et enclines aux erreurs humaines (par exemple, la compilation des listes des occurrences et des fréquences d'un texte) figure sûrement parmi les apports positifs de l'informatique.

Née en 1949, avec les expériences conduites par Roberto Busa sur l'œuvre de Saint Thomas, l'informatique appliquée aux sciences humaines s'avère aujourd'hui un conteneur trop généraliste, qui nécessite d'une délégation de certaines fonctions à des secteurs plus spécialisés. C'est le cas de la linguistique informatisée, *computational linguistics*, qui s'occupe de l'abstraction et conceptualisation des structures sous-jacentes à la langue et aux procès d'apprentissage, pour créer des formes qui puissent décrire n'importe quel procédé verbal. De cette discipline, on a tiré le schéma, reproduit dans la figure suivante², qui expose la représentation conceptuelle de ce qu'on fera surtout dans la dernière partie de ce travail : en partant de deux contes fantastiques de Gautier, on a établi un vocabulaire du fantastique qui peut être appliqué à tout conte appartenant à ce genre.

dall'altro più per la fantasia del nome che per i contenuti. E questo non aiuta certo gli studenti. Sull'abbondante produzione di proposte formative aveva ironizzato il ministro Letizia Moratti: «Scienze del fiore e del verde», «Informatica umanistica», «Tutela del territorio di interesse forestale» e via dicendo ».

² Exemple tiré de TONFONI, G., *Intelligenza Artificiale : automazione dei processi cognitivi e sistemi per la rappresentazione della conoscenza*, in « Lectures », n° 19, Bari, Edizioni dal Sud, 1986, p. 93.



On a fourni, à ce propos et toujours dans la dernière section de l'étude, une description de fantastique en partant de quelques exemples pratiques (*Omphale* et *La pipe d'opium* de Gautier pour le fantastique, mais aussi une exception qui devrait confirmer les règles qu'on s'est imposé, c'est-à-dire un conte naturaliste de Zola : *Le chômage*) ; on a créé un prototype de texte fantastique, qui doit prévoir un moment d'attente, un moment de doute, qui provoque certains effets chez le protagoniste ou spectateur etc. ; à travers un procès d'abstraction on a conceptualisé, enfin, la structure d'un vocabulaire qui, si appliqué à un texte fantastique, donne certains résultats différant de son application à un texte naturaliste.

En s'agissant d'un bref travail, qui ne se veut pas exhaustif ni au sujet de la linguistique informatisée ni aux égards d'un vocabulaire sémantique du genre fantastique, toutes les parties pourraient être approfondies et perfectionnées. L'étude voudrait néanmoins constituer un tremplin pour la mise en acte d'une méthodologie, préalable à l'utilisation pratique d'un logiciel d'analyse textuelle, à travers laquelle construire une méthode 'objective' (si jamais on arriverait à objectiver totalement une œuvre littéraire) pour saisir le fantastique à l'intérieur d'un texte.

1.1. Logiciels d'analyse textuelle et théorie de la critique textuelle informatisée

Lorsqu'on parle d'informatique textuelle, il faut distinguer entre une informatique de type 'documentariste', c'est-à-dire l'informatique qui gère les bases de données, qui fournit des informations concernant les textes en temps utile, et une informatique de type 'linguistique', qui utilise les informations de l'informatique documentariste pour tirer une synthèse des catégories et des structures présentes dans les textes. Il existe, aujourd'hui, un assez grand nombre de sites Internet qui fournissent des outils informatiques 'documentaristes', c'est-à-dire des textes numérisés et prêts, ou presque, pour être soumis à l'analyse du logiciel 'linguistique'. À l'égard des textes littéraires en langue française on peut consulter de nombreuses bases de données textuelles : Athena (<http://hypo.ge-dip.etat-ge.ch/www/athena/html/francaut.html>), l'ABU (la Bibliothèque Universelle du CNAM, <http://abu.cnam.fr/index.html>), Gallica de la BNF (<http://gallica.bnf.fr>), le CETE (Centre d'Édition de Textes Electroniques, <http://palissy.humana.univ-nantes.fr/CETE/CETE.html>) de l'Université de Nantes, le projet ARTFL (The American and French Research on the Treasury of the French Language, <http://humanities.uchicago.edu/orgs/ARTFL>) de l'Université de Chicago, la Bibliothèque Électronique de Lisieux (<http://www.bmlisieux.com>), la Littérature Francophone Virtuelle (<http://clicnet.swarthmore.edu/litterature/litterature.html>), Littératures et Compagnies (<http://paprika.iut.univ-aix.fr/Litt-and-Co>) etc. En ce qui concerne les logiciels d'analyse textuelle, qui utilisent le matériel fourni par les bases de données, il y en a aujourd'hui un grand nombre disponible gratuitement sur Internet, même s'ils ne sont que des versions limitées par rapport au produit original. À l'égard de logiciels qui analysent de préférence des textes en langue française, on peut rappeler Alceste, Lexico, Hyperpo, Occurrences, Sato, Tropes etc.

Considérée la disponibilité de ces outils, le chercheur doit créer une syntaxe, une morphologie et un lexique qui puisse travailler dans une nouvelle optique, beaucoup plus vaste par rapport à la manière traditionnelle du travail 'manuel'. Pour le reste, l'analyse des textes procède selon la volonté et l'usage de chaque chercheur : d'habitude on part d'une analyse syntaxique et, seulement à un deuxième moment, on arrive à une étude stylistique. Certes, l'ordinateur ne peut pas remplacer le chercheur et, tout seul, ne peut que rédiger une taxinomie des caractéristiques textuelles. Deux procès de

formalisation nécessaire à la critique du texte sont subséquentement désignés : un procès de logique mathématique, de relevé des statistiques des variantes textuelles, de leur distribution, etc. ; et un procès de logique interprétative, qui prévoit une pondération qualitative des variantes, de la ponctuation etc.³. On peut déléguer à la machine le dépouillement automatique des variantes, le relevé des discordances et concordances lexicales et d'autres segments de la chaîne verbale qui puissent être catégorisés : les occurrences, les formules, les dictionnaires des rimes et des homophones.

Toutes les fréquences à l'intérieur d'un texte sont intéressantes. Les formes les plus fréquentes sont d'habitude les conjonctions, les articles et les prépositions, qui peuvent rendre compte de la fonction d'un texte (argumentative, informative etc.). Les formes plus rares et les hapax sont hautement informatifs pour ce qui concerne la richesse du contenu d'un texte. Des différences consistantes, à ce propos, sont remarquables entre un texte rédigé au moyen d'une langue littéraire (qui aura vraisemblablement un grand nombre de séquences rare et hapax) et un texte écrit en une langue d'usage commun.

Ce procès de découverte des récurrences de certains segments linguistiques du texte, ressemble à une sorte de recherche génétique de segments du ADN⁴ : l'étude des récurrences peut porter à tracer les branches d'un hypothétique arbre généalogique, où l'on peut établir des lignes de filiation et parenté à l'intérieur d'un texte ou d'un *corpus* des textes. D'ici on comprend l'importance d'un facteur préparatoire à l'analyse : le *corpus* de textes doit être homogène ou, au moins, caractéristique d'une situation qu'on veut décrire à travers l'analyse. Le rôle du chercheur est donc, encore une fois, en premier plan : il faut avoir une hypothèse d'analyse, des critères bien définis et, possiblement, on devrait déjà avoir l'intuition de ce qui sera le résultat du travail de l'ordinateur. Si on ne répond pas à toutes ces exigences rigoureuses, l'analyse n'aura aucun sens scientifique. Malheureusement, dès que les logiciels d'analyse textuels sont facilement accessibles à tout le monde, le nombre de travaux de critique textuelle sans aucune valeur, a augmenté vertigineusement. Maintes fois, vue la quantité

³ Cf. RONCAGLIA, A., « Procedimenti formali e 'divination' nell'ecdotica », in *Lingua, letteratura, computer*, éd. Mario Ricciardi, Torino, Bollati Boringhieri, 1991, pp. 19-29.

⁴ LANA, Maurizio, « Testi stile frequenze », in *op. cit.*, pp. 30-45.

d'informations qu'on obtient sans aucun effort, on risque de donner des interprétations totalement arbitraires, qui ne tiennent pas compte du contexte littéraire et historique.

Le style d'un texte ou d'un *corpus* de textes (mais le discours est aussi valable pour plusieurs *corpora* de textes) peut être défini grâce à l'examen des fréquences. Le style est, pour Cesare Segre,

la totalité des traits formels qui caractérisent (dans l'ensemble ou à un moment particulier) la manière de s'exprimer d'une personne, ou la façon d'écrire d'un auteur [...] la totalité des traits formels qui caractérisent un ensemble d'œuvres, constitué sur des bases typologiques et historiques⁵.

Pour la définition du style, il est important de distinguer deux termes nécessaires pour la formulation d'une bonne interprétation des résultats de l'analyse : les mots-thèmes et les mots-clés. Les mots-thèmes ont une haute fréquence (en excluant les mots grammaticaux qui occupent toujours les premiers rangs des listes des fréquences) et contribuent à la définition de la thématization d'un texte, c'est-à-dire la tendance à répéter un petit nombre de mots-thèmes ; tandis que les mots-clés sont des mots-thèmes qui montrent un écart significatif par rapport à la langue commune et qui peuvent être reconnus scientifiquement en faisant recours aux listes des fréquences d'un auteur, d'un siècle, d'un courant. La thématization d'un texte s'oppose à la caractérisation, qui est la tendance à employer un grand nombre de mots détaillés, spécialisés, divers, chacun de fréquence faible. La caractérisation, correspondant toujours à un vocabulaire riche, est typique des récits fantastique de Gautier, qui fait recours à une variété lexicale énorme ; on ne s'occupera pas, toutefois, du style de Gautier dans cette étude. Ce qui est fondamental de se rappeler, afin d'obtenir des résultats valides, est que le compte des fréquences sans instruments statistiques de références n'a pas beaucoup de sens scientifique.

⁵ SEGRE, C., *Avviamento all'analisi del testo letterario*, Torino, Einaudi, 1985, p. 305. Notre traduction : « [lo stile] è l'assieme dei tratti formali che caratterizzano (in complesso o in un momento particolare) il modo di esprimersi di una persona, o il modo di scrivere di un autore [...] l'assieme dei tratti formali che caratterizzano un gruppo di opere, costituito su basi tipologiche e storiche ».

2. Le fantastique : un genre

2.1. Le genre

En considérant la définition de « genre » de Jean-Marie Schaeffer⁶ et en l'appliquant au domaine du « fantastique », on arrive à supposer que le « fantastique » soit un acte discursif, caractérisé par un cadre communicationnel et par sa réalisation. Le cadre communicationnel se définit à son tour par l'énonciation (qui parle ?), la destination (à qui parle-t-on ?) et la fonction (à quel effet parle-t-on ?) ; tandis que de l'acte discursif sera réalisé en termes sémantiques et syntaxiques (qu'est-ce qui est dit et comment ?). Au niveau de l'énonciation trois phénomènes jouent un rôle de différenciation générique : le statut ontologique de l'énonciateur, qui peut être réel (l'énonciateur effectif est toujours réel), fictif (lorsque l'énonciateur second, délégué par l'énonciateur effectif, est inventé par l'auteur) ou feint (lorsque l'énonciateur second est identifié à une personne ayant existé ou existant réellement) ; le statut logique (énonciation sérieuse ou fictive) et physique (énonciation orale ou écrite) de l'acte énonciatif ; et le choix des modalités de l'énonciation (narration ou représentation). Au niveau de la destination, on distingue entre un destinataire réel ou fictif, déterminé (par exemple dans les lettres) ou indéterminé (dans les blagues) et entre une destination réflexive (par exemple dans les journaux intimes) ou transitive (communication normale). Au niveau de la fonction, il existe une distinction entre la fonction ludique et sérieuse (presque tous les genres littéraires occidentaux rentrent au sein de la fonction ludique, sauf la biographie, l'histoire et, dans certains cas, la littérature épistolaire) et deux modalités d'intervention de la fonction : les actes illocutoires (assertifs, directifs, promissifs, expressifs, déclaratifs, selon la répartition de Searle) et les actes perlocutoires. À ce propos, le fantastique semble caractérisé par des actes perlocutoires, considéré qu'il vise, de la perspective d'un Caillois par exemple, à effrayer, à provoquer la peur et la perturbation chez le lecteur.

La réalisation sémantique et syntaxique de l'acte discursif s'avère essentielle dans le cadre d'une possible application des logiciels d'analyse textuelle. Les traits de

⁶ SCHAEFFER, J.-M., *Qu'est-ce qu'un genre littéraire ?*, Paris, Éd. du Seuil, 1989.

contenu du niveau sémantique (« sujet », « motif », « thème » etc.) jouent un rôle essentiel dans la formation de plusieurs noms de genre (par exemple la poésie d'amour, le récit de voyage, le récit de science-fiction etc.). La 'primarité' accordée au contenu n'est pas pourtant une simplification ; Jean-Marie Schaeffer écrit à ce propos :

L'importance accordé aux traits de contenu, même si ce n'est pas pour des besoins de simplification et de cohérence théorique (comme chez Hegel), n'a rien d'étonnant : après tout, les hommes communiquent, entre autres, pour se transmettre des informations, et la question du contenu du message, de la compréhension de son « à-propos-de » (*aboutness*), est donc tout à fait normalement l'aspect qui retient leur attention⁷.

Le niveau syntaxique, de son côté, comprend l'ensemble des éléments qui encodent le message, tels que des facteurs grammaticaux (la syntaxe au sens linguistique du terme), des facteurs phonétiques, prosodiques et métriques (surtout en poésie), des traits stylistiques (qui caractérisent un style élevé, moyen et bas, qui définit, à son tour, la 'grande' littérature de la littérature de masse) et des traits d'organisation macrodiscursive (qui permettent la distinction entre une réalisation narratologique, dramatologique etc.).

La définition de genre se démontre donc très articulée et, en fin de compte, elle revient à un choix subjectif, considéré qu'il existent des noms de genre relatifs à tous les niveaux du message qu'on vient d'énumérer. Cependant la difficulté de l'identité générique des œuvres littéraires ne dépend uniquement du fait que les textes sont des actes sémiotiquement complexes ; « elle est liée aussi au fait que les œuvres, tant écrites qu'orales, ont toujours un mode d'être historique. Tout acte de langage est contextuel et on n'accède à sa réalité pleine que si on peut l'ancrer dans ce contexte »⁸.

Le genre fantastique, au cours de l'histoire et surtout au XIXe siècle, a été souvent utilisé pour satisfaire les 'exigences alimentaires' des écrivains et principalement des poètes : il s'agissait d'une littérature populaire, qui réussissait à vendre et qui était recherchée par des nombreuses revues. Le fantastique, qui s'associe presque toujours à la forme narrative du récit (conte ou bref roman), facilement accessible à tout type de lecteur, caractérise généralement la première partie de la

⁷ *Ibid.*, p. 108.

⁸ *Ibid.*, p. 131.

carrière des auteurs qui n'ont pas encore connu le succès. Il devient donc plus légitime de tracer une liste de traits péculiers qui distinguent le genre fantastique, qui a un statut tout à fait comparable au rôle joué par le policier à nos jours, constamment en équilibre instable entre l'être compté parmi les rangs d'une littérature 'de train' et l'être considéré comme une partie fondamentale de la 'grande' littérature.

Le fantastique, étant donné la description de genre, est particulièrement déterminé par la fonction perlocutoire et par le niveau sémantique. Le niveau sémantique est certes plus saisissable au moyen d'un logiciel, puisqu'il s'agit d'une opération de dénotation directe à partir d'un 'dictionnaire' du fantastique qu'on doit mettre à point avant l'analyse même ; tandis que pour trouver la fonction perlocutoire à l'intérieur d'un récit fantastique, il faudrait 'communiquer' au logiciel comment la reconnaître.

2.2. Bref essai de caractérisation du fantastique

Étant donné que le « fantastique » est un genre littéraire, dont on analysera les détails au fur et à mesure qu'on avance dans la caractérisation du récit fantastique, on se propose d'utiliser un logiciel d'analyse textuelle (Tropes) pour voir s'il est possible de définir un *corpus* des règles valables pour l'individuation de cette catégorie littéraire spécifique. Avant d'employer pratiquement ce moyen technologique qui se prête surtout à une analyse sémantique et syntaxique, il faut tracer une liste schématisée des caractéristiques qui interviennent régulièrement dans tout récit fantastique. Bien évidemment on ne peut pas s'attendre à une occurrence systématique de tous les éléments 'typiques' dans tous les récits considérés fantastiques parce que, si cela se vérifiait, on se trouverait face à des textes sériels qui n'ont rien à voir avec la littérature. À cet égard Todorov écrit justement que chaque œuvre littéraire apporte au genre une certaine originalité, rendant les frontières mêmes du genre 'flottantes'. Mais cela n'empêche pas l'existence de traits distinctifs du fantastique qui peuvent être tirés surtout des champs sémantiques et lexicaux de l'œuvre. Cet argument détruit en part l'excellent travail de Todorov, qui a cherché à déplacer le champ définitoire du fantastique de la sphère assez superficielle du contenu (thèmes, sujets, motifs) vers des critères plus spécifiquement littéraires (la structure). Cependant Todorov n'ignore pas

l'importance de l'étude thématique du fantastique vu qu'il dédie, dans son *Introduction à la littérature fantastique*, un chapitre entier à cette question⁹.

On part en considérant les moments les plus difficiles à vérifier automatiquement à l'intérieur des textes fantastiques, c'est-à-dire avec les descriptions. Les descriptions jouent un rôle fondamental pour la préparation du scénario crucial du fantastique, du climax qui précède l'« apparition », comme Caillois aime la définir¹⁰, ou de l'« hésitation » todorovienne¹¹. Attendu que les récits fantastiques sont toujours assez brefs, les descriptions doivent être ponctuelles et laisser, au juste moment et au juste lieu, des fils en suspens, qui seront renoués seulement à la fin du conte ou du roman. Maintes fois, et surtout chez certains auteurs, parmi lesquels on peut citer Gautier, les descriptions sont très détaillées, riches en vocabulaire spécifique, parce que c'est à partir de l'anomalie des objets communs qui surgit le fantastique.

Or, même si on comprend l'importance de la fonction descriptive à l'intérieur des textes fantastiques, il résulte assez difficile saisir, au moyen d'un logiciel d'analyse textuelle, cette dimension. C'est pour cela qu'il s'avère nécessaire de constituer un vocabulaire du fantastique qui simplifie (bien que le mot « simplifier » s'adapte plus à l'ordinateur qu'à la littérature) la tâche de cette ressource avantageuse qui est le logiciel. Certainement la « simplification » pour la machine ne signifie pas une exemption du travail pour le critique, qui doit surveiller les résultats et les comparer avec son idée sur l'œuvre.

À la lumière de ces exigences de simplification et tenant compte de ce qu'on a dit sur la spécificité de certaines descriptions, les descriptions fantastiques comportent un grand nombre d'adjectifs, noms, adverbes indéfinis, tels que « inconcevable », « insignifiant », « ne sait quoi », « confus », « mystère », « quelque chose », « une figure », « une image », « quelqu'un », « étrange », « cela », « autre », etc.. Particulièrement intéressantes et 'typiques' se sont démontrées, au cours de l'analyse, les descriptions des lieux fantastiques. Au-delà des mots qui rentrent dans le champ

⁹ Cf. « Les thèmes du fantastique : introduction », in TODOROV, Tzvetan, *Introduction à la littérature fantastique*, Paris, Seuil, 1970, pp. 97-112.

¹⁰ Le thème de l'apparition revient dans plusieurs écrits de Caillois, entre autre : CAILLOIS, Roger, « Fantastique », in *Encyclopaedia Universalis*, vol. 6, 1968, p. 924 : « La démarche essentielle du fantastique est l' *apparition* ».

¹¹ TODOROV, T., *op. cit.*, p. 29 : « Le fantastique, c'est l'hésitation éprouvée qui ne connaît que les lois naturelles, face à un événement en apparence surnaturel ».

sémantique du non-défini, les descriptions auront la tendance à être plutôt argumentatives, dans le sens qu'elles doivent créer une situation différente par rapport à la réalité et qui n'est pas tout de suite acceptée par le protagoniste, sinon on se trouverait au milieu d'un récit « merveilleux ». Le discours fantastique tend « à faire admettre au récepteur une information qu'il ignore plus ou moins et aurait tendance à rejeter »¹².

Le fantastique, d'un point de vue plus générique et théorique, se caractérise, en outre, par trois moments fondamentaux : une phase qui précède la préfiguration du fantastique, la création de la suspense, le soupçon qui s'insinue dans le personnage (et en même temps chez le lecteur) qui pressent des anomalies dans le monde qui l'entoure ; une phase du saisissement du fantastique, du vraisemblable qui, comme on verra dans l'analyse ponctuelle de la structure qu'on a créée, est caractérisé par des verbes au passé simple, par la découverte de l'étrange, du bizarre ; une dernière phase des effets produits par le fantastique, qui part de l'éblouissement et arrive jusqu'à la folie. Toute cette structure doit être, dans le cas d'une analyse textuelle informatisée, reconduit à un niveau sémantique.

3. Est-ce qu'un logiciel d'analyse textuelle reconnaît le fantastique ?

3.1. Tropes

Chez les sceptiques grecs, le trope était un argument servant à prouver la nécessité de la suspension du jugement ; en linguistique, le trope est une figure rhétorique consistant à détourner un mot ou une expression de son sens propre.

Tropes est aussi le logiciel qu'on a utilisé dans ce travail. Tropes se propose une étude comparée de textes (principalement de textes historiques) à travers leur analyse sémantique et lexicométrique, mettant en place un examen sémantique très rapide et visuel, recourant aussi à l'utilisation d'un grand nombre de graphiques. Ce logiciel offre une réflexion, peut-être un peu trop restrictive, autour de notions et mots-clés, pouvant constituer une préparation intéressante aux méthodes de recherche dans l'univers numérique (en ligne ou hors ligne).

¹² DOLPHIN, C., *Méthodes de la statistique linguistique et vocabulaire fantastique de Malpertuis*, Genève, Slatkine, 1979, p. 33.

Tropes est particulièrement utilisé pour l'analyse du langage et surtout des récurrences lexicales dans les interventions politiques, pour saisir les nuances perlocutoires dans les programmes des candidats aux mêmes élections et dans les propos du même énonciateur sur une même situation mais pour des récepteurs différents, ce logiciel étant notamment efficace pour marquer les modifications et les adaptations de registres de langues. Dans cette liste d'applications s'adaptant parfaitement aux fonctions de Tropes, on doit inclure aussi les éditoriaux de journaux d'opinion différente sur les mêmes événements, les jugements et les appréciations comparés de témoins ou acteurs différents sur des événements identiques.

La version disponible gratuitement de Tropes V6 ES (Édition Spéciale) et téléchargeable du site de Acetic (www.acetic.fr), comporte certaines limites : elle ne permet pas de faire des copies de texte ou de graphe et les textes traités doivent être limités à 32.000 caractères (ce qui est cependant suffisant pour la plupart contes fantastiques). Une autre grande limite de ce logiciel, qui se fonde sur le rapprochement de textes, réside sur ce que la comparaison des documents ne peut être faite simultanément (on ne peut traiter et ouvrir qu'un seul fichier à la fois). Ce qui suppose de prendre des notes au fur et à mesure de l'investigation, texte par texte ou de remplir un fichier traitement de texte sous forme d'un tableau en deux colonnes.

Comme tous les logiciels d'analyse textuelle, Tropes exige un travail préliminaire sur le texte que seulement le critique littéraire, dans le cas des textes qu'on ira analyser, peut faire. Il faut prendre en considération des problématiques de codification numérique, d'homogénéité des critères d'examen et, surtout, on doit créer une table des valeurs à associer à une majeure ou mineure fréquence lexicale, à la présence d'un vocabulaire riche d'hapax ou des mots qui se répètent continûment. Il faut donc avancer des hypothèses et tirer profit de la technologie qui calcule au lieu de l'homme. En ce qui concerne les aspects les plus pratiques de l'emploi des softwares qui travaillent avec les sciences humaines, et donc avec des documents textuels écrits, il faudrait avant tout appliquer la règle du bon sens, c'est-à-dire on devrait prêter beaucoup plus d'attention aux critères qui précèdent et qui succèdent l'analyse automatisée d'un texte, plutôt qu'à l'analyse en elle-même.

Tout d'abord les textes nécessitent d'une codification en format lisible par le logiciel. Dans le cas de Tropes, les fichiers textes à traiter doivent être convertis en .txt

(document de texte) ou en .rtf (rich text format). Le texte, en outre, peut contenir des signes de ponctuation, des majuscules et des caractères spéciaux (comme des parenthèses, des nombres, des pourcentages, etc...), mais seules les lettres de l'alphabet et les caractères de ponctuation seront utilisés durant l'analyse. Si l'on veut qu'une suite de mots soit considérée comme un mot unique, on devra les lier avec le caractère souligné « _ » (par exemple : Cyrano_de_Bergerac).

Lorsque l'on veut entreprendre un travail comparatif, comme on le fera tout de suite avec les trois contes de Gautier, il est souhaitable de l'effectuer en faisant recours à une méthodologie cohérente et en l'appliquant à des textes de longueur, de style et d'époque comparables : en effet, on n'aura rien découvert si à la fin de l'analyse comparative d'un roman de Flaubert et d'un poème d'Ungaretti on obtient des résultats contrastants ; en s'agissant de deux siècles différents (le XIX^e et le XX^e), de deux formes d'écritures diverses (le roman et la poésie) et de deux styles totalement opposés (le descriptif de Flaubert et l'hermétique d'Ungaretti), l'issue sera facilement prévisible. Pour l'étude comparée des textes vaut aussi la règle d'un rigoureux établissement des propos d'analyse avant l'analyse même : l'examen automatique ne soulève pas le critique du devoir de lecture et d'interprétation du texte. L'idée générale à propos d'un œuvre doit naître indépendamment du logiciel, qui sert pour confirmer ou pour démentir certaines présuppositions formulées à l'avance.

Tropes effectue un diagnostic du style général du texte en fonction des indicateurs statistiques récupérés au cours de l'analyse. Le logiciel reconnaît quatre styles possibles : l'argumentatif (le essayer de persuader l'interlocuteur), le narratif (un narrateur expose une succession d'événements, qui se déroulent à un moment donné, en un certain lieu), l'énonciatif (le locuteur et l'interlocuteur établissent un rapport d'influence, révèlent leurs points de vue) et le descriptif (un narrateur décrit, identifie ou classifie quelque chose ou quelqu'un). Toujours en faisant recours à des indicateurs statistiques, Tropes exécute un bilan au sujet d'autres situations textuelles, telles que les mises en scène, les propositions remarquables, les univers de référence, les actants et les actés ; cependant ce type de résultats est modérément intéressant, étant donné que le diagramme à la base de la construction de ces unités d'analyse est assez superficiel pour une étude littéraire.

Pour surmonter l'obstacle du simplisme, l'utilisateur peut intervenir dans ce procès d'analyse automatisée. Il peut, par exemple, délimiter le seuil de détection des classes d'équivalents, c'est-à-dire qu'il peut décider le facteur de pertinence au-dessus duquel la fréquence des mots est prise en considération par le logiciel. L'utilisateur peut aussi créer des scénarios, conçus pour enrichir et filtrer les classes d'équivalents en fonction d'une stratégie d'analyse. Ces scénarios permettent de définir des classifications personnalisées, modifier ou restructurer les dictionnaires du logiciel, remplacer un thesaurus et personnaliser les fonctions de recherche d'informations et définir une grille d'analyse statistique pouvant être générée automatiquement lors de l'indexation d'une base documentaire. L'utilisation d'un scénario personnalisé est généralement indispensable pour effectuer une analyse correcte d'un texte et obtenir un examen sémantique particulier.

3.2. Définition d'un vocabulaire du fantastique : deux contes de Gautier

Théophile Gautier est connu aux critiques littéraires plutôt comme l'excellente plume des *Émaux et Camées* (première édition en 1852) ou du *Capitaine Fracasse*, qu'en tant qu'écrivain de contes fantastiques. Pendant toute sa carrière, cependant, Gautier fut particulièrement sensible à la veine fantastique et aux romans noirs, très diffusés à l'époque (les années '20 et '30 du XIX^e siècle). Dans « La Presse » du 21 mars 1852, Gautier écrit à propos de sa jeunesse et de sa lecture du *Diable amoureux* de Cazotte :

Ce conte nous a beaucoup frappé, et nous nous souvenons de l'avoir lu autrefois dans une vieille édition du temps, illustrée par la main de Cazotte lui-même d'eaux-fortes primitives imitant la naïveté des gravures de la bibliothèque bleue¹³.

Et dans « La Gazette de Paris » du 23 janvier 1872, il écrit au sujet du roman noir :

¹³ Cit. in GRASSO, L., *Percorsi labirintici nei racconti fantastici di Th. Gautier*, Palermo, Università degli studi di Palermo – Facoltà di Lettere e Filosofia, 1987.

Enfant, nous trouvions un charme inexprimable de curiosité et de terreur à suivre les héroïnes d'Anne [sic] Radcliffe dans leurs excursions nocturnes à travers le dédale de couloirs, de corridors, de passages secrets et de souterrains du château des Pyrénées et autres manoirs gothiques. Homme, notre goût n'a pas changé, et nous ne manquons pas une occasion de le satisfaire¹⁴.

À partir des années '30 Gautier, qui est désormais devenu membre du cénacle de Hugo, qui a déjà connu Nerval et le romantisme, s'adonne à l'écriture de contes fantastiques qui ressentent déjà le style des ses œuvres successives. Les contes fantastiques chez Gautier ne sont pas seulement le divertissement du jeune âge, mais recouvrent un grand laps de temps. En ce qui concerne les contes qui sont pris en considération dans ce travail, à cause de leur taille réduite plutôt que pour d'autres raisons plus spécifiquement littéraires¹⁵, *Omphale, ou La Tapisserie amoureuse* a été publié, pour la première fois, le 7 février 1834 dans le « Journal des gens du monde », tandis que le successif *La Pipe d'opium* paraît le 27 septembre 1838 dans « La Presse ».

Dans les contes fantastiques de Gautier le rêve et le sommeil sont constamment évoqués, ils constituent la justification d'une invraisemblance qui est toujours mise en scène comme une possible dimension onirique. Le narrateur de *Jettatura* dit : « on peut croire ou nier tout : à un certain point de vue, le rêve existe autant que la réalité »¹⁶. La sphère du rêve est à la base d'*Omphale* : « Je fis cette nuit-là un rêve singulier, si toutefois c'était un rêve »¹⁷. L'activité onirique se mélange inlassablement avec la vigilance de la veille, sans une ligne de démarcation nette. « Je m'éveillai ; du moins dans mon rêve il me sembla que je m'éveillais ». Les forces irrationnelles qui gouvernent les récits ne sont pas pourtant extérieures à l'homme : elles sont des projections mentales qui proviennent du profond de la conscience humaine, en envahissant la réalité extérieure¹⁸.

¹⁴ Cit. in *ibidem*.

¹⁵ L'édition spéciale de Tropes n'accepte pas de textes qui ont une extension supérieure aux 30.000 caractères.

¹⁶ GAUTIER, Théophile, « Jettatura », in *Contes fantastiques*, Paris, Flammarion, 1981, p. 427.

¹⁷ GAUTIER, T., « Omphale », in *Ibidem*, p. 108.

¹⁸ Cf. GRASSO, L., *op. cit.*

D'un point de vue générale de l'analyse, l'élément de la « porte » est un motif recourant à l'intérieur d'*Omphale* et de *La pipe d'opium* et il semble être un moment fondant du fantastique : une nouvelle dimension qui s'ouvre au regard du protagoniste, elle est le point de dépassement physique de la réalité au monde fantastique, qui est toujours réel, mais saisi dans son étrangeté, dans ses particularités qui éveillent la curiosité et l'inquiétude du narrateur. La « porte » physique est remplacée métaphoriquement par la « porte » du rêve, qui donne l'accès à la sphère onirique. De ce point de vue, l'hésitation qui caractérise le récit fantastique, pourrait se concrétiser chez le narrateur (et chez le lecteur) qui reste sur le pas de la porte, qui n'arrive pas à se décider si entrer à l'intérieur d'un monde merveilleux ou si sortir à la lumière de la réalité. Cette « porte » est représentée, dans *La Pipe d'opium*, par la fuite à un état second de conscience à l'aide de la drogue, qui provoque des visions hallucinées, partant des séquences vécues la veille. Un autre élément omniprésent dans les récits fantastiques de Gautier est la « femme », une femme d'une beauté perturbante, une femme dangereuse, qui cache dessous l'enveloppe charmante la laideur de la mort, de la pourriture. Une femme qui représente un autre monde.

Ce type d'analyse ne constitue pas, toutefois, un trait assez 'objectif' caractérisant de façon préférentielle le récit fantastique. Dans le conte naturaliste de Zola qu'on a pris en considération¹⁹, par exemple, les mots « porte » et « femme » sont même plus fréquents que dans les récits fantastiques de Gautier. On a, en conséquence, utilisé la fonction de Tropes qui permet la création de scénarios pour une analyse des classes sémantiques des textes, et on a intégré les classes sémantiques avec le lexique qu'on a retenu déterminant pour la formation d'une 'atmosphère' fantastique. L'édition spéciale de Tropes ne permet pas, malheureusement, d'imprimer ou de sauvegarder les nouveaux scénarios, ce qui empêche une analyse vraiment exhaustive et ce qui ne permet pas d'étendre la structure créée à un large *corpus* de textes.

Comme on peut dans l'annexe, dans la représentation du scénario caractérisant le fantastique, on a défini plusieurs champs sémantiques primaires (en ordre alphabétique) : attente, comportement, description du lieu (ombre/lumière sont un champ sémantique secondaire), doute/hésitation, effets du fantastique, enfer,

¹⁹ ZOLA, É., « Le chômage », in *Contes à Ninon*, Paris, Charpentier, 1879.

invraisemblable, perception (œil/vue secondaire), fonctions perlocutoires, peur, présupposition/préfiguration, refoulé, rêve/sommeil, verbes qui figurent souvent au passé simple, vraisemblable.

Après la mise en acte du scénario à l'aide du logiciel, on a trouvé assez insignifiants, au moins du point de vue du but qu'on veut rejoindre, les champs du refoulé et des actes perlocutoires. Par contre, l'expérience avec les autres catégories a prouvé une certaine cohérence dans la composition du tableau, comme on le verra pas à pas. Les carrés rouges indiquent les champs sémantiques connus par le logiciel, qui n'ont donc pas besoin d'être établis par l'utilisateur : par exemple, le carré rouge du « comportement » reconnaît, dans *Omphale*, des adjectifs et des substantifs qui reviennent à l'attitude : patience, prudence, innocence, impatience, dédain etc. Comme il est évident que les noms et les adjectifs du champ sémantique « comportement » ne sont pas tous intéressants de la même manière pour un discours finalisé au fantastique, on a spécifié, avec deux autres carrés rouges (les couleurs sont prédéterminées par le logiciel), les champs sémantiques « patience » et « prudence » qui limitent le nombre des résultats.

Toujours en considérant le cas du champ sémantique « comportement », il incluait 7 éléments au début (« précaution », « héros », « dédain », « naïf », « innocence », « joie », « impatience »), tandis qu'après la spécification il n'en comportait que deux (« précaution » et « impatience »). Les deux substantifs rentrent dans la phase qui précède le fantastique, la phase de préfiguration qu'on a délinée auparavant. Tout de même, le groupe sémantique « attente » (4 éléments lexicaux dans *Omphale*) et le groupe « présupposition/préfiguration » (21 éléments) sont inclus dans le moment pré-fantastique. Ce dernier groupe est particulier, puisqu'il a été construit avec un grand nombre de verbes (carré bleu) qui renvoient aux opinions, aux rapprochements et aux hypothèses que le protagoniste, ou d'autres personnages, se font d'une certaine situation. Qu'on prenne des exemples tirés d'*Omphale* : « Je me *hasardai* à regarder du côté d'Omphale, *soupçonnant* confusément qu'elle était pour quelque chose dans tout cela » ; « mais je *pressentais* vaguement que cela ne pouvait plaire au marquis » ; « Omphale *ressemblait* au charmant fantôme de la nuit comme un mort *ressemble* à un vivant » etc.

Certains groupes appartiennent à la sphère descriptive du fantastique : la « description du lieu » (9) qui est presque toujours caractérisée par des espaces clos, humides, froids et un contraste marqué entre lumière (artificielle) et ombre par exemple « la faible pénombre lumineuse qui flottait autour de la bougie ne servait qu'à rendre les ténèbres plus visibles » ; le champ sémantique de l' « enfer »²⁰ (8) ; la « perception » (28) où la majorité des éléments est représentée par le lexique de la vue et des yeux (24 sur 28).

Dans le tableau, des champs thématiques sont dédiés au « doute » (3 éléments), à la « peur » (10), au « rêve/sommeil » (15) et aux « effets suscités par le fantastique » (7) sur le protagoniste ou tout autre personnage. Certes, ces catégories sont très subjectives et sont susceptibles d'être réarrangées sans faire recours à des ressources scientifiques trop profondes. Un groupe est dédié aux verbes qui, dans les récits fantastiques analysés, ont la tendance à être au passé simple (17). Ils revêtent un rôle important, puisque le récit fantastique est caractérisé par une narration, généralement, à l'imparfait et les passages au passé simple signent une coupure où le fantastique peut s'amorcer, par exemple : « *J'entendis* les anneaux des rideaux de mon lit glisser en criant sur leurs tringles », « *Quand je me trouvais* dans cette belle chambre » etc. Le dernier groupe est dédié au champ du « vraisemblable » (15), qui appartient à la troisième phase du fantastique, c'est-à-dire à la définition de ce qui a été préfiguré comme fantastique, qui a produit des effets, des doutes et qui vient nommé (procédant à un niveau paradigmatique du texte) tel que faisant partie de la sphère d'un « vraisemblable positif » (4 éléments ; par « positif » on entend ce qui est tendant au réel) ou, par contre, d'un « vraisemblable hors norme » (7). Les adjectifs qui paraissent dessous les noms des groupes donnent l'idée du critère, ne pouvant pas tirer d'autres exemples des résultats de l'analyse à l'aide de la version spéciale de Tropes.

Les chiffres sans un contexte et sans des statistiques ont peu de sens dans le cadre de ce type d'analyse. Deux des trois textes qu'on a mis en parallèle appartiennent au genre fantastique (*Omphale* et *La pipe d'opium* de Gautier), tandis que le troisième conte appartient au naturalisme (*Le chômage* de Zola). On a choisi le troisième conte

²⁰ Exemple tiré d'*Omphale* : « Le diable, tranchons le mot, n'est-ce pas ? c'est cela que tu voulais dire ; au moins tu conviendras que je ne suis pas trop noire pour un diable, et que, si l'enfer était peuplé de diables faits comme moi, on y passerait son temps aussi agréablement qu'en paradis ».

parce qu'il avait des points de contact, au niveau sémantique, avec les textes de Gautier (la femme, la porte, l'humidité, une condition de malaise physique et psychologique) et parce que, à cause de cela, l'analyse résultait encore plus difficilement énigmatique.

STATISTIQUES ESSENTIELLES – GAUTIER, *OMPHALE*

	TEXTE ENTIER	VOCABULAIRE DU FANTASTIQUE	% FANTASTIQUE DANS TEXTE
MOTS	3429	144	4,2%

STATISTIQUES ESSENTIELLES – GAUTIER, *LA PIPE D'OPIMUM*

	TEXTE ENTIER	VOCABULAIRE DU FANTASTIQUE	% FANTASTIQUE DANS TEXTE
MOTS	2809	83	2,95%

STATISTIQUES ESSENTIELLES – ZOLA, *LE CHÔMAGE*

	TEXTE ENTIER	VOCABULAIRE DU FANTASTIQUE	% FANTASTIQUE DANS TEXTE
MOTS	1904	44	2,3%

Comme on peut voir, la différence du pourcentage de présence du vocabulaire fantastique à l'intérieur du *Chômage* et de *La pipe d'opium* est très petite ; cela peut être expliqué, comme on l'a dit, par le fait que le conte de Zola présente nombreuses ressemblances sémantiques par rapport aux autres récits. Mais on peut aussi déduire que le lexique utilisé n'est pas assez approprié pour saisir le fantastique à l'intérieur d'un texte et qu'il serait préférable, par exemple, d'appliquer un modèle *if-then* aux structures supérieures du texte (séquences).

Conclusion

L'informatique appliquée aux sciences humaines est un champ très vaste et son domaine d'action va, substantiellement, en deux directions complémentaires : d'un côté les grands courants du traitement électronique de textes - les banques de données, l'enseignement assisté par ordinateur, l'édition, qui comprend les multimédia ; de l'autre côté l'analyse herméneutique computerisée. L'analyse textuelle, à son tour, peut être focalisée à un *corpus* de textes et avoir comme but l'examen ponctuel des composants sémantiques et lexicales qui caractérisent la base documentaire, pour en tirer des considérations concernant le style d'un ou plusieurs auteurs, l'actualisation d'un certain genre littéraire, la fréquence des mots etc.. L'analyse textuelle, néanmoins, peut aussi être focalisée à la formulation de statistiques finalisées au saisissement des macrostructures qui règlent le fonctionnement de certains types de textes.

Le travail qu'on a entrepris au sujet du vocabulaire du fantastique voulait s'insérer plutôt dans la ligne macro-analytique, même si en pratique, considérée l'étroitesse du *corpus* et les limites de l'édition spécial du logiciel, il rentre à l'intérieur d'un discours sur la récurrence sémantique des textes.

Arriver à traiter les mots d'un texte en tant qu'éléments homogènes, semblablement à des nombres à l'intérieur d'une même opération de type mathématique, est certainement avantageux aux disciplines humaines et surtout aux banques de données, même si, à l'état actuel, les résultats de l'informatique appliquée aux lettres ne présentent pas, tous seuls, les conditions nécessaires à une cohérence scientifique impeccable. Les mots et les textes, au moins pour ce qu'on a constaté avec cette étude, se caractérisent par leur hétérogénéité qui peut, seulement en partie, être simplifiée au moyen de logiciels et d'autres outils informatiques. Il n'existe donc pas, pour le moment, la possibilité d'un remplacement de la figure du chercheur par des systèmes technologiques.

Bibliographie

- ATABEKIAN, Caroline D' : *Le logiciel Tropes : quel usage en cours de français ?*. <http://framanet.free.fr/logiciels/presentations/tropes>.
- BELLEMIN-NOËL, Jean, *Des formes fantastiques aux thèmes fantastiques*, in « Littérature », n° 2, mai 1971, pp. 103-118.
- *Dizionario di linguistica*, direzione Gian Luigi Beccarla, Torino, G. Einaudi, 1994.
- DOLPHIN, Colette, *Méthodes de la statistique linguistique et vocabulaire fantastique de Malpertuis*, Genève, Slatkine, 1979.
- GIGLIOZZI, Giuseppe,
 - *Il testo e il computer: manuale di informatica per gli studi letterari*, Milano, B. Mondadori, 1997.
 - *Introduzione all'uso del computer negli studi letterari*, a cura di Fabio Ciotti, Milano, B. Mondadori, 2003.
 - *Letteratura, modelli e computer: manuale teorico-pratico per l'applicazione dell'informatica al lavoro letterario*, Roma, EUROMA, 1993.
- GHIGLIONE, Rodolphe, « Le logiciel Tropes », in GHIGLIONE, R., *L'analyse automatique des contenus*, Paris, Dunod, 1998, chap. 2.
- GAUTIER, Théophile, *Contes fantastiques*, Paris, Flammarion, 1981.
- GRASSO, Luciana, *Percorsi labirintici nei racconti fantastici di Th. Gautier*, Palermo, Università degli studi di Palermo – Facoltà di Lettere e Filosofia, 1987.
- *La pratique des ordinateurs dans la critique des textes*, Paris, 29-31 mars 1978, édition Jean Irigoïn, Gian Piero Zarri, Paris, CNRS, 1979.
- *Lingua, letteratura, computer*, éd. Mario Ricciardi, Torino, Bollati Boringhieri, 1991.
- LANA, Maurizio, *L'uso del computer nell'analisi dei testi*, Milano, F. Angeli, 1994.
- LEBART, Ludovic, SALEM André, *Analyse statistique des données textuelles : questions ouvertes et lexicométrie*, préface di Christian Baudelot, Paris, Dunod, 1988.

- MORDENTI, Raul, *Informatica e critica dei testi*, Roma, Bulzoni, 2001.
- PASI, Carlo, *Théophile Gautier o il fantastico volontario*, Roma, Bulzoni, 1974.
- ORLANDI, Tito, *Informatica umanistica*, Roma, NIS, 1990.
- SCHAEFFER, Jean-Marie, *Qu'est-ce qu'un genre littéraire ?*, Paris, Éd. du Seuil, 1989.
- SEGRE, Cesare, *Avviamento all'analisi del testo letterario*, Torino, Einaudi, 1985.
- TODOROV, Tzvetan, *Introduction à la littérature fantastique*, Paris, Seuil, 1970.
- TONFONI, Graziella,
 - *Didattica del testo. Curricolo di formazione linguistica per gli insegnanti*, Teramo, Giunti, 1991.
 - *Sistemi cognitivi complessi. Intelligenza artificiale e modelli di organizzazione della conoscenza*, Paese (Treviso), PAGVS Edizioni, 1991.
 - *Intelligenza Artificiale : automazione dei processi cognitivi e sistemi per la rappresentazione della conoscenza*, in « Lectures », n° 19, Bari, Edizioni dal Sud, 1986, pp. 77-95.
 - *Paradoxes and censors*, in « Semiotica », vol. 60, Amsterdam, Mouton de Gruyter, 1986, pp. 247-257.
- *Tropes : un outil au service de l'analyse de textes historiques*. http://www.ac-rennes.fr/pedagogie/hist_geo/ResInternet/Tropes/Tropes-Aide.htm.
- VUILLEMIN, Alain, *Informatique et traitement de l'information en Lettres et Sciences Humaines*, préface G. Stamon, Paris, Masson, 1987.
- ZOLA, Émile, « Le chômage », in *Contes à Ninon*, Paris, Charpentier, 1879.

ANNEXE 1 : STRUCTURE SÉMANTIQUE DU FANTASTIQUE

