

Du vocabulaire spécifique à l'analyse stylistique : l'exemple de *Roderick Random*

Anne Bandry-Scubbi

EA 8211F Recherches sur le monde anglophone
Université de Strasbourg – France

Résumé

Cet article consiste en une étude du vocabulaire spécifique de *The Adventures of Roderick Random* (Smollett, 1748) par comparaison avec différents corpus de fiction britannique du XVIII^e siècle : l'ensemble des textes de fiction de Tobias Smollett, plusieurs romans des années 1740 et 1750 écrits par John Cleland, Henry Fielding, Sarah Fielding, Eliza Haywood, Charlotte Lennox et Samuel Richardson, ainsi que des autobiographies fictionnelles par Daniel Defoe, Jonathan Swift, Laurence Sterne. Plusieurs types de listes sont utilisées dans divers logiciels (*Hyperbase*, *Concordance*, *Tree-Tagger*, *Excel*) afin de confirmer ou d'infirmer une hypothèse stylistique : mettre en évidence ce qui dans l'écriture du premier roman de Smollett crée l'impression souvent notée par la critique d'un texte « énergique ».

Abstract

This paper provides an example of the use of textual imagery for stylistic analysis. It focuses on the specific vocabulary of *The Adventures of Roderick Random* (Smollett, 1748), by comparison within different corpora of eighteenth-century British fiction: Smollett's fictional works, novels from the 1740s and 1750s by John Cleland, Henry Fielding, Sarah Fielding, Eliza Haywood, Charlotte Lennox and Samuel Richardson, fictional autobiographies by Daniel Defoe, Jonathan Swift, Laurence Sterne. Several types of lists are used with different programmes (*Hyperbase*, *Concordance*, *Tree-Tagger*, *Excel*) to identify what creates the “stylistic energy” often noted by critics in Smollett's first novel.

Key-words: specific vocabulary, stylistics, discourse analysis, textual imagery, Smollett

1. Introduction

Après des analyses textuelles assistées par ordinateur de romans britanniques du dix-huitième siècle centrées sur les mots très fréquents dans le sillage de Burrows (Bandry, 1995 ; 2000 ; Bandry and Deconinck-Brossard, 1997), puis sur les spécificités selon la définition qu'en donnent Lebart et Salem (Bandry, 2001) et des listes pré-établies (Bandry-Scubbi and Deconinck-Brossard, 2005), le travail présenté ici utilise plusieurs types de listes afin de mettre en évidence certains traits stylistiques caractéristiques du premier roman de Tobias Smollett, *Roderick Random* (1748).

Plusieurs critiques ont formulé des hypothèses visant à expliquer la vigueur du style de Smollett, particulièrement dans *Roderick Random* (Stevick, 1967 ; Boucé, 1971 ; Grant, 1977). Datant des années 1970, ces travaux s'appuient sur l'analyse détaillée de courts passages du texte. Une approche quantitative permet de confirmer ou d'infirmer les généralisations qu'ils en tirent.

Dans ce but, plusieurs corpus de taille similaire ont été construits autour de ce texte, qui est ainsi comparé à dix-neuf autres romans. Le corpus 1740S est constitué de sept textes publiés dans les années 1740 et 1750, de longueur comparable (98000 à 224000 mots), écrits par des auteurs différents (trois femmes et quatre hommes), racontant les aventures de héros féminins ou masculins : *Pamela; or Virtue Rewarded* (Samuel Richardson, 1740), *The History of the Adventures of Joseph Andrews* (Henry Fielding, 1742), *The Adventures of David Simple* (Sarah Fielding, 1744), *The Adventures of Roderick Random* (Tobias Smollett, 1748), *Memoirs of a Woman of Pleasure* (John Cleland, 1748-49), *The History of Miss Betsy Thoughtless* (Eliza Haywood, 1751), *The Female Quixote or The Adventures of Arabella* (Charlotte Lennox, 1752). Afin de prendre en compte les contraintes liées à une narration à la première personne, un second corpus est composé de neuf autobiographies fictionnelles publiées entre 1719 et 1767 : 1stPERSON comprend *The Life and Adventures of Robinson Crusoe*, *The Farther Adventures of Robinson Crusoe*, *The Fortunes and Misfortunes of Moll Flanders*, *The History and Remarkable Life of the Truly Honourable Col. Jacque commonly call'd Col. Jack*, *Roxana*, *The Fortunate Mistress*, *Gulliver's Travels*, *The Adventures of Roderick Random*, *Memoirs of a Woman of Pleasure* et *The Life and Opinions of Tristram Shandy*. L'inconvénient d'avoir cinq textes sur neuf du même auteur (Daniel Defoe) est contrebalancé par l'avantage d'avoir un deuxième corpus de grande taille. L'ensemble des textes de fiction de Smollett constitue le corpus SMOLLETT, dans lequel les textes longs ont été divisés en deux pour éviter de trop grandes disparités de longueur. À ces trois corpus de plus d'un million de mots chacun s'ajoutent deux plus petits, constitués par *Roderick Random* : dans l'un le roman est divisé selon ses 69 chapitres et dans l'autre en grands épisodes ¹.

2. Une liste semi-automatique

S'inspirant de Stevick, Boucé évoque le « penchant [de Smollett] pour le style 'superlatif' » lors de l'analyse de passages précis (Boucé 1971 : 422). Une des caractéristiques de ce « style » consiste en « l'emploi fréquent d'intensifs », mais aucun des termes cités par le critique ne présente de fréquence relative élevée dans *Random* considéré comme une partie des corpus 1740s ou 1stPERSON. Une liste constituée d'intensifs, qu'ils soient « frustes » ou usuels (Boucé 1971 : 401, 405), montre même un déficit léger dans 1740s (14% des mots de la liste alors que le texte constitue 17% du corpus) ².

Ce « style superlatif » se manifeste aussi par le degré des adjectifs ³. Le lemmatiseur Tree-tagger permet de constater que *Random* comprend un peu moins d'adjectifs que les autres textes du corpus 1740S (5,74% des occurrences, la valeur la plus basse, alors que la valeur la plus forte de ce taux se trouve pour *Memoirs of Woman of Pleasure* avec 6,31%) ⁴. Si Tree-tagger répertorie les adjectifs superlatifs, il les identifie mal ⁵. En l'absence de résultats fiables établis par un lemmatiseur, il faut trouver d'autres moyens pour obtenir un décompte des superlatifs.

¹ Cet article est la version technique de l'analyse stylistique du roman de Smollett à destination de lecteurs spécialistes du dix-huitième siècle britannique (Bandry, 2009).

² Cette première liste n'a rien d'automatique mais est constituée de mots individuels entrés par la fonction LISTE d'Hyperbase.

³ Comme on le sait, le superlatif se forme en anglais par l'ajout du suffixe *-est* si l'adjectif est long et en le faisant précéder de *the most* ou *the least* s'il est court.

⁴ La richesse notoire du vocabulaire de ce texte est confirmée par le taux élevé de hapax legomena.

⁵ Certains des *most* répertoriés comme JJS par Tree-Tagger ne sont pas suivis d'un adjectif et tous les *most* précédés de *the* et suivis d'un adjectif ne sont pas répertoriés comme JJS.

On a donc établi une liste semi-automatique, à partir des termes se terminant en *-est* dans la base 1740S. Parmi les 262 vocables fournis par Hyperbase ont été gardés les 158 mots au superlatif. Le cumul de cette liste toilettée montre que 8% des adjectifs superlatifs en *-est* de 1740S se trouvent dans *Random* alors que ce texte constitue 17% du corpus. Le déficit est donc important.

Si l'on ajoute la proportion de ces superlatifs en *-est* à celle de *the most* et *the least*, le déficit de *Random* s'atténue mais demeure. La version d'Hyperbase dont je dispose ne pouvant prendre en compte les collocations ⁶, les données obtenues à partir d'une concordance de *most* et *least* ont été reportées dans un tableur et ajoutées à celles des superlatifs en *-est* pour *Roderick Random* et *Joseph Andrews* : on obtient une proportion de 0,14% des occurrences du texte de Smollett et 0,26% de celui de Fielding (respectivement 0,07% et 0,20% pour les seuls superlatifs en *-est*). Cette estimation de la proportion de superlatifs confirme le déficit de *Random*.

Le corpus constitué par *Random* divisé en ses 69 chapitres permet de constater que les variations sont faibles d'un chapitre à l'autre : aucun chapitre n'a un déficit significatif et seuls deux d'entre eux présentent un excédent de superlatifs en *-est* dépassant le seuil à 5%. L'examen détaillé du profil de ces deux chapitres consécutifs montre que les superlatifs y servent un but stylistique précis : les illusions perdues et la déchéance du poète Melopoyne racontées avec emphase à la première personne servent d'avertissement au héros (chapitres 63 et 64). La banalité des adjectifs employés par l'homme de lettres raté renforce la portée ironique de l'épisode (*best, highest, greatest*).

Observer un phénomène stylistique dans plusieurs corpus permet ainsi d'affiner l'analyse. Le repérage des parties de ce roman où les superlatifs sont plus fréquents amène à mettre à un jour l'un des procédés d'écriture de l'auteur pour un épisode précis. La comparaison de *Random* avec des textes contemporains y montre l'utilisation moindre d'une catégorie grammaticale, les superlatifs, contrairement à ce que laissaient attendre les études stylistiques précédentes. Il s'agit donc d'une illustration du principe de réalité philologique (Rastier, 2001 : 94).

3. Une liste pré-établie

Parmi les spécificités positives de *Random* dans les corpus 1740S et 1stPERSON figurent plusieurs pronoms et adverbes relatifs, notamment *which*, *who*, *whom* et *where*. Dans une approche alliant grammaire traditionnelle et linguistique de corpus, Biber fournit une liste de ces « WH-words » (Biber, 1988 : 223). Importée dans Hyperbase, une telle liste peut s'utiliser de différentes manières ⁷. On constate ainsi l'excédent des relatifs dans *Random* par rapport aux deux corpus de référence : 19% des occurrences dans 1stPERSON pour 16% du corpus, 19,5% dans 1740s pour 17% du corpus. Ces mots articulent les longues phrases complexes des premiers romans de Smollett. Leur répartition régulière dans *Random* montre qu'il s'agit d'un trait stylistique, sans lien avec le contexte d'épisodes particuliers, contrairement à l'utilisation des superlatifs par exemple : une analyse factorielle de cette liste dans la base où le roman est divisé en chapitres rassemble presque tous les mots au croisement des axes représentant les premier et deuxième facteurs. La fréquence décroissante de cette liste cumulée dans le corpus

⁶ La version 6.0 d'Hyperbase peut établir une concordance des « expressions » mais l'expérience a montré que cette liste de collocations n'est pas toujours complète. Il n'est en outre pas possible d'obtenir une représentation graphique de leur répartition dans le corpus.

⁷ Plusieurs listes sont fournies avec Hyperbase, mais elles concernent la langue française. Seule la liste PONCTUATION se transpose d'une langue à l'autre. Cependant la ponctuation instable des textes du dix-huitième siècle et sa transcription peu fiable dans les versions électroniques en rendent l'utilisation délicate.

SMOLLETT confirme la simplification de la syntaxe opérée au fil du temps par l'écrivain (détectée notamment par Boucé 393-94 — Fig. 1) :

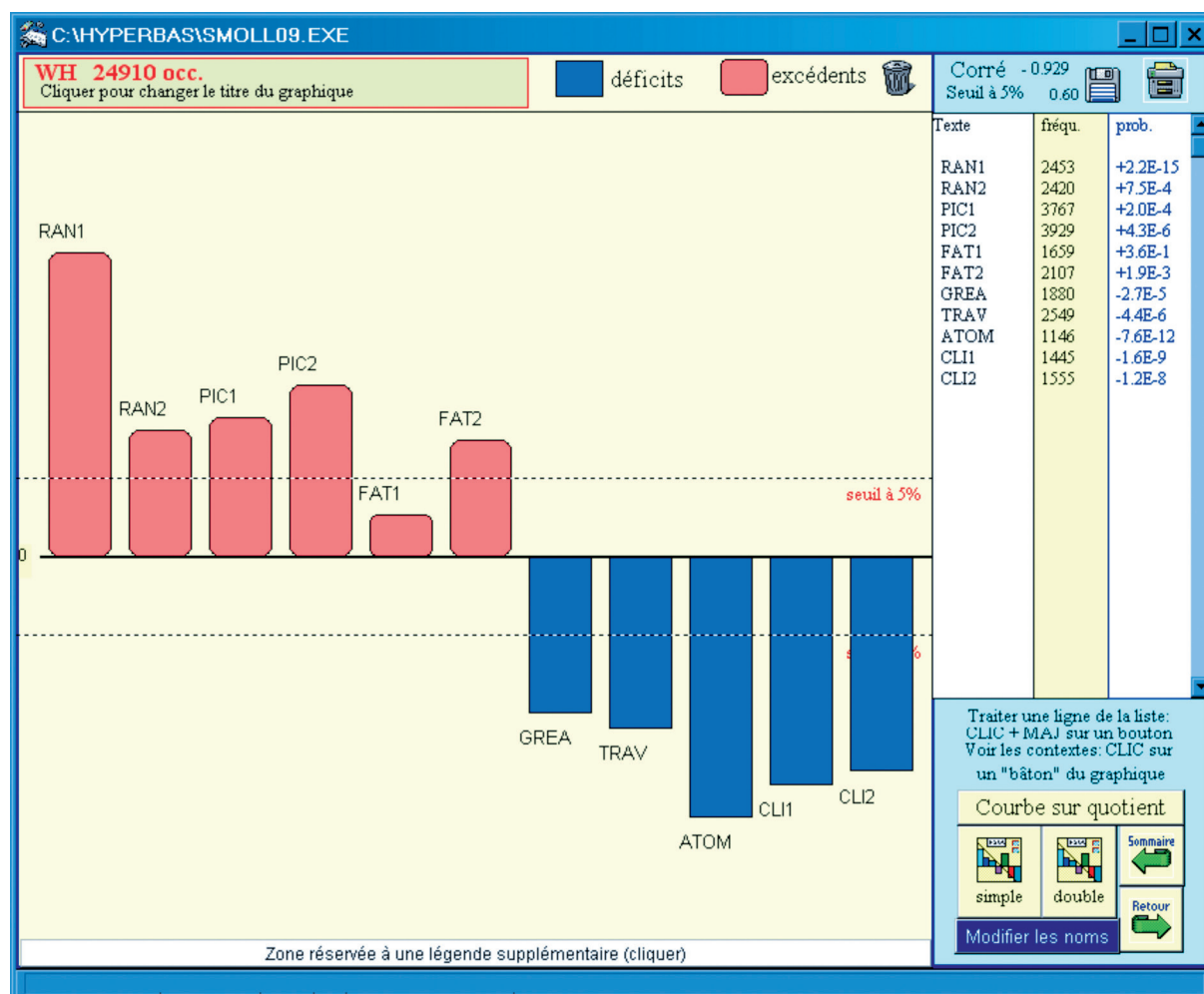


Figure 1 : Relatifs dans les textes de fiction de Smollett classés par ordre chronologique

La fréquence d'une liste cumulée peut être comparée à celle d'un mot très fréquent pour analyser un aspect stylistique. Contrairement à la proportion élevée des relatifs dans *Random*, celle de *and* est conforme à la moyenne du corpus 1740S. Cependant, si les fréquences de ce terme et celles de la liste WH- s'opposent dans la plupart des autres textes du corpus, elles se cumulent dans le roman de Smollett (Fig. 2).

Dans *Pamela*, Richardson pratique la juxtaposition aux dépens de la subordination, et donne ainsi un caractère réaliste aux lettres et au journal intime d'une servante dans lesquels une syntaxe complexe ne serait pas de mise.

Examiner l'utilisation d'un mot très fréquent dans plusieurs textes d'un auteur classés par ordre chronologique permet parfois de montrer à quel moment l'écrivain trouve sa cadence. J'avais pu établir que Sterne trouve le dosage de tirets qui semble le satisfaire à partir du deuxième tome de *Tristram Shandy*, l'utilisation intensive de ce signe de ponctuation constituant une sorte de signature de l'auteur (Bandry, 2006 : 47). L'évolution de la fréquence de *and* dans

les corpus RANDOM et SMOLLETT fait apparaître que le rythme d'utilisation de ce mot se stabilise à partir du deuxième tiers de *Random* ⁸.

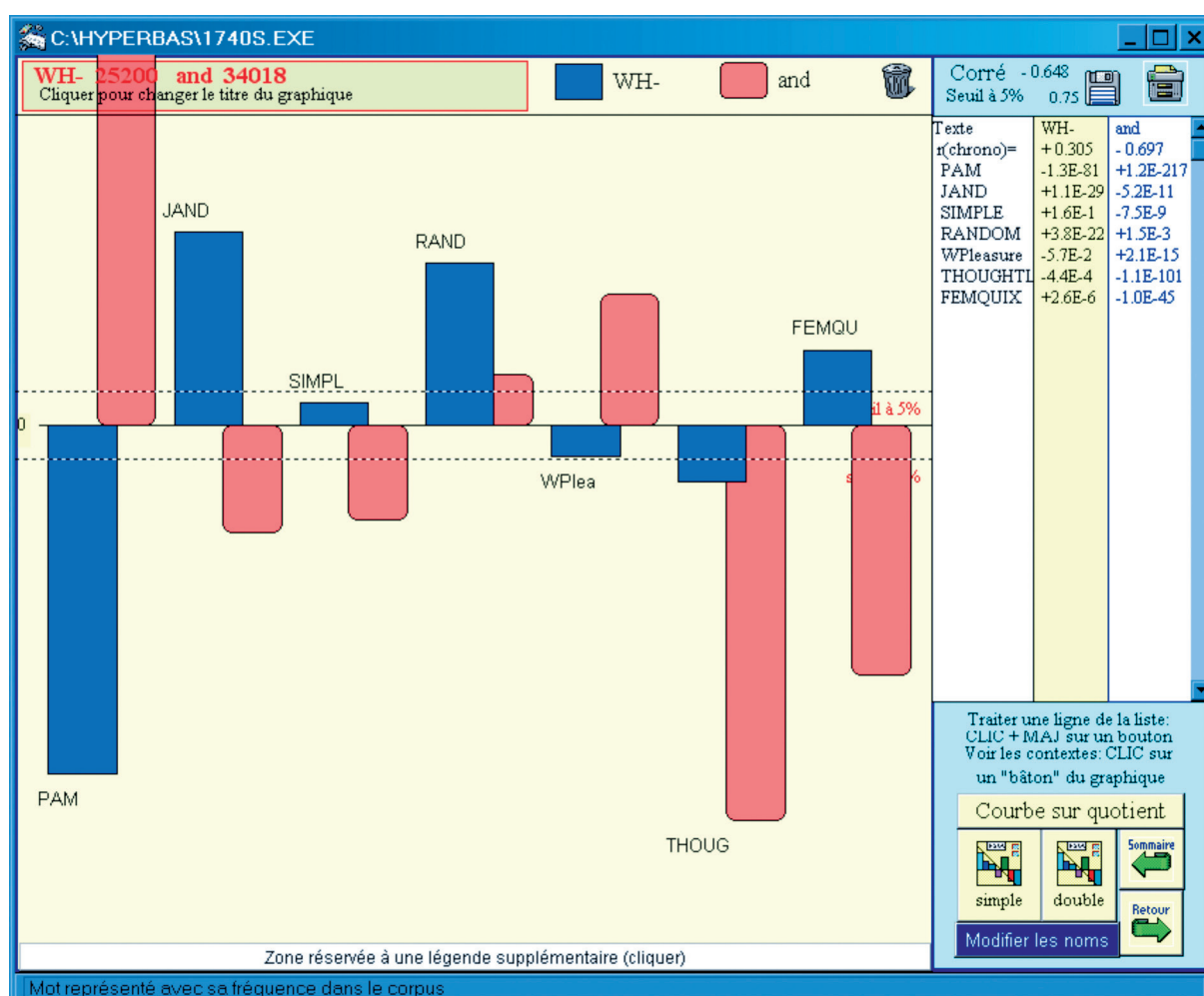


Figure 2 : Relatifs et and dans le corpus 1740S

4. Plusieurs listes construites autour d'une thématique : le temps

La vigueur de *Random*, l'impression d'être pris dans un tourbillon lors de la lecture du roman vient aussi, selon Stevick, d'un « écrasement du temps » (Stevick, 1967 : 714 ; Boucé, 1971 : 422). Contrairement aux intensifs et superlatifs examinés plus haut, toutes les expressions temporelles notées par les deux critiques sont en excédent par rapport au corpus 1740s : *no sooner than* (123 dans *Random*, 244 dans le corpus, soit 50%), *on/of a sudden* (13/38), *in a moment* (16/32), *in a trice* (5/12), *in a twinkling* (6/6), *on the spot* (4/7), *upon which* (116/167), *whereupon* (24/25). Plusieurs expressions temporelles figurent en outre dans la liste des spécificités positives de *Random* fournie par Hyperbase, dont l'adverbe *immediately* (32% des 575 occurrences de ce terme dans le corpus au lieu des 17% attendues) et l'adjectif *immediate* (12 des 43 utilisations soit 28% des occurrences).

⁸ Tous les 32 mots en moyenne, alors que Richardson y a recours tous les 24 mots dans *Pamela*.

Une liste de verbes, d'adjectifs et d'adverbes exprimant la vitesse permet de confirmer qu'il s'agit bien là d'un trait caractéristique de ce roman. Constituée à partir d'un dictionnaire de synonymes (*The Oxford Thesaurus*), après vérification de la pertinence des choix en parcourant la concordance de chaque terme dans le corpus, cette liste thématique confirme l'hypothèse. Les radicaux *hasten*, *hurry*, *immediate*, *precipit*, *quick*, *rapid*, *rush*, *speed*, *sudden*, *swift*, *veloc-* ont été retenus. Trois autres ont été éliminés en raison de leur signification dans le corpus ⁹. Toutes les formes prises individuellement ne sont pas en excédent, mais la liste agrégée, dénommée VITESSE, montre que cette thématique caractérise bien *Roderick Random* (Fig. 3) :

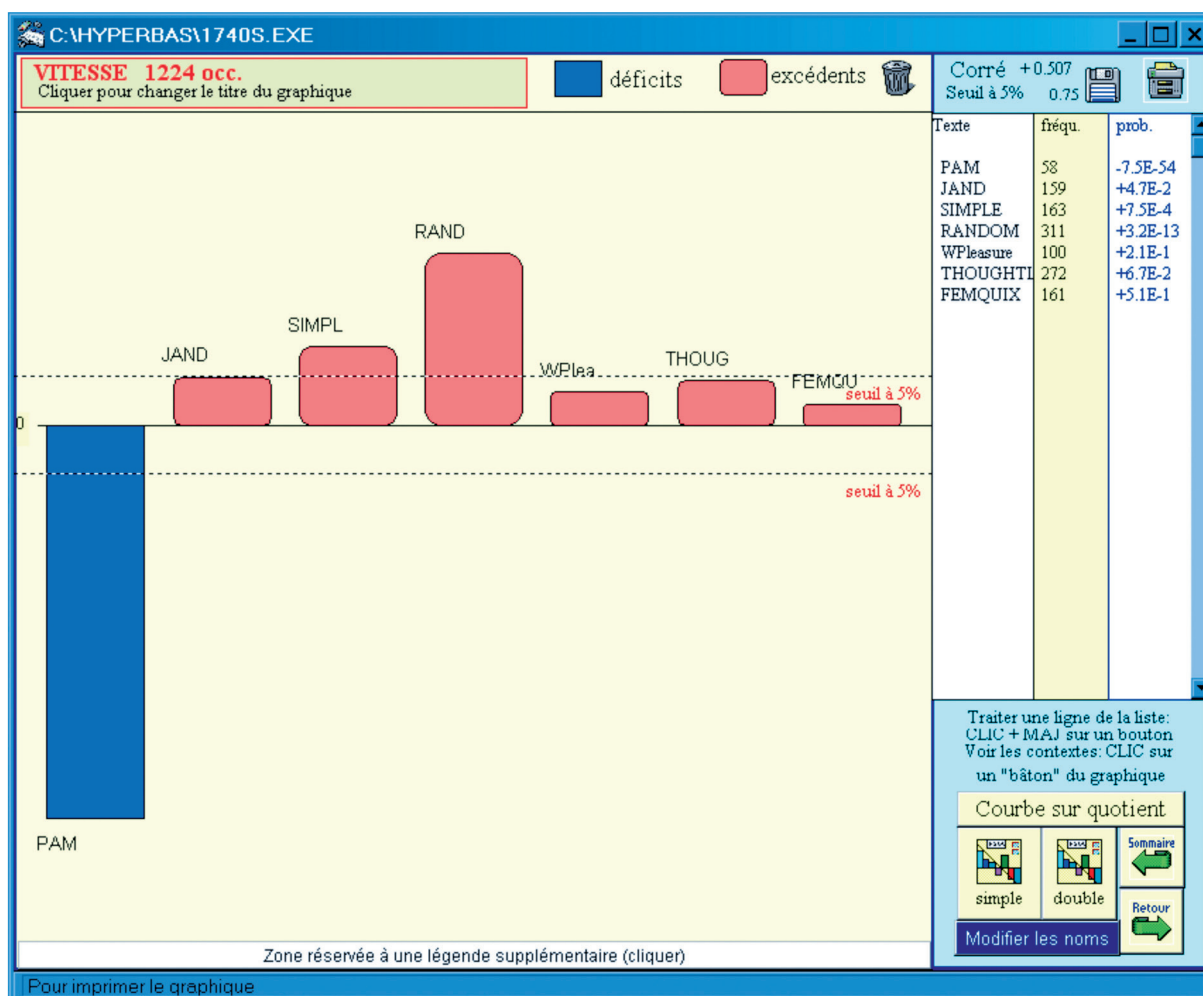


Figure 3 : Verbes, d'adjectifs et d'adverbes exprimant la vitesse dans le corpus 1740S

Si l'on combine la liste VITESSE avec les expressions temporelles analysées plus haut, l'on constate que *Random* comprend 35% de l'ensemble des mots et expressions de rapidité dans le corpus 1740s alors qu'il en constitue seulement 17% (618 des 1755 occurrences, dont 170 *immediately*). Les généralisations de Stevick et Boucé s'avèrent donc justifiées dans ce cas. Elles peuvent être développées.

⁹ *Dispatch* car il s'y trouve le plus souvent avec le sens d'«envoyer», *expedit-* parce qu'il désigne le plus souvent un voyage, une expédition et *fast* qui signifie «fermement» ou appartient à l'expression *fast asleep* («endormi»).

La présence d'adverbes de durée parmi les spécificités (*during, while*), ainsi que celle de *day* et d'*after*, amène à établir des listes d'indications temporelles plus précises : *year, month, week, day, hour, minute, seconds* et leurs dérivés, les parties de la journée (*morning, afternoon, evening, night*), l'indication des heures (*o'clock*). Les collocations classées par ordre de fréquence permettent de voir que *day* est le plus souvent précédé de *next* ('le jour suivant')¹⁰. On met ainsi en évidence l'utilisation plus marquée des divisions du temps et de leur succession dans le roman de Smollett que dans le reste du corpus et un intérêt moindre pour l'arrêt sur le présent (excédent pour *before* et *after*, déficit pour *now*, que préfèrent Pamela et Fanny Hill : ces héroïnes racontent leurs (més)aventures sexuelles dans une « écriture de l'instant »¹¹. La construction de plusieurs listes autour d'une même problématique permet donc de préciser la manière dont un auteur construit celle-ci.

5. Une liste unique construite pour une thématique : le corps

L'une des thématiques les plus évidentes de *Roderick Random* est la violence infligée aux corps : la renommée de roman vient en partie de la brutalité des premiers épisodes (la jeunesse en Écosse, le voyage et l'arrivée à Londres, l'engagement de force dans la marine et la cruauté de la vie en mer). Une liste des mots désignant les parties du corps a donc été établie à partir d'un inventaire de la physiologie humaine telle qu'elle me semblait apparaître dans la fiction britannique du dix-huitième siècle, complété par un relevé dans les index lexicométriques des différents textes des corpus 1740s et 1stPERSON. Afin d'étendre la comparaison à un plus grand nombre de textes, les fréquences relatives des mots du corps dans les deux bases 1740S et 1stPERSON ont été importées dans *Excel* (Fig. 4)¹² :

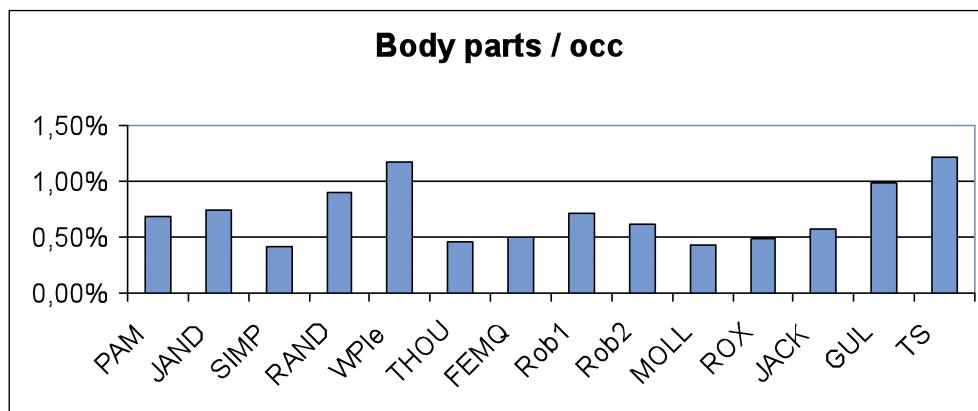


Figure 4 : Proportion des mots du corps dans le vocabulaire

L'examen des mots fréquents ou spécifiques liés à une thématique, associé à une réflexion sur la logique qui prévaut dans chaque texte, permet d'établir un profil des œuvres. Si 1740S est pris comme norme, *Random* comporte un tiers de mots du corps en plus que ne laisserait attendre sa longueur (22% des occurrences, 17% du corpus). Il se situe à la seconde place, derrière le

¹⁰ Le logiciel *Concordance* comporte cette fonctionnalité.

¹¹ « *Writing to the moment* », comme l'explique l'auteur de *Pamela* (Richardson, 1964 : 289).

¹² Ceci permet aussi de neutraliser l'effet grossissant des histogrammes d'*Hyperbase* dans lesquels « l'échelle des ordonnées est variable d'un graphique à l'autre [...] et tout l'espace du graphique est occupé, même si l'écart est faible en valeur absolue » (Brunet, 2006 : 32).

roman pornographique de Cleland, *Memoirs of Woman of Pleasure*, qui, de manière plutôt attendue, se réfère deux fois plus souvent au corps que les autres textes du corpus. Dans 1stPERSON, ces deux textes sont précédés par *Tristram Shandy* et *Gulliver*. Swift construit la satire sur la confrontation du corps de son héros à ceux d'êtres qui ne diffèrent de lui que par une caractéristique à la fois : la taille des lilliputiens ou des brobdnagiens, la nature des Houyhnhnms, chevaux rationnels, et des yahoos humanoïdes (Bandry, 2001). Sterne joue sur les ambiguïtés de *nose* pour ébranler le langage, puis sur les palpitations des cœurs et les contacts entre les mains pour s'amuser du sentimentalisme tout en y contribuant (Bandry, 2000). Smollett atomise le corps de ses personnages pour raconter coups, chocs et sensations, et accroît l'efficacité narrative de ces dernières en dissociant les perceptions sensorielles (Grant, 1977 : 115).

La liste des mots du corps appliquée à *Roderick Random* divisé en grands épisodes réserve une surprise. Si l'on s'attend à ce que la partie consacrée à la marine (NAVY) comporte le plus grand nombre de mots du corps (21,5% pour 17% du corpus), il est étonnant de premier abord que l'épisode de la prison (JAIL) présente un fort déficit (3% des mots de la liste pour 7% du corpus — Fig. 5) :

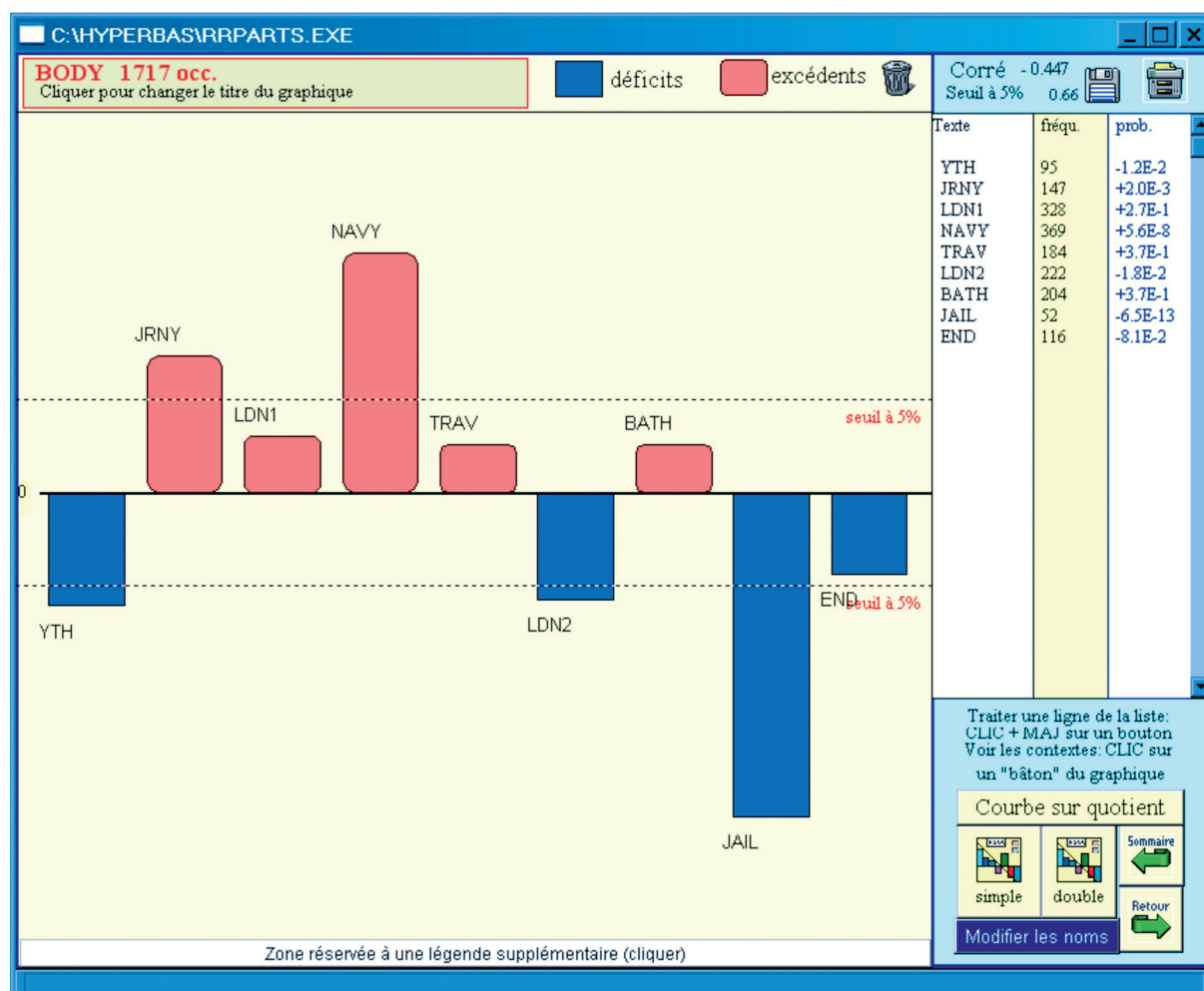


Figure 5 : Répartition des mots du corps (liste agrégée) dans Random divisé en grands épisodes

Les mots spécifiques à cette partie mènent à une explication, en faisant apparaître que c'est le corps du poète déchu qui prévaut et notamment sa barbe. L'observation de celui-ci fait mûrir le

héros, le préparant pour le dénouement qui en devient moins arbitraire ¹³. On a là un exemple où les données quantitatives font surgir une interprétation du texte.

La représentation des deux premiers facteurs d'une analyse factorielle des correspondances met en lumière une évolution dans le choix des mots du corps. Parmi les termes de fréquence élevée et selon le premier facteur, *hands* et *head* sont dans la même zone du graphique que les premières parties du texte où le héros doit frayer son chemin dans un monde hostile et brutal (la jeunesse, le voyage et la marine) tandis que *heart* et *eyes*, termes les plus fréquents dans les récits centrés sur des personnages féminins, se situent à proximité des derniers épisodes (ceux qui se déroulent dans la ville de Bath et les chapitres de la fin), dans lesquels le héros doit comprendre la logique du monde féminin et de la séduction pour mériter sa belle (Fig. 6) ¹⁴ :

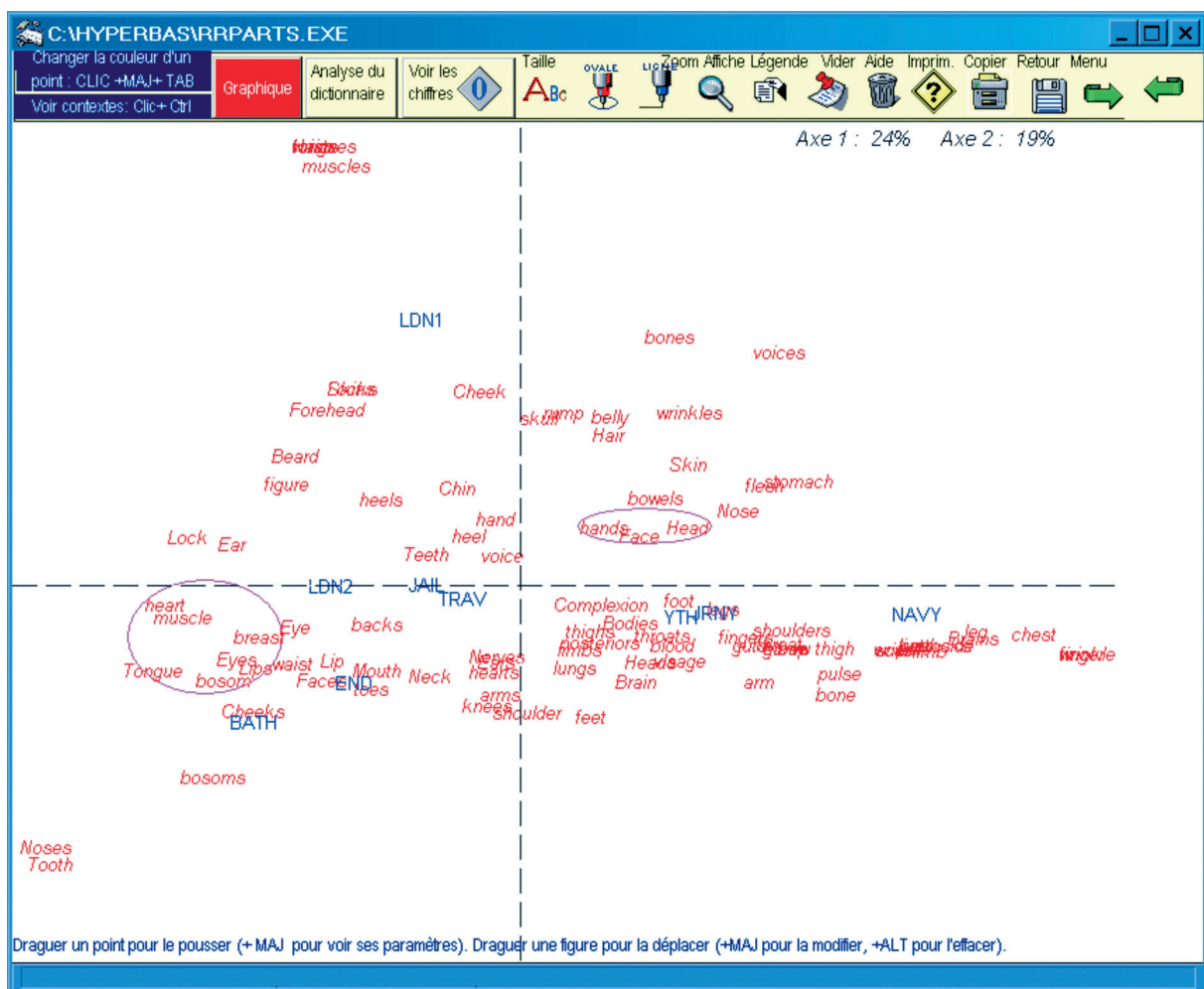


Figure 6 : Analyse factorielle des mots du corps dans Random divisé en grands épisodes

¹³ Smollett termine son roman par une série de coïncidences heureuses. Dans l'ensemble de ses textes de fiction, il attribue une barbe à plusieurs de ses héros à l'esprit dérangé, alors que le fidèle compagnon de Roderick, qui le sauve de la misère à plusieurs reprises, est barbier.

¹⁴ Les récits centrés sur des personnages féminins considérés dans cette étude sont Pamela, *Memoirs of a Woman of Pleasure*, *The History of Miss Betsy Thoughtless*, *The Female Quixote*.

6. Les collocations d'un terme spécifique : *my*

Comme *Roderick Random* est un roman autodiégétique, il a paru fructueux d'examiner l'utilisation des marques grammaticales de l'énonciation dans le corpus 1stPERSON. La fréquence relative de l'ensemble des pronoms personnels et adjectifs possessifs de la première personne du singulier dans *Random* ne présente ni excédent ni déficit. Cependant, le mot en tête de la liste des spécificités de *Roderick Random* dans le corpus 1740S est *my*, l'adjectif possessif de la première personne du singulier. Ce qui différencie ce roman des autres est l'utilisation que fait le héros du possessif de la première personne du singulier pour s'intéresser moins à lui-même qu'au monde tel qu'il s'organise en relation avec lui. L'examen des collocations de ce terme classées par ordre de fréquence décroissant met en évidence la manière dont Smollett organise le monde fictionnel autour de son héros.

Trois catégories se dessinent. Nombre de termes à la droite de *my* sont des mots du corps : d'abord abstraits comme *my heart* (47), *my soul*, *my mind*, puis plus concrets avec *my hand* (30), *my face* (21), *my blood* (20), *eyes*, *head*, *breast*, *back*, *bosom*, *neck*, *senses*, *tongue* (16 à 8 occurrences), *arms* (7 occurrences, différenciées de celles signifiant 'armes' par l'examen d'une concordance), et de manière plus dramatique, *my wounds* ('mes blessures', 9 fois). Il faut garder à l'esprit que le héros n'étant pas le seul à parler à la première personne, ces parties du corps ne lui appartiennent pas toujours. Les collocations montrent que les yeux et les mains sont plus souvent ceux d'un autre personnage masculin (*his* à gauche), mais que ces organes sont en général dirigés vers ou contre Roderick. L'examen du voisinage de ces mots confirme l'interprétation selon laquelle, dans ce roman, le corps n'apprend à fonctionner en société qu'après avoir pris des chocs et des coups (Douglas, 1995 : 49).

Une deuxième catégorie de termes précédés par *my* désigne l'état émotionnel, physique ou économique dans lequel se trouve celui ou celle qui les utilise : *my passion* (28 occurrences), *my situation* (27), *my fortune* (26) or *my misfortune* (18), *my business* (21), *my appearance* (18), *my condition* (16), *my fate* (14), *my happiness* (13). Certains des personnages, dont le héros, ont donc du recul par rapport à leur situation, argument qui tend à discréditer la thèse selon laquelle *Roderick Random* est un roman picaresque. Un examen détaillé des collocations peut donc permettre de débattre de la nature d'une œuvre ¹⁵.

Mais la catégorie la plus fréquente des termes précédés par *my* est constituée de noms de personnes : *my uncle* (79 occurrences), *my friend* (69) et *my friends* (18), *my father* (68 des 125 occurrences de ce substantif alors que le père de Roderick disparaît du chapitre 2 au chapitre 66) ¹⁶, *my grandfather* (32 des 36 occurrences de *grandfather*, regroupées là où on s'attend à les trouver, en début de récit), *my companion* (32), *my acquaintance* (15), *my dear* (29) mais aussi *my rival* (11). *My mistress* arrive en tête des désignations féminines (50 des 79 occurrences du nom) ; le substantif recouvre plusieurs statuts et personnages puisque Roderick se fait serviteur avant de tomber amoureux et ne limite pas ses visées matrimoniales à son premier amour. Comme Tristram Shandy (le second utilisateur de *my* après *Random* dans le corpus 1stPERSON), mais moins confortablement que lui, le héros de Smollett vit dans un monde peuplé de personnages. Dans les deux cas, le discours du héros-narrateur les relie à lui. Contrairement à Robinson (aussi un grand utilisateur de *my*, pas seulement sur son île), Roderick n'est pas un solitaire. Le fait que *his et with* se situent à un rang plus élevé que *my*

¹⁵ Il s'agit là d'une des questions les plus posées par la critique au sujet du roman de Smollett.

¹⁶ La concentration d'occurrences de *my* dans un chapitre particulier attire l'attention sur l'utilisation de cette collocation par un personnage féminin, Miss Williams.

dans la liste des spécificités de *Random* pour le corpus 1stPERSON montre bien que ce roman s'intéresse avant tout aux interactions du héros avec d'autres personnages. Un léger excédent du pluriel de la première personne confirme cette analyse ¹⁷.

En revanche, Roderick utilise le pronom sujet *I* nettement moins fréquemment que d'autres héros autodiégétiques. Une étude déjà ancienne avait mis en évidence la présence inhabituelle de ce terme en tête de la liste lexicométrique de *Moll Flanders* (Bandry and Deconinck-Brossard, 1997). Les personnages de Defoe ou l'héroïne de Richardson y ont recours beaucoup plus souvent que Roderick (2,5% des occurrences de l'histoire qu'il raconte sont des *I*, comme Gulliver, proportion qui monte à 4% ou plus pour Moll, Robinson et Pamela), mais sa fréquence d'utilisation dépasse celle de Tristram Shandy et de Fanny Hill (*Memoirs of a Woman of Pleasure* s'intéresse au moins autant aux partenaires de l'héroïne qu'à celle-ci). Les imitations d'un texte montrent combien il est difficile de percevoir la fréquence à laquelle un terme est utilisé : les plumitifs qui ont tenté de se faire passer pour l'auteur de *Tristram Shandy* ont eu recours au pronom trop fréquemment, alors que les continuateurs du *Voyage Sentimental* ne s'en sont pas assez servi (Bandry, 2000 : 134).

7. Imagerie textuelle

En croisant les requêtes et les données obtenues à partir de plusieurs corpus, en les considérant à l'échelle de textes entiers ou en les analysant en détail, on parvient à mettre en évidence certaines des lignes de force d'une œuvre, dans une démarche comparable à celle de l'imagerie médicale. Pour qu'une analyse textuelle assistée par ordinateur soit fructueuse, une connaissance approfondie du texte est indispensable (comme le médecin a besoin de connaissances anatomiques précises), afin d'être en mesure d'opérer plusieurs mouvements de va et vient entre le texte et les données fournies par l'ordinateur, de confronter les données quantitatives et l'observation. On peut alors estimer que l'utilisation de l'imagerie textuelle apporte une plus grande objectivité aux techniques de l'analyse textuelle traditionnelle.

Références

- Bandry-Scubbi A. (1995). Une étude stylostatistique de Sterne : méthodologie. In *Gestion Électronique de documents et nouvelles technologies en sciences humaines*. Paris : PU Sorbonne, pp. 79-99.
- Bandry-Scubbi A. (2000). Les livres de Sterne : suites et fins. *BSEAA XVII-XVIII*, 50 : 115-136.
- Bandry-Scubbi A. (2001). Gulliver et la machine à compter : une étude de spécificités. *BSEAA XVII-XVIII*, 53 : 145-57.
- Bandry-Scubbi A. (2006). *Tristram Shandy* à mots comptés. *RSEAA XVII-XVIII*, 63 : 41-61.
- Bandry-Scubbi A. (2009). *Roderick Random* amidst Eighteenth-Century Fiction: A Computer-aided Textual Analysis. *RSEAA XVII-XVIII*, 66 : 205-25.
- Bandry-Scubbi A. and Deconinck-Brossard F. (1997). On peut compter sur Moll [Flanders]. *BSEAA XVII-XVIII*, 45 : 171-90.
- Bandry-Scubbi A. and Deconinck-Brossard F. (2005). De la lexicométrie à la stylostatistique? Sterne et Swift : textes croisés. *Bulletin de Stylistique anglaise*, 26 : 67-85.
- Biber D. (1988). *Variation across Speech and Writing*. Cambridge : Cambridge University Press.

¹⁷ Defoe aime à raconter des histoires de communautés masculines : les plus fortes concentrations du pluriel de la première personne se trouvent dans *Colonel Jack* et surtout dans *The Further Adventures of Robinson Crusoe*.

- Boucé P.-G. (1971). *Les romans de Smollett*. Paris-Bruxelles-Montréal : Didier Érudition.
- Brunet E. (2006). *Hyperbase : Manuel de référence. Version 6.0 Windows*. Institut de linguistique française. Nice.
- Burrows J. F. (1987). *Computation into Criticism: A Study of Jane Austen's Novels and an Experimental Method*. Oxford: Clarendon Press.
- Douglas A. (1995). *Uneasy Sensations: Smollett and the Body*. University of Chicago Press.
- Grant, D. (1977). *Tobias Smollett: A Study in Style*. Manchester UP : Roman and Littlefield.
- Lebart L. and Salem A. (1994). *Statistique textuelle*. Paris : Dunod.
- Rastier F. (2001). *Arts et Sciences du texte*. Paris : PUF.
- Richardson S. (1964). *Selected Letters of Samuel Richardson*. Oxford : Oxford Clarendon Press.
- Stevick Ph. (1967). Stylistic Energy in the Early Smollett. *Studies in Philology*, 64, 5 : 712-19.

Logiciels

- Brunet E. *Hyperbase: Logiciel hypertexte pour le traitement documentaire et statistique des corpus textuels*, version Windows 6.0, INALF, 2006.
- Microsoft Office Excel 2003. Microsoft Corporation.
- Schmid H. *TreeTagger*. Institute for Computational Linguistics of the University of Stuttgart.
<http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/DecisionTreeTagger.html>.
- Watt R.J.C. *Concordance*. Version 3.3, 2009. <<http://www.concordancesoftware.co.uk/>>.

Sources primaires

- Cleland J. (1748-49). *Memoirs of a Woman of Pleasure*. Wiretap Electronic Text Archive. 29/10/2000.
 <<http://infomotions.com/etexts/literature/english/1700-1799/cleland-fanny-368.htm>>.
- Defoe D. (1719a). *The Life and Adventures of Robinson Crusoe*. Project Gutenberg E-text #521, 1996.
 11/01/2001.
 <<ftp://ftp.mirror.ac.uk/sites/metabolab.unc.edu/pub/docs/books/gutenberg/etext96/rbcru10.txt>>
- Defoe D. (1719b). *The Farther Adventures of Robinson Crusoe*. Project Gutenberg E-text #561, 1996.
 11/01/2001.
 <<ftp://ftp.mirror.ac.uk/sites/metabolab.unc.edu/pub/docs/books/gutenberg/etext96/rbcr210.txt>>.
- Defoe D. (1722a). *The History and Remarkable Life of Col. Jack*. *Eighteenth-Century Fiction Full-Text Database*, Chadwyck-Healey, 1996.
- Defoe D. (1722b). *The Fortunes and Misfortunes of the Famous Moll Flanders*. Project Gutenberg E-text #370, 1995. 12/05/1997.
 <<ftp://ftp.mirror.ac.uk/sites/metabolab.unc.edu/pub/docs/books/gutenberg/etext95/ollf10.txt>>.
- Defoe D. (1724). *Roxana, The Fortunate Mistress*. Project Gutenberg E-text #30344, 2009. 29/11/2009.
 <<http://www.gutenberg.org/files/30344/30344-8.txt>>.
- Fielding H. (1742). *Joseph Andrews [Electronic resource] / Henry Fielding*; edited by Martin C. Battestin. Oxford Text Archive. 12/03/2000. <<http://ota.ahds.ac.uk/>>.
- Fielding S. (1744). *The Adventures of David Simple*. Blackmask Online, 2002. 05/08/2009.
 <http://www.munseys.com/diskone/davidsimp.htm#1_1>.
- Haywood E. (1751). *The History of Miss Betsy Thoughtless*. *Eighteenth-Century Fiction Full-Text Database*, Chadwyck-Healey, 1996.
- Lenox C. (1752). *The Female Quixote or The Adventures of Arabella*; Girlebooks. 05/08/2009.
 <<http://girlebooks.com/ebook-catalog/charlotte-lennox/the-female-quixote/>>.

- Richardson S. (1740). *Pamela; or Virtue Rewarded*. *Eighteenth-Century Fiction Full-Text Database*, Chadwyck-Healey, 1996.
- Sterne L. (1759-67). *The Life and Opinions of Tristram Shandy Gentleman*. Edition électronique privée obtenue par Diana Patterson (1989-90).
- Smollett T. (1748). *The Adventures of Roderick Random*. Project Gutenberg E-text #4085, 2003. 04/06/2009. <<http://www.gutenberg.org/dirs/etext03/8rran10.txt>>.
- Smollett T. (1751). *The Adventures of Peregrine Pickle*. Project Gutenberg E-text #4084 (Text of the second edition, 1758), 2003. 14/08/2009. <<http://www.gutenberg.org/dirs/etext03/thdvn10.txt>>.
- Smollett T. (1753). *The Adventures of Ferdinand Count Fathom*. Project Gutenberg E-text #6761, 2006. 14/08/2009. <<http://www.gutenberg.org/files/6761/6761.txt>>.
- Smollett T. (1760-61). *The Adventures of Sir Launcelot Greaves*. Project Gutenberg E-text #6758. 2006. 14/08/2009. <<http://www.gutenberg.org/files/6758/6758.txt>>.
- Smollett T. (1766). *Travels through France and Italy*. Project Gutenberg E-text #2311. 2000. 14/08/2009. <<http://www.gutenberg.org/dirs/etext00/ttfai10.txt>>.
- Smollett T. (1769). *The History and Adventures of an Atom*. *Eighteenth-Century Fiction Full-Text Database*. Chadwyck-Healey. 1996. 03/03/2001.
- Smollett T. (1771). *The Expedition of Humphry Clinker*. Project Gutenberg E-text #2160. 2000. 14/08/2009. <<http://www.gutenberg.org/dirs/etext00/txohc10.txt>>.
- Swift J. (1726-35). *Gulliver's Travels*. *Gulliver's Travels by Jonathan Swift, based on the 1735 Faulkner edition*. 29/11/2009 <<http://www.jaffebros.com/lee/gulliver/ascii/gulliver.txt>>.

