

Marguerite Duras vue par la statistique lexicale
ATILF _ 4 février 2005
Véronique Montémont

1. Méthodologie

- 1.1. La structure de la langue
 - 1.1.1 Les flexions
 - 1.1.2. Polysémie
 - 1.1.3. Problèmes divers
- 1.2 La relativité : en fonction de quoi un corpus lexical prend-il son sens ?
 - 1.2.1. L'ensemble de l'ouvrage
 - 1.2.2 Les autres œuvres de l'auteur
 - 1.2.3 Les œuvres d'autres auteurs

2. Résultats

- 2.1 Les fréquences
- 2.2. La ponctuation
- 2.3. Le voisinage
- 2.4. La catégorisation
- 2.5. L'interrogation par expression

3. Problèmes

- 3.1. Les flexions
- 3.2. Le problème des mots-outils
- 3.3. Les registres de langue

4. Améliorations possibles

- 4.1. Interface graphique
- 4.2. Exportation des résultats
- 4.3. Reconnaissance des titres

C'est à titre d'amateur que je vais effectuer cette présentation : amateur dans la mesure où je ne suis ni linguiste, ni informaticienne, ni lexicographe. Mon projet aujourd'hui serait de vous raconter, en quelque sorte, les aventures d'une littéraire qui décide de se lancer dans une recherche avec l'outil Frantext, et la façon dont la manipulation de cet outil peut être perçue, justement, par un utilisateur littéraire qui ne possède pas d'expertise particulière quant au traitement informatisé de la langue. J'ai commencé à utiliser la statistique lexicale, à l'occasion de ma thèse. Je travaillais en effet sur 3 poètes différents, pour un total de 19 recueils, et je cherchais à dégager des récurrences pour corroborer des analyses thématiques, voire en ouvrir d'autres. J'ai donc numérisé tous les recueils, transféré leur contenu sous Word, ramené les textes à des listes, que j'ai à nouveau transférées sous Excel. J'ai ensuite affecté, avec un système de tableaux croisés, à chaque mot son nombre d'occurrences, puis rétrié les résultats avec la fonction tri d'Excel. Cette douloureuse expérience, qui a occupé un nombre non négligeable de mois de ma vie, me permet d'apprécier aujourd'hui d'autant mieux Frantext, qui met moins d'une minute pour extraire la totalité du vocabulaire d'un volume. Elle m'a aussi rendue sensible au rôle déterminant que peut avoir, dans une recherche, la statistique lexicale, notamment lorsque l'on souhaite isoler certaines isotopies particulières (la couleur, le métalangage, la spatialisation, etc...) et en faire une étude

spécifique. Je me propose aujourd'hui d'analyser à la lumière de cet outil trois romans de Marguerite Duras : dans un premier temps, j'exposerai la méthodologie utilisée pour dégager une liste de fréquences. J'en effectuerai ensuite une analyse partielle, avant d'exposer les problèmes rencontrés et de proposer quelques corrections.

Marguerite Duras naît en 1914 à Gia Dinh, en Indochine, dans une famille d'enseignants. Son père, très malade, meurt alors qu'elle est âgée de sept ans. Sa mère, qui a tenté de revenir en France, doit retourner à Saigon avec ses deux fils et sa fille en 1922. La famille vit dans une sorte d'entre-deux : pas assez riche pour être intégrée à la bourgeoisie coloniale, mais pas aussi pauvre que les autochtones. La mère, décidée à s'enrichir, achète avec ses économies une concession au Cambodge et en prend possession en 1928. Mais elle ignore qu'il faut, pour obtenir une bonne terre, soudoyer les employés du cadastre. Elle reçoit donc une concession inondée par la mer et incultivable. Cette expérience marque profondément Duras tout comme sa relation avec ses deux frères, l'un haï et l'autre adulé, dont le personnage de Joseph, du *Barrage*, est la synthèse. Marguerite Duras vient en France à l'âge de 18 ans et entreprend des études de mathématiques, de droit et sciences politiques, avant de travailler au Ministère des Colonies. En 1939, elle épouse Robert Antelme. Lorsque la guerre survient, tous deux s'engagent dans la Résistance ; Antelme est déporté à Buchenwald en 1944, ce que Duras raconte dans *La Douleur. Un barrage contre le Pacifique* est publié en 1950 : roman cru, âpre, désenchanté, qui dénonce la corruption de l'administration coloniale. Le livre est bien accueilli, mais manque le prix Goncourt : Marguerite Duras est persuadée que des raisons politiques — elle dit le livre « communiste » — sont la raison de cet échec. Elle continue à écrire et publie en 1958 *Moderato Cantabile*, roman bref, qui raconte comment une bourgeoise, Anne Desbarèdes, fascinée par un fait divers, vient boire au café avec un homme qu'elle connaît à peine, Chauvin. Le style s'affirme : des phrases brèves, nombre de répétitions, une syntaxe volontiers obsessionnelle. Les thématiques aussi : la confrontation des classes sociales, l'ennui, l'alcool. L'auteur continue à explorer un imaginaire de plus en plus ouvertement tourné vers la folie, la béance existentielle, la féminité, en particulier avec la trilogie qui regroupe *Le Ravissement de Lol V. Stein* (1964), *Le Vice-Consul* (1968) et *L'Amour* (1971), où apparaît le personnage récurrent d'Anne-Marie Strette. A partir de 1975, Marguerite Duras se lance dans la réalisation cinématographique (*India Song*, *Le Camion*) : son œuvre, assez confidentielle, a ses admirateurs, mais ne rencontre pas le grand public. Parallèlement à cela, elle mène une vie privée quelque peu tumultueuse et surtout, boit beaucoup. En 1984, elle publie aux éditions de Minuit *L'Amant*, roman qui puise dans la même inspiration coloniale et autobiographique que le *Barrage*. Mais cette fois, elle va plus loin dans la confiance et raconte sa liaison avec celui qu'elle appelle « Le Chinois ». Le roman est couronné par le Goncourt et connaît un immense succès populaire, faisant d'ailleurs l'objet d'une adaptation cinématographique. A partir de là, Duras, courtisée à l'envi par les médias, devient une sorte d'icône et joue largement du personnage qu'elle s'est construit, jusqu'à s'y enfermer. Les dernières années de sa vie sont assombries par l'alcool, les désintoxications et les rechutes, et son écriture, de plus en plus obsessionnelle, est fortement marquée par des tics sémantiques et syntaxiques. Marguerite Duras meurt en 1996 à l'âge de 82 ans.

1. Méthodologie

Frantext permet plusieurs types d'interrogations. La première démarche consiste à extraire la liste du vocabulaire et ses différences fréquences, qui peuvent être livrées par ordre alphabétique, croissant ou décroissant. Ces données constituent une ressource précieuse, mais nécessitent une méthodologie d'utilisation.

1.1. La structure de la langue

1.1.1 Les flexions

La première difficulté et la première limite d'une base de données lexicales tient au fonctionnement de la langue elle-même. La nature flexionnelle de celle-ci l'amène à utiliser un nombre n de formes, en fonction du contexte grammatical : genre, nombre, conjugaison. Ainsi, on va trouver dans la liste un certain nombre de forme fléchies, parfois nombreuses. Le cas simple est celui de l'adjectif : prenons l'exemple de *vert*. Le mot apparaît sous quatre graphies différentes, mais regroupe en réalité 18 occurrences :

vert	9
verte	2
vertes	5
verts	2

Evidemment, le fait d'opérer l'addition fait instantanément remonter le mot dans la liste des fréquences. Le cas est encore plus net en ce qui concerne les verbes, dispersés parfois sur quinze ou vingt formes, comme c'est le cas pour *voir* ou *aller*. Frantext peut rassembler ces formes grâce aux commandes **&c** et **&m**, qui permettent de regrouper les flexions du substantif, de l'adjectif et les formes conjuguées des verbes. L'avantage de ces fonctions, par rapport à une réduction manuelle, est également qu'elles autorisent le regroupement de formes parfois dispersées dans la liste alphabétique. Autant il est facile d'additionner les occurrences de mots qui se suivent, autant les risques d'erreur sont grands lorsque les radicaux sont phonétiquement éloignés. Par exemple, le verbe *voir* se retrouve à **ve-** (*verrions*), **vo-** (*voyions*), **vu** : dans *BCP*, les formes **verra** et **vu** sont distantes de 176 lignes. La commande de Frantext est encore plus utile dans le cas des verbes qui ont plusieurs bases, comme *aller* ou *avoir*, ou des substantifs dont le singulier et le pluriel sont deux mots différents (*œil/yeux*).

Mais cet outil pose des problèmes lorsqu'il existe des chevauchements entre les différentes formes. Par exemple, *suis* peut être rapporté soit au verbe *être*, soit au verbe *suivre*. Il faut alors demander à Frantext l'affichage des extraits. Dans *BCP*, la forme connaît 77 occurrences : l'examen exhaustif de la liste révèle qu'aucune forme ne se rapporte à *suivre*. Si l'on ne souhaite pas, pour chaque forme polyvalente, se livrer à cet examen exhaustif, on peut essayer de tourner la difficulté en prenant en compte le mode de fonctionnement syntaxique du terme, ce qui oblige cependant à procéder par la méthode des essais et des erreurs. Ainsi *vit* est le passé simple du verbe *voir* et présent du verbe *vivre*, qui revient 21 fois. Si l'on considère que *voir* se construit la plupart du temps avec un COD, on peut essayer d'interroger *vit* + déterminant, avec la formule suivante :

vit (unluneldeslallalleslcelcetlcttelceslsonlsalses)

Le résultat dégage six occurrences, dont voici deux exemples. Nous donnerons ceux-ci dans la présentation écran, de manière à ce que le lecteur se fasse une idée de la manière dont on peut visualiser les résultats d'une requête :

S267/ DURAS.M / UN BARRAGE CONTRE LE PACIFIQUE / 1950 page 80 / <i>PREMIÈRE PARTIE</i> Le caporal monta, vit le paquet, le regarda longuement, puis posa le riz sur la table et
--

commença à mettre le couvert.

[Retour en haut du document](#) Résultat 2 (Texte sous droits)

S267/ **DURAS.M** / UN BARRAGE CONTRE LE PACIFIQUE / 1950

page 128 / *PREMIÈRE PARTIE*

Quand Joseph

fut tout près, Suzanne tendit la main et au-dessus

des doigts tendus, Joseph **vit le** diamant.

[Retour en haut du document](#) Résultat 3 (Texte sous droits)

Mais en réalité, l'interrogation est inopérante, car elle n'englobe ni **vit + nom propre**, ni **vit + adverbe ou infinitif** : « Elle courut sur la véranda et **vit** descendre Joseph de la voiture », « Dès qu'il la **vit** mieux », « quand elle **vit** Suzanne apparaître ». Elle ne prend pas non plus en compte l'accusatif d'objet interne du type **vivre sa vie**. Mieux vaut donc en passer par l'examen exhaustif, qui révèle que toutes les occurrences se rapportent au verbe voir.

Le problème se pose dans des termes quelque peu différents lorsqu'un même mot appartient à deux classes grammaticales à la fois. Par exemple, le mot *été*, participe passé du verbe être et substantif qui désigne une saison, ressortit à une double classe. *Moderato cantabile* comporte 23 fois cette forme. Si on ne souhaite pas les examiner une par une, on peut chercher les emplois substantifs grâce à la commande **(unll')été**, qui fait ressortir six occurrences. Cette fois, le risque d'erreur est limité, car cette syntaxe n'est pas possible dans le cas du participe. Le résultat livre 6 occurrences, dont voici deux exemples.

L'été, les ouvriers de l'arsenal commencent à passer vers six heures. (*Moderato*, 62)¹

Souvent, vous regardez ces hommes qui vont à l'arsenal, surtout **l'été**, et la nuit, lorsque vous dormez mal, le souvenir vous en revient. (*Moderato*, 64)

Il suffit ensuite d'opérer la déduction : il reste 17 occurrences à imputer au verbe *être*. La manipulation prend quelques secondes, mais est de toute façon plus rapide que l'examen exhaustif des formes nombreuses, ce qui serait le cas par exemple dans *Un barrage contre le Pacifique*, où le terme revient 113 fois, (et pas une seule comme substantif). De la même manière, *or* peut être conjonction de coordination ou substantif ; dans ce dernier cas, il sera précédé soit par un déterminant, soit par une préposition : on peut donc grâce à la séquence **(unll'lcetlsonltonlmonlvotred'len)** essayer d'en localiser les emplois en tant que substantif. La recherche livre trois résultats positifs, ce qui correspond à la répartition grammaticale effective du terme. Plus largement, on peut donc envisager l'écriture de grammaires comportant ainsi des listes de déterminants fléchis (masculin, féminin, pluriel) et/ou de prépositions qui peuvent aider à retrouver les substantifs. Voici l'exemple de la grammaire **detmasc**, qui s'applique aux substantifs masculins et féminins au singulier

¹ Les références sont données dans les éditions suivantes : *Un barrage contre le Pacifique* [1950], Gallimard, « Folio », 2004 ; *Moderato cantabile* [1958], Editions de Minuit, « Double », 1980 ; *L'Amant*, Minuit, 1984.

SAUVER LA GRAMMAIRE

Vous pouvez modifier la grammaire ci-dessous avant de la sauvegarder.

```
detmasc :  
un|l'|le|ce|cet|notre|votre|son|d'|en  
defem :  
une|l'|la|cette|notre|votre|sa|son|d'|en
```

Il est donc possible, dans certains cas, d'optimiser la recherche grâce à une grammaire, mais celle-ci ne peut être définie qu'au cas par cas.

1.1.2. Polysémie

La deuxième difficulté de la réalisation de la réduction morphologique du corpus réside dans le fait que le même mot peut être polysémique, sans forcément changer de nature grammaticale. Cette fois, l'utilisation de structures combinant déterminants et prépositions ne peut pas nous aider à tourner le problème, et il faut examiner chaque terme : c'est l'exemple de **voile** (vestimentaire ou maritime) ou de **somme** (sommeil ou argent). Ce dernier cas est intéressant, car il peut relever chez Duras de deux isotopies aussi importantes l'une que l'autre dans le *Barrage contre le Pacifique* : le sommeil ou l'argent. Il n'est donc pas innocent d'imputer les occurrences à l'un ou l'autre terme.

D'autre part, certains mots changent de sens en passant du masculin au féminin, du singulier au pluriel : « *reste/restes* », « *second/seconde* » « *nouveau/nouvelle* ». Là encore, il faut aller y voir de plus près pour déterminer à quelle entrée sémantique les rattacher.

Enfin, lors de l'établissement de la liste, il faut également prêter attention aux apocopes : *phono/phonographe* ; *auto/automobile* ; *télé/télévision*. Il est important qu'elles restent distinctes lors de l'extraction du vocabulaire, car elles marquent un certain registre de langue. A partir de là se pose la question de savoir si on doit les réunir ou pas : si l'apocope est employée avec une connotation particulière (hargne, vulgarité, familiarité, ou plus simplement effet de mode), ou que le mot lui-même, apocopé, présente une légère nuance de sens par rapport à la forme complète, il faut envisager de lui laisser un place autonome dans la liste.

Une dernière difficulté, mais cette fois plus subtile, tient à la prise en compte du contexte. En effet, il n'est pas indifférent dans certains cas qu'un terme soit employé au singulier ou au pluriel. Par exemple, *Un barrage contre le Pacifique* comporte 29 fois le mot *enfant* au singulier et 90 fois au pluriel. Ce pluriel n'est pas neutre : dans le roman, les enfants sont le symbole le plus criant de la misère des habitants de la plaine. Parce qu'ils sont nombreux, ils affament leurs parents exténués ; eux-mêmes meurent par dizaines, pour avoir mangé des mangues vertes, ou écrasés sur la piste par des voitures de chasseur. Ils semblent à la fin être ramenés à de la matière proliférante et obstinée :

Vous ne le savez peut-être pas mais ici il meurt tellement de petits enfants qu'on les enterre à même la boue sous les rizières, sous les cases. [...] Ce qui fait que les terres que vous

convoitez et que vous leur enlevez, les seules terres douces de la plaine, sont grouillantes de cadavres d'enfants (BCP, 295-196)

L'accent mis sur le pluriel a donc ici une forte valeur symbolique : elle décrit un monde de pauvreté telle que la vie humaine ne peut y gagner le prix de son individualité. En conséquence, même si la réduction morphologique confond le singulier et le pluriel, les déséquilibres importants entre singulier et pluriel peuvent être relevés lors de l'établissement de la liste à des fins interprétatives ultérieures.

1.1.3. Problèmes divers

Les mots composés sont divisés par Frantext : par exemple *aujourd'hui*, *Pnomh Penh*. D'autre part, le logiciel ne fait pas toujours la différence entre le trait d'union et le tiret cadratin. Certaines éditions ne la font pas non plus, débutant par exemple les dialogues par des tirets, ce qui complique le travail de numérisation. Les deux formes sont donc conjointes et l'item *tiret* se trouve parfois à englober un signe de ponctuation, équivalent à une parenthèse, et un signe fonctionnel qui n'est pas une ponctuation. Une des façons de faire le tri consiste à interroger la base avec la séquence [— **+espace**], typographie qui ne concerne que les tirets cadratins.

Enfin, les formes contractées *au / des* sont aussi des prépositions et ne peuvent être réattribuées à la forme originale.

Une fois toutes ces opérations de réduction morphologique accomplies, il est possible d'établir un nouveau tableau de fréquences, qui aura l'avantage de supprimer les redondances sémantiques. On raisonnera alors en termes **d'unités sémantiques discrètes**. L'établissement de cette liste réduite est long et parfois fastidieux (*Du côté de chez Swann* comporte plus de 15000 formes à réduire) et peut être facteur d'erreurs (exemple de mes propres listes, avec un pourcentage d'erreurs qui va de 0,6 à 1,5%). Elle comporte clairement une part interprétative : il faut donc considérer le relevé de fréquences comme une aide à l'exégèse, mais pas forcément comme un outil pouvant être manipulé en dehors de tout cadre interprétatif et /ou théorique.

1.2 La relativité : en fonction de quoi un corpus lexical prend-il son sens ?

Le relevé des fréquences offert par Frantext ne peut avoir de valeur absolue. En effet, en soi, que le verbe *chanter* apparaisse 5 ou 35 fois, que les adverbes soient deux fois plus nombreux que les adjectifs ne signifie rien hors contexte. D'où la nécessité de créer justement un contexte d'appréciation. Celui-ci est articulé à un triple niveau.

1.2.1. L'ensemble de l'ouvrage

Pour voir si un mot est très représenté ou non, il faut évaluer son score relativement à celui de toutes les autres unités lexicales discrètes du texte. Une manière de le faire consiste à prendre, hors mots-outils, la valeur la plus élevée, à éliminer du compte tous les mots outils qui précèdent, et à diviser le nombre total d'unités restantes par dix, de manière à pouvoir classer les unités par déciles. La méthode a l'avantage de mettre en avant un plus ou moins fort taux de concentration du vocabulaire et de dégager une **zone critique de récurrence**, qui est approximativement celle des deux premiers déciles. Par exemple, *Un barrage contre le*

Pacifique comporte 4788 unités lexicales. Si l'on consulte le haut de la liste, on obtient le tableau suivant

de	3223
la	2673
être*	2626
avoir	2364
il	2033
elle	1954
à	1830
et	1830
le	1703
l'	1266
qu'	1211
les	1184
que	1170
en	1058
pas	1017
je	993
dire	942
d'	932
un	923
ce	918
ne	831
joseph	793
pouvoir*	792
une	766
se	738
lui	737
c'	725

Le premier item qui ne soit pas un mot outil est le prénom de l'un des protagonistes, Joseph. On va donc prendre cette valeur de départ pour comptabiliser le nombre total d'unités lexicales discrètes : en l'occurrence, 4766, soit des déciles de 476 mots. On constate ainsi la présence dans le premier décile de mots comme *voir*, *regarder*, qui dénotent une pulsion scopique, confirmée par *œil-yeux*, mais aussi d'*attendre* (maître mot de cette vie de pauvreté où rien ne passe). On trouve aussi la mention des maigres ressources dont on dispose pour s'occuper : *phonographe*, la passion du frère qui trompe son ennui en écoutant des disques usés, *cinéma*, où l'on oublie sa vie de misère. On repère également un important réseau lexical de l'argent : *argent*, *diamant*, *francs*. Mais on trouve aussi tous les éléments matériels qui forment le décor du désastre de la famille : *bungalow* (à la fois seul bien matériel et lieu du huis clos), *barrage*, *cadastre*. L'analyse de la liste passe donc par ce découpage, qui permet de déterminer à partir de quelle fréquence un mot peut prétendre à jouer un rôle déterminant sur la scène sémantique dans une œuvre donnée.

1.2.2 Les autres œuvres de l'auteur

L'intérêt de ce type de statistique est aussi de mesurer les évolutions ou les variations qui peuvent caractériser l'œuvre d'un écrivain, en comparant les données dans ses différents ouvrages. On peut ainsi opérer des observations stylistiques sur la taille des phrases, la ponctuation, la taille des œuvres les unes par rapport aux autres. Une étude du vocabulaire peut par exemple montrer comment une isotopie se fixe et se déploie. Nous nous intéresserons ici à celle de l'alcool. Dans *Un barrage contre le Pacifique*, elle est présente sous plusieurs formes :

- le verbe *boire* (85 occurrences)
- les noms de boisson : *champagne* (17) *whisky* (7), *cognac*, avec la marque *Martel* (4), *Martelperrier* (1). Une lecture des extraits contextuels nous montre que toutes ces boissons, sans exception, sont consommées hors de la maison : elles sont offertes par les riches, Monsieur Jo ou la maîtresse de Joseph, et ont un rôle vénéneux : égayer un incurable désespoir, endormir l'homme que l'on s'apprête à tromper. Les personnages du *Barrage*, à l'origine, ne savent pas boire, et ils apprennent l'ivresse des humiliés :

Les choses ne devenaient supportables que lorsqu'ils avaient suffisamment bu, surtout Joseph, pour négliger M. Jo, jusqu'à ne plus l'apercevoir. Comme aucun des trois n'avait, et pour cause, l'habitude du champagne, l'effet désiré ne se faisait pas attendre. Même la mère qui n'aimait pas précisément boire, buvait. Elle buvait, prétendait-elle, pour « noyer sa honte » (*BCP*, 90)

- les états liés, métaphoriquement ou non, à la consommation d'alcool : *saoul* (10), *saoulé* (2), *ivre* (1), *ivresse* (2)

Dans *Moderato Cantabile*, l'isotopie a évolué. A la suite d'un crime passionnel commis pendant que son fils prenait sa leçon de piano quelques étages plus haut, Anne Desbarèdes prend l'habitude de se rendre dans un café où elle s'enivre avec un ouvrier nommé Chauvin. Cette fois, plus rien à voir avec les alcools luxueux de la colonie : c'est le *vin* (55 occurrences) qui occupe tout l'espace. *Boire* représente 20 occurrences ; on trouve également *ivre* (5), *ivresse* (4), *ivrogne* (3). Quantitativement, rapportée au nombre total des mots du roman, l'isotopie est plus présente, plus implacable aussi. L'ardeur qu'Anne Desbarèdes met à s'enivrer est proportionnelle à son ennui de femme riche et oisive, et l'alcool vient tenter de combler un vide existentiel. Elle ne boit pas seulement au café, mais chez elle, sous son propre toit, lors d'un dîner, et son état crée le scandale.

Dans *L'Amant*, on retrouve 23 occurrences de *boire*, mais plus aucune de *vin*. Les désignations des boissons sont très peu représentées : 1 occurrence de *whisky*, 1 de *martel*. C'est le terme générique d'*alcools*, avec 6 occurrences qui prend le pas sur les hyponymes. Dans ce roman à forte coloration autobiographique, la question du rapport à l'alcool est abordée de front :

L'alcool a rempli la fonction que Dieu n'a pas eue, il a eu aussi celle de me tuer, de tuer. Ce visage de l'alcool m'est venu avant l'alcool. L'alcool est venu le confirmer. (*Amant*, 15)

L'évolution de l'isotopie confirme la maturation du regard de l'auteur sur sa propre histoire : celle du *Barrage* est plongée dans le contexte indochinois, et reconstituée avec une certaine minutie le décor de la vie coloniale, avec ses bars, ses alcools de prix. *Moderato* et *L'Amant*

projetent quant à eux plus crûment le personnage dans une relation passionnelle à l'ivresse, qui est une manière de se fuir.

1.2.3 Les œuvres d'autres auteurs

Enfin, l'œuvre d'un auteur ne peut être appréhendée sans tenir compte d'un contexte plus général comme celui du genre de l'ouvrage, de l'école, des influences subies, des auteurs contemporains. On ne compare pas un roman de Marguerite Duras à un recueil de poèmes de Charles Baudelaire ou à une chanson de geste médiévale. J'ai choisi pour ce travail quatre autres références externes, à savoir quatre romans du XXe siècle. D'abord *Du côté de chez Swann*, de Proust, dont l'écriture semble radicalement opposée à celle de Duras : l'opposition permet de voir quels sont les points de différence les plus saillants. *Les Choses*, de Georges Perec, contemporain de Duras, permettra de mettre en rapport deux imaginaires passablement éloignés, et de voir si les poétiques respectives des auteurs influencent leur manière de structurer grammaticalement leur texte. Enfin, *La Route des Flandres*, de Claude Simon, fournit un point de comparaison très pertinent, puisqu'il a été publié la même année que *Moderato Cantabile*. Nous y ajouterons *La Jalousie* de Robbe-Grillet, publié en 1957. En effet, Simon et Robbe-Grillet appartiennent à cette catégorie d'auteurs classés par leur théoricien Jean Ricardou, sous l'étiquette de « Nouveaux Romanciers », groupe auquel Duras a souvent été rattachée par la critique : nous verrons si ce rapprochement générique trouve son écho dans la comparaison des manières d'écrire.

2. Résultats

2.1 Les fréquences

L'analyse des fréquences permet de distinguer des récurrences significatives, une fois éliminés les mots-outils. Cela dit, il est difficile de conférer une valeur intrinsèque à une liste. La plupart du temps, un mot souvent utilisé sera la vedette d'une isotopie, dont il faut rassembler les différents éléments. Intéressons-nous à celle du sommeil dans les romans de Duras. Nous allons pour la constituer regrouper un réseau lexical, auquel nous adjoindrons le terme *pilule*², qui sont des somnifères dans le *Barrage*. Le résultat fait apparaître une isotopie fournie

	<i>Barrage</i>	<i>Moderato</i>	<i>Amant</i>	Total
Dormir	44	14	24	82
Endormir	14	3	5	22
Epuisement	0	0	0	0
Epuiser	8	0	0	8
Fatiguer	16	6	5	27
Exténué	3	3	7	13
Harassé	0	0	0	0
Las	1	0	0	1
Pilule	23	0	0	23
Rendormir	3	0	1	4
Torpeur	3	1	0	4

² Bon exemple de la nécessité de posséder une connaissance préalable du texte traité. En effet, *pilule*, chez un auteur comme Christiane Rochefort, et de façon plus générale, dans tous les textes postérieurs à 1960, peut aussi être rattaché à l'isotopie de la sexualité. Ce n'est absolument pas le cas ici.

Total	115	27	42	184
Total U.S.D.	4788	2027	2762	
% des U.S.D.	2,4 %	1,3 %	1,5 %	

Elle est plus importante dans le *Barrage* car le sommeil est étroitement associée à la figure maternelle, mais dans les trois cas, le mot relève du premier décile dans la liste des fréquences. Intéressons-nous maintenant à la valeur symbolique que peut revêtir le sommeil, en examinant les extraits

Toujours, dans les périodes difficiles de sa vie elle avait **dormi** comme ça. (*BCP*, 196)

Lorsque les barrages s'étaient écroulés, il y avait deux ans, elle avait **dormi** quarante-huit heures d'affilée. (*BCP*, 196)

Elle **dormait** une grande partie de la journée. (*BCP*, 249)

Le sommeil n'est ni un plaisir ni une fonction naturelle, mais un **échappatoire**. La mère dort pour s'abrutir, oublier, presque comme on meurt. Et l'on en trouve un écho et une confirmation dans *L'Amant*, qui revient sur ce sommeil de la mère : « Ou rien, ou **dormir**, mourir. » (*Amant*, 23).

L'interprétation des listes nécessite d'autre part, comme nous l'avons déjà mentionné, la prise en compte d'un contexte plus général, qui peut faire appel à des éléments de poésie, d'histoire, voire de biographie : par exemple, le mot *père* est sous-représenté chez Duras, ce qui renvoie à une réalité biographique, le décès précoce d'Henri Donnadiou. Enfin, la faible fréquence, la non-apparition d'un mot dans une liste peuvent parfois être tout aussi pertinents qu'une forte représentation. Dans *L'Amant*, le mot *amour* revient 28 fois, *désir*, 15, mais *tendresse*, 2 seulement, ce qui renseigne sur l'importance de la sexualité dans la relation que la jeune fille entretient avec « Le Chinois ».

2.3. La ponctuation

Frantext permet de quantifier la ponctuation. En divisant le nombre total de mots du corpus par le nombre de signes susceptibles de conclure une phrase (point, point d'exclamation, point d'interrogation, points de suspension³), on peut obtenir une longueur moyenne de la phrase. Chez Duras, sur les trois ouvrages considérés, la phrase fait entre 13 et 14 mots, ce qui est peu rapporté aux 22,9 mots de Perec, 38,3 mots/phrasede Proust et surtout aux 90 mots de Claude Simon ! En revanche, on retombe sur une fréquence semblable chez Robbe-Grillet. Il n'est donc pas possible d'extrapoler une syntaxe particulière au Nouveau Roman, compte-tenu de ces disparités. Le faible nombre de mots chez Duras est recoupé par le nombre peu élevé de subordinations. On remarque une diminution et /ou une disparition de la ponctuation exclamative, interrogative, qui accompagne une phrase de plus en plus assertive, voire sentencieuse. Nous en avons choisi un exemple tardif hors corpus, car il est symptomatique que la manière dont l'écriture s'est crispée autour de ses propres automatismes.

C'est dans une maison qu'on est seul. Et pas au-dehors d'elle mais au-dedans d'elle. Dans le parc il y a des oiseaux, des chats. Mais aussi une fois un écureuil, un furet. On n'est pas seul dans un parc. Mais dans la maison, on est si seul qu'on en est égaré quelquefois. C'est

³ Le point-virgule est exclus de cette catégorie, car il est le signe d'une pause forte, non d'une clôture phrastique.

maintenant que je sais y être restée dix ans. Seule. Et pour écrire des livres qui m'ont fait savoir, à moi et aux autres, que j'étais l'écrivain que je suis. (*Ecrire*, Gallimard, 1993)

2.3. Le voisinage

L'étude du voisinage est un outil précieux et très révélateur, plus peut-être que la liste des fréquences. En effet, elle permet de voir quels sèmes, mais aussi quels thèmes sont associés, ouvrant ainsi la voie à l'interprétation. Pour reprendre l'exemple du mot *mère* dans *BCP*, on voit à quel point l'expression de la détresse et du malheur est accentuée dès qu'on évoque cette figure tragique, victime de l'administration coloniale corrompue. Le voisinage, classé par ordre décroissant de fréquences, donne le résultat suivant :

736 mère
343 joseph
313 suzanne
133 ..
131 jo
64 bungalow
60 carmen
59 agosti
48 jours
47 voir
46 ram

45 chambre
41 caporal
37 jour
35 plaine
32 temps
31 concession
31 seule
31 soir
31 table
30 ans
30 barrages

Il est tout à fait révélateur : en tête, la mère et ses deux enfants, liés par l'enfer de la pauvreté, qui les oblige à coexister tous trois et à survivre tant bien que mal dans leur bungalow, pour former ce huis clos colonial dans lequel les émotions et le désespoir sont exacerbés. Jo est le quatrième membre du quatuor : le riche fils de négociant, amoureux de la fille, détesté par tous les trois, mais subtilement encouragé car la mère et la fille espèrent parvenir à lui soutirer quelque chose. *Bungalow*, *concession* et *barrages* font également partie du voisinage du mot *mère* : le lexique confirme ici la caractère obsessionnel de l'entreprise de la mère, qui s'accroche à ses enfants, sa concession incultivable et sa maison précaire au toit pourri, ses seuls biens. Les barrages ont été son espoir, sa folie : après leur effondrement, ils sont devenus l'objet d'une amertume dans cette ressassée.

Si l'on prend comme pivot « Suzanne », les résultats sont tout aussi significatifs :

921 suzanne
440 joseph
283 mère
245 ..
241 jo
83 carmen
69 agosti
61 auto
53 bungalow
52 voir
47 demanda

45 ram
43 air
42 chambre
42 temps
39 yeux
38 bague
37 seule
34 diamant
33 jour
32 regardait
31 faut

Suzanne est elle aussi est partie prenante du huis clos, mais plus proche lexicalement du *frère* que de la *mère*, ce qui correspond à leur proximité affective dans la diégèse. La concession et les barrages sont plus éloignés de ses préoccupations : en revanche, on note le thème de *l'auto* (qui désigne toujours celle de Monsieur Jo), de la *bague* et du *diamant*. L'auto correspond au désir d'évasion, à la fois physique (aller danser à Ram), sexuelle (rêver d'aller dans la garçonnière de Jo), et métaphorique (sortir de la pauvreté). Le diamant et la bague (il s'agit en réalité du même objet) sont offerts par Monsieur Jo : ils focalisent toute l'attention de Suzanne, passionnément avide d'argent, de luxe, de richesse. Le diamant est en quelque sorte le fantasme compensatoire, leur revanche contre un monde de riches qu'ils abhorrent et jalouent : « pour lui, une bague, c'est rien du tout » (162) dit Suzanne, mais pour eux, c'est l'objet d'une espérance ranimée, celle de rembourser les emprunts :

Car cette **bague**, elle était autant à eux maintenant et aussi difficile à reprendre que s'ils l'avaient mangée, digérée, et que si elle était déjà diluée dans leur propre chair. (*BCP*, 153)

L'étude de voisinage se révèle donc pertinente pour dégager le degré de relation entre les personnages, les valeurs symboliques ou métaphoriques attachées à certains objets.

2.4. La catégorisation

La catégorisation permet de poser les bases d'une analyse stylistique, en donnant un aperçu d'ensemble des préférences grammaticales de l'auteur. Là encore, pour donner du sens aux chiffres, il est nécessaire de les rapporter à une unité commune, en l'occurrence le nombre total de mots utilisés, et en même temps de les comparer à d'autres auteurs, pour en dégager la possible singularité. Chez Duras, le verbe est fortement représenté, avec 11% environ des occurrences, ce qui est un score plus élevé que celui de Proust (8,9%) et beaucoup plus élevé que chez Perec (7,4) ou Claude Simon (7,1). L'écriture durassienne est en effet peu réflexive, volontairement focalisée sur la description ou les gestes et les dialogues de ses protagonistes : la mère/Suzanne, Anne Desbarèdes/Chauvin. Chez cet auteur très proche du cinéma, les personnages sont la plupart du temps décrits en action ou en séquences temporelles fragmentées comme des plans, par des suites de verbe :

Ca a été long. Ca a duré sept ans. Ca a commencé nous avions dix ans. Et puis nous avons eu douze ans
Ca a duré tout cet âge, sept ans. Et puis enfin l'espoir a été renoncé. Il a été abandonné.
Abandonnées aussi les tentatives contre l'océan ? (*Amant*, 71)

La comparaison nous montre également que l'écriture durassienne est assez peu conjonctive : le score des conjonctions (coordination et subordination confondues), rapporté au total des mots du livre, est de 3,4% dans le *Barrage*, 2,3 dans *Moderato Cantabile*, 2,4 dans *L'amant*, ce qui est faible, si on les rapporte au score de 4% chez Proust ou 5,4% dans *La route des Flandres*. En revanche, ce score est proche de celui de Perec (2,8%) ou Robbe-Grillet (2,3%). Dans ces deux derniers cas, comme chez Duras, il s'agit en effet d'une écriture qui est plutôt une écriture de l'observation et du constat que celle de la péripétie. Il ne se passe rien dans la plaine, rien dans la vie d'Anne Desbarèdes, rien non plus dans celle de Lol. V Stein. Duras ne cherche pas à établir des causalités, mais pose des séries de descriptions, de gestes parfois mécaniques, qui semblent sans lien les uns avec les autres : c'est justement ce vide existentiel, cette béance, cet ennui qu'interrogent ses livres. Grammaticalement, le faible score des conjonctions s'explique par une syntaxe qui privilégie l'asyndète, voire l'anacoluthie :

Elle se leva, se leva avec lenteur, fut levée, réajusta une nouvelle fois sa veste. Il ne l'aida pas. Elle se tint en face de lui encore assis, ne disant rien. Les premiers hommes entrèrent au café, s'étonnèrent, interrogèrent la patronne du regard. (*Moderato*, 35)

Cette absence de liaisons devient d'ailleurs un trait presque caricatural de son écriture. Dans les deux parodies de Patrick Rambaud, *Virginie Q.* et *Mururoa mon amour*⁴, le trait est d'ailleurs souligné avec force, ce qui indique bien qu'il est perçu comme caractéristique de l'écriture de Duras :

Il le fait. le noir est sombre dans la pièce. C'est un peu comme les yeux quand on les ferme. Ca provoque de la peur. Elle a peur (*Virginie Q.*, 60).

Duras aime également, sur le plan syntaxique, user de la répétition et de l'énumération. Là encore, Frantext peut nous aider à essayer d'en retrouver des exemples, en particulier des énumérations verbales. Par exemple, une recherche par **&e(g=V), &e(g=V), &e(g=V)**, qu'on peut répéter avec des substantifs ou des participes, livre des résultats pertinents.

Il se prêtait à toutes les fantaisies de la mère, empierrait, plantait, transplantait, taillait, arrachait, replantait (*BCP*, 248)

La mère était devenue farouche, silencieuse, jalouse (*Amant*, 139)

Sur la quinzaine de concessions de la plaine de Kam, ils avaient installé, ruiné, chassé, réinstallé et de nouveau ruiné et de nouveau chassé, peut-être, une centaine de familles » (*BCP*, 27)

La grammaire s'accorde donc à l'écrasante présente de la concession, à la récurrence d'un vieux fantasme, celui d'y faire pousser quelque chose. Mais elle insiste aussi sur l'aspect toujours recommencé de l'exploitation de l'homme par l'homme, décrite comme un cycle sans fin. Autre structure fétiche de Duras, la postposition, que l'on peut rechercher par **&e(g=A)&e(g=Per)&e(g=V)**.

jalouse elle est (*Amant*, 31).

et de faire tant de détours pour l'attraper, lui, il ne pourra jamais (*Amant*, 48).

dévergondés on était (*Amant*, 58).

Cette tendance s'accroît de livre en livre, au point de devenir une sorte d'estampille stylistique, qui consiste à rejeter le substantif en fin de phrase. On peut rechercher la structure grâce au catégoriseur : une phrase comme « on se débrouillerait pour que ça nous le **fasse, le bonheur** » (*BCP*, 45) peut être schématisée par **&e(g=V),&e(g=D) &e(g=S)**. Si l'on ne demande que les deux derniers termes, on peut repérer ainsi 118 séquences postposées parmi lesquelles certaines sont de véritables anaphores :

Je parle souvent de mes frères comme d'un ensemble, comme elle le faisait elle, notre mère. (*Amant*, 71)

Nous sommes ensemble, elle et nous, ses enfants (*Amant*, 21).

La mort, en chaîne, partait de lui, l'enfant (*Amant*, 137-138).

⁴ Patrick Rambaud, sous le nom de Marguerite Duraille, *Virginie Q.*, Balland, 1998 ; *Mururoa mon amour*, J-C. Lattès, 1996.

Cette anaphore grammaticale possède une double vertu : souligner le mot qu'elle postpose, et créer une dynamique du ressassement tout à fait caractéristique de l'écriture de l'auteur et repérée comme telle par Patrick Rambaud : « Droite dessus le tabouret, elle regarde le moustique, dans ses yeux à lui, ses yeux de moustique » (*Mururoa*, 37). En outre, le parodiste a bien pointé, dans cet exemple, une autre habitude durassienne : un usage assez particulier de la préposition *dans*, souvent utilisée d'une façon un peu étrange, ce que confirme l'interrogation des textes par l'opérateur **dans (lelalles) &e(g=S) de** : « dans le sang de sa bouche », « dans le silence de l'enfant » (*Moderato*, 15), « dans l'amour de la mère » (52), « il est dans un amour abominable » (*Amant*, 50), « dans l'émotion de la mort de notre mère » (95). On remarquera que la structure permet de décrire l'entité (*amour, silence, sang*) comme un continuum dans lequel les personnages s'immergent d'une manière presque physique, ce qui en accentue le caractère pesant ou abominable. Frantext permet donc, une fois repéré un trait sémantique, de le schématiser, le repérer et isoler les séquences où il apparaît pour étudier son fonctionnement. Il permet également de quantifier sa récurrence, ce qui va déterminer le fait qu'il puisse être oui ou non considéré comme caractéristique stylistique de l'auteur.

2.5. L'interrogation par expression

La possibilité d'interroger la base non seulement sur un mot, mais aussi sur une expression permet d'observer le jeu des pronoms et des déterminants. Nous allons pour ce faire reprendre l'exemple du terme *mère*.

	<i>la mère</i>	<i>leur mère</i>	<i>notre mère</i>	<i>ma mère</i>
<i>L'Amant</i>	41	1	11	114
<i>Un Barrage contre le Pacifique</i>	572	2	0	12

Certes, le caractère autobiographique de *L'Amant* suffit à expliquer le basculement vers la première personne « *ma mère* ». En revanche, la forme concurrente *la mère* au lieu de *leur mère*, ou encore du prénom de la protagoniste (jamais mentionné), est étonnante. Elle l'est encore plus dans le *Barrage*, où elle représente un pourcentage massif des occurrences : elle peut être vue comme volonté de desserrer le lien affectif pour mieux figer le personnage dans l'icône de la *mater dolorosa*, mère universelle et anonyme de toutes les douleurs.

L'interrogation par expression permet également de repérer d'autres tournures qui peuvent devenir des marques de fabrique stylistiques. Duras, on l'a vu, affectionne les anacoluthes, les constructions légèrement déséquilibrées, les tournures syntaxiques postposées parfois étranges. On en remarque peu dans le *Barrage*, mais l'un d'elle est déjà là : *en allée*. On retrouve ensuite ce syntagme à plusieurs reprises dans le corpus⁵

1 - Et tout d'un coup, la tête ballante, la couche entrouverte, complètement **en allée** dans le lait du sommeil, elle flotta, légère (*Barrage*, 142).

2- Modéré et chantant, dit l'enfant en **allé ou ?** (*Moderato*, 16)

3- Lol était drôle, moqueuse impénitente et très fine bien qu'une part d'elle-même eût été toujours **en allée** loin de vous et de l'instant. (*Le Ravissement de Lol V. Stein*)

⁵ Nous avons ici ajouté un exemple extrait du *Ravissement de Lol V. Stein*, en cours de numérisation au moment où a été fait l'exposé, et un autre extrait de *La Douleur*.

4 - Il n'avait jamais l'air de rien voir, toujours **en allée** au cœur de l'absolue bonté. (*Douleur*, 52)

5 - Toujours **en allé** au cœur de l'absolue douleur de la pensée. (*Douleur*, 52)

Ce petit comparatif a son importance : il montre d'une part la persistance de ces traits syntaxiques, que l'on retrouve à vingt-cinq ans d'intervalle, trop marqués et trop visibles pour être l'effet du hasard : il y aurait donc chez l'auteur un véritable souci stylistique, occulté par l'apparente déstructuration de son écriture. Cette hypothèse est corroborée par le fait qu'à l'occasion, on constate de franches ruptures stylistiques, de quasi « à la manière de », comme si l'auteur cherchait à indiquer à la fois sa maîtrise de la phrase classique et le fait que son refus en soit délibéré.

Lentement, la digestion commence de ce qui fut un saumon. Son osmose à cette espèce qui le mangea fut rituellement parfaite. Rien n'en troubla la gravité. L'autre [le canard] attend, dans une chaleur humaine, sur son linceul d'oranges. (*Moderato*, 69)

D'autre part, on remarque de façon générale chez Duras une tendance à exagérer le désordre syntaxique, par l'utilisation de constructions ouvertement asymétriques, inversées, voire incomplètes. Le phénomène commence à s'affirmer dans *Moderato Cantabile*

A l'avoir vu, on ne peut s'empêcher, n'est-ce pas, c'est presque inévitable ? (*Moderato*, 44)

Mais sans doute ce soir-là crut-elle son étonnement comme à lui-même renouvelé (*Moderato*, 26)

Il ne cesse de s'accroître ensuite, ce que confirme l'examen hors corpus.

C'est le lieu où plus tard me tenir une fois le présent quitté, à l'exclusion de tout autre lieu (*Amant*, 104).

Et moi, quand je vous la raconte, comme dans la vérité, dans ma vérité, c'est ce que je viens de vous dire, ce qui a été vécu entre la mouche et moi et qui ne se prête pas encore à rire (*Ecrire*, 40).

L'évolution stylistique va donc vers une obscurité de plus en plus grande : absence totale de liens logiques entre les phrases, et même entre les membres de la phrase, tournures étranges, découses, passifs très marqués. Ces ruptures accompagnent, épousent plus exactement une prise de distance avec les formes classiques du roman : peu de descriptions, des décors statiques, des personnages parfois réduits à une initiale, une ombre, comme celui, emblématique, de Lol. V. Stein. Perturber l'ordre grammatical, qui est aussi un ordre social, politique, permet d'explorer dans l'œuvre un territoire dangereux, celui de la folie, l'écriture devenant le lieu où tout se défait. Comme l'auteur le dit à Xavière Gauthier dans *Les parleuses* : « Evidemment, dans le *Barrage*, je ne voulais pas raconter tout. Je voulais que ce soit harmonieux. On m'avait dit : "Il faut que ce soit harmonieux". C'est beaucoup plus tard que je suis passée à l'incohérence »⁶. (p. 139) Après son Goncourt, Duras, durant les nombreuses émissions et interviews télévisées où elle a été conviée, a usé et abusé de ce style de pythonisse sentencieuse, indissociable de son personnage.

3. Problèmes

⁶ Marguerite Duras, Xavière Gauthier, *Les parleuses*, Minuit, 1974, p. 139.

3.1. Les flexions

Il n'est pas possible d'effectuer une véritable analyse des fréquences sans avoir opéré préalablement une réduction de toutes les formes fléchies, opération on l'a vu longue et fastidieuse. Compte tenu de la puissance de l'outil d'interrogation Frantext, dans quelle mesure pourrait-on envisager que l'étude des fréquences se fasse en intégrant les listes, qui permettent de repérer les formes fléchies, et en y soumettant tout le corpus ? L'extraction ainsi obtenue comporterait toujours de formes ambiguës, qui demanderaient à être examinées, mais permettrait une réduction significative de l'effort de l'utilisateur.

3.2. Le problème des mots-outils

La charge lexicale est inégalement répartie sur les différentes catégories grammaticales : certains mots-outils comme les auxiliaires, les prépositions, certaines conjonctions et adverbes sont si couramment employés qu'ils en deviennent sémantiquement vides ou peu significatifs⁷. Il est donc délicat de les mettre sur le même plan que des substantifs, des verbes ou des adjectifs à la charge sémantique plus importante : un *que* ou un *ane* portent pas en soi de sens, alors que l'apparition de *barrage*, *alcool*, *désir*, en aura toujours chez Duras. La question est complexe, mais peut-être pourrait-on imaginer la création de grammaires « restrictives », mises à disposition des utilisateurs, qui permettraient d'exclure de la liste des fréquences un certain nombre de termes, comme les conjonctions de coordination et de subordination, l'auxiliaire être et avoir, les déterminants. Peut-être est-il possible d'arriver soi-même à ce résultat en écrivant une série de grammaires, mais leur mise en place nécessite une capacité de raisonnement logique et un minimum de compétences informatiques qui peuvent paraître décourageantes pour un utilisateur qui n'est pas familiarisé avec l'informatique.

3.3. Les registres de langue

La liste classe les mots selon des critères grammaticaux, numériques (leur fréquence) ou alphabétiques. Elle ne prend pas en compte leur sémantisme. Or, le *Trésor de la Langue Française* répertorie le sens des mots, mais également le marquage de leur registre de langue. Un tel marquage, qui permettrait par exemple d'extraire des listes de vocabulaire « soutenu » ou « vulgaire » pourrait réellement avoir son utilité. Dans *Un Barrage contre le Pacifique*, publié rappelons-le en 1950, les mots vulgaires interviennent à plusieurs reprises, non sans intentions. Il sont, en dernier recours, l'ultime défouloir de la violence et de la frustration rentrée. Par exemple, le verbe *foutre* revient plus d'une quinzaine de fois, mais la plupart du temps dans une acception bien précise : *foutre le camp*. Malheureusement, Frantext a refusé la requête **&cfoutre**, ce qui oblige à ruser en interrogeant par « le camp ». On se rend compte alors du caractère litannique de l'obsession du départ, répété presque terme à terme par Joseph et sa sœur

- Si tu ne la fermes pas, **je fous le camp**, dit Joseph. (16)
- Si **tu fous le camp**, t'auras raison, dit Suzanne (17)
- Si je vends le phono, **je fous le camp** et en vitesse (32)
- Quinze jours sans phono et **je fous le camp** d'ici (32)
- On **fout le camp**, dit-il (98)
- S'il **foutait le camp**, j'irais rejoindre Joseph (152)

⁷ Ils sont d'ailleurs écartés par Frantext de l'étude de voisinage.

Le dire de la sorte, plutôt que « je m'en irai », ou « partons », c'est en même temps dire sa lassitude, son dégoût. Le mot *merde* mérite lui aussi un petit examen : 35 occurrences, ce qui lui vaut une place dans le quatrième décile. Or Duras n'est pas Céline et n'est pas coutumière de ce registre. Le terme a deux utilisations possibles. Dans le premier cas, il est employé comme insulte par Joseph

- Merde, dit Joseph, il ne veut même plus y aller (16)
- Merde, dit Joseph, demain, on ira à Ram (30)
- Merde, cria Joseph, ce coup-ci, il est crevé (36)
- Merde, quelle bagnole, dit Joseph (42)

Mais il est aussi, lors d'une soirée de beuverie avec Monsieur Jo, étroitement associé au terme *mer*, lors d'une énième évocation des barrages. Les deux matières se confondent, sont assimilées l'une à l'autre grâce au jeu de l'homophonie

- C'est de la mer, le Pacifique, dit Suzanne
- C'est de la merde, dit Joseph [...] C'est de la merde, parfaitement, dit Joseph, de la merde ou de la flotte, c'est comme vous voudrez. Et nous on est là à attendre comme des cons que la merde se retire.
[...]
- Si c'était de la merde, dit Agosti, dans le fond du bar, ce serait mieux...
- Du riz de merde, dit Joseph en riant de nouveau, ce serait mieux que pas de riz du tout. (BCP, 59)

De façon générale, tout le vocabulaire ordurier est associé à leur échec : *il n'y a rien à bouffer* (160), *les crabes nous les [barrages] ont bouffé* (52) ; l'auteur évoque en outre « ce bordel colossal qu'était la colonie » (198), « la saloperie des agents de Kam » (195). Le terme *crever* revient 16 fois : au début, il s'applique au cheval, une vieille haridelle que Joseph a achetée et qui rend l'âme quelques jours plus tard : « il creva ». Mais assez rapidement, le mot concerne les êtres, qui dans cette plaine cambodgienne sont réduits au rang d'animaux affamés : « il préférerait se laisser crever » (16), « si je ne crève pas assez vite » (292). La mère utilise en particulier le terme pour s'adresser aux agents cadastraux : « On est les quatrièmes sur cette concession. Tous ruinés ou **crevés**. Et eux ils s'engraissent » (314) Ici, la séparation des registres de langue, chez un auteur qui se passe très bien du registre vulgaire lorsqu'elle le souhaite, a un double intérêt : susciter une dimension pathique (la colère, le dégoût), mais aussi alimenter une dénonciation virulente. *Le Barrage contre le Pacifique* est un roman politique, qui met à nu les rouages d'une administration coloniale pourrie, et qui ne prend pas de gants lexicaux pour le dire.

4. Améliorations possibles

4.1. Interface graphique

- Dans l'interface **graphique** d'entrée, le bouton « soumettre le formulaire » est en bas à gauche et le bouton « effacer le formulaire » en bas à droite. Machinalement, on a tendance à cliquer à cet endroit et à effacer sa propre requête.

<input type="button" value="Aide"/> Sélection par auteur	<input type="button" value="Aide"/> Sélection par titre
<input type="text" value="duras"/>	<input type="text" value="un barrage contre le Pacifique"/>
Sélection par date	<input type="button" value="Aide"/> Sélection par genre
<input checked="" type="radio"/> Date indifférente <input type="radio"/> A la date précise DATE1 <input type="radio"/> Avant DATE1 <input type="radio"/> Après DATE1 <input type="radio"/> Entre DATE1 et DATE2 DATE1: <input type="text" value="0"/> DATE2: <input type="text" value="2000"/>	<input type="checkbox"/> correspondance <input type="checkbox"/> récit de voyage <input type="checkbox"/> éloquence <input type="checkbox"/> roman <input type="checkbox"/> mémoires <input type="checkbox"/> théâtre <input type="checkbox"/> pamphlet <input type="checkbox"/> traité <input type="checkbox"/> poésie <input type="checkbox"/> essai
<input type="button" value="ENREGISTRER LA SELECTION"/>	<input type="button" value="EFFACER LE FORMULAIRE"/>

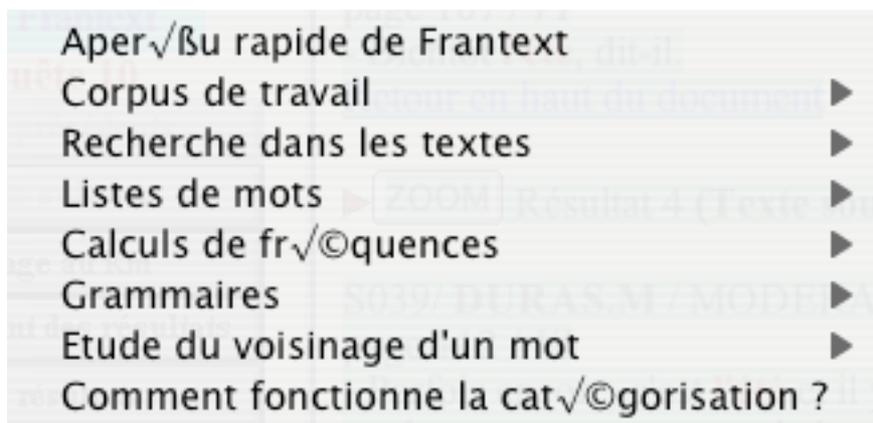
- Si l'on fait une recherche bibliographique sur un auteur
 - les œuvres de cet auteur et ses traductions sont confondues. Un utilisateur distrait, ou qui fait confiance à Frantext et s'appuie sur le corpus fourni, peut donc être amené à englober dans sa recherche...une œuvre qui n'est pas de l'auteur. Exemple : la recherche sur Koltès donne

Visualisation du corpus de travail par ordre alphabétique des auteurs (4 textes.)

N°	Cote	Auteur	Titre	Date	Genre	Edition
1	S596	<KOLTÈS:Bernard-Marie>	Combat de nègre et de chiens suivi des Carnets	1983	prose, théâtre	Paris : Les Éditions de Minuit, 1996.
2	S799	<KOLTÈS:Bernard-Marie>	Quai ouest	1985	prose, théâtre	Paris : Les Éd. de Minuit, 2001.
3	S413	<KOLTÈS:Bernard-Marie>	Dans la solitude des champs de coton	1986	prose, théâtre	Paris : Les Éditions de Minuit, 1999.
4	S651	<KOLTÈS:Bernard-Marie>	Le Conte d'hiver [trad.]	1988	prose, théâtre	Paris : Les Éd. de Minuit, 1996.

— il n'y a pas possibilité d'isoler directement un corpus de travail en cochant les œuvres de l'auteur que l'on souhaite examiner (ce qui repose le problème de l'intégration des traductions dans la liste des œuvres de l'auteur)

- Affichage des résultats (ou comment vivre dangereusement avec un Macintosh)
L'affichage des résultats ne fonctionne pas toujours sous Safari, où les dysfonctionnements sont aléatoires. Il fonctionne correctement sous Explorer, mais avec un délai d'attente de quinze secondes à chaque survol de menu et un problème d'affichage des caractères dans les menus.



En revanche, la vitesse de réponse est toujours rapide, voire instantanée.

4.2. Exportation des résultats (vivre dangereusement avec un Macintosh, 2)

Les résultats s'exportent sous la forme d'un fichier .exe qu'il ne m'a pas été possible dans un premier temps d'exporter sous Mac. En réalité, j'ai appris que ce fichier pouvait s'ouvrir à l'aide d'un navigateur. Il pourrait être intéressant d'envisager une exportation sous .doc ou .pdf ; à défaut, une notice explicative indiquant comment traiter le fichier, en particulier pour les utilisateurs de Mac.

4.3. Reconnaissance des titres

Dans la version catégorisée, la reconnaissance des titres est limitée : il faut soumettre la requête de façon assez précise. Par exemple, le titre *Un barrage contre le Pacifique* a été saisi avec une coquille *Un barrage contre la Pacifique* : il est donc impossible d'y accéder en saisissant le titre correct ! Le moteur ne reconnaît pas les requêtes suivantes :

- Barrage pacifique
- Duras / Barrage pacifique
- Barrage contre le Pacifique
- Duras / Le barrage contre le Pacifique
- Duras / Moderata cantabile

Serait-il possible d'étendre la capacité de reconnaissance du moteur de recherches (surtout si la requête est croisée avec le nom de l'auteur) ?

D'autre part, prévoir peut-être une relecture systématique de l'entrée « **Titre** » lors de la saisie

4.3 Dysfonctionnements et problèmes liés à la consultation

La catégorie infinitif ne fonctionne pas (renseignements pris, l'équipe travaille à la correction de ce problème).

Les majuscules ne sont pas reconnues : par exemple, une étude de voisinage lancée avec « Suzanne » n'a pas donné de résultats

Une séquence .. [deux points consécutifs] apparaît régulièrement dans la liste de la ponctuation. L'origine en a été détectée : il s'agit de l'interprétation d'un ancien marquage, qui fera l'objet d'un nettoyage.

Conclusion

Frantext est un outil d'une puissance exceptionnelle, qui permet des recherches très poussées sur un corpus. Son principal atout est la possibilité d'interroger par flexions, catégories et de créer des grammaires, qui permettent d'appréhender la réalité syntaxique d'un auteur. L'étude de voisinage est également un outil à forte valeur informative. Les requêtes sont satisfaites dans un délai record, compte tenu de l'ampleur des corpus, et la base économise au chercheur un temps précieux de recherches fastidieuses, de vérifications, de listage.

Cela dit, on n'y trouve bien que ce que l'on cherche, au moins dans les grandes lignes. En effet, les fréquences ou les listes d'occurrences ne dispensent pas d'une connaissance approfondie du corpus et interviennent pour valider ou informer des hypothèses formées après lecture.

D'autre part, un mot, et même un corpus peuvent difficilement être saisis de manière isolée et nécessitent la mise en place de structures qui permettent de donner aux fréquences un poids relatif : structuration du corpus en déciles, pour dégager une zone de signifiante, comparaison avec les autres œuvres, et comparaisons avec des corpus externes.

Enfin, la prise en compte du contexte biographique, génétique, thématique, sémantique est absolument nécessaire à donner du sens aux résultats lexicaux, qui sont une donnée brute et dont la lecture ne se peut concevoir qu'inscrite dans une démarche d'interprétation.