



HAL
open science

Cardiogrammes

Étienne Brunet

► **To cite this version:**

Étienne Brunet. Cardiogrammes. Rastier, François. L'analyse thématique des données textuelles. L'exemple des sentiments, Didier Érudition, pp.25-52, 1995, Études de sémantique lexicale, 978-2-8646-0244-6. hal-04681648

HAL Id: hal-04681648

<https://hal.science/hal-04681648v1>

Submitted on 3 Sep 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Etienne BRUNET

CNRS, INaLF, Nice

CARDIOGRAMMES

1 - Si le cœur a ses raisons que la raison ne connaît pas, on imagine mal que le calcul puisse rendre compte des méandres du sentiment, dont il est plus éloigné encore que la raison. C'est pourtant l'entreprise ici tentée. Il ne nous échappe pas que l'objet visé est de ceux qui précisément échappent à la mesure. Quel mètre-étalon peut-on proposer en effet pour l'intensité des sentiments ? Quel spectre colorimétrique pour leur tonalité ? Où s'arrêter dans la zone de l'infrarouge ou de l'ultraviolet ? Quelles nuances estimer discriminantes ? De même que chaque langue a sa propre gamme des couleurs qu'on peut difficilement transposer dans la langue voisine, de même chaque individu a ses propres goûts, physiques ou moraux, son échelle des valeurs et sa palette des sentiments, que ne partage pas nécessairement la conscience du voisin. Ce que l'on partage ce sont les mots qui désignent ou évoquent les choses du sentiment. Mais la réalité vécue et sentie qui se cache derrière ces prête-mots diffère d'une conscience à l'autre.

En réalité cet embarras est commun à toute démarche qui s'aventure sur le pont du langage pour mettre en communication les mots et les choses, même si le pont est moins dangereux lorsqu'on a affaire à des objets palpables ou, à l'opposé, à des concepts précis et définitoires. Comme rien n'est si flottant que les sentiments, plutôt que de s'enliser dans les sables mouvants, on s'abstiendra de toute tentative visant à appréhender les affects en eux-mêmes. On ne prendra en considération que leur expression verbale, sans s'inquiéter de savoir si cette expression est un masque ou un leurre. Notre rôle ne veut être ici que celui d'un appareil enregistreur, qui réagit seulement à l'apparition et à la récurrence des mots, comme est sensible aux rythmes électriques la machine qui reproduit l'électrocardiogramme. Quant à interpréter les symptômes et les courbes et à décider de l'intervention, c'est l'affaire du clinicien ou du chirurgien. Dans le domaine ici abordé, l'intervention — qui au reste n'a rien d'urgent — incombe à la psychologie, à la pragmatique ou à l'histoire des mentalités.

2- On s'en tiendra donc aux mots. Mais quels mots ? Certains termes viennent immédiatement à l'esprit, comme *l'amour, la joie, la terreur*. Et il est facile d'allonger la liste, tant la langue fournit de variétés pour un domaine longuement exploré. La difficulté est de s'arrêter. Car la liste est ouverte et inépuisable, sinon infinie. Non seulement l'analyse du cœur découvre sans cesse des recoins cachés, des nuances inaperçues ou inavouées, des éclairages nouveaux projetés par le progrès des connaissances et l'évolution des sensibilités (ainsi le paysage intérieur a été largement modifié par la psychanalyse). Mais surtout l'inflation lexicale s'exerce particulièrement dans ce domaine où la vivacité d'un sentiment s'exprime mal par un mot usé. On a volontiers recours aux ressources de la suffixation pour raviver les couleurs du sentiment. Ainsi Péguy propose *abandonnement* pour signifier une déréliction plus profonde que le simple abandon. Le plus souvent c'est par le mélange des couleurs de base qu'on produit la teinte nouvelle et l'on utilise alors la combinatoire phraséologique, et la panoplie des philtres correcteurs, en particulier ceux de l'adjectif. Certains sentiments même, fort bien catalogués, n'ont pas de dénomination propre et ne peuvent se dire que par des expressions complexes ou des périphrases. Ainsi, pour prendre un exemple parfaitement lexicalisé, le complexe d'Œdipe pose un problème à la statistique lexicale qui ne sait guère traiter efficacement que les unités simples, c'est-à-dire graphiques, du vocabulaire.

3- Reste la dernière difficulté, qui tient à la polysémie et qu'on retrouve chaque fois qu'on additionne des occurrences d'une même forme. Que vaut l'addition lorsque les éléments sont disparates, même s'ils revêtent le même habit ? Quand on cumule l' *amour* du jeu, l' *amour* du prochain et tant d'autres *amours*, licites ou interdits, procède-t-on à l'intersection des sèmes communs, ou bien s'agit-il de la réunion des sèmes divergents ? Le même effectif peut dans le premier cas recouvrir un sous-ensemble vide, dans le second un sur-ensemble

débordant. Malgré leur précision, les chiffres obtenus sont-ils autre chose que des ballons gonflables, sensibles à la pression alternée de l'extension et de l'intension ? Nul n'a jamais répondu définitivement à cette objection majeure — qui n'est pas propre au langage et qui obère pareillement l'ensemble des sciences humaines et même la plupart des sciences d'observation. Avant de conclure à la répétition d'un phénomène — ce qui permet de conclure à quelque loi —, comment être sûr que les conditions sont rigoureusement semblables et qu'il s'agit bien de deux occurrences d'un même événement plutôt que de la coïncidence de deux séries causales ? La science, attentive à ce risque, s'en protège par l'expérimentation. Mais l'expérimentation est peu praticable dans les sciences humaines, particulièrement en matière littéraire, et l'ambiguïté qui s'attache aux manipulations lexicales ne peut se dissoudre que dans le pari des

grands nombres et le jeu des grands espaces. Aveugle dans les espaces restreints, la statistique n'est pas plus utile qu'une boussole dont on se servirait pour trouver la porte de la chambre ou la sortie de la maison. Mais à l'extérieur, elle peut aider à découvrir les perspectives et à dégager l'horizon.

-I- Les données

1 - C'est pourquoi nous avons situé d'emblée notre observatoire là où l'horizon s'ouvre le plus largement et où s'offre aux yeux l'immensité des textes, c'est-à-dire au sommet du Panthéon littéraire qu'est FRANTEXT. Nul besoin de détailler une fois de plus l'ampleur du panorama. Trois mille textes sont là rassemblés, dont chaque page est ouverte (il y en a 500 000) et chaque mot repérable et lisible (il y en a 150 millions). Mais n'est-ce pas trop ? Comme les époques et les genres y sont mêlés, ne risque-t-on pas la confusion quand l'objet d'étude — des sentiments tout aussi mêlés — est lui-même impur ? Aussi avons-nous suivi sagement la consigne de François Rastier en n'envisageant qu'un seul genre et un seul siècle. Le corpus de référence est donc limité aux romans entre 1830 et 1970 (du moins à ceux qui se trouvent dépouillés à l'INaLF à l'heure présente, soit plus de 600 textes représentant près de 40 millions d'occurrences¹). Comme le roman est en grande faveur durant cette période, la limitation n'est pas trop sévère et la représentativité reste assurée, puisque 167 romanciers différents se trouvent recensés dans le corpus. Quant à la fourchette chronologique, elle reste assez large (140 ans) pour que les phénomènes évolutifs y soient perceptibles. Mais comme elle correspond à un stade où l'orthographe est stabilisée, on n'a pas à craindre les perturbations qu'engendre l'inconstance de l'écriture dans les siècles passés et qui ruinent la fiabilité des dénombrements, particulièrement lorsqu'intervient un signe diacritique.

2 - Circonscrire le corpus est aisé et se justifie sans difficulté. Mais délimiter la liste des formes étudiées est chose plus délicate. Là encore nous suivrons François Rastier, non pas seulement parce qu'il est confortable —

quoique un peu
hypocrite — de substituer la subjectivité d'un tiers à la sienne
propre, mais parce
que la procédure utilisée par Rastier s'est entourée de
précautions. Dans un
premier temps ont été retenus comme sentiments les formes
associées directement
à l'expression *sentiment de* ou *sentiment d'*. En réalité la
détermination a une
signification variable, depuis l'équivalence {*sentiment de colère* =
le sentiment
qu'est la colère) jusqu'à la dépendance du génitif objectif où
l'action du verbe
sentir subsiste dans le substantif dérivé (*sentiment de danger* =
sentir un danger).

Tableau 1. Environnement spécifique de la thématique du sentiment dans le corpus *Romans 1830-1970*

(à partir d'une liste initiale de 120 mots)

a : fréquence dans l'ensemble des phrases où apparaît l'un des mots retenus

b : fréquence dans le corpus de référence (littérature des XIX^e et XX^e siècles)

c : écart réduit

<i>a</i>	<i>b</i>	<i>c</i>	<i>a</i>	<i>b</i>	<i>c</i>		
25117	34692	amour	405.4	1356	1723	amertume	99.4
12358	16003	peur	296.4	1545	2240	sympathie	97.4
10383	13670	joie	268.4	1738	2805	enthousiasme	96.2
10980	16558	bonheur	253.1	1208	1486	indignation	95.7
10662	15881	plaisir	251.4	1023	1133	lassitude	94.0
8263	11043	sentiment	237.5	1301	1789	éprouvait	92.5
6237	8330	envie	206.5	1008	1138	timidité	92.2
5980	8310	douleur	197.1	1490	2292	dignité	92.0
5051	6188	pitié	196.3	1022	1172	exaltation	92.0
4633	5371	honte	194.6	1417	2159	ambition	90.3
4340	4805	tendresse	193.7	1332	2019	ardeur	87.9
4075	4594	surprise	185.6	1175	1620	pudeur	87.8
4389	5751	orgueil	175.3	1364	2112	embarras	87.7
3884	4711	haine	173.2	747	740	humiliation	86.0
4946	7304	courage	172.2	677	665	compassion	82.3
3323	3586	angoisse	172.2	887	1108	impuissance	81.2
3716	4509	désespoir	169.4	784	933	humilité	78.7
3724	4691	crainte	165.6	569	573	panique	74.3
4722	7154	amitié	165.5	2341	6514	religion	74.2
5893	10567	honneur	164.8	812	1114	indulgence	73.2
4135	5717	espoir	164.4	545	583	gratitude	70.1
5413	9918	foi	155.6	527	560	complicité	69.2
3672	5190	tristesse	152.7	622	764	soumission	68.8
3932	5909	confiance	151.8	758	1111	extase	67.8
3711	5602	horreur	147.0	635	814	cruauté	67.6
2093	2076	jalousie	143.8	522	581	insouciance	67.0
2904	3817	trouble	142.3	600	934	incertitude	57.9
2305	2644	remords	138.1	618	1062	responsabilité	55.0
3354	5197	solitude	137.4	459	664	inspirait	53.2
3015	4445	respect	134.6	553	925	bienveillance	53.0
2435	3144	affection	131.8	482	730	ferveur	52.9
2464	3282	terreur	130.0	265	261	répulsion	51.4
2344	3269	admiration	123.1	3435	16458	sorte	51.3
2441	3589	aise	119.6	263	259	camaraderie	51.2
2505	3995	ennui	116.5	337	418	nostalgie	50.3
1679	1995	gaieté	115.3	2045	8125	sentait	49.4
1277	1215	rancune	115.2	2470	10934	propre	48.0
1568	1820	épouvante	113.1	295	370	aversion	46.7
1767	2393	satisfaction	108.9	264	305	exaspération	46.5
1786	2504	fureur	107.1	353	553	tact	44.2
1322	1467	malaise	106.8	394	682	éprouva	43.7
1744	2447	vanité	105.8	402	717	fraternité	43.2
1881	2835	estime	104.8	7172	47935	cœur	43.0
1458	1818	ironie	104.3	391	746	éprouvais	40.6
1591	2129	effroi	104.1	281	431	mécontentement	40.0
1196	1300	Stupeur	102.9	1372	5741	profonde	38.2

304	536	patriotisme	37.9	1040	4950	vague	28.5
178	234	paternité	35.2	763	3232	sentit	28.1
192	272	culpabilité	34.9	416	1327	causait	27.7
939	3661	souffrance	34.2	8115	65419	vie	27.6
1384	6580	passion	32.9	1164	5981	sentiments	27.2
2926	17993	visage	32.3	1321	7100	donnait	27.1
156	212	rébellion	32.3	224	507	indéfinissable	27.1
224	406	ressentait	31.9	454	1557	douloureuse	27.1
187	308	instinctive	31.1	257	646	inspirer	26.7
211	396	solidarité	30.2	710	3059	physique	26.6
1167	5565	expression	30.1	212	476	indicible	26.5
2784	17653	regard	29.8	300	851	soudaine	26.1
199	369	témoignait	29.6	417	1425	infinie	26.0
1259	6337	cri	29.1	103	145	vergogne	25.6
867	3819	curiosité	28.6	566	2289	certitude	25.5
294	739	exprimait	28.6	194	435	ressentir	25.4

Dans ce dernier cas où la différence s'estompe entre sensation et sentiment, la construction s'étend démesurément, le verbe *sentir* pouvant régir de si nombreux objets. Pour écarter ces associations fugitives où le cœur n'est pas véritablement impliqué, on a fixé un seuil de fréquence², qui en outre a permis de filtrer et de rejeter les rares cas où l'on a affaire à un génitif subjectif et où le sentiment est synonyme de jugement (*sentiment de spécialiste* = jugement porté par un spécialiste).

3 - On obtient en définitive une liste de 120 unités. Avant de l'accepter, il convient de la tester, ce qu'on peut faire en consultant le corpus de FRANTEXT. Une procédure — nouvelle et puissante — du logiciel d'interrogation permet de connaître l'environnement spécifique d'un mot ou d'une liste de mots. On obtient la liste, triée selon ses vœux, des formes le plus souvent associées au mot pris pour pôle (ou aux éléments de la liste primitive) à l'intérieur de la même phrase³. La nouvelle liste contient nécessairement la liste de départ puisque celle-ci fait l'objet d'une recherche systématique mais elle propose aussi des éléments annexes qui font partie du thème et auxquels on n'avait pas d'abord pensé. L'examen d'une telle liste (voir tableau 1) donne lieu à des repêchages justifiés (par exemple : *apaisement, avarice, bonté, chagrin, consolation, contentement, découragement, dévouement, douceur, dureté*, etc.). Précisons que la procédure de FRANTEXT délivre les fréquences absolues des mots trouvés dans l'entourage de ce qu'on cherche et que de telles fréquences ne sont rigoureusement interprétables que si on les convertit préalablement en une mesure probabiliste. Celle que nous avons choisie, l'écart réduit, sert de critère de classement au tableau 1.

On remarquera que les différences ne sont pas considérables avec le classement qu'on pourrait faire sur la fréquence absolue⁴. Mais la correspondance

n'est vérifiée que pour la première centaine de sièges qui sont quasiment réservés par les conditions de l'expérience. Au delà apparaissent les formes dont la présence n'est pas imposée. C'est là qu'on repère les sentiments injustement oubliés, qui voisinent avec des verbes affectifs ou expressifs⁵ (*éprouvait, inspirait, sentait, éprouva, éprouvais, ressentait, témoignait, exprimait, sentit, causait, donnait, inspirer, ressentir, etc.*), mais aussi beaucoup d'adjectifs qui caractérisent l'intensité ou la nature de l'affection (*profonde, instinctive, vague, indéfinissable, douloureuse, physique, indicible, soudaine, infinie, inexprimable, sourde, profond, poignante, fraternelle, filiale, plein, feinte, atroce, insurmontable, enfantine, vif, affreuse, folk, inconsciente, inexplicable, absolue, irrésistible, sournoise, etc.*). Il n'y a pas lieu de s'étonner de trouver tant de formes féminines parmi les adjectifs, car la majorité des termes qui définissent un sentiment appartiennent — comme c'est le fait plus général des abstractions — au genre féminin. Mais l'abondance des superlatifs montre la tendance hyperbolique de la littérature romanesque, où les sentiments atteignent d'emblée leur intensité maximale en échappant à l'analyse. Dans l'univers du cœur le romancier manque d'outils descriptifs alors que sa trousse est complète dans l'univers physique : d'où tant de notations vagues (*vague, sourd, sournois*) ou négatives (sur le modèle *indicible* ou *inexprimable*) où s'avoue l'impuissance à décrire.

-II-La phrase

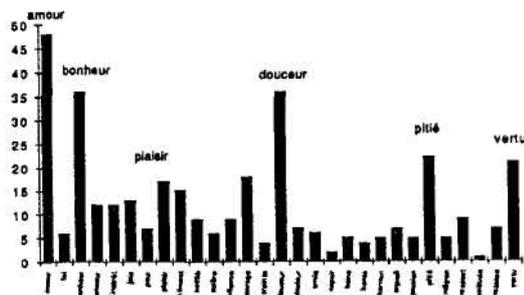
Avec les retouches qu'on vient d'apporter, la gerbe des sentiments contient désormais 165 espèces. Mais ces espèces qui constituent le champ sémantique y sont-elles semées régulièrement ? Ou voit-on s'opérer des alliances privilégiées qui accaparent un coin du terrain, tandis que d'autres regroupements auraient le champ libre ailleurs ? S'il en est ainsi, peut-on y opérer un tri et observer les affinités et les incompatibilités ? Plusieurs approches sont possibles, dont l'une prend pour base l'unité large du roman, et l'autre la cellule étroite de la phrase. Dans la première perspective, la contrainte est lâche puisque l'on ne considère pas la distance qui s'établit sur le terrain entre deux variétés étudiées. La cooccurrence n'y est appréciée qu'au niveau global du texte et n'implique aucunement des relations de voisinage. C'est en revanche la cohabitation dans le même espace restreint qui est mesurée dans la seconde perspective et que nous voulons tenter. Ce faisant, c'est une sorte de combinatoire ou syntaxe sémantique qu'on essaie de préciser. En dehors des contraintes de la syntaxe, comment les mots se marient-ils ? Par affinité ? par complémentarité ? par opposition ? Les

synonymes ont-ils tendance à se donner la main ? Le marquage mutuel des antonymes est-il plus étroit ? Quelle sorte de couple voit-on le plus souvent dans la phrase : celui qui apparie les semblables, comme la *pitié* et la *compassion* ou celui qui apparie les contraires, comme le *vice* et la *vertu* ?

1 - Là encore la réponse est dans FRANTEXT. La même fonction du logiciel d'interrogation peut être sollicitée pour étudier l'environnement immédiat de chaque variété de sentiment. Chaque interrogation produit une longue liste dont on extrait les 165 éléments étudiés. En théorie il faudrait procéder à 165 traitements pour constituer les 165 colonnes du tableau et l'on obtiendrait un tableau carré représentant le croisement de notre liste avec elle-même. Comme le corpus est considérable et la procédure de recherche onéreuse, on a réduit le nombre de colonnes aux 10 unités les mieux représentées puis aux 29 premières. Comme les résultats de ces deux essais n'ont guère varié, on n'a pas poussé plus avant l'expérience. Les données à exploiter sont donc celles du tableau 2 qui croise les 29 sentiments les plus fréquents (ayant plus de 5 000 occurrences de 1830 à 1970, tous genres mêlés) avec les autres éléments de la liste. Les deux groupes, celui qui forme les colonnes et celui qui constitue les lignes, ont d'ailleurs un poids comparable : respectivement 281 313 et 274 380 occurrences.

Le score obtenu par chaque espèce se lit dans la ligne ou colonne marginale, là où sont calculés les totaux. Ce résultat dépend certes de la fréquence du mot dans le discours mais il mesure aussi la cote de popularité de l'espèce concernée dans la société des sentiments. *L'amour* avec 3 790 suffrages l'emporte parmi les colonnes pour la première raison, mais c'est la seconde qui hisse le mot *sentiment* au deuxième rang (2 048). À l'inverse la *religion* (256 bulletins) est fort délaissée et le score médiocre obtenu par *l'intérêt*, *V honneur*, la *confiance*, le *respect* et la *vertu* (500 votes chacun environ) témoigne de la moindre intégration de ces mots dans le champ sémantique. Des observations semblables pourraient être faites à partir des éléments répartis sur les lignes du tableau. Mais abandonnant la marge, pénétrons à l'intérieur de ce tableau en suivant une colonne de bout en bout. Choisissons la seconde qui est celle de la *foi* et dont les effectifs croissent brusquement quand cette vertu théologique croise les deux autres (*espérance* 50, *charité* 47). Ailleurs les rencontres de la foi sont épisodiques (quelques unités) sauf si le sentiment prend une coloration religieuse (*certitude* 20, *enthousiasme* 16, *humilité* 13, *ferveur* 10).

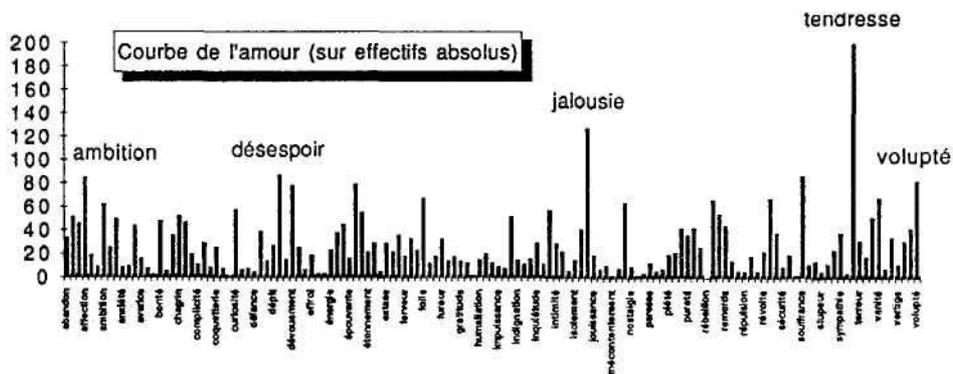
Graphique 3. Courbe de bonté



Si l'on suit une ligne, par exemple celle de la *bonté*, la lisibilité est tout aussi immédiate. La *bonté* assemble ce qui lui ressemble : la *douceur*, *l'amour*, le *bonheur*, *la pitié*, la *vertu*, (figure 3). Ces deux exemples invitent à conclure que le principe qui préside aux rapprochements des mots dans la chaîne du discours est celui de la ressemblance sémantique. Plus nombreux sont les sèmes partagés et plus grande serait l'attraction mutuelle.

L'exemple de *l'amour* qu'on lit en première colonne est loin d'être aussi clair. Le graphique ci-dessous met en exergue les liens privilégiés que ce terme entretient avec des variantes proches comme *tendresse* ou *volupté* mais aussi avec des sentiments rivaux comme *l'ambition*, ou des avatars dégradés comme le *désespoir* et la *jalousie*.

Graphique 4.

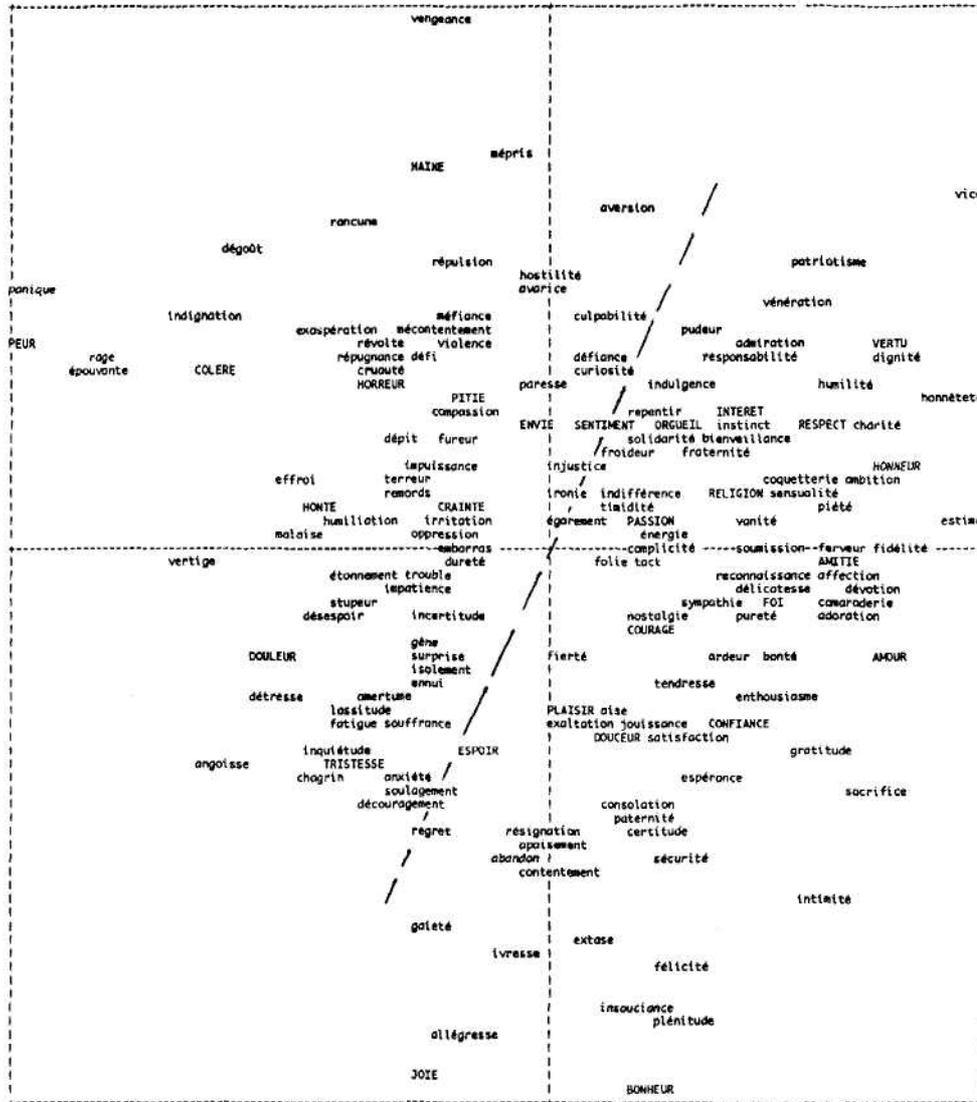


2 - 11 faut donc se méfier des sondages trop rapides et envisager le tableau dans son intégralité. Il s'agit de radiographier cette tapisserie où les fils des sentiments parcourent la chaîne et la trame. Comme l'ordre alphabétique a présidé à la disposition des lignes et des colonnes, la tapisserie prend un aspect artificiel

et bariolé où n'apparaît rien qui soit immédiatement perceptible. La lisibilité se ferait jour si l'on déplaçait les lignes et les colonnes de telle façon que les sentiments qui vont ensemble se massent aux deux coins de la diagonale principale (où seraient concentrés les effectifs les plus forts). On n'a pas essayé de réaliser cette échelle de Guttman, bien que certaines techniques, dont l'une purement visuelle et manuelle, permettent d'en optimiser l'élaboration. Car comme c'est devenu l'usage en pareille situation, on aura recours à l'analyse factorielle, et particulièrement à la méthode de "correspondance" de J.-P. Benzecri, laquelle permet l'interprétation simultanée des individus et des variables, en l'occurrence des 165 « sentiments » répartis sur les lignes et les colonnes du tableau 2. On en profitera pour se libérer des perturbations que l'inégalité de poids peut engendrer dans les effectifs absolus. Les données du tableau 2 seront d'abord converties en écarts réduits puis translétés dans la zone positive en gardant leurs distances respectives, avant d'être proposées à l'analyse. Le résultat apparaît sur la page suivante.

Une évidence s'impose : le tableau 5 obéit à la technique du camaïeu. Pas de heurts de couleurs, partout des dégradés. Dans l'enceinte où chaque sentiment doit choisir sa place, chacun se dirige vers ceux qui ont la même couleur en vertu du principe : qui se ressemble s'assemble. Ainsi le quadrant supérieur gauche est tout entier rempli de teintes sombres, de sentiments violents où se manifestent tous les degrés de la *PEUR* (*crainte, effroi, terreur, épouvante, panique*), de la *HONTE* (*malaise, embarras, humiliation, remords, impuissance*), de la *COLÈRE* (*irritation, dépit, défi, indignation, exaspération, rage, fureur, révolte*), de *Y HORREUR* (*répugnance, répulsion, dégoût, aversion*), et de la *HAINE* (*mépris, rancune, cruauté, hostilité, violence, vengeance*). On cherchera en vain dans cet enfer un sentiment positif ou seulement neutre. Tout y est noir et sanglant, à l'exception des deux anges *pitié* et *compassion* qui, comme Eloa, ont suivi Satan dans sa chute. À l'opposé, le bas du graphique représente un Éden où tout est *paix* (*apaisement, consolation, résignation, insouciance, sécurité, certitude*), *PLAISIR* (*aise, jouissance, satisfaction, contentement, abandon, douceur*), *JOIE* (*gaieté, allégresse, ivresse, exaltation*) et *BONHEUR* (*félicité, plénitude, extase*)⁶. Visiblement dans le court espace de la phrase les romanciers cultivent les relations de bon voisinage, même s'il est possible qu'à des niveaux de segmentation supérieurs, ces relations se détériorent et qu'apparaissent alors des tensions, des oppositions, des antithèses. Sans doute avons-nous là l'effet d'une loi de composition qui est analogue à la règle des trois unités du théâtre classique et qui s'appelle pareillement l'unité de ton. On a le soupçon que ces effets harmoniques se prolongent au delà de la phrase, ayant remarqué à plusieurs reprises (chez Proust, Zola et Hugo) que les phrases ne sont pas autonomes et qu'un rythme plus large les traverse tour à tour comme la houle soulève les

Graphique 5. Analyse factorielle des relations entre sentiments



vagues. On pourrait le prouver en utilisant une possibilité de FRANTEXT qui autorise l'examen des phrases précédente et suivante. Mais le long traitement qu'on vient d'accomplir serait à reprendre.

3 - L'examen attentif du graphique 5 révèle des faits plus complexes qu'une simple distribution linéaire des sentiments, des plus sombres aux plus clairs. Cette opposition qui correspond au premier facteur suit l'axe vertical et met en regard la gauche où la mer d'inimitié occupe tout l'espace et la droite où se ramifie le cours du Tendre. Mais une autre tension qui correspond au second facteur se superpose à la première en opposant le haut et le bas, c'est-à-dire les sentiments qui, positifs ou négatifs, ont un caractère moral et social (on les trouve dans la partie supérieure) et ceux qui n'engagent que l'individu et qui prennent position au bas du graphique. Ainsi parmi les éléments négatifs on distinguera ceux qui mettent autrui en cause (ils sont en haut autour de la *HAINE* et nous venons de les détailler) et ceux où l'individu déplore seulement son inconfort (*gêne, trouble, inquiétude, incertitude, vertige, étonnement, surprise, stupeur*), sa *SOLITUDE* (*isolement, ennui*), sa *TRISTESSE* (*lassitude, fatigue, amertume, chagrin, regret*), et sa *DOULEUR* angoissée (*souffrance, angoisse, découragement, désespoir, anxiété, détresse*). De même au salut individuel et égoïste qui se protège au bas du graphique à droite, s'oppose la morale sociale qui développe ses beaux sentiments, en haut et à droite, autour de la *vertu*, du *respect* et de *l'honneur* (*dignité, honnêteté, vénération, estime, admiration, responsabilité, solidarité, bienveillance, indulgence, fraternité, charité*). Ces valeurs sont aussi celles de la *RELIGION* qui prend position à cet endroit, non loin de la *FOI* (*piété, pureté, ferveur, dévotion, adoration*). Mais l'amour de Dieu utilise souvent les mêmes mots que l'amour humain et cette zone qui est proche de l'axe central à mi-chemin entre la société et l'individu est celle de la *CONFIANCE* (*espérance, gratitude, reconnaissance, sacrifice, bonté*), de *l'AMITIÉ* (*sympathie, camaraderie, complicité, fidélité, délicatesse*) et de *l'AMOUR* (*affection, soumission, tendresse, ardeur, folie*). Reste une plage incertaine, près de l'origine des axes. Elle est fréquentée par les indécis, dont le signe est ambigu, comme *l'orgueil, l'intérêt*, la *passion* et le *sentiment* lui-même.

Au total les deux principes de division ne sont pas totalement orthogonaux et indépendants. Quelques éléments positifs (comme *la joie, l'ivresse* et *l'allégresse*) franchissent, à gauche et en bas, l'axe des X et inversement cette même frontière est dépassée, en haut à droite, par quelques éléments peu recommandables comme *la défiance, l'aversion* et le *vice* (ce dernier sans doute attiré par la *vertu* toute proche⁷). C'est pourquoi il serait peut-être opportun d'opérer une légère rotation des axes, comme indiqué en pointillé sur le graphique. La diagonale ainsi créée ferait une décantation plus pure entre l'huile des sentiments doux et le vinaigre des réactions acides.

-m-

L'évolution

Cette distinction proposée par l'analyse quantitative entre les bons et les moins bons sentiments va être utile pour apprécier, sous ce rapport, l'évolution de la production romanesque de 1830 à 1970.

Mais dans un premier temps reprenons la liste sans faire de distinction. Les 165 éléments sont regroupés dans un même ensemble dont on suit la progression au cours du temps. En choisissant une périodicité de cinq ans on obtient une série de 28 observations. FRANTEXT offre de grandes facilités pour de telles recherches et s'il ne délivre pas directement toutes les informations contenues dans le tableau 6, du moins permet-il de les reconstruire⁸. Les hasards de la publication et de la sélection font que les tranches n'ont pas la même étendue, le rapport étant de 1 à 20 entre les extrêmes (voir dernière colonne du tableau 6). Mais le calcul opère les pondérations nécessaires et le test révèle presque partout des écarts significatifs (un seul est au-dessous du seuil à 5 %). Cela n'est pas pour surprendre, étant donné la taille du corpus étudié (40 millions de mots) et la quantité d'observations recueillies (260 672 occurrences⁹). La loi des grands nombres joue ici son rôle et nous éviterons de crier au miracle devant un événement que le hasard ne devrait pas produire. Dans le domaine du langage et généralement des comportements humains ce serait un grand hasard que le hasard fût seul en cause.

Tableau 6. L'évolution des sentiments

N	Tranche	Réduit	Réel	Théo.	Taille
1	1830	35.79	13608	10084	1524558
2	«35	44.27	15568	11020	1666032
3	1840	7.23	5353	4854	733818
4	1845	2.94	12277	11963	1808527
5	1850	0.67	4311	4268	645208
6	1855	-10.19	7587	8512	1286801
7	1860	-18.41	9527	11453	1731496
8	1865	-3.10	7496	7765	1173868
9	1870	-20.35	8032	10030	1516375
10	1875	-17.44	6270	7786	1177045
11	1880	-1.64	13530	13717	2073748
12	1885	13.75	12747	11317	1710843
13	1890	9.08	10209	9347	1413043
14	1895	5.92	4569	4189	633297
15	1900	-1.65	5780	5906	892821
16	1905	3.24	8826	8532	1289850
17	1910	11.35	11520	10387	1570234
18	1915	-7.29	6592	7202	1088782
19	1920	7.90	16659	15700	2373474
20	1925	13.63	16816	15186	2295804
21	1930	-11.30	12700	14001	2116615
22	1935	-7.47	18734	19743	2984731
23	1940	-11.16	3369	4076	616196
2	1945	-11.08	12214	13467	2035867
25	1950	-2.95	5167	5382	813573
26	1955	-6.23	846	1047	158312
27	1960	-25.09	2649	4276	646415
28	1965	-18.31	7716	9465	1430932

La courbe 7 représentée ci-dessous traduit graphiquement la série des écarts réduits. L'orientation du sentiment est à la baisse : les premières tranches ont l'excédent le plus élevé, et les dernières le déficit le plus accusé. Le coefficient de corrélation qui mesure la pente descendante est formel : avec une valeur de -0,47 le seuil de 1 % est atteint. Du romantisme au nouveau roman on assiste donc à une décrue, et presque à un assèchement du sentiment.

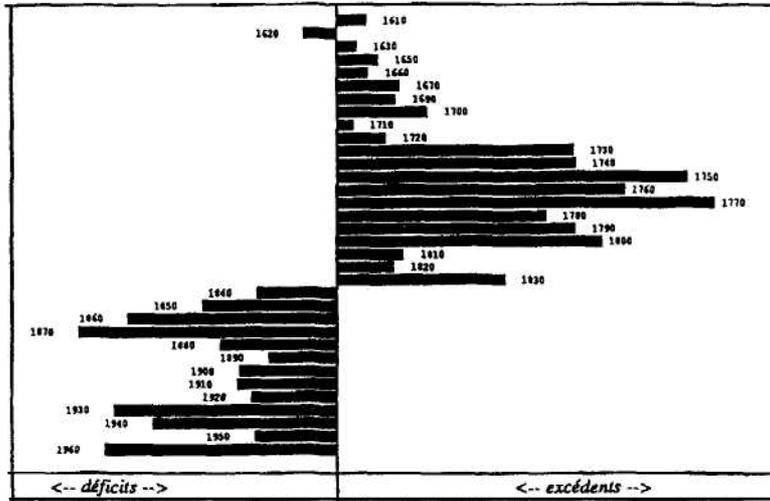
Mais cette orientation générale n'est pas régulière. Le cours du sentiment tout au long de 140 ans se resserre ou se répand au gré de l'histoire des sensibilités. La crue du romantisme est suivie d'une période aride qui coïncide avec le réalisme et le naturalisme. Puis à partir de 1885 les sources du sentiment retrouvent la vigueur, au moment où le renouveau du spiritualisme irrigue la pensée et la littérature françaises. Cela se prolonge jusqu'à Proust. Après quoi le roman français hésite à s'abreuver aux sources psychologiques, et cherche d'autres voies, du côté de l'engagement politique, de l'invention poétique, de la description ou de la célébration des choses, ou parfois du simple jeu verbal.

Graphique 7. Courbe de l'évolution du sentiment



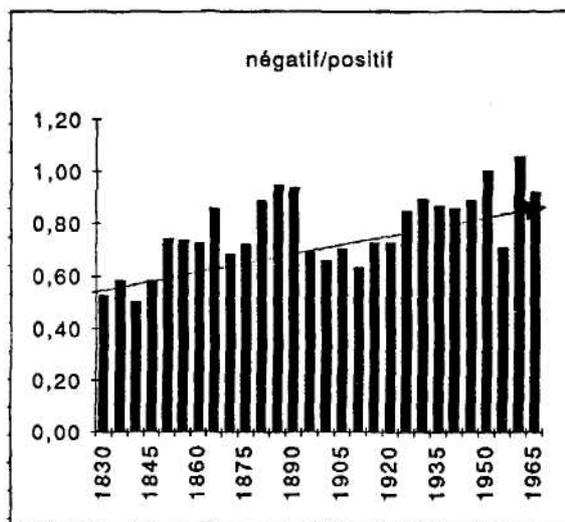
Est-ce que cette courbe en S reproduit ces mêmes méandres dans les siècles passés ? La réponse est dans le graphique 8 qui remonte la chronologie jusqu'en 1600 et ajoute au recensement 300 autres romans. Comme les tranches sont deux fois plus longues (10 ans) un phénomène de lissage se produit qui arrondit les angles. Les mouvements courts qu'on vient de décrire subsistent mais ils sont noyés dans une perspective d'ensemble qui offre une grande netteté : toutes les tranches (sauf une) sont largement excédentaires jusqu'en 1830, toutes sont déficitaires après cette date. Ainsi le mouvement de décrue qu'on avait observé, malgré les dents de scie, dans les deux derniers siècles se trouve considérablement amplifié lorsque le regard s'étend sur quatre siècles (coefficient -0,52, probabilité 0,001). On a longtemps affirmé que le roman français était porté à la psychologie. Cette originalité semble se réduire au cours du temps.

Graphique 8. Évolution sur quatre siècles



Si la part du sentiment s'amenuise ainsi dans le roman, est-ce que sa coloration se maintient ? Cultive-t-on les mêmes espèces aujourd'hui qui avaient la faveur des romantiques ? Un romancier du XX^e siècle a prétendu qu'on ne faisait pas de bonne littérature avec les bons sentiments. Or l'évolution du roman semble donner raison à Gide, puisque les romanciers modernes tendent à préférer les sentiments noirs. Si l'on établit le quotient entre la part sombre du sentiment et sa face claire (en suivant la ligne de démarcation établie par la diagonale dans la figure 5), on observe un assombrissement progressif, malgré une rémission provisoire à la fin du XIX^e siècle¹⁰. Voir graphique 9.

Graphique 9. Assombrissement du sentiment



-Iv-

Les auteurs

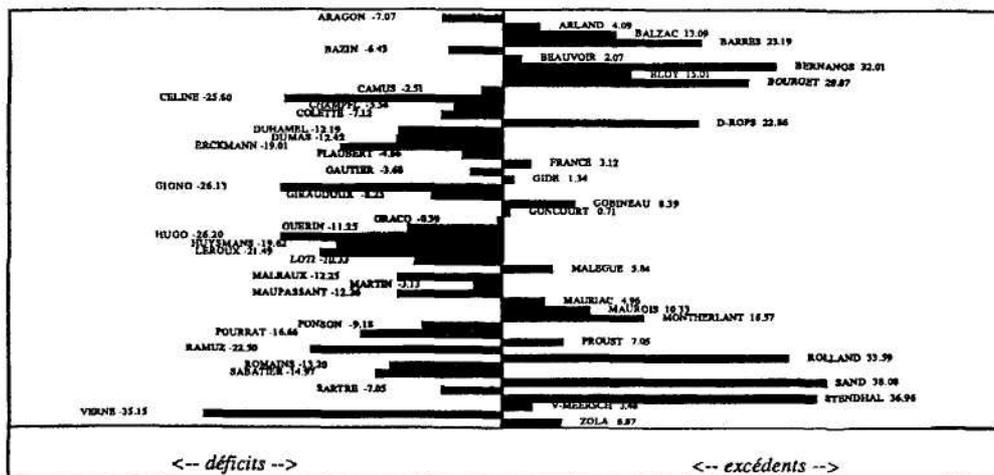
Les faits de chronologie qu'on vient d'aborder ne relèvent pas d'une sorte de gravitation universelle, d'un courant irrésistible auquel l'écrivain ne saurait échapper. La dérive du temps ne se voit guère qu'en se retournant quand on mesure le chemin parcouru. Chaque romancier — cela est vrai aussi de tout artiste et de tout savant — avance de son côté pas à pas, comme le filet d'eau progresse goutte à goutte sur la terre sèche. En fin de compte cela produit le lit du fleuve ou le sens (rétrospectif) de l'histoire. Bien loin que le temps, en dehors des effets peu profonds de la mode, engendre et pilote les écrivains, ce sont les écrivains — et les créateurs plus généralement — qui forment les jalons et permettent le repérage du cours du temps. Aussi l'étude chronologique qui précède resterait abstraite si aucun de ces jalons n'était localisé. On aurait mesuré un effet sans aller jusqu'à la cause. Il importe donc de compléter et d'expliquer l'évolution en isolant les textes et les auteurs qui participent au mouvement.

Fort heureusement une fonction de FRANTEXT¹¹ permet pour un mot donné (ou une liste de mots) de connaître la répartition de ce mot par auteurs à l'intérieur du corpus considéré. Les résultats sont fournis dans la même présentation que ceux de la chronologie, à savoir : fréquence réelle, fréquence relative et histogramme. Nous opérerons sur ces données les mêmes transformations que précédemment, pour aboutir à des écarts réduits¹². Une correction supplémentaire a été apportée, qui filtre les auteurs pour ne retenir que ceux dont l'importance est reconnue par les chiffres et dont les œuvres ont dans la base une étendue minimum. Comme le corpus réunissait trop de romanciers — dont certains peu connus et mal représentés — on a fixé la barrière à 200 000 occurrences et seuls 51 auteurs ont été retenus sur les 167 recensés, soit un tiers. Mais cet effectif réduit au tiers réunit les trois quarts des occurrences du champ sémantique du sentiment (190 046 sur 260 672), si bien que l'essentiel de l'information est conservé. Au reste les calculs ont été faits en prenant en compte la totalité des auteurs et des occurrences¹³. Le résultat apparaît dans le graphique 10.

L'axe vertical y sépare les auteurs où le sentiment déborde (à droite) et ceux qui, par pudeur ou désintérêt, s'en éloignent (à gauche). On prendra garde que les romanciers ici représentés ne bénéficient pas de l'exhaustivité. Le corpus par exemple ne considère qu'un seul roman de Aragon (*Les Beaux Quartiers*), 2 de Céline (*Voyage au bout de la nuit* et *Mort à crédit*), 5 de Colette, 3 de Bloy. En revanche Bernanos compte 12 titres, Bourget 16, Balzac 15, et Zola 22. Il arrive que la sélection ne donne qu'une image partielle d'un auteur : ainsi les 4 romans choisis pour Giono appartiennent à la première période (1929-1931) qui est

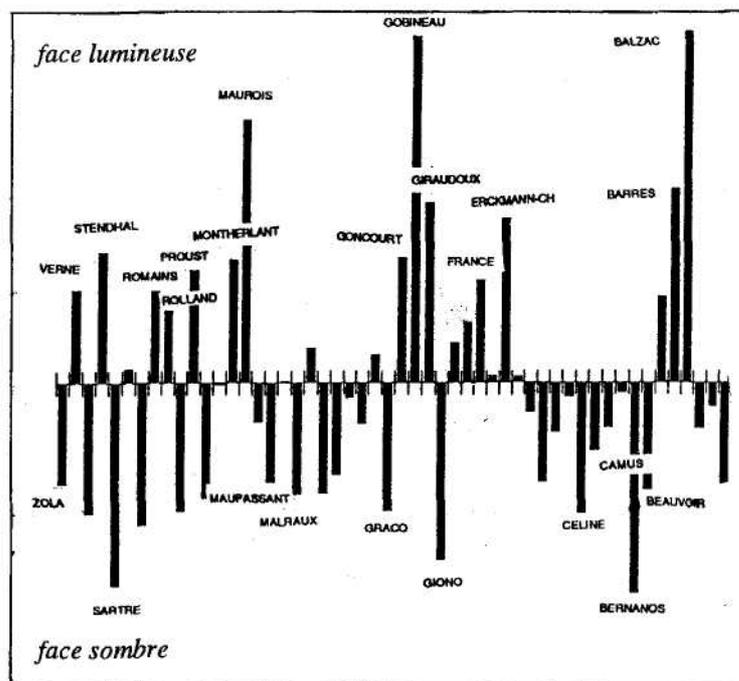
lyrique et épique. Avec cette réserve, la courbe n'en est pas moins révélatrice des tendances. La tradition du roman psychologique est illustrée par Stendhal qui se classe en tête. La part des sentiments y est prépondérante. Elle est forte aussi chez Balzac, Sand, Bourget, Barrés, Bloy, Bernanos, Romain Rolland, Montherlant, Mauriac et Proust, tous ces écrivains relevant à des degrés divers du courant spiritualiste. Les romanciers réalistes ou naturalistes sont moins soucieux d'états d'âme et dans la zone négative on trouve Champfleury, Flaubert, Maupassant, Erckmann-Chatrian, Huysmans (mais Zola, qu'on attendait à cet endroit, leur a faussé compagnie). Dès que l'intérêt majeur est porté ailleurs, à la guerre, aux réalités sociales ou politiques, au débat d'idées ou à la description des choses, des paysages ou des gens, dès que le ton se fait poétique, épique, philosophique, critique, voire comique, la part des sentiments s'amenuise. C'est pourquoi la liste de gauche est fort hétérogène, étant constituée de romanciers qui, à ces titres divers, rompent avec la tradition psychologique. Et aux romans réalistes viennent se joindre les romans épiques (Hugo, Giono, Céline), poétiques (Giraudoux, Colette, Gracq), régionalistes (Pourrat, Ramuz), politiques (Malraux), existentialistes (Camus, Sartre), ou les romans d'aventure (Dumas, Gautier, J. Verne). Si toutes les époques ont des représentants dans les deux camps, on voit que les romanciers modernes sont plus souvent dans le camp de la rupture. Et cette tendance serait sans doute mieux marquée si quelque adepte du nouveau roman figurait dans les données.

Graphique 10. Histogramme du sentiment chez les romanciers

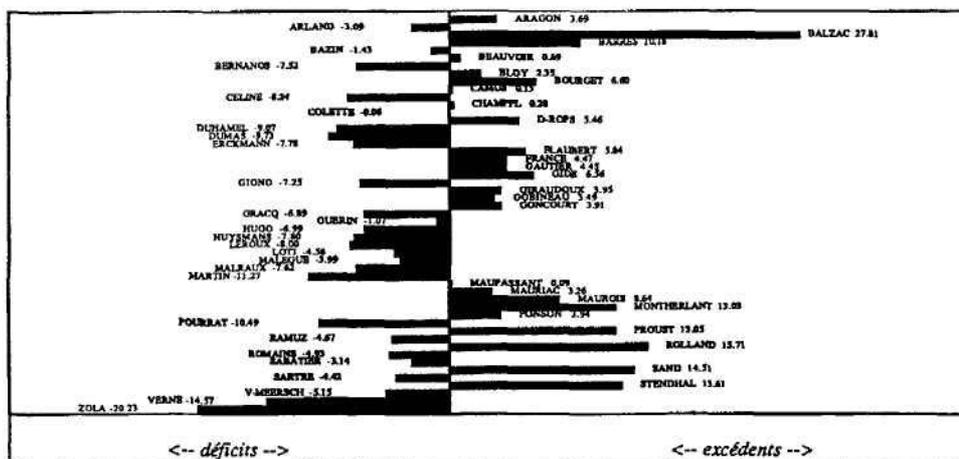


Bien entendu il n'est pas de romans où n'éclate, à un moment ou à un autre, l'expression d'un sentiment, d'autant que la ferveur religieuse, l'ambition politique, l'ardeur poétique, ou le désespoir métaphysique sont aussi des sentiments, à l'instar de la passion amoureuse. Mais celle-ci domine dans le roman et le sentiment s'affadit quand le cœur n'y est plus. Toutefois les occurrences du sentiment se comptent par milliers, même chez les romanciers réticents. Et les effectifs sont partout suffisants pour qu'on procède à un quotient qui mette en relation les « bons » et les « mauvais » sentiments (la séparation entre les uns et les autres restant celle de la barre oblique du graphique 5). Le graphique 11, établi à partir de ce rapport, distingue deux zones. En haut, dans la sphère positive, se tiennent les auteurs « optimistes » qui montrent la face lumineuse des sentiments. Derrière Balzac, leur chef de file, se rangent Stendhal, Jules Verne, Barrés, R. Rolland, J. Romains, Giraudoux, Montherlant. En bas, la zone négative, où régne le malheur, la peur et la violence, est fréquentée par les romanciers plus sensibles au mal. Zola est ici où on l'attend, avec Maupassant (Flaubert est neutre). Sont là aussi, en compagnie des existentialistes (Sartre, Beauvoir et Camus), les romanciers qui partagent une idée tragique, quoique différemment tragique, de l'existence : Céline, Bernanos, Giono (du moins celui qui se trouve dans Frantext) et Malraux¹⁴.

Graphique 11. La coloration du sentiment chez les romanciers



Si l'on renâcle devant la dichotomie trop manichéiste qu'on vient de mettre en œuvre, on peut rendre à chaque sentiment sa liberté et à chaque romancier son indépendance. Faisant table rase de tout présupposé, on peut reconstituer un tableau de contingence en cherchant les 165 formes chez les 51 romanciers. En tenant compte de l'étendue, on peut alors dresser le profil de l'emploi d'un mot parmi les auteurs, ou le profil d'un auteur particulier à travers les sentiments dont il fait mention. Plus de 200 histogrammes peuvent ainsi être détaillés, selon qu'on isole une ligne (un sentiment) ou une colonne (un romancier). On donnera un exemple de chaque espèce. Celui du graphique 12 concerne la forme la plus sollicitée : *Y amour*. Balzac l'utilise plus que les autres (plus que les autres écrivains et aussi plus qu'il n'utilise les autres formes du sentiment¹⁵). Zola y a recours le moins souvent. Entre ces deux extrêmes, les romanciers choisissent leur camp : Stendhal, Proust, G. Sand, R. Rolland, Montherlant prennent le parti de l'amour, et plus mollement Mauriac, Giraudoux, Gide, Bourget, Flaubert. Par contre le mot ne trouve guère l'occasion de s'employer, pour des raisons diverses, sous la plume de J. Verne, Dumas, Martin du Gard, Giono, Duhamel, Malraux, Céline et Sartre.

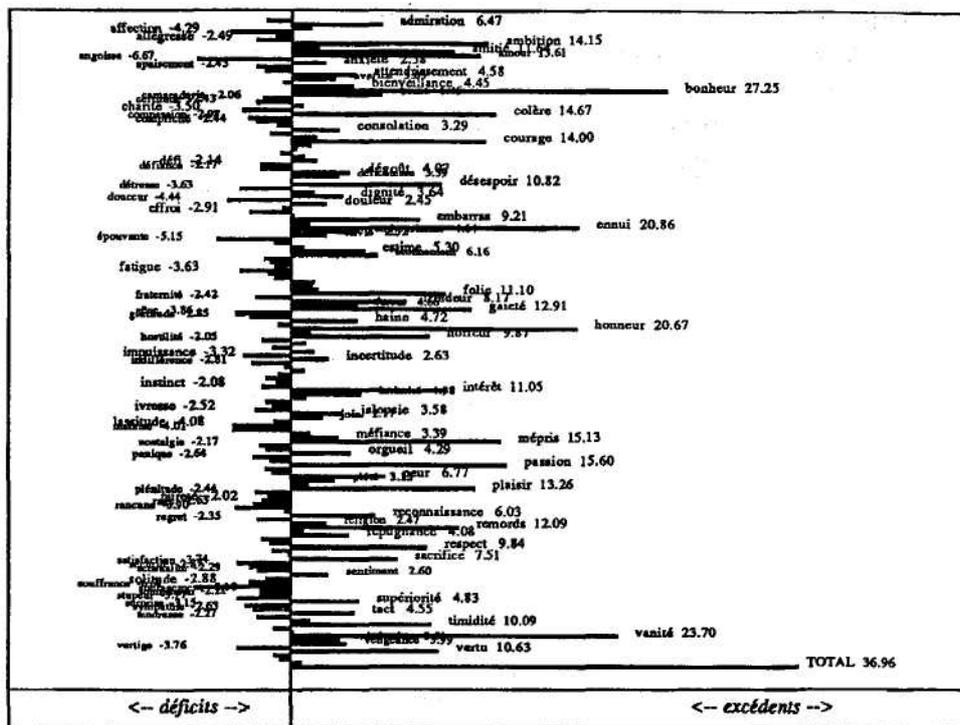


Graphique 12. Courbe du mot *amour*

Stendhal nous fournira le second exemple (figure 13). Quoique *l'amour* et *l'amitié* arrivent en bon rang dans la thématique stendhalienne, c'est au *bonheur* que revient la première place. Et ici la statistique confirme une observation qui n'a pas échappé à la critique. Mais *l'honneur* se hisse au même niveau, l'un

n'allant pas sans l'autre. La *virtus* stendhalienne est faite *déplaisir* et de *passion*, de *gaieté* et de *folie*, de *courage*, de *sacrifice* et de *vertu*, d'*ambition* et de *vanité*, d'*orgueil* et de *timidité*, de *respect*, d'*estime* et de *dignité*, de *froideur*, de *mépris* et de *colère*. L'action se mêle au sentiment et le suscite, rien n'étant pire que *l'ennui*.

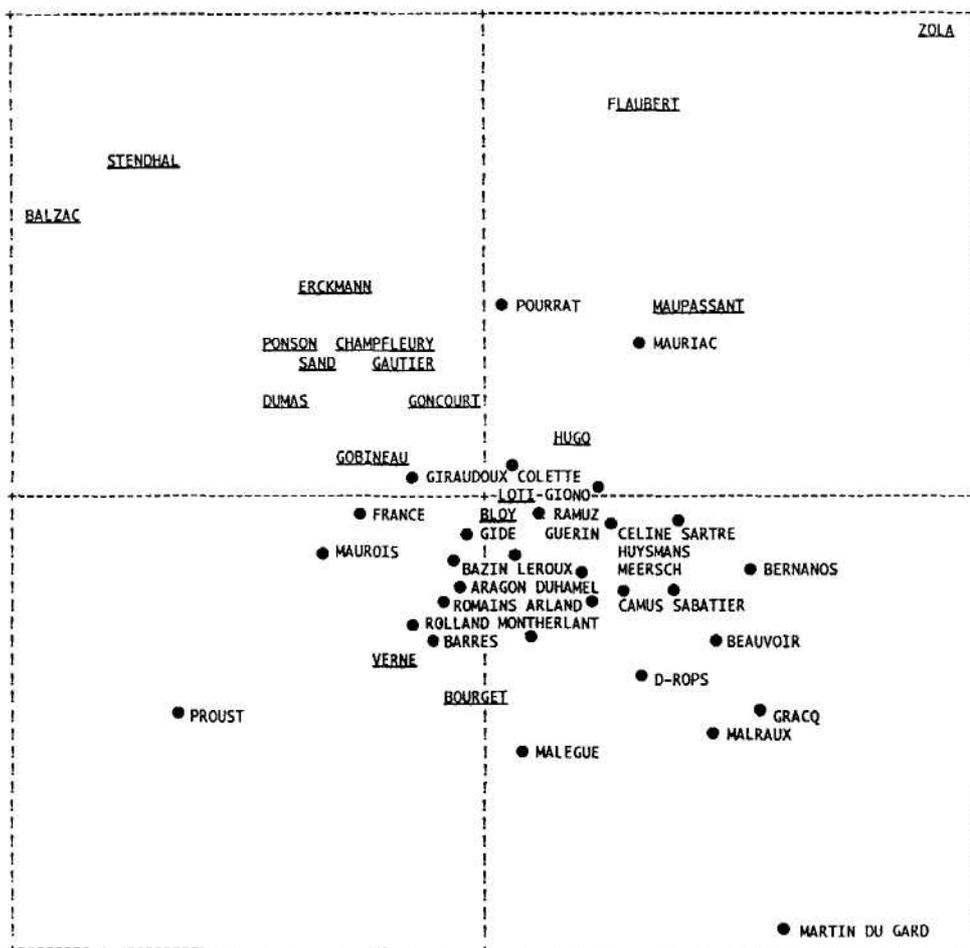
Figure 13. Le profil de Stendhal



La place nous manque pour rendre compte des autres profils : celui de Proust met en avant la *jalousie*, le *plaisir* et la *souffrance*. Chez Hugo c'est *l'anxiété*, *l'épouvante*, *l'effroi*, *l'égarement* et la *stupeur*, mais aussi *l'extase*, *l'adoration*, la *fraternité* et la *pitié*. Chez Gracq le roman secrète une angoisse sourde et provoque *malaise*, *gêne* et *vertige*... On peut toutefois s'élever à une vue panoramique en recourant à l'analyse factorielle. Dans celle qu'on a représentée ci-dessous (figure 14), cinquante romanciers prennent place. Les fortes personnalités s'isolent aux quatre coins du graphique : Balzac et Stendhal en haut

à gauche, Flaubert, Maupassant et Zola en haut à droite, Proust en bas à gauche et Martin du Gard en bas à droite¹⁶. On admirera en particulier la ligne de fuite qui enchaîne les uns aux autres, et dans la succession des écoles, les grands romanciers du XIX^e siècle, de Balzac à Zola. On connaît les sentiments d'estime qui ont uni Stendhal et Balzac d'une part, Flaubert et Zola d'autre part. On est presque reconnaissant à la machine qui, incapable elle-même de sentir, a su préserver les couples en dénouant les fils du sentiment. Mais en réalité ce qui unit les grands romanciers du XIX^e, c'est moins leur amitié personnelle, que

Figure 14. Analyse factorielle du sentiment chez les romanciers

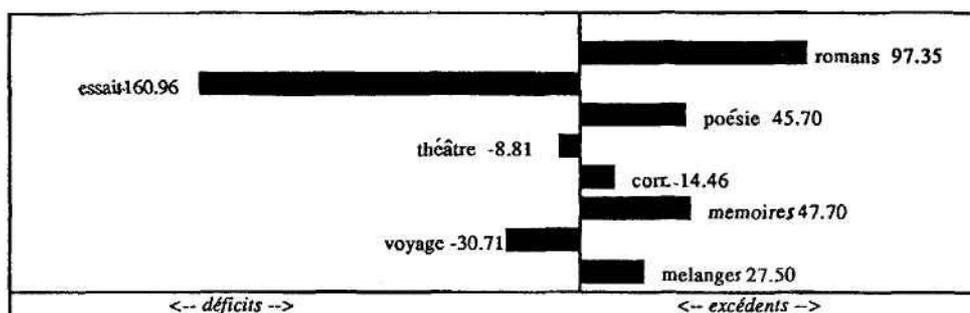


l'appartenance à la même époque et le partage d'une même sensibilité diffuse dans l'esprit du temps. Ce qui tend à le prouver c'est l'opposition entre le haut et le bas du graphique. Presque tous les auteurs installés dans la moitié supérieure ont publié avant 1900 ; on les a soulignés d'un trait. Tous ceux ou presque qui ont pris place dans la moitié inférieure appartiennent au XX^e siècle ; on les a marqués d'un point. Il y a de rares exceptions : d'une part Pourrat et Mauriac sont restés attardés en deçà de la frontière ; leur inspiration et leur technique n'ont en effet rien qui puisse s'opposer à la tradition. D'autre part Jules Verne qui est le seul romancier de son espèce et qui laisse au sentiment une part des plus réduites, s'est situé au hasard sur l'échiquier, n'étant pas concerné par ces histoires de cœur. Mis à part ce léger flottement, la chronologie est respectée. C'est donc que le sentiment évolue et que le cœur des sociétés est instable comme celui des individus.

-IV- Les genres

Parvenu au terme de notre étude, voilà que le doute nous assiège sur le bien-fondé du point de départ. Était-il vraiment judicieux de limiter l'observation au roman ? Certes il fallait se conformer à la règle du jeu et à l'unité du volume. Et on avait *a priori* le sentiment que dans le champ romanesque la matière étudiée était à la fois plus pure et plus abondante. Essayons pourtant, *a posteriori*, de justifier cet *a priori*. Et puisons de nouveau dans l'urne de FRANTEXT, en acceptant que tous les genres y soient mélangés, mais en gardant la même fourchette chronologique, de 1830 à 1970, et la même liste de 165 formes. En réalité bien loin de confondre les genres, on va les introduire l'un après l'autre dans un nouveau tableau de données où une colonne sera dévolue à chacun d'eux (les 165 lignes restant les mêmes que précédemment). Le roman peut dès lors être confronté aux autres genres littéraires. Pour délimiter ces genres — qui se mêlent souvent dans la réalité des textes — on doit nécessairement s'appuyer sur un codage initial qui a été réalisé à Nancy et qui distingue 8 classes : le roman, le théâtre, la poésie, la correspondance, les mémoires, les récits de voyage, les mélanges littéraires et les essais. Les essais sont pourvus d'un second niveau de classement qui permet un catalogue selon les matières ou domaines abordés et que nous ignorerons. Ce codage est nécessairement approximatif¹⁷ et nous n'entreprendrons pas de justifier l'appartenance de tel ou tel texte à la rubrique où on l'a casé. De plus les classes sont d'importance très inégale — par la faute non des codeurs mais des écrivains qui cultivent plutôt les genres qui ont du succès, principalement le roman (pour un tiers) et l'essai (pour un autre tiers). Tous les

autres genres se disputent le dernier tiers, et là encore le partage est léonin, les mémoires s'arrogeant la meilleure part. On a même renoncé à accorder le droit de cité au genre oratoire qui n'a guère de représentant dans la période étudiée. Bien entendu l'inégalité des classes a imposé un traitement statistique approprié qui a fait appel à l'écart réduit et dont le résultat apparaît dans le graphique 15.



Graphique 15. Distribution du sentiment dans les genres

Nul ne s'étonnera que les sources du sentiment soient taries dans le désert des essais. Les rares espèces qu'on y rencontre (partie droite du graphique 16) sont desséchées comme du bois mort. Elles ont un caractère juridique et abstrait qui fait douter que le cœur soit véritablement intéressé : *certitude, incertitude, impuissance, dureté, énergie, instinct, intérêt, culpabilité, supériorité, responsabilité, sécurité, solidarité*. Même le mot *jouissance* apparaît décharné, car on a toute raison de penser qu'il s'agit de la jouissance d'un droit. La polysémie est à mettre en cause ici, le même mot pouvant relever de registres forts opposés suivant le contexte. Le déficit dans les récits de voyage n'est pas inattendu. Si l'on met à part les voyages de noces (mais les acteurs ne prennent guère la plume dans cette circonstance), le voyage polarise l'activité et l'attention sur les choses et les gens qu'on rencontre et avec qui des relations approfondies — et des sentiments — n'ont guère le temps de se développer. Le seul sentiment qui en voyage dépasse le seuil de l'indifférence est le sentiment de *solitude* (écart réduit +5). Voir partie gauche du tableau 16.

été éliminées. C'était la recommandation écologique de François Rastier, malheureusement peu applicable aux grands corpus. Quand les déchets ne sont pas proprement filtrés à l'entrée, on a toutes les chances de les retrouver à la sortie.

En revanche le cœur déborde dans la correspondance (écart réduit = 14,46), sans que l'étalage des sentiments implique nécessairement leur sincérité. On retrouve ici beaucoup des expressions verbales précédentes qui s'inscrivent dans des formules convenues et qui utilisent pareillement les mots *honneur, bonheur, joie, plaisir, peur, regret, envie*. S'y ajoutent les protestations *d'amitié, de sympathie, de dévouement, de reconnaissance, d'affection, d'estime, d'admiration*. Si le malheur du correspondant (le sien propre ou celui de son interlocuteur) explique en certaines circonstances l'emploi du vocabulaire de la commisération (*chagrin, tristesse, sympathie et consolation*), le climat épistolaire est généralement rassurant et exclut les nuages sombres et les sentiments noirs. Un filtrage s'y opère qui retient les sentiments de bon ton et les propos de bonne compagnie, en ravivant au besoin les couleurs grâce au lustrage de l'hyperbole.

Les mélanges littéraires, qui contiennent une réflexion critique sur la littérature, auraient pu se fondre dans le groupe indistinct des essais. Ils s'en détachent pourtant fortement avec un écart nettement positif (écart réduit - 27,5). Que fait en effet le critique littéraire, sinon expliciter les sentiments que provoque la lecture d'un texte, lequel est lui-même le plus souvent imprégné de sentiments ? Le matériau lexical de la critique — et cela vaut pour la peinture et la musique — est essentiellement fait de notations subjectives qui appartiennent à l'ordre du cœur. Même lorsque l'œuvre elle-même — peinture abstraite, nature morte ou paysage extérieur — refuse l'épanchement, le critique aura du mal à éviter le vocabulaire affectif et pour définir ou qualifier l'impression produite par le tableau, il parlera de violence, de froideur ou de sérénité.

Dans la poésie, qui ignore la démarche secondaire de l'analyse, le sentiment constitue le matériau primaire (écart réduit = +45). Certes il y a des poèmes descriptifs, narratifs, épiques où l'expression des sentiments humains ne trouve pas place, comme ces temples ou ces mosquées d'où toute figure humaine est bannie. Mais sauf lors du bref épisode parnassien, la poésie française, du romantisme au surréalisme, n'a guère cultivé la froideur marmoréenne et le sentiment y affleure partout, débordant chez Hugo, chuchotant chez Verlaine.

C'est pourtant dans le roman que les vaisseaux du cœur se dilatent le plus largement. Et la figure 17 en est une claire illustration. On y voit les sentiments irriguer la moitié droite qui est celle des excédents, avec de rares transfuges qui s'égarent dans la zone gauche et qui ont partie liée avec le genre des essais, en sorte que les figures 16 et 17 se correspondent comme le négatif et le positif du même cliché. Deux raisons expliquent cette inondation romanesque. La première

NOTES

Exactement 606 textes et 39 408 268 occurrences. Mais tant de précision n'est pas nécessaire, la définition des unités n'étant pas des plus rigoureuses. Pour des raisons qui tiennent aux limitations techniques des débuts du dépouillement, les « textes » sont parfois des morceaux de textes qui ne doivent pas dépasser 100 000 occurrences. Ainsi la *Recherche du temps perdu* compte pour 23 « textes », alors qu'en suivant les indications de Proust il fallait opter pour 7 ou pour 1. Quant aux « occurrences », il ne s'agit pas seulement des mots, mais aussi des signes de ponctuation.

Cela n'a toutefois pas suffi pour écarter certaines constructions fréquentes qui associent le sentiment à la *présence*, à la *réalité*, *l'existence*, au *devoir*, à la *puissance*, à la *force*, et où l'expression *sentiment de* ou *d' + substantif* met en œuvre le génitif objectif, sur le modèle *sentiment de danger*. En voici le relevé trié pour le corpus des romans.

16 terreur	18 certitude	18 crainte	18 souvenir
19 dégoût	19 désir	19 devoir	20 admiration
20 reconnaissance	20 sécurité	21 douleur	21 malaise
22 justice	22 supériorité	23 respect	24 colère
25 jalousie	25 tristesse	26 curiosité	28 horreur
30 peur	30 puissance	30 réalité	31 existence
32 dignité	32 honte	33 orgueil	36 bonheur
36 joie	38 pitié	40 impuissance	40 présence
44 force	45 solitude	65 amour	1841 sentiment

En de tels cas, il était difficile de faire figurer les mots simples *réalité*, *présence*, etc., parmi les sentiments à part entière, la grande majorité de leurs emplois échappant au thème qui nous préoccupe. Mais il est vrai qu'on a affaire à un continuum et que la ligne de partage ne va pas sans quelque arbitraire.

La fenêtre découpée dans le texte n'est pas nécessairement la phrase. Ce peut être un ensemble de phrases ou un ensemble de mots adjacents dont le nombre est choisi par l'utilisateur.

La référence ou norme de comparaison est fournie par l'ensemble des textes de Frantext appartenant au X^e et au XX^e siècles. La fréquence correspondante se lit dans la seconde colonne du tableau 1. La fréquence des mêmes mots dans l'ensemble des phrases retenues apparaît dans la colonne 1. Prendre garde que la même phrase apparaît autant de fois qu'y figure l'un des mots cherchés, ce qui gonfle les effectifs et accentue les écarts. Ainsi la tête de liste qui a 34 692 occurrences dans le corpus d'ensemble, en compte 17 204 dans le sous-ensemble des romans, et, compte tenu des doublons, 25 117 dans l'ensemble des phrases isolées.

La construction transitive y est de loin la plus fréquente, tout en recouvrant deux cas opposés, selon que le sujet est le siège (*éprouver*) ou la cause (*inspirer*) du sentiment. Le corpus romanesque s'y reconnaît aisément à la surabondance de l'imparfait et de la troisième personne.

Les noms de sentiments en majuscules correspondent à ceux qu'on trouve en colonne (il y en a 29). Les minuscules désignent les 135 autres, répartis sur les lignes du tableau.

La paire *vice-versu* est l'un des seuls couples antithétiques que l'analyse ne désunisse pas.

Le logiciel d'interrogation s'engage sur la voie de la fréquence relative, alors que nous préférons le calcul de l'écart réduit. Mais on peut aisément passer d'une procédure à l'autre.

9. Qu'on ne s'étonne pas de la disproportion entre ce total et celui qu'on lit dans le tableau 3 et qui est dix fois moindre (26 808). Ce dernier comptabilisait les cooccurrences (avec diverses contraintes) alors que le présent calcul s'intéresse aux simples occurrences (sans contrainte).

10. Quant à la distinction entre individuel et social, le quotient a été aussi calculé pour chaque tranche, avec un résultat décevant qui ne laisse deviner aucune tendance.

11. On se surprend à penser que certains lecteurs, peu familiers de Frantext, ignorent peut-être le nom du démiurge qui a conçu et mis en œuvre cette immense base de données, qui n'a pas d'équivalent au monde en matière littéraire. Le logiciel brille de mille feux et porte le nom de *Stella*. Et le démiurge qui l'a créé — il faut au moins un démiurge pour les étoiles — s'appelle Jacques Dendien.
12. L'exploitation proprement statistique de FRANTEXT réclame des opérations spécifiques qu'on a regroupées dans un logiciel (THIEF) adapté au standard Apple. Rien n'empêche de proposer aussi ce logiciel à la nonne IBM. Mais il serait sans doute plus opportun d'intégrer des compléments statistiques dans le logiciel STELLA — ce qui est à l'étude.
13. Quoiqu'il soit un peu tard pour s'expliquer là-dessus en fin d'article — et en doutant de la nécessité de détailler des procédures proposées par Charles Muller il y a belle lurette —, précisons à l'intention de ceux qui abordent la linguistique quantitative que l'écart réduit se calcule comme suit (à partir de la première ligne du tableau 10 consacrée à Aragon et compte tenu que le corpus des romans a une étendue de 39 408 386 occurrences et que le champ sémantique du sentiment a un effectif de 260372) :

$$\frac{-263}{\dots}$$

$$\text{fréquence théorique du sentiment chez Aragon} = \frac{260372 * 221444}{39408386} = \dots$$

$$\text{écart absolu} = 1202 - 1465 = -263$$

$$\text{écart réduit (forme simplifiée)} = \dots = -6,87$$

(la formule exacte qui fait intervenir la probabilité complémentaire aboutit à un résultat très voisin : -6,89)

14. L'étude des sentiments chez les romanciers n'est ici qu'esquissée, avec une optique assez grossière. On trouvera dans le présent volume des études plus approfondies et mieux focalisées sur un auteur ou un sentiment particulier.
15. Aussi bien dans l'absolu que proportionnellement.
16. Comme l'analyse a été faite sur les écarts réduits l'effet de taille ne joue pas — qui habituellement ramène vers le centre de gravité les effectifs les plus lourds. On constate plutôt l'effet inverse qui est d'éloigner du centre les écarts les plus forts. Car les écarts réduits ont tendance à augmenter quand agit la loi des grands nombres. La perspective se trouve inversée : au lieu de constituer la norme indistincte, près de l'origine des axes, les gros effectifs deviennent discriminants et attirent l'attention sur les marges qu'ils atteignent plus rapidement, comme ces nébuleuses qui fuient aux confins de l'univers et dont la vitesse est proportionnelle à leur masse.
17. En particulier on a du mal à accepter l'autonomie des « mélanges » qui sont en réalité des essais, la critique littéraire étant seulement un domaine technique parmi d'autres.

$$V1465$$

(la formule exacte qui fait intervenir la probabilité complémentaire aboutit à un résultat très voisin : -6,89)

14. L'étude des sentiments chez les romanciers n'est ici qu'esquissée, avec une optique assez grossière. On trouvera dans le présent volume des études plus approfondies et mieux focalisées sur un auteur ou un sentiment particulier.
15. Aussi bien dans l'absolu que proportionnellement.
16. Comme l'analyse a été faite sur les écarts réduits l'effet de taille ne joue pas — qui habituellement ramène vers le centre de gravité les effectifs les plus lourds. On constate plutôt l'effet inverse qui est d'éloigner du centre les écarts les plus forts. Car les écarts réduits ont tendance à augmenter quand agit la loi des grands nombres. La perspective se trouve inversée : au lieu de constituer la norme indistincte, près de l'origine des axes, les gros effectifs deviennent discriminants et attirent l'attention sur les marges qu'ils atteignent plus rapidement, comme ces nébuleuses qui fuient aux confins de l'univers et dont la vitesse est proportionnelle à leur masse.
17. En particulier on a du mal à accepter l'autonomie des « mélanges » qui sont en réalité des essais, la critique littéraire étant seulement un domaine technique parmi d'autres.

