



**HAL**  
open science

## La phrase de Proust. Longueur et rythme.

Étienne Brunet

► **To cite this version:**

Étienne Brunet. La phrase de Proust. Longueur et rythme.. Travaux du cercle linguistique de Nice, 1981, 3, pp.97-117. hal-01438683

**HAL Id: hal-01438683**

**<https://hal.archives-ouvertes.fr/hal-01438683>**

Submitted on 17 Jan 2017

**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

LA PHRASE DE PROUST  
LONGUEUR ET RYTHME

Longueur et complexité de la phrase proustienne font l'objet principal de l'étude de Conrad Bureau : *Linguistique fonctionnelle et stylistique objective*. En mesurant dans chaque phrase de Combray les niveaux de subordination, Conrad Bureau obtient les mêmes résultats que nous avons acquis à travers les mots grammaticaux qui interviennent dans la subordination (entendue au sens large) : subordonnants, relatifs, prépositions. Certes notre étude systématique diffère de la sienne sur bien des points : les unités distinguées ne sont pas les mêmes (1), les méthodes statistiques ont gagné en

(1) Nous nous en tenons au mot graphique, suivant la tradition lexicographique et lexicologique, alors que Bureau prend pour base de calcul des unités plus larges, les syntagmes. Le critère graphique manque parfois de pertinence linguistique, mais il a l'avantage pratique d'une grande sécurité, alors que la séparation des syntagmes est source constante d'embarras et d'inconstance et que les relevés changent d'un lecteur à l'autre et pour le même lecteur d'une lecture à l'autre.

puissance, l'enquête en étendue et la comparaison en portée (2), mais les conclusions ne changent pas sur la complexité de la phrase proustienne. Et il en va de même de la longueur de la phrase.

Bureau observe que les 1686 phrases qu'il a recensées dans *Combray* ont en moyenne 18 syntagmes. Puis il fait la comparaison avec des extraits de Gide et de Le Clézio, où il constate une moyenne deux ou trois fois inférieure (respectivement 8 et 6 syntagmes par phrase). Il est donc amené à conclure que la phrase de Proust est relativement longue, ce qui s'accorde avec le sentiment subjectif du lecteur (3) et avec le concert unanime des commentateurs (4). Le recensement intégral de la *Recherche* n'apporte point là-dessus de révélation extraordinaire. Et nous ne pouvons que confirmer un trait qui n'a échappé à personne, pas même au plus inattentif des lycéens : la phrase de Proust

••••

(2) L'enquête de Bureau se limite à *Combray* et, pour la comparaison, à 500 phrases de Gide et de Le Clézio - ce qui est considérable pour un relevé manuel mais ne donne peut-être pas une assise assez large au jugement.

(3) On comprend mal toutefois les réserves formulées p.130 : "A l'encontre des réactions subjectives de la plupart des lecteurs et des critiques de Proust, on peut établir objectivement que les phrases relativement courtes sont plus nombreuses puisqu'elles constituent 62% de l'effectif total - relativement, c'est-à-dire par rapport à l'indice témoin, seul point de comparaison valable." Raisonner ainsi c'est confondre médiane, mode et moyenne et négliger l'asymétrie de la distribution, qu'on retrouverait pareillement chez tous les écrivains et qu'on constate précisément chez les auteurs étudiés par Bureau : Gide et Le Clézio (respectivement 62% et 63%). Cette dissymétrie imposée par la nature des choses livre donc peu d'information sur la spécificité de la phrase proustienne - que le lecteur a le droit, même subjectivement, d'estimer longue, du moment qu'il la compare à celle des autres écrivains. Cette référence externe, quoi qu'en dise Bureau, constitue "le seul point de comparaison valable."

(4) A la fin de son ouvrage l'auteur passe en revue les principaux critiques qui ont parlé de la période proustienne. Et il s'amuse à faire un florilège des épithètes identiques et, dit-il, conventionnelles que la phrase de Proust a suscitées : *longue, complexe, compliquée, interminable, ou encore dédale, labyrinthe, monstre*. Le Bidois, Léo Spitzer, Curtius, Léon Guichard, Ramon Fernandez, Léon Pierre-Quint et Gilles Deleuze sont successivement sollicités pour donner leur voix à l'architecteur de Riffaterre. En fin de compte l'auteur fait chorus lui aussi, mais sur un ton plus "scientifique". Et notre ordinateur, qui ne se pique pas d'originalité, dit à peu près la même chose. Voir Bureau ouvrage cité, pp.220-224.

est longue. Certes il est des phrases courtes dans la **Recherche** et Bureau a raison de les souligner. Peut-être leur brièveté même fait leur éclat, puisqu'elles s'opposent au rythme habituel et que cet écart accuse leur relief. Mais il en est de si longues qu'il faut chercher longtemps dans notre littérature pour en trouver l'équivalent. Bureau relève dans **Combray** une phrase de 222 syntagmes (5). Nous avons vainement fouillé les romans de Giraudoux pour trouver une phrase qui ait autant de mots qu'on a ici de syntagmes. La plus longue qu'on ait trouvée dans **Juliette aux pays des hommes** en a 190, 151 dans les **Aventures de Jérôme Bardini**, et 168 dans **Eglantine**. Si l'on cumule les dix phrases les plus longues de ces trois romans, on obtient des effectifs de 1389, 1167 et 1067 respectivement, alors que le total s'élève à 1011 syntagmes dans **Combray** - ce qui converti en mots représente plus du double. Même dans les siècles antérieurs, où la tradition rhétorique était plus favorable à la phrase longue, le record de Proust est difficile à battre si ce n'est occasionnellement quand l'auteur veut produire un effet (6). Or il ne semble pas que Proust veuille produire

.....

(5) Cette phrase s'étend sur les pages 7 et 8 de l'édition de la Pléiade. Elle compte 518 mots graphiques.

(6) Voici un exemple tiré de **La vie de Marianne** (édition Garnier, p.486), où Marivaux s'amuse visiblement, non seulement à tirer des phrases en longueur, mais à prolonger à l'infini la cascade des relatifs "Dès le lendemain, elle alla loger dans le **château** qu'elle le pria sans façon de lui laisser libre le plus tôt qu'il pourrait, et **dont** il sortit huit jours après pour s'en retourner chez lui, fort honteux du peu de succès de ses respects et ses **courbettes, dont il** vit bien qu'elle avait deviné les motifs, et **qui** n'avait servi qu'à la faire rire, sans compter le **chagrin qu'il** eut de me laisser dans le **château, où** le bonhomme **Villot, qui** connaissait cette dame, m'avait amené depuis cinq ou six jours, et **où** je plaisais, **où** mes façons ingénues réussissaient auprès de Mme **Dursan, qui** commençait à m'aimer, **qui** me caressait, à **qui** je m'accoutumais insensiblement, **que** je trouvais en effet bonne et franche, avec **qui** j'étais le lendemain plus à mon aise et plus libre que la veille, **qui** de son côté prenait plaisir à voir qu'elle me gagnait le coeur, et **qui** pour surcroît de bonne fortune pour moi, avait trouvé au château un **portrait qu'on** avait fait d'elle dans sa jeunesse, à **qui** il est vrai que je ressemblais beaucoup, qu'elle avait mis dans sa chambre, qu'elle montrait à tout le monde." En une phrase à effet, Marivaux peint une scène de comédie où l'héritier éconduit cède la place à une rivale inattendue. Il y a là une sorte d'humour grammatical dans l'amoncellement des relatifs (on en compte 19) et dans la multiplication des niveaux (7 antécédents différents dont l'un, **château**, se répète en écho à trois endroits de l'échelle). Or cette phrase compliquée à plaisir et par jeu n'a que 203 mots, soit deux fois moins que la plus longue de **Combray**.

un effet de style. L'idée de battre des records lui est très étrangère et s'il n'ignore pas la spécificité de son style - que les critiques soulignaient à l'envi et sur laquelle il fut maintes fois sommé de s'expliquer - il est loin de s'en vanter. Il avoue seulement ne pouvoir écrire autrement, et que la nature des choses qu'il dit l'oblige à s'exprimer ainsi. "Je suis bien obligé, lit-on dans sa *Correspondance*, de tisser des longues soies, et si j'abrégais mes phrases, cela ferait de petits morceaux de phrase mais pas des phrases" (7).

## I

1. Après avoir d'emblée souligné une conclusion évidente et reconnue par Proust lui-même, est-il utile de détailler les faits observés ? On les trouvera dans les tableaux 42 et 43, où les différents signes de ponctuation ont été recensés. Trois seulement ont été pris en compte pour le dénombrement des phrases : le point, le point d'interrogation et le point d'exclamation. Sans doute ces signes ne sont-ils pas toujours séparateurs de phrases : il arrive que les signes affectifs ne portent que sur un mot, par exemple une interjection, et que la phrase, un instant suspendue, reprenne son cours. Le point lui-même est ambigu puisqu'il sert à l'abréviation, à la siglaison et à la suspension. Mais dans le cas de Proust la désambiguïsation du point n'est pas trop malaisée : tout d'abord Proust proscrit la siglaison dont on ne trouve qu'un seul exemple dans toute la *Recherche*. Encore Proust ne le place-t-il dans la bouche de Mme Verdurin que pour ridiculiser ce tic de langage : "Et dans tous ses récits revenait tout le temps le G.Q.G. ("j'ai téléphoné au G.Q.G."), abréviation qu'elle avait à prononcer le même plaisir qu'avaient naguère les femmes qui ne connaissaient pas le prince d'Agrigente, à demander en souriant quand on parlait de lui et pour montrer qu'elles étaient au courant : Grigri ?". En second lieu Proust utilise très peu l'abréviation pour les prénoms ou les noms de ses personnages. Il s'y résout parfois pour évoquer une silhouette fugitive dont le nom importe peu ou pour désigner un individu qui sert d'exemple à un comportement général - et dans ce rôle c'est la variable

••••

(7) *Correspondance générale*, t.IV, p.201.

X qui s'impose le plus souvent à son esprit et l'on trouve sous sa plume le comte X, le prince de X, l'ambassadeur de X, soit 34 occurrences dans la *Recherche*. Mais Proust se refuse au mystère facile qui s'attache aux initiales dans la littérature populaire et les romans policiers, et l'intention parodique n'est pas douteuse dans la lettre maladroite qu'Aimé envoie au narrateur et où Albertine est désignée 9 fois par l'initiale (8). Au total les abréviations ne comptent que 85 occurrences dans la *Recherche* (9), à quoi il faut ajouter les 2116 occurrences de M. pour Monsieur. Restent les points de suspension dont les exemples sont fort rares dans la *Recherche* et cette exclusive est là aussi volontaire, comme Proust s'en explique clairement à Jacques Rivière au moment de la publication dans la N.R.F. d'un extrait des *Jeunes filles en fleur* : "Vous mettriez aussi, comme j'ai horreur des points de suspension : *Les points de suspension ne sont pas de l'auteur mais de la Rédaction de la Revue qui remplace par eux les passages que la brièveté du numéro l'oblige de supprimer.* (...) Je ne préfère pas des astérisques aux points de suspension. Simplement je voudrais ôter à ces points de suspension le caractère littéraire par lequel certains écrivains, à la suite de Loti, croient indiquer l'ineffable, et montrer simplement qu'il s'agit d'un manque de place" (10). Qu'il s'agisse du rejet de l'abréviation, et des points de suspension, et aussi du refus des lignes blanches, le goût de Proust le porte à proscrire les sous-entendus éloquentes et les silences appuyés.

2. Cela facilite grandement notre tâche puisqu'il suffit d'enlever les 2116 occurrences de M. pour obtenir approximativement le nombre de

••••

(8) t. III, p.515. En un autre passage Proust s'en prend à l'ambivalence de l'abréviation, Gilberte nouvellement mariée dissimulant sous une initiale son origine juive : "elle ménagea quelque temps la transition en signant G.S. Forcheville. La véritable hypocrisie dans cette signature était manifestée par la suppression bien moins des autres lettres du nom de Swann que de celles du nom de Gilberte. En effet, en réduisant le prénom innocent à un simple G. Mlle de Forcheville semblait insinuer à ses amis que la même amputation appliquée au nom de Swann n'était due aussi qu'à des motifs d'abréviation." t.III, p.587.

(9) En voici le détail pour chacune des lettres : 14 A, 2 B, 1 C, 10 E, 9 G, 1 H, 1 L, 1 N, 2 P, 2 Q, 3 S, 1 T, 1 W, 34 X, 2 Y et 1 Z.

(10) *Correspondance avec J. Rivière*, p.32 et p.38 (Cité par J. Milly).

points-fins de phrase. Pour faire bonne mesure nous retrancherons les 85 abréviations et quelques centaines de points de suspension et l'on gardera  $38177 - 2500 = 35677$  occurrences.

En admettant que les séparateurs de phrase se confondent avec les ponctuations fortes, on aurait donc dans la Recherche une moyenne de :

$$\frac{1267069}{(35677 + 2756 + 2448)} = 30,994$$

(nb.mots) (point) (excl.) (inter.)

soit près de 31 mots par phrase, ou 3 lignes par phrase dans l'édition de la Pléiade (ou encore 14 phrases par page dans la même édition) (11). Cette moyenne élevée ne prend sa valeur que par comparaison avec celle des autres écrivains. Voici les éléments dont nous disposons et qui ont

	mots phrases moyenne		
Rousseau (Emile)	257154	9280	27,71
Chateaubriand	1398984	62919	22,23
Giraudoux (romans)	412268	19971	20,64
Corpus XIX-XX	70273552	4611432	15,24
Prose littéraire de 1893 à 1926	12216571	914130	13,36

été obtenus dans les mêmes conditions :

Les monographies disponibles sont encore rares mais les auteurs que nous rapprochons de Proust sont suffisamment représentatifs pour qu'on puisse affirmer que la phrase de Proust l'emporte en longueur sur toutes les autres. On remarquera que Proust va nettement plus loin dans ce domaine que Chateaubriand dont la phrase a pourtant une relative ampleur, mais, trop pauvre en subordonnants, elle n'a pas une ossature assez puissante pour soutenir de vastes pans de discours. La complexité et la longueur de la phrase sont en effet deux éléments liés. La rhétorique fait de Rousseau le suivant immédiat de Proust (au moins dans l'*Emile*) et c'est du côté du XVIIIe et du XVIIe que Proust peut trouver des concurrents. Car la phrase depuis 1789 tend à raccourcir comme nous l'avons constaté dans le corpus du T.L.F. Il est d'autant plus significatif que Proust ait résisté à ce courant, et sa phrase est

•••••

(11) Si l'on tient compte des noms propres, la moyenne est plus élevée encore (32,04).

ainsi deux fois plus longue que celle des prosateurs de son temps (moyenne 13,36). Il est vrai que cette moyenne enveloppe des oeuvres de théâtre où les répliques ont nécessairement un rythme plus rapide.

3. On pourrait supposer que le déficit en ponctuations fortes doit être compensé par quelque excédent des ponctuations faibles, point et virgule et deux points. Or il n'en est rien. Le tableau 42 montre que les écarts sont là aussi très significatifs (respectivement -37,65 et -44,13). Si donc on avait modifié le critère de segmentation en acceptant ces deux signes comme séparateurs de phrases, la phrase proustienne ainsi définie aurait gardé ses caractéristiques - qui ne tiennent donc nullement à l'arbitraire de notre définition de la phrase. On observera toutefois que le déficit en ponctuations faibles est moins considérable, ce qui apparaît mieux si on raisonne sur les pourcentages :

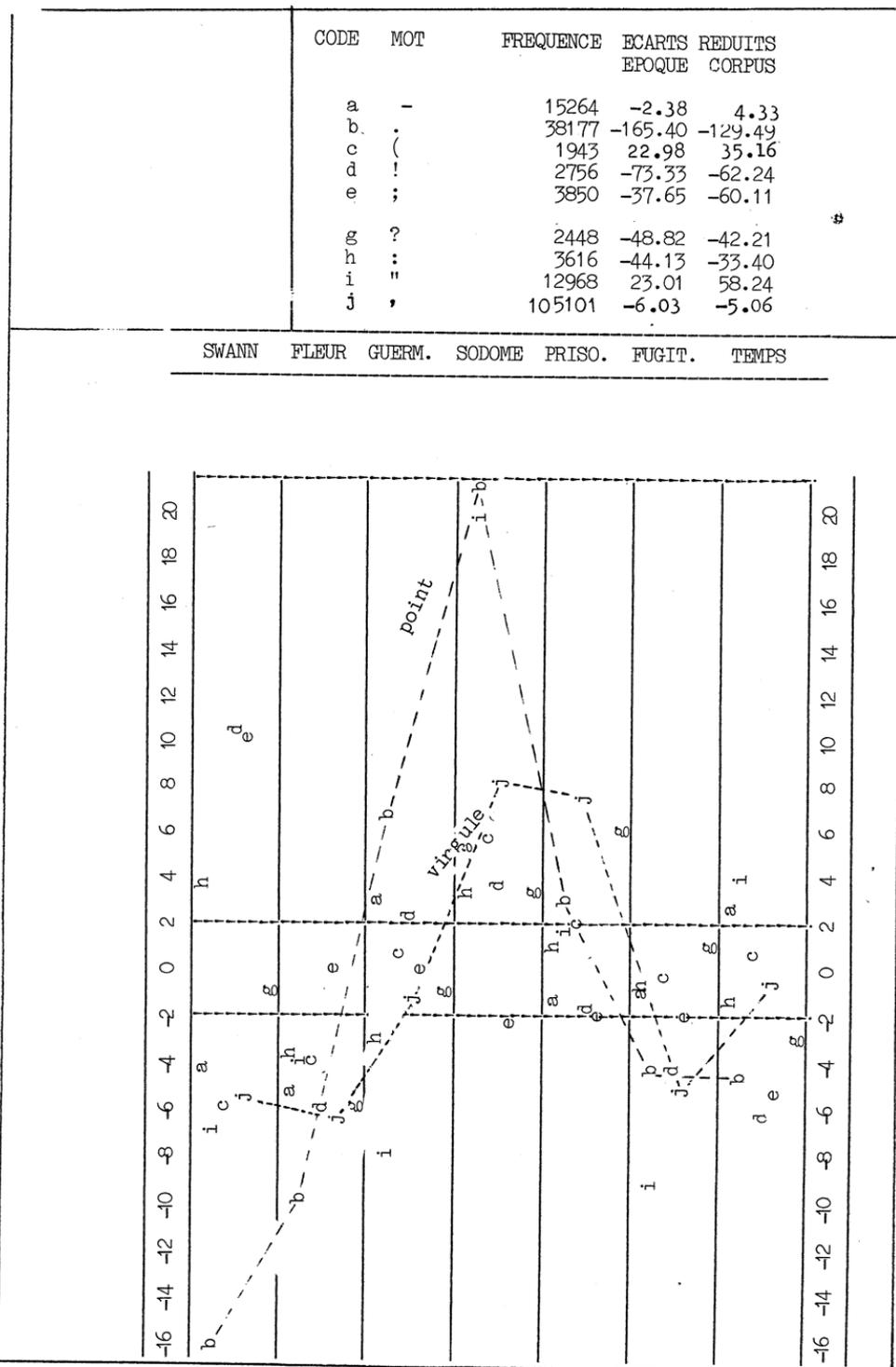
	Effectif réel dans la <u>Recherche</u> a	effectif théor. d'après l'époque b	Rapport a/b
Point	38177	83403	0,46
exclamation	2756	9535	0,29
interrogation	2448	6040	0,41
point et vir- gule	3850	6786	0,57
deux points	3616	7148	0,51
virgule	105101	106968	0,98
parenthèses	1943	1192	1,63
guillemets	12968	10713	1,28
trait d'union	15264	14924	1,02

Alors que les signes forts (12) ont moins de la moitié des effectifs attendus, les signes faibles atteignent ou dépassent 50%. Quant au plus léger des signes de ponctuation, la virgule, c'est le seul qui trouve grâce aux yeux de Proust et le dosage des virgules dans la *Recherche* est à peu près le même que dans la prose littéraire de l'époque (98% du total théorique). Cette inégalité de traitement réservé aux signes

••••

(12) C'est le plus violent de tous, le point d'exclamation, dont le pourcentage est le plus bas (29%).

FIGURE 42 . COURBE DES SIGNES DE PONCTUATION .



faibles et aux signes forts ne permet pas de dire que Proust éprouvait pour la ponctuation une parfaite indifférence. André Ferré dans sa note sur le texte de son édition de la Pléiade insiste sur la rareté de la ponctuation dans la *Recherche*, et l'explique par un désintéret de Proust : "Quant à la ponctuation, dont Proust ne se souciait guère, nous avons souvent dû en rétablir le minimum essentiel pour l'intelligibilité du texte" (13). Plutôt que d'invoquer l'indifférence de Proust, nous chercherions l'explication dans l'inachèvement du manuscrit et aussi dans la volonté, plusieurs fois exprimée dans sa correspondance, de présenter des blocs de texte compacts, en évitant les trous, les blancs, les alinéas et même les fins de phrases (14). La fluidité de la virgule permet de l'intégrer au texte ainsi conçu, dont elle marque le rythme sans le rompre.

4. Proust n'est pas si dédaigneux de la ponctuation puisque certains symboles sont en excès sous sa plume. Il s'agit de la parenthèse et des tirets (à quoi il faut ajouter aussi les guillemets). Remarquons tout d'abord que ces symboles vont par paire, et qu'au lieu d'amener une rupture, ils n'introduisent qu'une interruption, une suspension provisoire, le plus souvent une addition, un élargissement, une précision, parfois une restriction ou un repentir de la pensée. Ce décrochement visuel correspond souvent à un second plan de la pensée, presque à un commentaire ou à un contrepoint. Et en effet ces ajouts coïncident parfois dans le manuscrit à des bécquets marginaux, nés d'un retour sur le texte. Et c'est pourquoi Leo Spitzer a pu dire : "mais comment Proust crée-t-il linguistiquement cette dimension de profondeur spirituelle ?

....

(13) Tome 1, p.XXXVI. Nous espérons que le nombre de signes ajoutés par André Ferré constitue bien un minimum, sinon la présente étude perdrait beaucoup de son intérêt.

(14) Inversement Proust met parfois des points, là où la syntaxe devrait l'interdire, et notamment il lui arrive de couper principale et subordonnée quand la phrase lui paraît trop longue. C'est le cas de la construction *de sorte que*, que Proust place parfois en tête de phrase, comme s'il s'agissait de l'adverbe *ainsi*. On trouve deux exemples de ce type pages 554 et 555 de la *Fugitive*.

TABLEAU 43 : LA PONCTUATION DE PROUST. EFFECTIFS.

	Swann	J. F.	Guer.	Sodo.	Priso.	Fugit.	T.Reptr.	Total
Virgule	14067	17365	19486	19000	14575	8688	11920	105101
Point	4213	5825	7691	8072	5169	3123	4084	38177
Point et virgule	751	662	729	604	462	306	336	3850
Deux points	583	538	608	691	497	308	391	3616
Interrog.	328	310	446	479	423	229	233	2448
Exclam.	573	356	566	543	330	178	210	2756
Guillemets	1533	2059	2091	3050	1756	847	1632	12968
Parenthèse	182	267	380	428	288	167	231	1943
trait d'un.	1941	2371	3011	2854	1929	1305	1853	15264
tiret	839	864	1276	874	568	310	312	5043

LA PONCTUATION DE PROUST. ECARTS REDUITS.

	Swann	J.F.	Guer.	Sodo.	Priso.	Fugit.	T.Reptr.
Virgule	-5,00	-5,39	-1,79	9,46	8,39	-5,81	-0,84
Point	-16,27	-9,80	6,95	21,65	3,15	-4,09	-4,46
Point et virgule	10,02	0,08	0,28	-1,28	-1,84	-1,81	-5,26
Deux points	3,83	-3,62	-2,99	3,41	1,34	-0,54	-1,16
Interrog.	-0,74	-5,89	-0,68	3,41	6,30	1,02	-2,96
Exclam.	10,42	-5,89	2,39	3,80	-1,60	-4,29	-6,28
Guillemets	-6,90	-3,84	-7,68	19,83	1,84	-9,02	4,16
Parenthèse	-5,80	-3,98	0,90	5,92	2,39	-0,28	0,64
trait d'un.	-4,29	-5,29	3,07	5,64	-1,32	-0,98	2,78
tiret	5,58	-0,03	11,91	0,66	-3,66	-6,59	-11,69

Principalement par la parenthèse" (15). Ces parenthèses ne sont pas toutes des incidentes (16), comme le remarque C. Bureau, et leur rattachement grammatical au reste de la phrase est le plus souvent régulier. Leur nécessité n'est donc pas syntaxique mais rythmique.

5. Si les parenthèses abondent dans la *Recherche*(17) Proust leur préfère souvent les tirets qui jouent le même rôle mais, semble-t-il, sur un registre plus discret. Il y a dans la parenthèse un isolement plus étroit, avec la nécessité de fermer quand on a ouvert. Le tiret signale plutôt un léger retrait et la répétition n'en est pas obligatoire lorsque la phrase s'achève sans remonter au niveau supérieur. C'est pourquoi il n'a pas été possible de dégrouper les paires de tirets et le nombre que nous indiquons, 5043, cumule les tirets qui signalent

(15) *Etudes de style*, p.452. Jean Milly aborde lui aussi cette question dans la *Phrase de Proust* p.258, en la reliant à une tendance générale de Proust à l'enchâssement qui se manifeste aussi dans la subordination : "La multiplication des parenthèses et des déterminants correspond à un mode de construction du récit qui enchâsse les épisodes les uns dans les autres (...) Dans la phrase comme dans le récit, nous trouvons un mode de développement de type télescopique : tout élément est susceptible de dérouler ses implications, par les procédés subordinatifs."

(16) Proust est cependant le premier à établir une équivalence entre incidente et parenthèse, comme en témoigne ce passage : "Dans certaines familles menteuses un frère venu voir son frère sans raison apparente et lui demandant **dans une incidente**, sur le pas de la porte, en s'en allant, un renseignement qu'il n'a même pas l'air d'écouter, signifie par cela même à son frère que ce renseignement était le but de sa visite, car le frère connaît bien ces airs détachés, ces mots dits comme **entre parenthèses**, à la dernière seconde, car il les a souvent employés lui-même." *La Prisonnière*, t. III, p.110.

(17) Le mot *parenthèse* l'est également avec une fréquence de 15 et un écart réduit de +2,79. Mais il s'agit presque toujours de l'expression *entre parenthèses* que Proust place souvent dans la bouche des bavards. En une occasion le mot est employé à contretemps par Françoise pour désigner la parenté : "il reste toujours le respect qu'on doit à la parenté" I, p.154. A ce cuir de Françoise, répond à la fin de la *Recherche* la savoureuse maladresse d'Aimé qui confond guillemets et parenthèses : "Aimé qui avait un certain commencement de culture voulait mettre Mlle A. en italique ou entre guillemets. Mais quand il voulait mettre des guillemets il traçait une parenthèse, et quand il voulait mettre quelque chose entre parenthèses il le mettait entre guillemets." *La Fugitive*, note p.515.

le commencement d'une incidente, et ceux, quand ils existent, qui en signalent la fin. L'effectif des parenthèses au contraire est celui des parenthèses ouvrantes, qui correspond exactement à celui des parenthèses fermantes. Si donc on veut les comparer aux tirets (ou aux guillemets) il convient d'en doubler d'abord le nombre :  $1943 \times 2 = 3886$ , lequel reste encore inférieur à celui des tirets. Il n'est pas très facile de pousser plus loin l'analyse parce que les tirets remplissent deux fonctions, dont la première est analogue à celle de la parenthèse et la seconde fort différente, puisqu'il s'agit de signaler le début d'une réplique. Au surplus les données manquent là-dessus dans le corpus du T.L.F.

6. Comme les parenthèses, les guillemets et les traits d'union abondent dans le texte de Proust. Les premiers dénoncent la présence du dialogue. Mais ils servent aussi aux citations et particulièrement aux expressions empruntées aux personnages et caractéristiques de leur pensée ou de leur langage. Ils désignent ainsi les "cuirs" de Françoise ou du maître d'hôtel, les calembours de Cottard, les anglicismes d'Odette, les hyperboles de Mme Verdurin ou les platitudes de Norpois, etc... Comme Flaubert, Proust a son bêtisier et les guillemets servent à épingle ses trouvailles. Les deux fonctions des guillemets se rejoignent en fait dans le goût du langage. Brillant salonnière, brillant dialoguiste, Proust a le don de la parole mais ce causeur est aussi un observateur du langage et même un linguiste à sa façon (18). L'excédent considérable des guillemets (écart réduit = 23) rend compte globalement de ces traits. Quant aux traits d'union, nous avons affaire là encore à un signe ambigu, qui peut être le lien des mots composés, mais aussi une particularité purement graphique qui accompagne l'inversion du sujet, le *t* euphonique (*comment va-t-il?*) et beaucoup de locutions adverbiales ou prépositives (*peut-être, au-dessus*). La proportion de ce signe dans la *Recherche* n'appelle pas de commentaire : on en trouve 2% de plus par rapport à l'époque, mais 3% de moins par rapport au corpus entier du T.L.F.

(18) Les remarques sur la langue et l'histoire de la langue ne manquent pas dans la *Recherche*, comme celle-ci que provoque la confusion des parenthèses et des guillemets : "C'est ainsi que Françoise disait que quelqu'un restait dans ma rue pour dire qu'il y demeurerait et qu'on pouvait demeurer deux minutes pour rester, les fautes des gens du peuple consistant seulement très souvent à interchanger - comme fait d'ailleurs la langue française - des termes qui au cours des siècles ont pris réciproquement la place l'un de l'autre." *La Fugitive*, p.515, note.

1. Les difficultés que nous venons de rencontrer quand il fallait comparer Proust à une norme extérieure venaient d'une certaine inconstance dans la saisie des données au cours de 15 années de dépouillement à Nancy et d'un flottement initial dans la définition des signes de ponctuation (19). Ces embarras s'évanouissent lorsqu'il s'agit de comparer Proust à lui-même, c'est-à-dire d'opposer les différentes parties de la *Recherche*. L'ambiguïté des signes demeure certes, mais la saisie ayant été faite d'un coup, on est assuré de la constance des critères. Les tableaux 42 et 43 dans leur moitié inférieure, montrent le mouvement des signes de ponctuation à l'intérieur de l'oeuvre. Cette fois c'est la *Recherche* dans son ensemble qui constitue la norme, selon la procédure déjà mise en oeuvre dans les chapitres précédents. Les courbes du graphique 42 laissent apparaître une grande dispersion, la zone non significative (entre -2 et +2) étant peu fréquentée (22 points sur 70). Proust a donc changé sa ponctuation pendant la rédaction, sa phrase est plus longue dans les premiers textes mais aussi dans les derniers, et plus courte dans les livres intermédiaires, de *Germantés* à la *Prisonnière*. Le point peut suffire à rendre sensible cette évolution dans une courbe en cloche qui culmine au niveau de *Sodome* (voir graphique 42). Si l'on réunit les ponctuations fortes, les indications du point sont confirmées :

	Swann	J.F. Guerm.	Sodome	Priso.	Fugit.	Temps	
. + ? + !	5114	6491	8703	9094	5922	3530	4527
écart réduit	-12,82	-12,08	6,97	22,08	4,05	-4,68	-6,47

En réalité les ponctuations faibles augmentent aussi au milieu de la *Recherche*, c'est donc la segmentation dans son ensemble qui adopte alors un rythme plus court. Au lieu de s'opposer au point la virgule accompagne son mouvement, mais en l'amortissant.

••••

(19) Il ne faut pas oublier que l'objectif n'était pas de constituer une banque de données linguistiques mais de fournir des exemples aux rédacteurs du dictionnaire.

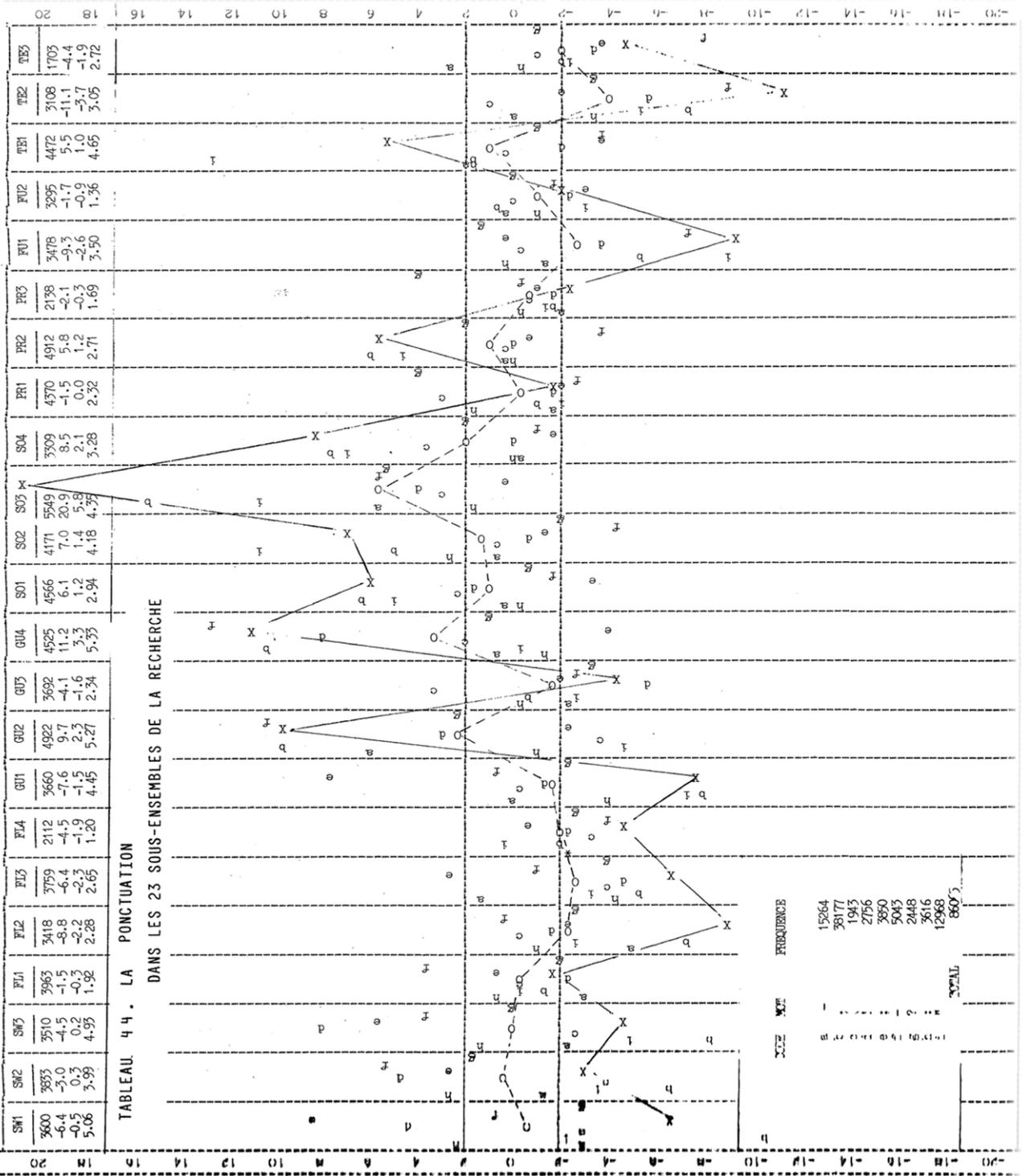


TABLEAU 44. LA PONCTUATION  
DANS LES 23 SOUS-ENSEMBLES DE LA RECHERCHE

SM1	SM2	SM3	FI1	FI2	FI3	FL4	GU1	GU2	GU3	GU4	S01	S02	S03	S04	FR1	FR2	FR3	FU1	FU2	TE1	TE2	TE3
3600	3655	3510	3963	3418	3759	2112	3660	4922	3692	4525	4566	4171	5549	3709	4370	4912	2138	3478	3295	4472	3108	1703
-6.4	-5.0	-4.5	-1.5	-8.8	-6.4	-4.5	-7.6	9.7	-4.1	11.2	6.1	7.0	20.9	8.5	-1.5	5.8	-2.1	-9.3	-1.7	5.5	-11.1	-4.4
-0.5	0.3	0.2	-0.3	-2.2	-2.3	-1.9	-1.5	2.3	-1.6	3.3	1.2	1.4	5.8	2.1	0.0	1.2	-0.3	-2.6	-0.9	1.0	-3.7	-1.9
5.06	3.99	4.95	1.92	2.28	2.65	1.20	4.45	5.27	2.34	5.33	2.94	4.18	4.35	3.28	2.32	2.71	1.69	3.50	1.36	4.65	3.05	2.72

XXZ	ME	FREQUENCE
1	15264	
2	38177	
3	1943	
4	2796	
5	3850	
6	5043	
7	2448	
8	3616	
9	12968	
TOTAL	8605	

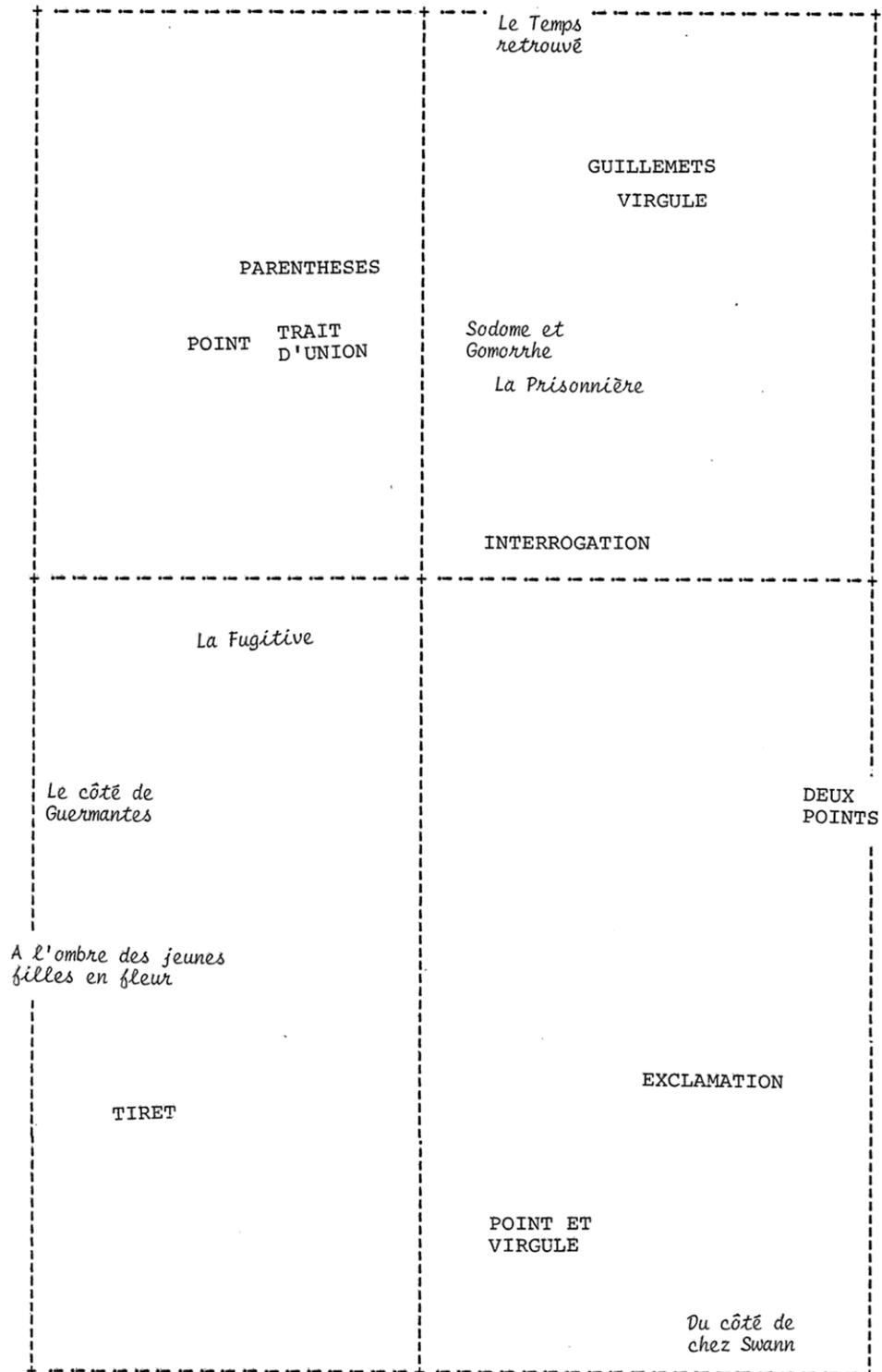
2. Le mouvement d'ensemble est plus sensible dans la figure 44 qui donne une analyse plus fine et plus détaillée de la distribution. La courbe en traits pleins - qui représente le cumul des occurrences de tous les signes de ponctuation - monte graduellement jusqu'à *Sodome* pour redescendre ensuite. Au sommet, c'est-à-dire dans la troisième partie de ce texte, tous les signes sont dans la zone positive et de la même façon ils sont tous dans la zone négative au niveau de la deuxième partie des *Jeunes Filles*, où se situe le point le plus bas. Malgré les valeurs très diverses des signes de ponctuation, il y a donc une loi commune qui les repousse ou les attire ensemble dans certaines situations de discours. Pour mettre ce fait en évidence et pour réduire l'influence écrasante du point, on a dressé la courbe pondérée en utilisant pour ordonnée l'écart réduit moyen de neuf signes de ponctuation (la virgule trop majoritaire ayant été écartée). C'est le dessin en pointillés de la figure 44. Le mouvement est amorti, mais il reste très net.

3. En dehors de ce mouvement parallèle qui anime l'ensemble du système des ponctuations, certaines fluctuations apparaissent quand par exemple deux signes sont concurrents. C'est le cas des tirets (pour certains de leurs emplois) et des parenthèses, le choix de Proust s'exerçant de plus en plus en faveur des secondes, comme le montre la suite chronologique du quotient **tiret/parenthèse** :

Swann J.Filles Guerm. Sodom. Priso. Fugit. Temps  
2,31 1,62 1,68 1,02 0,99 0,93 0,68

4. La vue synthétique des mouvements internes des ponctuations est obtenue par l'analyse factorielle de la figure 45, qui distribue les textes en deux moitiés : en bas les premiers textes, **Swann**, **Jeunes Filles** et **Guermantes**, en haut les derniers **Sodome**, **La Prisonnière**, le **Temps retrouvé** (**La Fugitive** ne prend pas parti et se tient près de l'origine des axes). Du côté de **Swann** les ponctuations atténuées qui conviennent au récit, point-virgule et deux points. De l'autre, le point, la virgule et la parenthèse qui s'adaptent mieux au procès explicatif. Le point d'exclamation rejoint le premier groupe et le point d'interrogation hésite près de la ligne de démarcation. Nous retrouvons là, semble-t-il, les deux tendances du style de Proust, plus poétique au début et plus démonstratif ensuite.

FIGURE 45. ANALYSE FACTORIELLE DES SIGNES DE PONCTUATION



Reste un point à élucider qu'il faut bien aborder lorsqu'on parle de rythme. Comment les phrases réagissent-elles les unes aux autres ? Leur succession est-elle aléatoire ou obéit-elle à une loi, à une périodicité plus ou moins régulière ? C. Bureau n'a guère abordé cette question mais il donne les éléments qui permettent de l'étudier. Nous trouvons en effet dans l'appendice de son ouvrage (pp.234-253) la longueur (en nombre de syntagmes) des 1686 phrases recensées dans Combray. Cet échantillon est suffisant pour notre étude où nous cherchons à savoir si les phrases longues ont tendance à se regrouper, et les phrases courtes de leur côté, ou si au contraire l'auteur a soin de les mêler pour casser le rythme, ou si enfin la séquence des phrases est telle que le hasard peut en rendre compte. Il existe dans la panoplie statistique des armes sophistiquées pour traiter des données sérielles, c'est-à-dire soumises à l'ordre du temps. Nous n'utiliserons cependant que des outils relativement simples déjà mis en oeuvre dans les textes de Giraudoux et Rousseau (20).

1. Le premier est un calcul de variance un peu particulier puisque la variable n'est pas le nombre de syntagmes dans la phrase, mais la différence de longueur entre cette phrase et la précédente. La pondération **est** obtenue par la variance classique et le résultat se situe autour de la valeur 2 dans le cas de l'hypothèse nulle. Rappelons le nom de cette méthode : différences quadratiques moyennes successives et la formule.

$$\eta = \frac{\sum_{i=1}^{N-1} (x_{i+1} - x_i)^2}{\sum_{i=1}^N (x_i - \bar{x})^2}$$

Dans une première approche nous avons soumis au calcul 16 tranches de 100 phrases en observant la valeur prise par le rapport  $\eta$ . Voir partie haute du tableau 46. Le fait majeur est que le résultat est toujours inférieur à 2 (sauf une fois) et que le seuil à 2% (1,54) est franchi 4 fois. On a l'indication d'une tendance, d'autant plus nette que 16 tranches témoins

(20) Voir l'Index de l'*Emile*, pp.579-583.

TABLEAU 46 . LA PERIODICITE DE LA PHRASE PROUSTIENNE

TRANCHES ISOLEES							
	$\sum_{i=1}^{N-1} (x_{i+1} - x_i)^2$	$\sum_{i=1}^N (x_i - \bar{x})^2$	a/b	tirage aléatoire	$\sum x_i x_{i+1} - \bar{x} \sum x_i$	c/b	tirage aléatoire
	a	b			c		
1- 100	96154	51461	1.868	2.367	2592	0.050	-0.185
100- 200	38982	20371	1.914	1.851	870	0.043	0.078
200- 300	24440	14536	1.681	1.934	1986	0.137	0.034
300- 400	14574	9487	1.536	1.825	856	0.090	0.086
400- 500	38624	27011	1.430	1.795	7507	0.278	0.099
500- 600	67471	36373	1.855	1.983	2381	0.065	-0.018
600- 700	43366	27873	1.556	1.930	6100	0.219	0.055
700- 800	21105	15126	1.395	1.910	4361	0.288	0.040
800- 900	32111	17112	1.876	2.071	1052	0.061	-0.045
900-1000	17293	12233	1.414	1.945	3330	0.272	0.013
1000-1100	34948	19452	1.797	2.024	1786	0.092	0.017
1100-1200	41573	20692	2.009	2.339	-199	-0.010	-0.171
1200-1300	32930	16886	1.950	2.123	91	0.005	-0.056
1300-1400	31547	17607	1.792	1.948	1729	0.098	0.027
1400-1500	18276	9378	1.949	1.942	6	0.001	0.023
1500-1600	38204	22387	1.707	1.765	2715	0.121	0.123
TRANCHES CUMULEES							
100	96154	51461	1.868	1.619	2592	0.050	0.161
200	135136	72164	1.873	1.858	3794	0.053	0.070
300	159576	86939	1.836	1.846	6019	0.069	0.077
400	174150	97409	1.788	1.926	7858	0.081	0.036
500	212774	124420	1.710	1.914	15365	0.123	0.035
600	280245	161371	1.737	1.968	18324	0.114	0.016
700	323611	190492	1.699	1.958	25672	0.135	0.018
800	344716	205706	1.676	2.028	30121	0.146	-0.016
900	376827	222971	1.690	2.038	31326	0.140	-0.020
1000	394120	236503	1.666	2.030	35955	0.152	-0.015
1100	429068	256089	1.675	2.033	37875	0.148	-0.017
1200	470641	276924	1.700	2.011	37819	0.137	-0.005
1300	503571	294976	1.707	1.999	39076	0.132	0.000
1400	535118	312852	1.710	1.991	41074	0.131	0.004
1500	553394	322711	1.715	1.996	41561	0.129	0.002
1600	591598	348385	1.698	1.997	47563	0.137	0.002

constituées par tirage aléatoire ont donné des valeurs beaucoup plus proches de 2 et réparties au-delà comme en-deçà de cette valeur. Mais les faits de périodicité deviennent plus incontestables lorsqu'on cumule les tranches et que les effectifs considérés croissent de 100 à 1600. Voir partie basse du tableau 46. Le seuil de signification à 2% s'élève dans le même temps de 1,54 pour 100, à 1,79 pour 500, à 1,85 pour 1000 et enfin à 1,89 pour 1600. Or les valeurs du coefficient R ne cessent de s'abaisser et franchissent le seuil dès qu'intervient la 4ème tranche pour atteindre finalement la valeur 1,69 très en dessous de la limite 1,89 (21). C'est dire qu'à l'échelle d'un texte complet le phénomène de la périodicité est indubitable. Nous l'avions déjà observé dans un texte de haute rhétorique, l'Emile : tantôt dans un mouvement large la phrase longue y appelait la phrase longue, tantôt les phrases courtes s'y précipitaient ensemble en une succession rapide. Or cette houle qui varie l'amplitude des vagues du discours, nous l'observons partout, non seulement chez Rousseau et Proust mais aussi dans les 500 phrases témoins que C. Bureau a recensées chez Gide et Le Clézio. Le tableau ci-dessous précise les valeurs obtenues pour des tranches de 100 à 500 mots.

TABLEAU 47 : LA PERIODICITE. COMPARAISON DE QUELQUES ECRIVAINS.

	COMBRAY		Gide	Le Clézio	ROUSSEAU				
		tirage aléatoire	(La porte étroite, extrait)	(La Guerre extrait)	Emile Livres I à V				
100	1,87	2,34	2,11	1,89	2,03	1,73	1,93	1,30	1,53
200	1,87	2,06	1,97	2,03					
300	1,84	2,08	1,93	1,94					
400	1,79	2,07	1,85	1,66					
500	1,71	1,98	1,84	1,58	1,70	1,58	1,64	1,64	1,60

••••

(21) Le tirage aléatoire des phrases, dans le même temps, converge rapidement vers la valeur 2 que fournissent les tables. La colonne 4 du tableau 46 s'approche de très près de cette valeur ( $r_i = 1,997$ ) quand le tirage porte sur 1600 phrases. Précisons qu'un générateur de nombres aléatoires fournit l'adresse d'une phrase au hasard et le calcul est exécuté comme s'il s'agissait de la phrase suivante.

Au niveau 500, on voit que partout le seuil significatif à 2%, soit 1,79, est dépassé, parfois largement. Gide toutefois fait exception (1,84) - ce qui n'étonnera pas de la part d'un écrivain qui cultive le déséquilibre, la dissymétrie, la rupture, et qui a peu d'attrait pour les régularités trop attendues de la rhétorique.

4. Les résultats qui précèdent sont confirmés par une autre méthode. Pour contrôler le déterminisme, c'est-à-dire l'influence exercée par un élément d'une série sur l'élément suivant, on peut utiliser le test dit de l'autocorrélation. Le principe est assez semblable à celui du test précédent et l'on retrouve au dénominateur la même pondération : la variance. Mais le numérateur fait intervenir non pas la différence entre deux éléments consécutifs mais leur produit comme dans le coefficient de Bravais-Pearson. Aussi bien le résultat est-il toujours situé entre -1 et +1 et sa valeur est interprétée à l'aide de tables appropriées, suivant le nombre d'éléments considérés. La formule fait intervenir un paramètre h qui est le décalage entre l'élément i et l'élément associé i + h. Dans notre cas h vaut 1 puisque nous associons chaque phrase à la suivante :

$$R_h = \frac{\sum x_i x_{i+h} - \bar{x} \sum x_i}{\sum (x_i - \bar{x})^2}$$

Le seuil bilatéral à 2% a comme limite gauche 0,22 pour N = 100, 0,10 pour N = 500, 0,078 pour N = 1000 et 0,058 pour N = 1600. On voit que le seuil est atteint 4 fois sur les 16 tranches de 100 mots soumises au calcul séparément et que la valeur de Rh est toujours positive (elle serait négative si chaque phrase s'opposait à celle qui la précède et à celle qui la suit). Dans les tranches cumulées (bas du tableau 46, 6e colonne), le seuil à 2% est atteint dès la phrase 500 et au bout de 1600 phrases la valeur obtenue 0,137 est très au-delà des possibilités du hasard. Les conclusions coïncident donc exactement avec les enseignements de la méthode précédente. La phrase proustienne recèle une périodicité, un rythme non aléatoire et c'est à bon droit qu'on peut l'appeler période.