



HAL
open science

Les paragraphes, un autre rythme du texte

Claude Patricia Tardif

► **To cite this version:**

Claude Patricia Tardif. Les paragraphes, un autre rythme du texte. 9e Congrès Mondial de Linguistique Française (CMLF), Jul 2024, Lausanne (UNIL), Suisse. pp.06012, 10.1051/shsconf/202419106012 . hal-04645691

HAL Id: hal-04645691

<https://hal.science/hal-04645691>

Submitted on 11 Jul 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.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Les paragraphes, un autre rythme du texte

Claude Patricia Tardif

Université Paris 8

claudeclpro@gmail.com

Résumé. Cette contribution propose une nouvelle approche de la visualité du texte littéraire, sans tenir compte ni de sa disposition spatiale ni de « l'énonciation éditoriale ». Mais alors que reste-t-il de la forme visuelle du texte ? Il reste les paragraphes et, dans une moindre mesure, les chapitres, qui la modèlent en fonction de leurs longueurs. Ils donnent au texte sa morphologie, invariable quelle que soit l'édition. Ainsi, nous nous intéressons aux paragraphes, non en tant qu'unités textuelles autonomes mais pris dans la linéarité du texte. La méthode consiste à mesurer chaque paragraphe d'un texte, puis, à partir de cette suite numérique, nous réalisons des visualisations qui exposent un rythme du texte jusqu'alors inexploré en fonction de la longueur de ses paragraphes. Partant de la dimension visuelle du texte, nous en arrivons à sa dimension temporelle. Nous verrons que ce rythme correspond à la fréquence des changements, de sens pour les paragraphes narratifs et de locuteurs pour les paragraphes dialogaux. Enfin, nous verrons que cette forme dynamique a des impacts sur la lecture et participe du sens.

Abstract. This contribution proposes a new approach to the visibility of the literary text, without considering its spatial arrangement or the "editorial enunciation". But then what remains of the form of the text? There remain the paragraphs and, to a lesser extent, the chapters, which shape it according to their lengths. They give the text its morphology, invariable whatever the edition. We are therefore interested in paragraphs, not as autonomous textual units but taken within the linearity of the text. The method thus consists of measuring each paragraph of a text, then, from this numerical sequence, we realize visualizations which expose a hitherto unexplored rhythm of the text according to the length of its paragraphs. Starting from the visual dimension of the text, we arrive at its temporal dimension. We will see that this rhythm corresponds to the frequency of changes, of meaning for the narrative paragraphs and of speakers for the dialogic paragraphs. Finally, we will see that this dynamic form has impacts on reading and participates in meaning.

La forme visuelle du texte littéraire a longtemps été étudiée sous l'angle de sa disposition spatiale, assimilée à l'image du texte telle qu'analysée par Emmanuël Souchier dans *L'image du texte pour une théorie de l'énonciation éditoriale* (1998), Liliane Louvel dans *Texte Image. Images à lire, textes à voir* (2002), Annette Béguin-Verbrugge dans *Images en texte, images du texte* (2006) et Anne-Marie Christin dans *L'image écrite ou la déraison graphique* (2009).

Cette contribution propose une autre approche. D'une part, nous considérons que l'image du texte ne peut être que celle de sa totalité, en un regard, non d'une ou plusieurs pages. D'autre part, nous ne tenons compte ni de la disposition spatiale ni de « l'énonciation éditoriale »¹, lesquelles sont rarement décidées par l'auteur ou l'auteure. Mais alors, que reste-t-il de la visualité du texte ? Il reste les paragraphes et, dans une moindre mesure, les chapitres qui, visuellement, rythment le texte en fonction de leurs longueurs. Ce sont ces marques visuelles, provenant pour la plupart de l'auteure ou de l'auteur, qui donnent au texte sa morphologie. Nous nous intéressons donc aux paragraphes non en tant qu'unités textuelles autonomes, mais en tant qu'enchaînements pris dans la linéarité du texte. Nous nous rapprochons ainsi de la conception de Franck Neveu de l'image du texte.

L'image textuelle peut être minimalement décrite par deux ordres de configuration, linéarité et tabularité [...] au niveau des marquages les plus courants : les faits de *modulation* ([...] italique, gras, soulignement) et les faits de *segmentation* ([...] structure volumétrique des paragraphes, gestion des alinéas et des espaces, etc.) Bref, il s'agit de la signalétique qui est au service du repérage physique de la structure du texte » (2000/2014, p. 202).

Nous étudions ainsi la forme du texte et non sa mise en forme que nous dissociions, bien que les paragraphes et les chapitres participent également de la mise en forme du texte. Nous montrerons qu'ils ne sont pas *stricto sensu* des éléments graphiques ou esthétiques car ils obéissent principalement à une autre logique, selon laquelle « la pensée [...] commande le découpage »² (Castille, 2012, p. 150). « Le paragraphe est sans doute une des unités les plus fortes dans la construction des textes » (Coirier *et al.*, 1996, p. 192), piste que Jean-Michel Adam « suit résolument » (2018, p. 38).

Nous verrons que cette approche visuelle du texte littéraire hors « énonciation éditoriale » permet de découvrir un rythme jusqu'alors ignoré, spécifique de l'écrit.

Précisons limiter l'usage de « texte » à sa matérialité écrite, qui le distingue de l'oralité, et utiliser le terme « littéraire » pour désigner les romans, les nouvelles et les contes, à l'exception des textes tels que *La Maison des feuilles* de Mark Z. Danielewski (2002).

1 Les paragraphes, la morphologie du texte

En regardant un texte, nous observons des espaces blancs plus ou moins rapprochés qui créent un rythme visuel. Ces espaces correspondent aux alinéas qui marquent un changement de paragraphe. Un paragraphe correspond au texte compris entre deux alinéas (« à la ligne »). Bien que selon Roger Laufer « la ponctuation proprement dite se limite à la phrase » (1986, p. 75), l'alinéa est considéré comme un signe de ponctuation dans son acception extensive. Jacques Drillon l'inclut dans son *Traité de la ponctuation française* (1991). « Il [le paragraphe] signifie : quelque chose vient de se terminer. Une autre commence » (1991, p. 442). Nina Catach considère les alinéas comme un signe de ponctuation intérieur au texte, qui « dépend malgré tout dans une grande mesure de l'auteur » (1980, p. 19). Toutefois, près de quinze ans plus tard, elle précise *a contrario* que, pour la ponctuation, les auteurs « s'en sont remis alors et s'en remettent encore aux "spécialistes", par modestie ou par paresse » (1994, p. 44). Elle cite, à titre d'exemple, George Sand qui, en défendant sa ponctuation d'écrivaine, a été rabrouée par les imprimeurs. On ne sait pas si elle considérait ses paragraphes comme de la ponctuation. Nous estimons qu'à cette époque, par ponctuation, on entendait celle de la première approche présentée par Nina Catach qui correspond à la « ponctuation proprement dite » de Roger Laufer.

Selon les termes de Jean-Michel Adam, le paragraphe est « précisément une unité propre à l'écrit et doit donc être, avec l'étude de la ponctuation scripturale, un objet d'étude en soi » (2018, p. 17). Remontons à l'origine étymologique du terme « paragraphe ».

« Paragraphe » est emprunté au bas latin *paraphrasis*, « marque pour distinguer les différentes parties d'un exposé », lui-même emprunté au grec *paragraphe*, proprement « ce qui est écrit à côté ». Le mot s'applique au signe de ponctuation³ qui signale la fin d'un développement et désigne, en grec tardif, la section d'un texte terminée par un tel signe, qui correspond à une unité du développement.

Le mot a dès les premières attestations le sens de « section d'un texte en prose délimitée par des passages à la ligne. (Rey, 2001, p. 2428)

Par « paragraphes », nous entendons les paragraphes alinéaires, comprenant à la fois les paragraphes narratifs et dialogaux, visuellement repérables sans lire le texte. Nous utilisons ce terme « paragraphe alinéaire » – bien que cela puisse sembler redondant car un paragraphe est *de facto* compris entre deux alinéas – de la même manière que Jean-Michel Adam utilise le terme de « paragraphe graphique » (2018, p. 54), afin d'attirer l'attention sur le fait qu'un paragraphe n'est pas forcément un « groupe d'idées » (Drillon, 1991, p. 438), mais peut également être une graphie volontaire visant à mettre en relief une information. Un paragraphe peut être un mot, une phrase.

Afin de tenir compte de l'ensemble du texte et de préserver sa linéarité, nous choisissons de ne pas distinguer les paragraphes narratifs des paragraphes dialogaux, ce qui créerait deux groupes textuels discontinus.

1.1 La forme visuelle du texte déterminée par la longueur de ses paragraphes

La démarche qui fonde la présente contribution consiste en premier lieu à regarder un texte, sans le lire. Nous observons que c'est surtout au niveau des paragraphes qu'une segmentation textuelle s'opère, laquelle crée un rythme visuel en fonction de leurs longueurs, et que, dans une moindre mesure, une division textuelle s'effectue au niveau des chapitres, sauf s'ils sont petits, donc nombreux, auquel cas le rythme visuel s'effectue davantage à leur niveau. Quant à la phrase, elle n'est visible qu'à la lecture.

Dès lors, nous ne tiendrons pas compte des éléments graphiques de l'édition, tels que les marges, la disposition du texte dans la page et les blancs typographiques, qui relèvent d'une « mise en livre » du texte, en s'écartant *in fine* un peu du texte lui-même. Nous retenons uniquement les retours à la ligne, que l'on voit dans les logiciels de traitement de texte par le symbole du *pieds de mouche* (« ¶ ») et qui marquent la fin d'un paragraphe. Ce sont les paragraphes qui donnent au texte sa forme propre que nous appellerons « morphologie », car celle-ci ne varie pas d'une édition à une autre qui peut donner des visualités du texte différentes. Nous évitons ainsi toute confusion entre la forme et l'image du texte. Cependant, pour les textes antérieurs au XIX^e siècle, cette morphologie du texte peut varier selon l'édition. Ainsi, pour l'édition ancienne chez Claude Barbin de 1678, *La Princesse de Clèves* comprend 89 paragraphes, alors que, pour l'édition moderne chez Antoine Adam de 1966, le texte en comprend 129 (Arabyan, 1994, p. 124).

La morphologie du texte ne correspond ni à la forme du « texte premier » ni à celle du « texte second » énoncés par Emmanuël Souchier pour qui « le "texte premier" n'est autre que le texte de l'auteur à proprement parler. [...] Il est supposé être à l'origine du "texte pluriel" publié. [...] La fonction du "texte second" consiste à donner à lire le "texte premier" » (1998, p. 144). Il conclut que « l'énoncé de cette "énonciation" n'est donc pas le texte (le discours de l'auteur), mais la forme du texte, son image » (1998, p. 145). La morphologie du texte se rapproche plutôt de « la signalétique qui est au service du repérage physique de la structure du texte et que sollicite la nature fondamentalement visuelle. [...] Cette iconicité [...] est le texte lui-même, diversement spatialisé » (Neveu, 2000/2014, p. 202).

Relevons, de prime abord, que les paragraphes forment des blocs de lignes. Or, ces blocs ou ces masses résultent de la mise en page, d'éléments spatiaux extérieurs au texte tels que le format de la page et la largeur des marges qui le forcent à aller à la ligne. Sans les contraintes spatiales de la page, le paragraphe se poursuit sur une seule ligne continue jusqu'à sa fin.

Afin d'illustrer notre propos, prenons *Le Horla* de Guy de Maupassant. Réduisons la taille de police de caractères au maximum, une façon d'éliminer les contraintes spatiales de la page. Nous obtenons la morphologie du texte suivante.

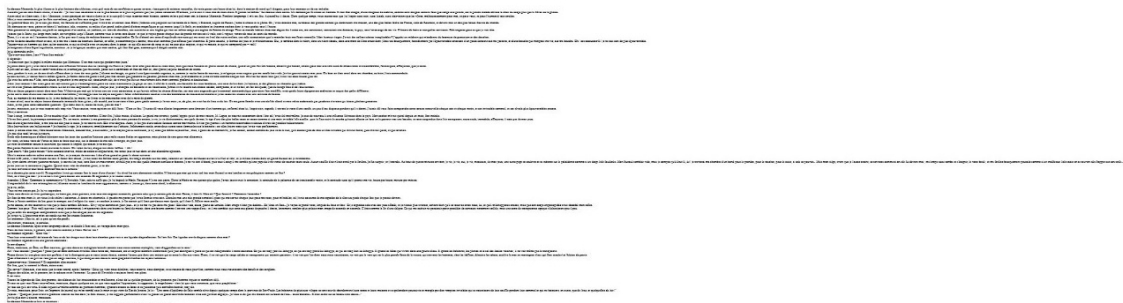


Figure 1. Capture écran du texte *Le Horla* dont la taille de police de caractères a été réduite à 1

On constate que, indépendamment de la mise en page et de l'édition en général, le texte se compose d'autant de lignes que de paragraphes.

1.2 Les paragraphes et non les chapitres : une différence de nature

Nous avons vu que le rythme visuel s'effectue surtout au niveau des paragraphes et, dans une moindre mesure, des chapitres. Parfois, lorsque ces derniers sont très courts et donc nombreux, il s'effectue à leur niveau. Aussi, dans un premier abord, nous nous sommes également intéressés aux chapitres, malgré leur nature fondamentalement différente. Si tous deux constituent des marques visuelles et des unités de sens, le paragraphe est un segment textuel tandis que le chapitre est une division textuelle. Il existe une distinction substantielle entre un segment et une division : le segment a partie liée, tandis que la division est une partie.

Les chapitres structurent le texte de façon évidente. Ils sont pensés, réfléchis par l'auteur ou l'auteure, tandis que les paragraphes sont des formes sensibles, suivant le mouvement de la pensée. Ils ne sont pas systématiquement réfléchis. On notera d'ailleurs qu'il existe des termes pour l'acte qui consiste à créer des chapitres, par exemple « chapitrage » et « chapitrer », et aucun, sinon peu usités, pour celui qui consiste à créer des paragraphes.

Le chapitre peut fonctionner comme une unité textuelle autonome, comme un « épisode » selon Max Vicedo dans son article « *Salammô*, un roman par épisodes » dans lequel il précise :

Chaque chapitre devient un tableau, complet et autosuffisant, dont les marques de divisions matérielles, telles que le titre ou le blanc, sont soutenues par cette logique de fragmentation. Le chapitre peut alors être regardé comme une partie distincte de l'ensemble pouvant fonctionner et survivre par lui-même, et acquiert par là le statut d'*épisode*, au sens antique du terme. (2011, p. 250)

Il tempère néanmoins. « *Salammô* apparaît comme une suite de tableaux, peints par un artiste, liés entre eux car étant chacun partie d'un même sujet » (2011, p. 252). Ugo Dionne, quant à lui, traite de la fonction rythmique des chapitres par les « interstices » qu'ils créent telles « des aires de repos, des stations où [le lecteur] peut "reprendre son souffle" » (2008, p. 251). Il parle même de « rupture chapitrale ». Certains auteurs, tels Georges Perec et Régis Jauffret, jouent d'ailleurs avec cette autonomie du chapitre et considèrent que leurs ouvrages respectifs, *La Vie mode d'emploi* et *Promenade*, peuvent être lus comme s'il s'agissait de puzzles aux multiples combinaisons sans altérer le sens.

Philippe Hamon ouvre les actes du colloque du projet *ANR Chapitres* ainsi : « le chapitrage est toujours senti comme un "agencement" plus ou moins arbitraire, toujours senti comme une forme imposée de l'extérieur [...], comme l'intervention extérieure surplombante et régissante d'une *autorité* » (2017, p. 12). Les chercheurs de ce projet conçoivent le chapitre comme une « division naturelle », un « ordonnancement », une « instance démarcative », une « frontière », une « coupe », une « coupure », un « fragment kaléidoscopique », une « brisure », un « fractionnement d'un donné linéaire (une lignée) », un « dispositif », un « agencement plus ou moins arbitraire », une « structure », un « bloc de sens », un « opérateur de la distinction », une « unité de construction ».

La conception du chapitre pour ces chercheurs opère au niveau de la coupure, de la pause, de la forme passive, de la rupture et de la division. Or, pour qu'il y ait pause, il faut une continuité et une dynamique. Notre approche, plus adaptée au paragraphe, est celle du flux, du souffle et de l'enchaînement. Pour reprendre les termes du début, le chapitre divise, tandis que le paragraphe segmente, plus ou moins rapidement en fonction de sa longueur, tout en préservant l'idée du lien.

S'intéresser à la façon dont les segments textuels (les paragraphes) s'enchaînent revient à s'intéresser à la configuration temporelle du texte, où les paragraphes constituent des durées variables selon leur étendue ou leur longueur. Cette approche temporelle du texte à partir de sa morphologie n'a jusqu'alors pas été explorée. Nous constatons que certains chercheurs du projet *ANR Chapitres* ont perçu cette dimension temporelle du chapitre, qui peut s'appliquer au paragraphe.

Le chapitre [...] peut être abordé non tant comme une structure ou une « division », que comme une étendue, selon une perspective non plus verticale (le chapitre comme unité intermédiaire en une plus grande et une plus petite, ou comme élément emboîté), mais horizontale (le chapitre comme durée, comme espace d'action ou de réflexion). [...] Une

telle perspective n'est évidemment pas sans poser une difficulté de mesure. (Daunais, 2017, p. 281)

La longueur des chapitres crée un rythme plus ou moins rapide ou lent. (Introduction du livre, p. 20)

On se demandera ici ce que le rythme des chapitres fait à la lecture. (Introduction du chapitre cinq, p. 277)

Le roman progresse en effet en faisant alterner de manière imprévisible des échanges qui tourment court (« abrégez ») et de longs segments narratifs qui racontent des bouts d'histoires (« poursuivez », « décrivez », « détaillez ». (Piéguy, 2017, p. 305)

Étant donné l'impact limité des chapitres sur le rythme visuel et leur nature réfléchie – expliquant ainsi la régularité parfois quasi métronomique de leurs longueurs, introuvable pour les paragraphes –, notre contribution se limitera ultérieurement à l'étude des paragraphes.

2 Une méthode de visualisation de la longueur des paragraphes et des chapitres pris dans leur enchaînement

Nous avons élaboré une méthode en deux phases : la première consiste à mesurer chacun des paragraphes et des chapitres du texte sélectionné ; puis, à partir de ces mesures, à savoir de cette suite numérique – autant de nombres que de paragraphes ou de chapitres dans le texte – la seconde phase consiste à réaliser des visualisations en fonction de la longueur des paragraphes ou des chapitres. Ce processus de visualisation se base sur le logiciel *Narra 2.0* développé afin, d'une part, d'automatiser les mesures, quelle que soit l'édition, et, d'autre part, de pouvoir numériquement les utiliser pour en extraire des données statistiques et pour en réaliser des œuvres digitales, sur un large corpus textuel.

2.1 Mesurer la longueur des paragraphes et des chapitres d'un texte avec le logiciel *Narra 2.0*

Commençons par citer le linguiste et philologue George K. Zipf à l'origine d'une loi mathématique éponyme⁴, car il a joué un rôle pionnier en analyse statistique du langage. Il considérait que « tant que les moyens de mesurer les phénomènes d'un domaine donné n'ont pas été trouvés, on ne peut ni faire de ce domaine une science exacte ni étudier sa dynamique avec un degré de précision acceptable » (1935/1974, p. 20) et que sa « méthode de travail est l'application des principes statistiques au phénomène observable du flux du discours » (1935/1974, p. 16).

La méthode numérique ici mise en œuvre consiste tout d'abord à mesurer la longueur des paragraphes et des chapitres des textes littéraires afin de comparer objectivement leurs longueurs d'un texte à l'autre. En offrant « une sorte d'indépendance par rapport au sujet pensant » (Crosby, 2003, p. 222), cette méthode élimine l'interprétation. Selon François Dagognet, « par principe c'est la mesure seule qui permet l'intelligence de la chose même [...] Finalement la mesure (penser d'ailleurs vient de "peser" [du latin *pendere* signifiant "peser", "apprécier"]) relève moins du simple reflet que de la réflexion » (1993, p. 158-159). Les mesures sont effectuées avec le logiciel applicatif *Narra 2.0* développé à cet effet, en nombre de caractères (ou signes). L'interface d'accueil permet à l'utilisateur de choisir entre la longueur des paragraphes ou celle des chapitres. L'interface suivante permet de choisir entre un texte long, tel qu'un roman, pouvant comporter plus de 10 000 paragraphes, ou un texte court, comme une nouvelle ou un conte, comportant en général entre 100 et 1000 paragraphes. Ce texte doit être sous format ePub ou Word.

Ainsi, en entrée, il est nécessaire de fournir un texte que le logiciel va parcourir.

En sortie, pour les paragraphes, il génère :

- un tableau comprenant les mesures de chaque paragraphe – autant de nombres que de paragraphes –, ainsi que des données statistiques : des valeurs brutes, des indicateurs de tendance centrale, un indicateur de dispersion ;
- un histogramme en fonction de la longueur des paragraphes du texte sélectionné (Figure 2) ;

- une courbe des fréquences (Figures 3 et 4).

Pour les chapitres, il génère :

- un tableau comprenant les mesures de chaque chapitre – autant de nombres que de chapitres –, ainsi que des données statistiques : des valeurs brutes, des indicateurs de tendance centrale, un indicateur de dispersion, un indicateur de forme ;
- un histogramme en fonction de la longueur des chapitres du texte sélectionné (Figure 5).

Prenons l'exemple de *Notre-Dame de Paris* de Victor Hugo. Outre le tableau des mesures que le logiciel génère, ci-dessous sont présentées les captures des interfaces de résultat.

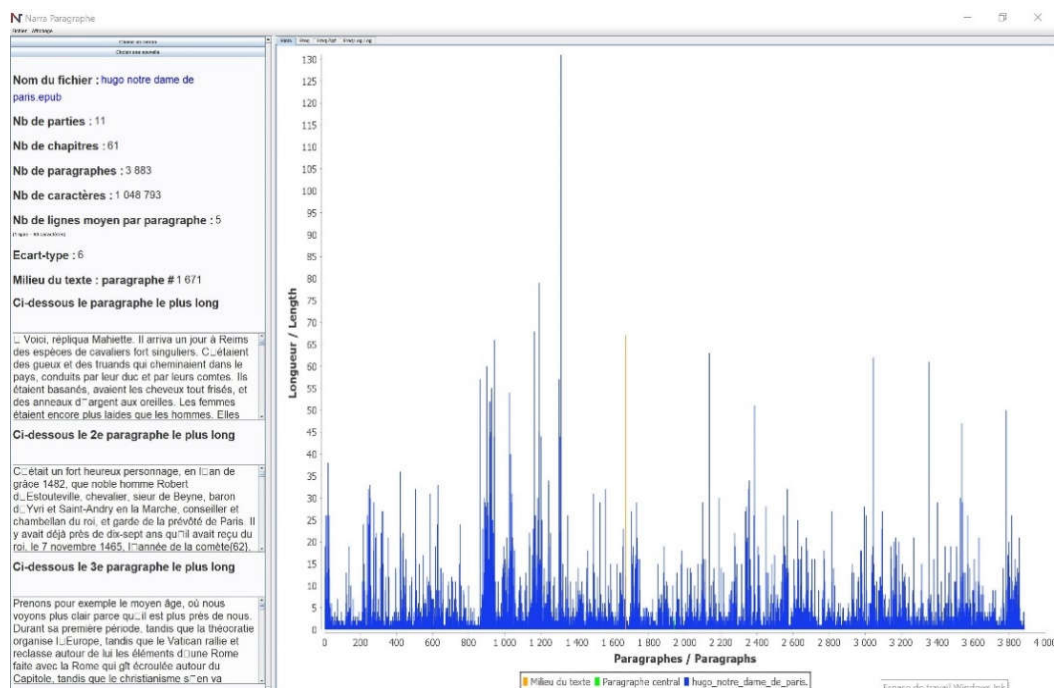


Figure 2. Capture écran de l'interface du logiciel *Narra 2.0*.

Histogramme en fonction de la longueur des paragraphes de *Notre-Dame de Paris*.

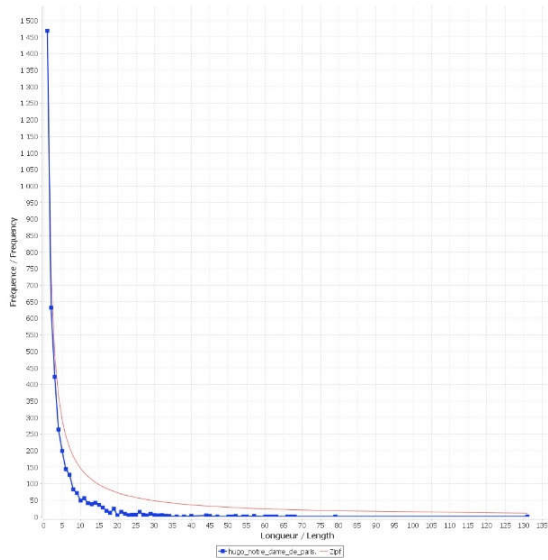


Figure 3. Capture écran de l'interface du logiciel *Narra 2.0*. Courbe de fréquence en fonction de la longueur des paragraphes de *Notre-Dame de Paris*

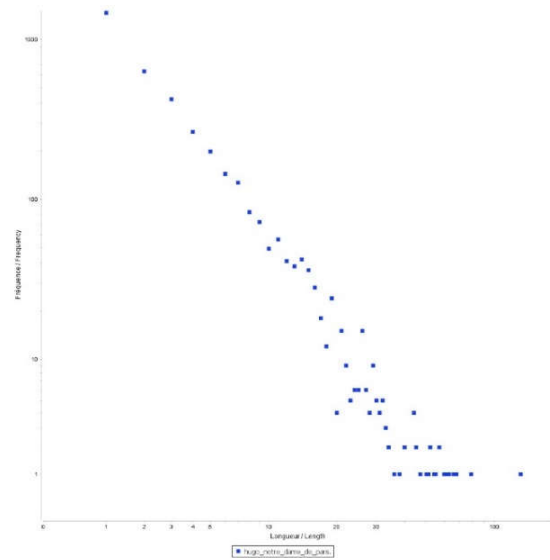


Figure 4. Capture écran de l'interface du logiciel *Narra 2.0*. Courbe log-log de fréquence en fonction de la longueur des paragraphes de *Notre-Dame de Paris*

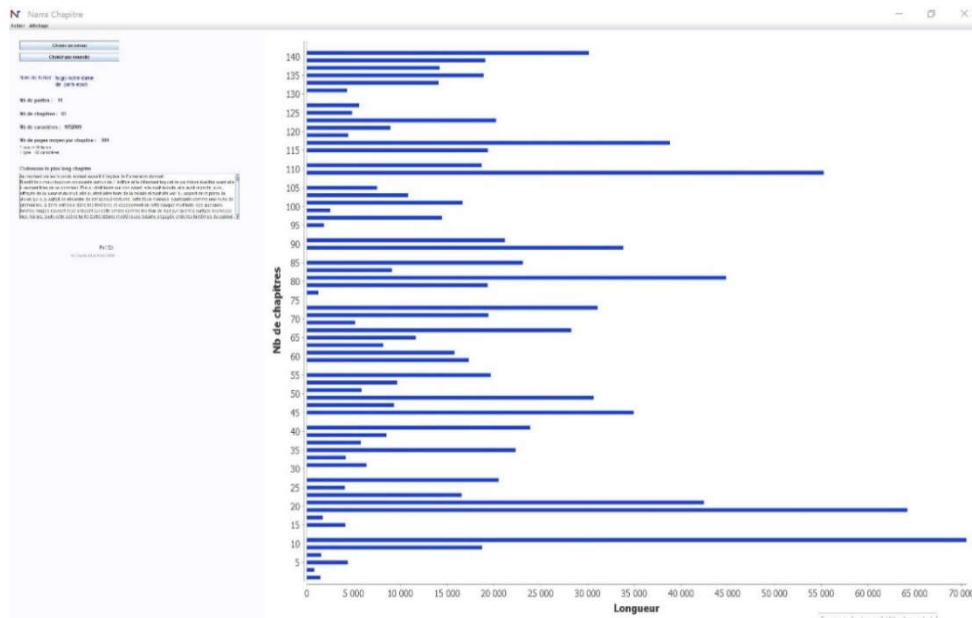


Figure 5. Capture écran de l'interface du logiciel *Narra 2.0*. Histogramme inversé en fonction de la longueur des chapitres de *Notre-Dame de Paris*. Le premier chapitre se situe en haut jusqu'au dernier en bas. Le logiciel a « considéré » par erreur les 2 dernières notes comme des chapitres.

Ces figures sont la partie visible de l'interface du logiciel. Mais le tableau qu'il génère – comprenant les mesures de chaque paragraphe (autant de nombres que de paragraphes) ainsi que les données statistiques

mentionnées plus haut – est si fondamental que nous l'appelons *Matrice*, car il permet d'exploiter ces données, d'effectuer des comparaisons entre les textes d'un même auteur ou entre les auteurs et constitue la base du processus de visualisation.

2.2 Visualiser les longueurs successives des paragraphes par un processus numérique

À partir de cette *Matrice* ou suite de nombres (ou mesures de la longueur des paragraphes d'un texte), un autre processus numérique a été développé pour réaliser un ensemble de visualisations pour un même texte. Chaque type de visualisations, tel que *Sismogramme*, *Allover* ou *Étoile*, a nécessité le développement d'un algorithme spécifique. Elles permettent de visualiser la longueur de tous les paragraphes d'un texte en un regard et, ainsi, un rythme du texte.

Les visualisations sont réalisées en utilisant un code couleur basé sur le spectre de la lumière visible, partant du court/rouge pour aller vers le long/bleu. Une couleur se définit par sa longueur d'onde, notée λ . Les couleurs sont en effet des ondes électromagnétiques visibles (propagation d'énergie d'un point à un autre de l'espace) véhiculées par le photon, particule vectrice du rayonnement électromagnétique (ou force), qui se déplace à la vitesse de la lumière dans le vide, notée v , à près de 300 000 km/s. L'onde lumineuse « rouge » est la plus rapide du spectre visible. Elle arrive en premier sur la rétine de l'observateur, car sa longueur d'onde est la plus grande et la distance que l'onde « rouge » parcourt en un temps donné t est grande. À l'inverse, l'onde lumineuse « bleu » est plus lente, car sa longueur d'onde est plus petite et la distance que l'onde « bleu » parcourt en un temps donné t est plus petite. Sur ce fondement, le code couleur commence donc par le rouge pour un paragraphe court (temps bref), passe par le jaune-vert pour un paragraphe de moyenne longueur et se termine par le bleu pour un paragraphe long (temps long). Ainsi, chaque couleur caractérise une longueur de paragraphe précise. La couleur devient donc un langage en soi.

Quelques œuvres produites, dont le titre porte toujours celui du texte, sont visibles sur ce [site](#). De plus, une application Web nommée [Narra Horizon](#) a également été développée pour 250 textes courts, des nouvelles et des contes.

Bien que les œuvres présentent des similitudes avec la *datavisualisation*, nous estimons qu'elles ne peuvent y être assimilées. Comme son nom l'indique, cette dernière repose sur des *data*. Dès lors que l'on a un enchaînement de longueurs ou de durées variables non permutable, la notion de temps intervient inévitablement. À cet égard, une analogie peut être établie avec la musique constituée, entre autres, d'une suite de durées et de hauteurs (fréquence). Les spectrogrammes sont-ils de la *datavisualisation* ? Non, ils ne sont pas une visualisation de données relatives à l'œuvre sonore ; ils sont une visualisation de l'œuvre musicale, autrement dit de la composition musicale, dont les éléments s'enchaînent sans pouvoir être ni permutés ni dissociés. Il en va de même avec le texte. Ainsi, les diagrammes et les œuvres réalisées sont des visualisations de l'œuvre littéraire, autrement dit, de la composition écrite dont les segments ont été mesurés en respectant leur ordre d'apparition. Les mesures prennent la totalité du texte en respectant la linéarité du texte. Elles sont un instrument de transposition du texte qui permet de le condenser afin de le visualiser en un coup d'œil. Si, au contraire, nous avançons que les œuvres se basent sur les données du texte, il serait juste de penser qu'elles sont réalisées à partir d'éléments tels que la fréquence de mots, la répartition des accents, etc. Or, elles ne sont pas créées à partir d'éléments rassemblés pour former un ensemble, mais à partir d'un ensemble séquencé qu'est le texte. Certes, *stricto sensu*, cela est exact, puisque chaque mesure des paragraphes ou des chapitres constitue une donnée et que les graphiques et les œuvres créés se fondent sur une suite de nombres, donc de données. Cependant, cela revient à s'arrêter aux nombres, alors que ce sont leurs rapports qui comptent. Nous en revenons à l'idée du lien. Nous travaillons sur la façon dont la narration se déploie, dont les segments se lient les uns aux autres, plus ou moins vite selon leurs longueurs ou leurs durées, c'est-à-dire sur l'enchaînement du texte, sa dynamique, à l'antipode des données du texte. En d'autres termes, c'est le processus qui nous intéresse, non les unités en elles-mêmes.

3 Résultat : un autre rythme du texte mis au jour

Les diagrammes et les œuvres visuelles réalisés à partir de la mesure des paragraphes agissent comme des révélateurs ; ils rendent visible un rythme. Mais, quelle est la nature de ce rythme ? Parmi les multiples définitions existantes, nous en retenons deux.

La première est celle qui prévalait chez les Présocratiques, telle que décrite par Émile Benveniste dans son étude étymologique du mot « rythme » (1966). Il est dérivé de *rhein* « couler » et désigne « la forme distinctive, l'arrangement caractéristique des parties dans un tout » (1966, p. 330), « littéralement "manière particulière de fluer" » (1966, p. 333). Or, selon Benveniste, Platon en donne une nouvelle acception, toujours actuelle, en l'associant au nombre et à la mesure au détriment des notions d'agencement spécifique et continu. « Cette "forme" est désormais déterminée par une "mesure" et assujettie à un ordre. [...] Et c'est l'ordre dans le mouvement » (1966, p. 334). « On atteint le "rythme", configuration des mouvements ordonnés dans la durée » (1966, p. 335).

La seconde conception du rythme est celle de Paul Fraisse : « *le rythme comme structure dynamique* » (1974, p. 94). Par « structure », il entend « un jeu de brèves et de longues » (1974, p. 88), une combinaison de durées différentes, l'organisation d'une série successive d'éléments. Par contre, nous écartons l'idée de périodicité qu'il avance en ce « qu'il n'y a pratiquement jamais rythme sans des retours isochrones plus ou moins mis en valeur au niveau de la perception » (1974, p. 113). À ces notions de « structure » et de « retour », Pierre Sauvanet y ajoute « le mouvement » et définit le rythme comme « tout phénomène, perçu ou agi, auquel on peut attribuer au moins deux des qualités suivantes : structure, périodicité, mouvement » (1992, p. 238).

3.1 Du visible au temporel : une composition

De la dimension spatiale du texte, nous sommes passés à sa dimension temporelle, puisque les nombres ici ne caractérisent pas des quantités, mais des longueurs liées les unes aux autres, et une succession d'éléments liés les uns aux autres (les paragraphes) dans un ordre non permutable advient nécessairement dans le temps. Le texte passe d'un temps étalé de l'œuvre écrite, page après page, imprimée ou numérique, à un temps condensé, par la mesure, des diagrammes et des œuvres visuelles. Notons que l'étymologie du mot « nombre » le relie à la mesure et au rythme. « Nombre vient du latin *numerus* "partie d'un ensemble classée à son rang". [...] En outre, *numerus* a servi à rendre toutes les acceptions techniques du grec *arithmos* (arithmétique), "mesure, rythme", "nombre grammatical", "foule, quantité" » (Rey, 2001, p. 2258).

En mesurant des longueurs chaînées, nous mesurons des durées. L'unité de temps retenue est de 60 caractères, soit, pour simplifier et saisir immédiatement sa dimension, environ une ligne. Par conséquent, la durée d'un paragraphe de 120 caractères, soit environ 2 lignes, est de 2 temps, etc. Cette équivalence est possible car le texte est une succession de segments et que ces segments sont continus selon un ordre déterminé. On obtient ainsi une succession de durées variables chaînées, c'est-à-dire un rythme, de « *rhuthmos*, au sens de "modalité d'un accomplissement" » (Michon, 2018, p. 25), de « manière particulière de fluer ».

Outre, l'histogramme de l'interface graphique du logiciel (Figure 1), à partir du tableau des mesures de la longueur des paragraphes d'un texte, un sismogramme est généré avec un algorithme spécifique. Les sismogrammes sont semblables aux histogrammes. La seule différence étant que les lignes sont disposées à leur base pour l'histogramme, tandis qu'elles le sont en leur centre pour le sismogramme. Pour les textes courts comme les nouvelles, les sismogrammes sont plus « lisibles ».

Prenons l'exemple du sismogramme *Accident nocturne* de Patrick Modiano (Figure 6). Remarquons qu'il peut porter à confusion. En effet, contrairement à ce qu'on peut penser, son centre sur l'axe horizontal ne correspond pas au milieu du texte, lequel se situe vers le 100^e paragraphe, alors que le texte comprend 350 paragraphes. Cela s'explique car une ligne d'un pixel est placée pour chaque paragraphe. Seule la



Figure 6. Sismogramme
Accident nocturne

hauteur de la ligne varie selon la longueur du paragraphe. Dès lors, le centre du sismogramme correspond au paragraphe central, le $175^{\circ}(350/2)$. Vu que les paragraphes du début sont beaucoup plus longs que ceux de la seconde partie, il est normal que le milieu du texte (en nombre de signes ou de mots) se situe à un tiers du texte, vers le 100° paragraphe.

Au regard du sismogramme, une structure en deux ou trois temps apparaît. Elle ne correspondant à aucune division apparente dans le texte. Deux temps, si l'on considère les deux grandes amplitudes comme formant une unité, avec un premier temps au rythme ample et étendu, constitué de longs paragraphes, et un second temps au rythme saccadé et rapide, constitué de paragraphes courts. Trois temps, si l'on considère que les deux grandes amplitudes se distinguent, dont la séparation correspond à une courte liste de noms et d'adresses suivie, peu après, par un petit dialogue.

On observe que le changement radical entre les paragraphes longs et les paragraphes courts se produit exactement au milieu du graphique, soit au niveau du paragraphe central, c'est-à-dire aux trois quarts du texte. Cette rupture visuelle résulte du changement de nature des paragraphes : narratifs, puis dialogaux. Il est intéressant de relever qu'elle coïncide à un moment de basculement dans le récit, celui de la seconde rencontre tant recherchée.

En ce qui concerne l'histogramme *Notre-Dame de Paris* (Figure 1), une structure se dégage également, laquelle ne correspond, aussi, à aucune division apparente du texte. Elle se compose en trois temps. Un premier temps, avec un ensemble de paragraphes assez courts, un deuxième comprenant un grand nombre des paragraphes longs, y compris le plus long du texte, et un troisième composé d'une alternance de paragraphes courts et longs. Cette deuxième partie inclut les livres 3, 4, 5 et 6 du roman ainsi que les deux paragraphes les plus longs. La question se pose de savoir si ces deux paragraphes distinctifs, notamment le plus long, sont des moments clefs dans la narration. Si le premier l'est, le second en revanche ne semble pas être important. Le premier, en effet, est une description qui, par la suite, permet au lecteur d'établir le lien entre les personnages et de comprendre l'intrigue.

Cette méthode permettrait de mettre au jour, outre un rythme du texte selon la longueur de ses paragraphes, une structure compositionnelle. Jean-Michel Passerault et David Chesnet le suggèrent : « la rupture de paragraphes pourrait donc constituer un des indices permettant l'identification des différentes parties de textes » (1991, p. 160). Michel Charolles considère que les séquences de paragraphes participent des plans d'organisation textuelle. « Les marques de séquentialité sont l'indice d'une activité *métadiscursive* chez celui qui les utilise. Elles dénotent un travail explicite d'organisation de l'énonciation visant en particulier à faciliter la tâche de l'interprétation » (1988, p. 9). Dans une étude sur « les modèles d'organisation thématique des paragraphes et entre les paragraphes », Guy Achard-Bayle et Ondřej Pešek conclut en premier lieu que « l'analyse a confirmé l'hypothèse qui dit que la structure textuelle matérielle a pour but de faciliter la *visibilité* des unités du plan du texte. Chapitres, paragraphes, pieds de mouche, lettrines : tous ces éléments délimitent des segments qui correspondent aux parties de la structure compositionnelle et thématique du texte » (2020, p. 28). Dans une autre étude, Ondřej Pešek souligne qu'au « niveau formel, la structure compositionnelle du texte est fortement corrélée à la division en paragraphes » (2022, p. 122). « Il est important de souligner que si du point de vue superficiel, les paragraphes constituent une suite linéaire, au sein de la structure, ils se regroupent en entités compositionnelles d'une complexité variable, formant ainsi des blocs hiérarchisés » (2022, p. 123). Cette dimension compositionnelle des paragraphes en tant que structure segmentale actionnelle et thématique, telle qu'une introduction, une présentation de thème, une explication, participe de la structure globale du texte.

Il conviendrait néanmoins d'étudier un large panel de textes et d'effectuer le recoupement avec la narration afin de savoir si cette structure est l'indice d'épisodes dans le récit et/ou si elle correspond à des types de séquences : narrative, descriptive, argumentative, explicative et dialogale.

Enfin, les figures 3 et 4 permettent d'émettre l'hypothèse de l'applicabilité de la loi de puissance, celle de Zipf déjà mentionnée. Elles montrent que plus les paragraphes sont longs moins ils sont fréquents selon un certain coefficient. Le graphique d'une étude réalisée par le laboratoire Litlab sur un corpus réduit de romans de la seconde moitié du XIX^e siècle, dont les paragraphes ont été repérés manuellement, confirme cette hypothèse (Algee-Hewitt et al., 2015, p. 7).

En outre, les données statistiques produites par le logiciel figurant dans le tableau des mesures restent très limitées et mériteraient d'être traitées avec le logiciel R sur un large corpus, comme cela a été mis en œuvre pour l'analyse d'*Un Atlas de textes littéraires, avec une classification en fonction de la longueur* (Tardif, 2021).

Enfin, plus généralement, la démarche gagnerait à être développée pour une étude contrastive de textes d'un même auteur ou autrice, entre auteurs, entre genres littéraires, entre textes littéraires de différentes époques, sachant que le genre littéraire est lié à une époque, comme l'a démontré Matthew L. Jockers (2013, p. 85) qui affine une étude de Franco Moretti, voire entre des textes littéraires et non littéraires (philosophiques, historiques, journalistiques, etc.).

3.2 La longueur des paragraphes dans la linéarité du texte, un rythme ignoré...

Si « le "corps" du texte est sa dimension graphique et matérielle, souvent occultée par sa dimension verbale, linguistique, dont il ne peut pourtant être dissocié : l'écriture en tant que matérialité participe à l'écriture en tant qu'invention verbale » (Dahan-Gaida, 2023, p. 299), nous avons mis en évidence que le texte possède également une dimension temporelle. Bien que l'étymologie du mot « texte » est emprunté au latin *textus* « tissu, trame, enchaînement », du verbe *texere* « tisser », qui « récuse le dogme de la linéarité » (2023, p. 298), nous nous intéressons précisément à cette linéarité, au sens d'enchaînement de segments de longueurs et de durées variables. Le texte advient au fil de ses développements. Il se déploie dans le temps. C'est cette actualisation du texte, ce processus dynamique, que nous exposons. « Le livre n'échappe pas à sa nature successive [...] il se découvre dans le temps ; il obéit à des rythmes, à des mouvements, à des cadences » (Rousset, 1962, p. XIII). Le rythme du texte découvert, au gré de la longueur de ses paragraphes, est l'un de ces rythmes.

En effet, nous considérons les paragraphes dans leur rôle séquentiel. C'est la nature même du paragraphe que de découler du précédent en préparant le suivant tout en étant une entité en soi. Harald Weinrich l'exprimait déjà ainsi :

Le déroulement textuel, oral ou écrit [...] C'est évidemment un déroulement dans le Temps. [...] C'est en ce sens qu'on parle de "flux" informationnel. [...] Tout signe linguistique faisant partie d'un texte est entouré par un avant et un après textuel. Information préalable et information à venir contribuent à sa détermination. D'une manière générale, nous pouvons parler de Temps du texte [Textzeit]. (1973, p. 67-68)

Cependant, une question se pose : qui de l'enchaînement ou de la coupure est ici l'objet d'étude ? S'intéresser aux paragraphes, c'est s'intéresser à ce qui les met en mouvement dans un incessant va-et-vient entre une fin et un recommencement. Parce que ce sont principalement les paragraphes qui donnent au texte sa morphologie, en fonction de leurs longueurs, nous nous sommes intéressés à eux.

Ce rythme du texte résulte pour partie de la manière d'écrire de l'écrivain ou de l'écrivaine et pourrait être considéré comme une composante du style. La notion de style est appréhendée comme « l'ensemble des traits distinctifs [...] comme une façon particulière de dire les choses » (Monière et Labbé, 2002, p. 2), soit comme « une manière d'être distinctive » (Cressot, 1974, p. 304). Vu que ces trois chercheurs y incluent la longueur ou le « volume » des phrases, il est possible de considérer que la longueur des paragraphes y participe également. On notera également que John B. Colby constate aussi que « *the length of a paragraph depends a great deal on the current style* » et que « *philosophical works and journals tend to have longer paragraphs than the more popular writings do* » (1971, p. 13). Enfin, dans le chapitre « de la grammaire

au style » de son ouvrage *La stylistique*, Joëlle Gardes-Tamine consacre une section au paragraphe, concluant qu'il « joue un rôle dans le texte en le structurant ».

Lors du colloque *Qu'est-ce que le style ?* (Molinié et al., 1994), le style est non seulement considéré comme une marque distinctive d'un auteur ou d'une autrice mais aussi comme le reflet de la « personnalité » de son époque et du genre littéraire. Le préliminaire, qui synthétise les actes du colloque, s'ouvre ainsi : « le style est dans l'écriture, l'inscription d'une individualité, ou bien la marque d'un genre ou d'une époque » (Martin, 1994, p. 12). Jean-Michel Adam l'exprime en ces termes : « il s'agira donc de passer d'une conception du style comme écart, au fait de style pris dans l'unité organique d'une œuvre spécifique, historiquement située » (1994, p. 17).

Toutefois, nous avons montré que la longueur des paragraphes dépend également du genre littéraire. En effet, l'étude précédemment mentionnée portant sur un corpus de 1 255 textes de 162 auteurs et autrices (Tardif, 2018) a permis de démontrer l'existence d'une corrélation entre la longueur des paragraphes et le genre littéraire ou la période d'écriture (Tardif, 2021), notamment pour les romans de science-fiction, d'horreur et de fantasy. En outre, un auteur tel que Tolstoï, qui a écrit des textes de nature différente, présente une grande diversité de longueur des paragraphes selon ses textes. Reste à savoir si le genre littéraire contraint la longueur des paragraphes au point de prédominer la façon dont l'auteur ou l'autrice segmente son texte. Une étude menée par Carlos Maciel, Damon Mayaffre et Laurent Vanni sur un corpus de discours politiques français et brésiliens atteste que « le genre transcende la langue et les auteurs » (2022).

Selon John B. Colby, la longueur des paragraphes dépend du style actuel (1971, p. 13). Il constate que, dans les écrits plus anciens, la longueur des paragraphes était bien plus grande et qu'elle variait en fonction de l'audience : les paragraphes des écrits populaires étaient beaucoup plus courts que ceux des traités philosophiques. En conséquence, la morphologie du texte résulte d'une dynamique d'écriture, qui dépend non seulement de l'autrice ou de l'auteur, mais aussi du genre littéraire et de l'époque durant laquelle le texte a été écrit.

Jean-François Castille propose deux conceptions du paragraphe. La conception rhétorique, qui, selon lui, prévalait jusqu'au XIX^e siècle, selon laquelle le paragraphe est « un marqueur empathique d'articulation de sens » (2012, p. 152). Il ne résulte donc pas du choix de l'autrice ou de l'auteur. Avec Victor Hugo, la conception poétique se développe, selon laquelle le paragraphe joue un rôle de « blanc qui fait signe ». Il résulte désormais du choix de l'auteur. Flaubert ou Maupassant, qu'il prend pour exemple, l'utilisent « pour signaler un changement de point de vue » (2012, p. 157). Dans la première conception, les paragraphes sont des marqueurs du flux de la pensée et donc involontaires. Dans la seconde, ils sont des signes de ruptures indépendants du sens, utilisés volontairement. Il est intéressant de noter que ces conceptions reprennent, entre autres, celles d'Henri Meschonnic du rythme rhétorique et du rythme poétique. Cette conception dichotomique du paragraphe pose la question de savoir qui les détermine et ce qui les détermine, permettant de comprendre la nature de ce rythme mis au jour. Nous savons que la longueur des paragraphes dépend du genre littéraire et de l'époque d'écriture. Nous allons voir que les paragraphes sont non seulement des marqueurs thématiques, donc liés au sens, ou des marqueurs de locuteurs dans les dialogues. Leur longueur dépend ainsi de la manière dont les thèmes ou les répliques sont développés.

3.3 ... qui transcrit la fréquence des changements

« Les repères de paragraphes sont informatifs » (Stark, 1988, p. 299). Mais quelles informations délivrent-ils ? La visualisation de la longueur des paragraphes successifs d'un texte expose la fréquence des changements et leur répartition dans le texte, changement de locuteurs pour les paragraphes dialogaux et changement de « thème » – au sens de combinaison ou d'interaction de thèmes (le plus souvent de deux ou trois) comme l'a prouvé le laboratoire Litlab (Algee-Hewitt et al., 2015, p. 8) qui, par ailleurs, avait pour objectif de départ d'étudier le style à l'échelle du paragraphe. Leur étude a démontré que le paragraphe narratif est par nature une unité thématique.

Si les paragraphes nous permettent de "voir" des thèmes, c'est parce que les thèmes, n'"existent" pleinement qu'à l'échelle du paragraphe. Cette question ne relève pas de l'épistémologie, mais plutôt de l'ontologie » (2015, p. 21).

Si on utilise la modélisation thématique pour analyser la littérature, *le paragraphe fournit une unité plus efficace que tout autre segment "mécanique"*, et que cette unité devrait donc être utilisée de préférence par les chercheurs. [...] Le paragraphe, qui concentre le matériau thématique au sein d'un espace limité, est *l'habitat textuel des thèmes* (2015, p. 8).

Ils ont également montré qu'un changement de paragraphe est un indice de discontinuité thématique, plus qu'un autre segment textuel, notamment de 82 mots et de 200 mots, mais pas de façon systématique. Des études antérieures l'avaient déjà montré (Piérard & Bestgen, 2006), (Ji, 2008), (Stark, 1988) et (Bond & Hayes, 1984). Un changement de paragraphe résulte d'un changement de point de vue, de temps ou de lieu (Arabyan, 1994, p. 98) ou d'une mise en évidence d'un élément du texte.

Guy Achard-Bayle et Ondřej Pešek considèrent que « le dispositif de marquage typographique, qui participe de la matérialité textuelle, ne se réduit pas à une fonction esthétique » (2020, p. 4). Ils mentionnent Frantisek Daneš pour qui « le paragraphe est plus qu'une "unité graphique" » (2020, p. 8).

C'est le principe thématique qui détermine d'une manière décisive la segmentation opérée (2020, p. 20). La disposition graphique du texte en reflète la structure compositionnelle ("rhétorique") qui est ainsi signalée et donc communiquée au récepteur du texte. Si le paragraphe est incontestablement une unité graphique, il l'est aussi pour des raisons de contenu et de composition (2020, p. 30).

Ils concluent que le paragraphe est, certes une unité graphique, mais il est aussi une unité sémantique (rhétorique ou thématique) et compositionnelle.

Ainsi, la longueur des paragraphes narratifs dépend de la manière dont l'autrice ou l'auteur développe ses thèmes. Plus sa façon d'écrire est descriptive et explicative, plus les paragraphes seront longs, donnant ainsi un rythme ralenti, posé. À l'inverse, s'il passe d'un thème à un autre et que les dialogues sont nombreux, dont les répliques des locuteurs sont généralement courtes, le rythme du texte sera saccadé et rapide. Ce qui explique l'influence du genre littéraire sur la longueur des paragraphes. En somme, plus la fréquence des changements – de thèmes ou de locuteurs – est élevée, plus le rythme est rapide, et inversement. En d'autres termes, étudier l'enchaînement des paragraphes revient à analyser comment les thèmes ou les répliques se développent tout au long du texte. Le passage d'un paragraphe à un autre signale au lecteur ou à la lectrice que quelque chose change. « C'est une autre façon de lui [le destinataire] dire : "je change de sous-thème, et je vous invite à en tenir compte". [...] L'ampleur des parties et des sous-thèmes est une pure affaire de décision de la part du scripteur ; la longueur des paragraphes sera l'expression de ces décisions » (Le Ny, 1985, p. 133). En résumé, la longueur des paragraphes montre comment les thèmes et les répliques se développent, durent, donc s'enchaînent.

Soulignons toutefois, que la pratique des paragraphes fluctue selon les époques. Prenons l'exemple de *Micromégas* de Voltaire où les paragraphes sont très longs. On pourrait en conclure à une absence de dialogues dans l'œuvre et à une surabondance de scènes descriptives. Or, ses dialogues en discours direct sont intégrés dans un seul paragraphe, pratique courante jusqu'au XVIII^e siècle. « Les premiers romans s'en tiennent encore, timidement, à la première personne, et les romans par lettres à deux personnages qui s'expriment seuls. Mais les premiers récits romanesques (Marivaux, Diderot) s'animent et passent vite du monologue aux dialogues » (Catach, 1994, p. 41). Dès lors, il serait plus exact de dire que la longueur des paragraphes montre comment les thèmes, les répliques – ou les dialogues, si le texte a été écrit avant le XIX^e siècle –, se développent, durent et s'enchaînent.

Pour conclure, les œuvres visuelles réalisées permettent d'observer la fréquence des changements et leur répartition dans le texte. « Les changements sont la trame du temps [...] nous estimons souvent le temps par la quantité de changements qui s'y produisent » (Fraisie, 1974, p. 326). Il est important de souligner que ce temps du texte n'est pas lié à la vitesse narrative que Gérard Genette définit comme le rapport entre la durée de l'histoire, en mois ou année par exemple, et une longueur de texte, en lignes et pages (1972).

La vitesse ou le rythme du texte s'en tiennent ici à ce qu'il nous donne à voir, c'est-à-dire des rapports de longueurs, équivalentes à des durées.

4 Discussion, les impacts de ce rythme et les perspectives

Ce rythme visuel du texte a un impact tant sur la vitesse de lecture que sur l'importance perçue du discours. Les langues possèdent des marques (ou systèmes de marques) [...] permettant de fournir à un destinataire potentiel des instructions interprétatives propres à favoriser la compréhension (Charolles, 1988, p. 4).

4.1 Un rythme délié du sens ?

Une critique du rythme mérite d'être mentionnée. Henri Meschonnic et, par la suite, Valérie Beaudoin, auraient démontré que les longueurs textuelles n'apportent aucune information quant au rythme, lequel ne peut s'appréhender en dehors du sens. Meschonnic critique d'ailleurs vivement l'approche numérique du rythme poétique.

La passion du nombre, principe pythagorien – « les choses ne sont que l'apparence du Nombre » – présupposant un rapport inexplicable inexplicé entre le nombre et le sens, établissait également une relation étrange entre le nombre et la passion [...] L'imposition mathématicienne de l'ordre et des contraintes formelles est un maintien de l'ordre, une censure et une dénégation du désordre. C'est pourquoi le rythmicien-mathématicien refuse le sujet et l'inconscient. Le primat des proportions sur le chaos, la possession du continu par-delà le discontinu maintiennent le théologique : l'unité. Le rythme est l'hypostase du retour. Il garantit l'identité. Il rassure. Selon Pius Servien, les rythmes sont « les seuls amis de l'homme » [...] Traiter les formes comme des nombres, c'est en éliminer le sens. (Meschonnic, 2009, p. 573)

On pourrait en conclure que la présente démarche, visuelle, puis numérique, participe de cette recherche d'ordre qui fige. L'approche numérique serait *in fine* une conception structurale du rythme, le réduisant à une forme régulière reconnaissable. Le nombre viendrait ainsi couper le langage de la vie et du sujet. Or, notre approche du texte n'a pas pour objet de percevoir un ordre, encore moins des régularités. Elle vise, au contraire, à percevoir la façon dont le texte évolue dans le temps, son mouvement au sens de « manière particulière de fluer », de « perception d'une forme dynamique », donc au sens premier du terme de « rythme », sur cette force de vie qui l'anime. Rappelons qu'Henri Meschonnic s'est intéressé à la poésie et qu'en poésie la métrique et la structure numérique peuvent « compter » au point de primer. C'est cela qu'il dénonçait, cette dictature du nombre qui veut régenter le rythme et qui prétend le créer.

Débutée avec la forme visuelle des textes littéraires, la présente recherche est très vite devenue une recherche sur leur forme temporelle, un temporel libre. La morphologie des textes est une trace visuelle de l'agencement spécifiant un discours, de sa configuration temporelle. Le nombre est ici un enregistrement de ses différents temps. Il ne régente pas le rythme du texte. Il n'édicte pas de règles, comme pour un poème ou pour le vers classique étudié par Valérie Beaudoin. Il n'est pas coercitif. Le nombre est un moyen de mesurer les segments textuels et devient, à son tour, visuel, sous une autre forme.

La critique de Meschonnic envers l'approche numérique du rythme – poétique – est intéressante en ce qu'il soutient une conception du rythme qui est l'instanciation du sujet dans son discours. « Parce que le rythme [...] peut être pris comme la structuration d'ensemble de tous les signifiants, il est l'inscription du sujet dans l'ensemble de l'œuvre comme système de valeurs de langage, à travers le sens » (2009, p. 363). Et c'est parce que le sujet s'y inscrit que l'on entendra dans l'œuvre écrite son auteur. Il n'y a pas de rythme hors du sens selon Meschonnic, car le rythme ne peut être saisi hors du sujet. Or, dans cette recherche, le sujet n'est pas nié. En effet, nous considérons que cette inscription du sujet dans l'œuvre ne passe pas uniquement par le sens. Elle passe aussi par ses respirations et ses pauses, dans la façon dont ses segments textuels s'enchaînent. En admettant que l'inscription du sujet ne passerait que par le sens, étant donné que le paragraphe est une unité thématique, qu'il est donc lié aux inflexions de sens, l'étude des paragraphes en fonction de leur longueur permet d'examiner comment l'auteur ou l'autrice déploie ses unités de sens. C'est

à partir de la morphologie du texte que les visualisations réalisées révèlent le flux du discours écrit, et c'est en recouvrant le sens que l'on peut percevoir la dynamique du texte, sinon l'on est dans le sens. Par le nombre, on retrouve la vie du texte, son mouvement, son rythme. George K. Zipf, qui a analysé la longueur des mots et leur fréquence d'usage selon une méthode statistique, estimait qu'étudier la dynamique du langage, le flux du discours, par une méthode statistique permet d'étudier « les forces sous-jacentes qui suscitent et dirigent l'expression linguistique » (1935/1974, p. 16).

Nous avons vu que ce rythme du texte littéraire est déterminé par la fréquence des changements, de lieu, de point de vue, de temps, de sujet, de « thème », d'une mise en évidence d'un élément et, selon les époques, de locuteur ou de dialogue. En conséquence, ce rythme est lié au sens puisque les marques de paragraphes indiquent souvent une discontinuité thématique et/ou, à partir du XIX^e siècle, une discontinuité des répliques.

On notera que les marques de paragraphes ne résultent pas seulement de ces changements. Ils peuvent venir d'un souci formel, à savoir « des contraintes rythmiques [visuelles] » peuvent motiver « le scripteur à répartir ses paragraphes en masses harmonieuses » (Bessonnat, 1988, p. 94). Celui-ci viendrait ainsi modifier la longueur de ses unités thématiques. Ce à quoi nous pourrions ajouter que, si cette contrainte formelle s'exerce, le scripteur le fera en fonction d'unités cohérentes.

Les marques de paragraphes varient d'une époque à l'autre. Avant la naissance de la ponctuation, les mots étaient accolés. Ensuite, les paragraphes étaient marqués par des *pieds de mouche* (« ¶ ») sans retour à la ligne pour des raisons d'économie de papier. Le passage du paragraphe à lettrine au paragraphe à alinéa a lieu au XV^e siècle avec l'invention de l'imprimerie en 1454. Les paragraphes tels que nous les connaissons se sont stabilisés au XIX^e siècle. C'est aussi à cette période que les écrivains semblent davantage imposer leurs alinéas.

4.2 Les impacts de ce rythme du texte

Les études des paragraphes tendent à les définir et à déterminer leurs fonctions en tant qu'unité textuelle. Peu ont été menées sur leur enchaînement, autrement dit, sur le rythme visuel que leurs longueurs génèrent. Nous allons voir pourtant que ce rythme du texte influe sur le temps de lecture, l'importance perçue du discours et, plus largement, participe du sens.

Le rythme mis au jour en fonction de la longueur des paragraphes de textes littéraires influence la réception du texte. Selon Umberto Eco, « il est tout aussi important d'étudier comment le texte (une fois produit) est lu et comment toute description de la structure du texte doit être, en même temps, la description des mouvements de lecture qu'il impose » (1985, p. 8) et, selon Jean-Michel Adam,

[le paragraphe] découle de ces possibilités des rythmes aux effets sémantiques non négligeables : quand une succession de courts paragraphes fragmente les actions ou événements d'un récit, cela crée un rythme haletant et une dramatisation qui peut même aller jusqu'à une certaine ironie ; en revanche, le regroupement d'actions et d'événements divers au sein d'un seul long paragraphe tend à unifier, uniformiser voire enliser le vécu des personnes ou personnages représentés. (2018, p. 59-60)

La lisibilité fondée sur une certaine idée des capacités cognitives des lecteurs potentiels est un critère aussi important que la volonté de mettre en valeur un événement, de créer un certain rythme textuel avec les paragraphes de transition, des paragraphes brefs et d'autres plus longs, de séparer ou non le discours direct et la narration, les moments descriptifs ou explicatifs. (2018, p. 180)

Daniel Bessonnat assigne au paragraphe trois fonctions : (i) faciliter la lecture en « [signalant] au lecteur qu'il va passer à une unité ultérieure », (ii) « combattre l'organisation linéaire du texte en tant qu'instrument de programmation de la lecture », (iii) « dialoguer avec l'interlocuteur fictif » (1988, p. 85-87). Il distingue également le paragraphe descriptif, ample quand il s'agit de commenter et d'expliquer, du paragraphe narratif – bref, quand il s'agit de dramatiser, d'accentuer la solennité du propos.

Enfin, Pierre Coirier, Daniel Gaonac'h et Jean-Michel Passerault attribuent au paragraphe deux fonctions. La première, rhétorique, sert à « marquer une frontière de contenu [...] ou une étape dans une argumentation » (1996, p. 193). Cette fonction est similaire à la conception du paragraphe de Jean-François Castille. La seconde sert à marquer l'importance d'un élément. En ce qui concerne cette dernière fonction, ils estiment que « la taille du paragraphe, de ce point de vue, doit être prise en compte (Hofmann, 1989) [...] Un paragraphe long, au contraire, peut être utilisé pour conduire le lecteur à élaborer une représentation sémantique unique relativement à des faits ou concepts multiples » (1996, p. 193).

Plus précisément, ce rythme du texte a un impact sur le temps de lecture. Une étude menée auprès de 21 étudiants montre que les repères de paragraphes n'affectent pas la vitesse globale de lecture dans un contexte où les lecteurs disposaient d'un temps suffisant pour relire (Stark, 1988, p. 298). Cependant, Heather A. Stark précise ne pas exclure la possibilité qu'à l'intérieur de ce temps de lecture global il y ait des effets locaux sur les phrases. Une étude ultérieure, réalisée auprès de 48 étudiants, confirme partiellement cette supposition. Le temps de lecture aux frontières des paragraphes varie. En effet, à la fin des paragraphes, le temps de lecture s'allonge, en d'autres termes, la vitesse de lecture ralentit, compensée par une vitesse plus élevée au début des paragraphes, soit une lecture accélérée (Passerault & Chesnet, 1991).

De plus, ce rythme du texte agit de manière plus subtile ; il participe du sens. « Sa pertinence au plan des traitements cognitifs est pourtant bien attestée » (Coirier *et al.*, 1996, p. 192). Il a un impact sur l'intensité du discours perçue, qu'il peut renforcer ou réduire. « Les repères de paragraphes affectent les idées considérées comme importantes » (Stark, 1988, p. 275). Ils « font penser aux lecteurs que le début des paragraphes est important » (1988, p. 300). En ce qui concerne les écrits techniques, la longueur des paragraphes n'a aucune incidence sur le niveau d'expertise et la qualité de l'écrit perçu ou sa compréhension (Markel *et al.*, 1992).

Dans la préface du livre de Donald F. McKenzie, Roger Chartier, historien du livre, de l'édition et de la lecture, souligne que « *"forms effects meaning"* : contre toutes les définitions uniquement sémantiques des textes, McKenzie rappelle avec force la valeur symbolique des signes et des matérialités » (1991, p. 7). Ce qui caractérise un texte est le « cours [des] idées [...] et la façon de les lier les unes aux autres » (Chartier, 2010, p. 685), que l'on peut comprendre comme l'enchaînement des unités de sens.

« Nous soutenons que l'établissement de paragraphes est une activité non triviale, intrinsèque à la production de textes qui ne peut aucunement être réduite à une sorte d'intervention éditoriale subalterne » (Pešek, 2022, p. 122). Nina Catach mentionne George Sand qui « défendait sa ponctuation d'écrivain » (1994, p. 44). Rappelons que l'on ignore si elle considérait l'alinéa comme une ponctuation. Daniel Bessonnat cite Zola qui écrit « j'ai la faiblesse de tenir à mes alinéas » (1988, p. 96). Ce même auteur déplorait la lecture de *l'Assommoir* en feuilleton car des phrases y sont enlevées et des alinéas ajoutés⁵.

Ricard Ripoll i Villanueva affirme que « l'alinéa devient ainsi un recours essentiel à la mise en valeur du tissu textuel qui va créer le sens » (1996, p. 343). Il développe le concept de portée qu'il appelle « force » et précise qu'un « premier aspect de la force est donc celui des relations internes qui s'établissent dans un texte » (1996, p. 244). La façon dont les paragraphes s'enchaînent est essentielle car elle influence l'intensité perçue du sens. Par exemple, « un paragraphe court sera plus fort s'il est inséré dans un texte qui privilégie les paragraphes longs » (1996, p. 244). À l'inverse, si un paragraphe très long se trouve au milieu d'un ensemble de paragraphes très courts, l'impact du long paragraphe sera renforcé.

En plus d'agir sur la réception du texte, tant sur le temps de lecture que sur la perception de l'intensité du discours, nous émettons l'hypothèse que la longueur et la durée des paragraphes ont également une influence sur le rythme perçu du récit. Afin de connaître avec précision les effets de la longueur des paragraphes, cette étude mérite d'être élargie et approfondie.

4.3 Perspectives

Les pistes de développement sont multiples. Nous en privilégierons ici deux qui nous semblent prioritaires.

La première concerne les particularités visuelles, notamment celle d'un très long paragraphe dans un ensemble de paragraphes très courts. Nous pensons qu'elles sont le signe de moments clefs dans la narration ou qu'elles pourraient contenir ce pourquoi l'œuvre a été écrite. Paul Fraisse considère que, dans un groupe rythmique qui est en fait une succession, applicable aussi bien aux formes temporelles que spatiales, un son faible peut devenir « l'accent subjectif parce qu'il est sans doute l'élément différenciateur » (1974, p. 82). Toutefois, « en prosodie, quand il y a un jeu de longues et de brèves, l'accent est toujours placé sur une longue. [...] puisque l'accent intensif allonge la durée apparente du son et qu'un léger allongement de la durée joue le même rôle que l'accent » (1974, p. 83).

Sachant que, dans notre démarche, les paragraphes sont des durées, du moins possèdent-ils une durée, à propos de la durée d'une note, Gaston Bachelard cite un extrait de l'ouvrage *Histoire de la langue musicale* du compositeur Maurice Emmanuel : « En principe... l'intensité se trouve liée à la longueur, en ce sens que, de deux éléments de durée inégaux, c'est le plus long qui est réputé fort. La longueur et la force sont connexes : c'est, en rythmique primitive, une sorte de nécessité. Dans la versification rythmique, la force appellera la longueur » (1936/2013, p. 119). Il ajoute que « c'est l'intensité qui donne la durée et que la durée – encore une fois – n'est qu'une conséquence » (1936/2013, p. 120). Afin de savoir si les ruptures visuelles en termes de longueur des paragraphes, surtout les longs paragraphes distinctifs, sont des moments clefs, le recouplement avec le récit devra être effectué sur un large corpus.

Ensuite, lorsque les histogrammes et les sismogrammes présentent des parties distinctes, comme nous l'avons vu avec *Notre-Dame de Paris* et *Accident nocturne*, un recouplement systémique avec la narration devra également être effectué afin de déterminer si ces différences montrent une structure compositionnelle globale du texte autre que celles mises en évidence par la ponctuation au sens élargie.

La seconde piste de développement privilégiée tend vers une étude mathématique. En plus de vérifier l'applicabilité de la loi de Zipf à la longueur des paragraphes, comme le suggèrent les figures 3 et 4, il serait intéressant d'explorer si les autrices et les auteurs privilégient certains *patterns* ou motifs – au sens d'une récurrence marquée d'une suite de longueurs de paragraphes, donc de nombres – qui se retrouveraient fréquemment dans un ou plusieurs de leurs textes. La recherche de motifs pourra être étendue aux séquences de paragraphes non consécutifs. Cette opération complexe et ses développements relèvent de l'intelligence artificielle. Si de telles régularités existent, le rythme du texte, au sens où nous l'entendons de "manière particulière de couler", devra être reconsidéré dans le sens « d'un mouvement cadencé » (Benveniste, 1966, p. 335).

Par la suite, la question se pose de savoir s'il existe une distribution spécifique aux textes les plus célèbres (notion qu'il faudra cerner) comme l'a découvert le neuroscientifique Daniel Levitin de l'université de McGill pour les œuvres musicales classiques.

Enfin, à une échéance plus éloignée, une étude pourrait être menée afin de savoir si la manière dont les auteurs segmentent le texte participe du style.

Références bibliographiques

- Achard-Bayle, G., & Pešek, O. (2020a). Modèles d'organisation thématique des paragraphes et entre les paragraphes, à l'épreuve de la Rhetorique de Cicéron. *Discours. Revue de linguistique, psycholinguistique et informatique. A journal of linguistics, psycholinguistics and computational linguistics*, 26. <https://doi.org/10.4000/discours.10794>
- Adam, J.-M. (1994). Style et fait de style : Un exemple rimbaldien. In G. Molinié & P. Cahné, *Qu'est-ce que le style ? : Actes du colloque international* (p. 15-43). Presses universitaires de France.
- Adam, J.-M. (2018). *Le paragraphe : Entre phrases et texte*. Armand Colin.
- Algee-Hewitt, M., Heuser, R., & Moretti, F. (2015). On Paragraphs. Scale, Themes, and Narrative Form – Stanford Literary Lab. *Stanford Literary Laboratory*. <https://litlab.stanford.edu/LiteraryLabPamphlet10.pdf>

- Arabyan, M. (1994). *Le paragraphe narratif : Étude typographique et linguistique de la ponctuation textuelle dans les récits classiques et modernes*. L'Harmattan.
- Bachelard, G. (1936). *La dialectique de la durée*. PUF.
- Béguin-Verbrugge, A. (2006). *Images en texte—Images du texte*. Presses universitaires du Septentrion.
- Benveniste, É. (1966). *Problèmes de linguistique générale, I*. Gallimard.
- Bessonnat, D. (1988). Le découpage en paragraphes et ses fonctions. *Pratiques*, 57(1), 81-105. <https://doi.org/10.3406/prati.1988.1475>
- Bond, S. J., & Hayes, J. (1984). Cues people use to paragraph text. *Research in the Teaching of English*, 18.
- Castille, J.-F. (2012). La poétique du paragraphe dans l'écriture réaliste de Flaubert et Maupassant. *Questions de style*, 9, 147-159.
- Catach, N. (1980). La ponctuation. *Langue française*, 45(1), 16-27. <https://doi.org/10.3406/lfr.1980.5260>
- Catach, N. (1994). *La ponctuation : Histoire et système*. Presses universitaires de France.
- Charolles, M. (1988). Les plans d'organisation textuelle : Périodes, chaînes, portées et séquences. *Pratiques*, 57(1), 3-13. <https://doi.org/10.3406/prati.1988.1468>
- Chartier. (2010). *Résumés annuels*. <https://www.college-de-france.fr/site/roger-chartier/resumes.htm>
- Christin, A.-M. (2009). *L'image écrite ou La déraison graphique*. Flammarion.
- Coirier, P., Gaonac'h, D., & Passerault, J.-M. (1996). *Psycholinguistique textuelle : Une approche cognitive de la compréhension et de la production des textes*. A. Colin.
- Colby, J. B. (1971). Paragraphing in Technical Writing. *Technical Communication*, 18(2), 13-16.
- Colin, C., Conrad, T., & Leblond, A. (2017). *Pratiques et poétiques du chapitre du XIXe au XXIe siècle (1-1)*. Presses universitaires de Rennes.
- Cressot, M. (1974). *Le Style et ses techniques : Précis d'analyse stylistique* (L. James, Éd.; 8e éd. mise à jour / par Laurence James, 1-1). Presses universitaires de France.
- Crosby, A. W. (2003). *La mesure de la réalité : La quantification dans la société occidentale, 1250-1600* (J.-M. Mandosio, Trad.). Éd. Allia.
- Dagognet, F. (1993). *Réflexions sur la mesure*. Encre marine.
- Dahan-Gaida, L. (2023). *L'art du diagramme : Sciences, littérature, arts*. Presses universitaires de Vincennes.
- Daunais, I. (2017). La nature « moyenne » des chapitres. In C. Colin, T. Conrad, & A. Leblond (Éds.), *Pratiques et poétiques du chapitre du XIXe au XXIe siècle* (p. 279-288). Presses universitaires de Rennes.
- Dionne, U. (2008). *La voie aux chapitres : Poétique de la disposition romanesque*. Éd. du Seuil.
- Drillon, J. (1991). *Traité de la ponctuation française*. Gallimard.
- Eco, U. (1985). *Lector in fabula : Le rôle du lecteur ou La coopération interprétative dans les textes narratifs* (M. Bouzaher, Trad.). Grasset, Librairie générale française.
- Fraisse, P. (1974). *Psychologie du rythme*. Presses Universitaires de France.
- Genette, G. (1972). *Figures III*. Éditions du Seuil.
- Ji, S. (2008). What do paragraph divisions indicate in narrative texts? *Journal of Pragmatics*, 40(10), 1719-1730. <https://doi.org/10.1016/j.pragma.2007.11.010>
- Jockers, M. L. (2013). *Macroanalysis : Digital methods and literary history*.
- Laufer, R. (1986). L'énonciation typographique : Hier et demain. *Communication & Langages*, 68(1), 68-85. <https://doi.org/10.3406/colan.1986.1762>
- Le Ny, J.-F. (1985). Texte, structure mentale, paragraphe. In *La notion de paragraphe* (p. 129-136). Ed. du Centre national de la recherche scientifique.

- Louvel, L. (2002). *Texte/Image : Images à lire, textes à voir*. Presses Universitaires de Rennes.
- Maciel, C., Mayaffre, D., & Vanni, L. (2022). Corpus non alignés et ADT. Essai de comparaison entre les présidents français et brésiliens de l'ère contemporaine. In *JADT2022* (p. 568-575). Vadistat press et Edizioni Erranti. <https://hal.science/hal-03858547>
- Markel, M., Vaccaro, M., & Hewett, T. (1992). Effects of Paragraph Length on Attitudes Toward Technical Writing. *Technical Communication*, 39(3), 454-456.
- Martin, R. (1994). Préliminaire. In G. Molinié & P. Cahné, *Qu'est-ce que le style ? : Actes du colloque international* (p. 9-13). Presses universitaires de France.
- McKenzie, D. F. (1991). *La bibliographie et la sociologie des textes* (M. Amfreville, Trad.). Éd. du Cercle de la librairie.
- Meschonnic, H. (2009). *Critique du rythme : Anthropologie historique du langage*. Verdier.
- Michon, P. (2018). Sommes-nous en train d'assister à l'émergence d'un nouveau paradigme scientifique : Le paradigme rythmique? In *Rythmanalyse(s) : Théories et pratiques du rythme ontologie, définitions, variations* (p. 25-42). Jacques André éditeur.
- Molinié, G., & Cahné P. (1994) *Qu'est-ce que le style ? : Actes du colloque international*. Presses universitaires de France.
- Monière, D., & Labbé, D. (2002). Essai de stylistique quantitative Duplessis, Bourassa et Lévesque. *Vie JADT*, 561-569.
- Neveu, F. (2000/2014). De la syntaxe à l'image textuelle. Ponctuation et niveaux d'analyse linguistique. *La Licorne*, 52, 201-215.
- Passerault, J.-M., & Chesnet, D. (1991). Le marquage des paragraphes : Son rôle dans la gestion des traitements pendant la lecture. *Psychologie Française*. <https://hal.archives-ouvertes.fr/hal-03338075>
- Pešek, O. (2022). Structure textuelle globale et représentation computationnelle. Un défi pour l'analyse textuelle des discours. *Écho des études romanes*, 18, 117-135. <https://doi.org/10.32725/eer.2022.008>
- Piégy, N. (2017). La nature « moyenne » des chapitres. In C. Colin, T. Conrad, & A. Leblond (Éds.), *Pratiques et poétiques du chapitre du XIXe au XXIe siècle* (p. 303-313). Presses universitaires de Rennes.
- Piérard, S., & Bestgen, Y. (2006). A la pêche aux marqueurs linguistiques de la structure du discours. *Journées internationales d'Analyse statistique des Données Textuelles - JADT*, 2, 749-757.
- Rey, A. (2001). *Dictionnaire historique de la langue française : Contenant les mots français en usage et quelques autres délaissés, avec leur origine proche et lointaine* (Nouvelle éd. augmentée par Alain Rey). Dictionnaires Le Robert.
- Rousset, J. (1962). *Forme et signification : Essai sur les structures littéraires de Corneille à Claudel*. J. Corti.
- Sauvanet, P. (1992). « Le » rythme : Encore une définition ! (p. 233-243). Centre culturel international de Cerisy Éd. l'Harmattan.
- Souchier, E. (1998). L'image du texte pour une théorie de l'énonciation éditoriale. *Les cahiers de médiologie*, 6(2), 137-145. <https://doi.org/10.3917/cdm.006.0137>
- Stark, H. A. (1988). What do paragraph markings do? *Discourse Processes*, 11(3), 275-303. <https://doi.org/10.1080/01638538809544704>
- Tardif, C. P. (2018). *Un Atlas de spectres de textes littéraires*. <https://hal.science/hal-02986259>
- Tardif, C. P. (2021). *Une prédominance des paragraphes très courts et une corrélation entre la longueur des paragraphes avec le genre littéraire ou la période d'écriture : Les résultats de "Un Atlas des spectres de textes littéraires"*. <https://hal.science/hal-03146960>
- Vicedo, M. (2011). Salammbô, un roman par épisodes. In S. Triaire & P. Victorin (Éds.), *Deviser, diviser : Pratiques du découpage et poétiques du chapitre de l'Antiquité à nos jours* (p. 241-257). Presses universitaires de la Méditerranée.

Villanueva, R. R. i. (1996). Le texte littéraire et la mise en paragraphes. *La lingüística francesa: gramática, historia, epistemología.*, 2, 339-346.

<https://dialnet.unirioja.es/servlet/articulo?codigo=4034332>

Weinrich, H. (1973). *Le temps : Le récit et le commentaire*. Ed. du Seuil.

Zipf, G. K. (1935). *La psychobiologie du langage : Une introduction à la philologie dynamique* (S. Deflandre & M.-R. Delorme, Trad.). Retz-C.E.P.L.

¹ Terme d'Emmanuel Souchier

² Si Jean-François Castille limite cet usage « rhétorique » du paragraphe à une période antérieure au XIX^e siècle, nous verrons qu'il est celui qui domine encore largement. En effet, il a été démontré que le paragraphe est une unité thématique.

³ Avant la naissance de la ponctuation, les mots étaient accolés. À la naissance du paragraphe, celui-ci était marqué par un signe typographique (¶) sans retour à la ligne pour des raisons d'économie de papier. Les paragraphes tels que nous les connaissons sont nés au XVI^e siècle.

⁴ Zipf découvrit que la fréquence d'emploi (f) du n -ième mot le plus employé est inversement proportionnelle à son rang (r) : f varie en fonction de $1/r$. Ainsi, les mots les plus courts sont les plus utilisés, inversement les plus longs sont les moins utilisés. Il trouva que le mot le plus fréquent dans un échantillon apparaît en moyenne une fois tous les dix mots, le deuxième mot le plus fréquent une fois tous les vingt mots..., le n -ième une fois tous les $10 \times n$ mots. Dans le roman *Ulysses* de James Joyce, il observa que le mot le plus courant (*the*) apparaît 8 000 fois, que le deuxième (*of*) apparaît deux fois moins, que le troisième trois fois moins, ..., que le centième cent fois moins, etc. Par exemple, le douzième mot apparaît $1/12^e$ fois le premier. Cette loi empirique et contestée pendant des décennies, Benoît Mandelbrot la théorisa en 1950 et démontra qu'elle était un cas particulier d'une loi de puissance (*power law*) plus générale, appelée loi de Mandelbrot. Elle s'applique à tout ensemble d'éléments discrets, c'est-à-dire discontinus et énumérables, telles la taille des villes en nombre d'habitants, la répartition des richesses, etc.

⁵ Émile Zola, « lettre à Ludovic Halévy », 24 mai 1876.