



HAL
open science

Découvrir et visualiser un rythme du texte littéraire en fonction de la longueur de ses paragraphes grâce aux arts visuels numériques

Claude Patricia Tardif

► To cite this version:

Claude Patricia Tardif. Découvrir et visualiser un rythme du texte littéraire en fonction de la longueur de ses paragraphes grâce aux arts visuels numériques. *Humanistica* 2024, Association francophone des humanités numériques, May 2024, Meknès, Maroc. hal-04588653

HAL Id: hal-04588653

<https://hal.science/hal-04588653>

Submitted on 27 May 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 4.0 International License

Découvrir et visualiser un rythme du texte littéraire en fonction de la longueur de ses paragraphes grâce aux arts visuels numériques

Claude Patricia Tardif
Université Paris 8
claudecldpro@gmail.com

Résumé

L'étude de la forme visuelle des textes littéraires a longtemps été assimilée à celle de leur mise en page et de leur disposition spatiale. Dans cette étude exploratoire, nous nous intéressons à la morphologie des textes modelée par la longueur des paragraphes. La méthode développée est un processus de visualisation qui débute par la mesure de chaque paragraphe d'un texte avec un logiciel dédié. Puis à partir de cette suite de nombres, des œuvres visuelles numériques sont réalisées. Elles rendent visible un rythme du texte en fonction de la longueur de ses paragraphes et, dans une moindre mesure, de ses chapitres. Nous prendrons l'exemple de *La Petite Fille qui aimait Tom Gordon* (LPF) de Stephen King (2002).

1 Introduction

Le projet *Narra* que nous présentons propose une approche novatrice de la littérature en explorant un aspect du texte souvent délaissé, sa forme visuelle hors « énonciation éditoriale » (Souchier, 1998, p. 144). La méthode utilisée se situe à la croisée des humanités numériques et des arts visuels et rappelle celle soutenue par Paul Valéry d'une continuité, voire d'une intrication, entre les sciences et les arts en ce que « le graphique est capable du continu » (Valéry, 1919-1957, p. 137).

Ce projet se fonde sur la thèse selon laquelle la forme est une source de connaissance, y compris celle des textes littéraires. En étudiant la forme du texte *via* la mesure et les arts visuels, nous rendrons visible une force, une dynamique qui la modèle, nous révélerons un rythme du texte jusqu'alors inexploré, propre à l'écriture. La forme du texte est considérée comme le témoin ou le résultat d'un processus temporel.

Par « forme du texte », nous entendons la forme visuelle des textes indépendamment de la disposition spatiale, de la mise en page, de l'édition

en général, à savoir de « l'énonciation éditoriale ». Elle est façonnée par leurs paragraphes – narratifs et dialogaux – en fonction de leur longueur et, dans une moindre mesure, par leurs chapitres. Afin de la distinguer de la visualité du texte, souvent assimilée au texte mis en page, nous préférons utiliser le terme de morphologie.

Dans le cadre de cet article, bien que les chapitres déterminent aussi la morphologie du texte, nous ne retenons que les paragraphes, car ce sont surtout ces derniers qui visuellement rythment le texte selon leur longueur. En outre, non seulement les paragraphes préservent l'idée du lien, alors que les chapitres relèvent de la division ou de la coupure (Colin et al., 2017), mais aussi, les paragraphes ne sont pas systématiquement réfléchis, contrairement aux chapitres qui sont « pesés » par l'auteur ou l'autrice. Enfin, précisons que c'est moins les paragraphes qui sont étudiés en tant qu'unité textuelle que leur lien les uns aux autres et leurs rapports de longueur (ou de durée). Jean-Michel Adam constate que le paragraphe suscite encore « peu d'intérêt des linguistes » (2018).

L'usage du terme « texte » se limite ici à sa matérialité écrite et celui de « littéraire » aux romans, aux nouvelles et aux contes, à l'exception des textes tels que *La Maison des feuilles* de Mark Z. Danielewski (2002).

2 Travaux liés et corpus de travail

Nous présenterons d'abord la démarche, qui s'appuie sur la visualité des textes « hors énonciation éditoriale », puis le corpus des textes littéraires étudiés.

2.1 Travaux liés

La présente démarche, à la fois numérique et artistique, s'inscrit dans le champ des arts visuels qui, outre leur fonction esthétique vers un spectateur désintéressé, notamment depuis Emmanuel Kant (contemplation esthétique

désintéressée), acquièrent une fonction heuristique, notion développée par Stephen Wright, proche de la notion d'art cognitif qu'explore Stephen Wilson (2010) au sujet de l'art hybride et expérimental mixant les pratiques et brouillant les frontières entre art, science et technologie.

La démarche consiste à condenser un événement temporel, ici un texte, en un visuel (ou image), et cette transposition rend visible une propriété passée inaperçue ou méconnue d'un phénomène qui se déroule dans le temps. L'œuvre visuelle opère ainsi comme un révélateur qui permet de mieux connaître l'élément à partir duquel elle a été réalisée.

Parmi les artistes et les designers dont la démarche s'inscrit dans cette fonction heuristique des arts visuels, citons Mark Lombardi et ses diagrammes finement dessinés à la main (Casemajor Loustau, 2013), Julien Prévieux et ses diagrammes de Voronoï (Prévieux, 2011–2015), Fabienne Kilchör et ses diagrammes réalisés à partir de l'*incipit* de dix nouvelles de l'écrivain Heinrich von Kleist (Nehrlich et Kilchör, 2014), Michael J. Ruocco et ses compressions de films en un visuel (Ruocco, 2013), le collectif Grégory Castéra, Yaël Kreplak et Franck Leibovci et leur projet « des récits ordinaires » réalisé à partir de conversations dont le sujet était une œuvre visuelle (Castéra et al., 2014), enfin le collectif Peter Richards, Susan Schwarzenberg, Scott Snibbe, Tomas Apodaca et Amy Balkin de l'agence Stamen Design et leurs vidéos et images réalisées à partir des trajets (*via* leurs positions GPS) d'une compagnie de taxis à San Francisco (Richards et al., 2005).

2.2 Le corpus de travail

Notre présentation du projet *Narra* se limite au texte de Stephen King, *La Petite Fille qui aimait Tom Gordon* paru en français en 2002. Il convient de noter que, quelle que soit la langue utilisée, le résultat est identique, car, ce sont les rapports de longueurs (ou de durées) qui comptent.

Cependant, le corpus de travail est large. Par exemple, une étude a porté sur un corpus de 1 255 textes de 162 auteurs et autrices d'époques et de genres différents (pour un aperçu complet, voir le tableau en annexe de l'étude Tardif (2021), dont nous mentionnerons les résultats. Il se compose surtout (72 %) d'œuvres publiées entre 1850 et 2000. À l'inverse, 4 % du corpus sont des œuvres publiées avant 1800 et 14 % après 2000.

Ceci s'explique, d'une part, pour des raisons historiques – la pratique du roman moderne et l'usage du paragraphe, y compris pour les dialogues, tel que nous les connaissons, se sont développés à partir du XIX^e siècle – et, d'autre part, pour des raisons d'accès aux ouvrages les plus récents plus complexe.

En outre, une application Web, développée dans un objectif principalement visuel, a porté sur un corpus de 250 textes courts, des nouvelles et des contes, de 47 auteurs et autrices d'époques et de genres différents de pays occidentaux et russes.

3 Une méthode transversale, à la croisée des humanités numériques et des arts visuels

L'étude de la morphologie du texte se base sur une méthode développée en deux phases : d'abord, mesurer ce qui est observé, soit la longueur des paragraphes, puis, visualiser ce qui est mesuré.

3.1 Mesurer la longueur des paragraphes d'un texte avec un logiciel dédié

Narra 2.0 est un logiciel applicatif qui mesure la longueur des paragraphes d'un texte en nombre de caractères ou signes¹. À titre indicatif, il mesure également la longueur des chapitres.

En entrée, un texte, sous format ePub ou Word, qu'il va parcourir, est nécessaire. En sortie, 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 et un indicateur de forme ;
- un histogramme en fonction de la longueur des paragraphes du texte sélectionné (fig. 1) ;
- une courbe des fréquences (fig. 2).

3.2 Visualiser ces longueurs successives par un processus numérique élaboré

À partir de ces mesures, soit de la suite de nombres que génère le logiciel, un processus de visualisation numérique a été conçu. Il permet d'obtenir un ensemble d'œuvres visuelles pour un même texte. Chaque type de visuel (par exemple, les figures 3, 4, 5 et 6) a nécessité le développement d'un algorithme spécifique.

1. La notice est disponible en ligne : <https://narra.hypotheses.org/188>.

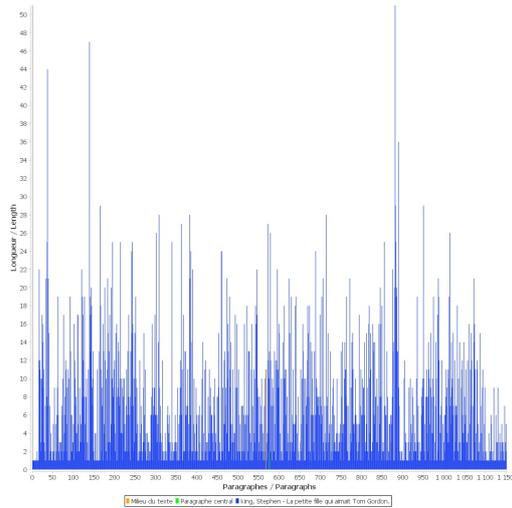


FIGURE 1 – © Narra 2.0. Histogramme de *LPF*.

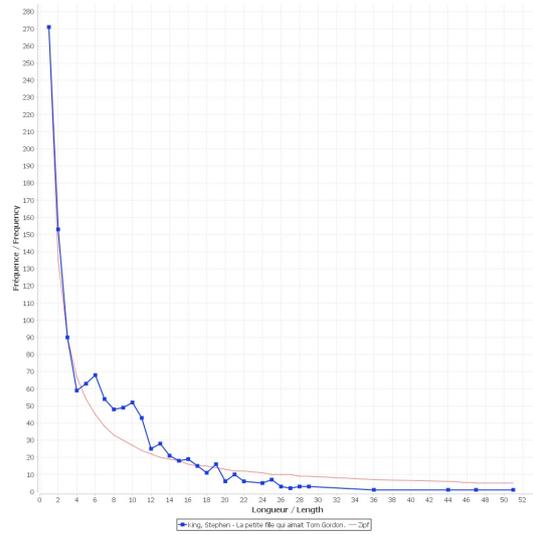


FIGURE 2 – © Narra 2.0. En bleu, la courbe des fréquences de *LPF*, en rouge celle de la loi de Zipf.

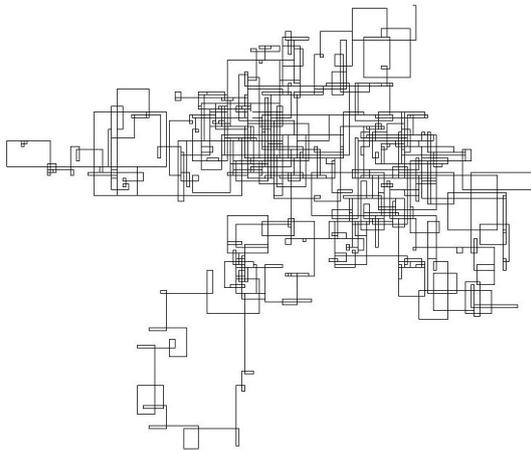


FIGURE 3 – © Claude CLd, *Plan-masse de LPF*.



FIGURE 4 – © Claude CLd, *Sismogramme de LPF*.

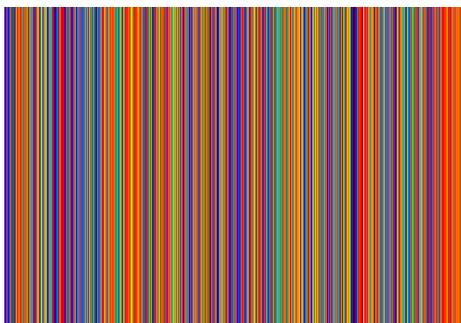


FIGURE 5 – © Claude CLd, *Allover de LPF*.

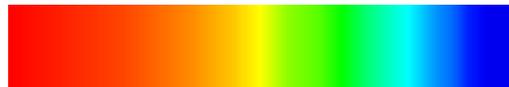


FIGURE 6 – © Claude CLd, *Spectre de LPF, en fréquence*.

Pour la réalisation des œuvres visuelles, un code couleur, basé sur le spectre de la lumière visible,

a été défini de sorte qu'à la vue des couleurs l'on sache immédiatement la longueur des paragraphes

du texte sélectionné. Il commence par le rouge pour un paragraphe court, passe par le jaune-vert pour un paragraphe de moyenne longueur, jusqu'au bleu pour un long paragraphe. La couleur est ainsi un langage en ce qu'elle caractérise une longueur de paragraphe précise.

Prenons la figure 5. Une ligne verticale d'un pixel de large correspond à un paragraphe de *La petite fille qui aimait Tom Gordon*, allant du premier paragraphe à gauche au dernier à droite. À la fin, on observe une série de lignes formant un bloc rouge, constituée en fait d'une succession de lignes rouge. Cela signifie que le texte se termine par une série de paragraphes très courts, que l'on retrouve dans la Figure 1 et 4).

On observe (fig. 1 et 4) une diversité en termes de longueur de paragraphes, mais sans véritable structure. Tous les romans ne présentent pas une telle « uniformité ». En outre, bien que la construction des deux graphiques soit identique – seule la disposition des lignes diffère, sur leur base pour l'histogramme et au centre des lignes pour le sismogramme –, nous constatons que l'histogramme met en évidence les longs paragraphes distinctifs et le sismogramme les séries de paragraphes très courts. Chaque type de visualisation souligne une caractéristique distincte du texte.

La figure 3, également composée en fonction de la longueur des paragraphes en effectuant un quart de tour à chaque nouveau paragraphe, met en avant l'expansion du texte. Plus les paragraphes sont longs, plus le tracé occupera un espace élargi, et inversement.

La figure 6, à l'instar de la figure 2, s'appuie sur la fréquence des paragraphes selon leur longueur, sans tenir compte de leur ordre d'apparition dans le texte. Elle montre la proportion des paragraphes de 1 ligne, 2 lignes, etc. Avec cette méthode, une étude 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.

Enfin, pour 250 textes courts, des nouvelles et des contes, une application Web nommée *Narra Horizon*² a été mise en ligne.

La présente méthode, qui part de textes littéraires sans les lire, fournit un objet mathématique –

une suite numérique – exploitable pour une étude approfondie ainsi que des œuvres visuelles qui sont non seulement à voir mais aussi à lire dans le sens où elles signifient et renvoient au texte. Nous allons voir qu'elles mettent au jour une caractéristique du texte inexplorée.

4 Résultat : la découverte d'un rythme du texte visualisable

Les graphiques et les œuvres de *Narra* opèrent comme un révélateur d'une propriété du texte littéraire, un rythme du texte, qu'elles rendent visibles. Elles permettent d'observer tout le texte de manière à saisir les rapports de longueurs ou de durées instantanément, difficilement saisissables autrement. Ce rythme du texte, non du récit, résulte de la fréquence des changements qui s'y produisent.

4.1 Un rythme du texte inexploré...

« Qu'il s'agisse de physiciens ou d'économistes, il faut convaincre les chercheurs que l'analyse morphologique peut être plus révélatrice qu'une analyse en termes de forces » (Thom et al., 1983, p. 113). Aussi, en partant de la forme, il est possible de remonter à la dynamique qui l'a créée par déploiement. Ce processus s'applique au texte littéraire.

En effet, en mesurant la longueur des paragraphes, nous obtenons une suite de nombres, soit, dans notre étude, des rapports de durées, donc un rythme. Selon Pius Servien (1930, p. 195), le rythme est indissociable du nombre. « Une méthode pour aborder scientifiquement l'étude des faits esthétiques, c'est de s'attacher à l'examen de leur trame numérique, les rythmes. Une méthode générale pour étudier ces rythmes, c'est de les saisir toujours au moyen d'une notation numérique ». Les visuels réalisés à partir de cette suite numérique permettent de visualiser comment les paragraphes s'enchaînent dans le temps, plus ou moins rapidement selon leur longueur.

Si l'étymologie du mot « texte » est emprunté au latin *textus* du verbe *texere* « tisser » – qui « récuse le dogme de la linéarité » (Dahan-Gaida, 2023, p. 298) –, nous nous intéressons néanmoins à cette linéarité du texte, 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 et c'est cette actualisation du texte, ce processus dynamique, que nous exposons.

2. <https://www.claude-cld.com/narra-horizon>.

« 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 au gré de la longueur de ses paragraphes découvert est l'un de ces rythmes.

4.2 ... déterminé par la fréquence des changements

Nous allons voir que ce rythme du texte littéraire en fonction de la longueur de ses paragraphes montre la fréquence des changements qui s'y produisent.

En mesurant la longueur des paragraphes, nous mesurons combien de temps dure ou se développe une réplique pour les paragraphes dialogaux ou un « thème » pour les paragraphes narratifs – au sens de combinaison de thèmes (Algee-Hewitt et al., 2015, p. 8). Si des études antérieures avaient déjà établi que les paragraphes narratifs sont des unités thématiques (Bond et Hayes, 1984; Stark, 1988; Bestgen et Piérard, 2006; Ji, 2008), leur étude l'a prouvé. Ils ont également démontré qu'un changement de paragraphe est un indice de discontinuité thématique, plus que tout autre segment textuel, du moins sont-ils un indice 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.

Ainsi, les réalisations visuelles 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 » (Fraissee, 1974, p. 326).

5 Conclusion et perspectives de recherche

Les arts visuels et graphiques jouent le rôle de révélateur en rendant visible une dynamique à l'œuvre que l'on peut saisir en un regard.

5.1 Rendre visible une force à partir de la forme

Les sismogrammes obtenus rappellent ceux des ondes sismiques aux amplitudes variées. Le texte est impulsion, souffle, flux qui rappelle le passage d'un état à un autre, un état présent découlant du précédent et contenant déjà le suivant. Ainsi la forme du texte est le produit de cette force qui le modèle. Ce rythme du texte s'en tient à ce que le texte nous donne à voir, à savoir des rapports de longueurs chaînées ou de durées des

paragraphes. « À partir des apparences des objets ; Léonard en réduit, ou tente de réduire, les caractères morphologiques à des systèmes de forces » (Valéry, 1919-1957, p. 130). Gilles Deleuze attribue cette fonction à l'art en général : « en art, et en peinture comme en musique, il ne s'agit pas de reproduire ou d'inventer des formes, mais de capter des forces. [...] La célèbre formule de Klee "non pas rendre le visible mais rendre visible" ne signifie pas autre chose » (2002, p. 57).

5.2 Perspectives de recherche

La méthode mise en œuvre revient à condenser le texte, *via* la mesure, donc les nombres, et les visualisations, multiples à partir d'une suite numérique.

Afin d'affiner ce travail, une étape supplémentaire consistera à déterminer si les longs paragraphes distinctifs sont des indices de moments clefs dans le récit, s'il existe des motifs mathématiques qui constituent les « accords » de l'auteur ou de l'autrice au même titre que ses préférences lexicales, si la longueur des paragraphes participe du style et si les visualisations révèlent une structure compositionnelle du texte autre que celle de ses parties et de ses chapitres.

Les autres pistes de développement consistent à étudier l'évolution d'un auteur ou d'une autrice, à vérifier l'applicabilité de la loi de Zipf à la longueur des paragraphes (Zipf, [1935] 1974) – ce que montre la fig. 2 –, loi qui implique de considérer l'ensemble du texte, à savoir tous ses paragraphes, tant narratifs que dialogaux, enfin, à savoir s'il existe une distribution spécifique aux textes les plus célèbres comme l'a découvert le neuroscientifique Daniel Levitin de l'université de McGill pour les œuvres musicales classiques (Levitin et al., 2012).

Bibliographie

- Jean-Michel Adam. 2018. *Le Paragraphe : entre phrases et texte*. Armand Colin, Paris.
- Mark Algee-Hewitt, Ryan Heuser, et Franco Moretti. 2015. *On paragraphs. scale, themes, and narrative form*. Pamphlet, Literary Lab, 10.
- Marc Arabyan. 1994. *Le Paragraphe narratif : étude typographique et linguistique de la ponctuation textuelle dans les récits classiques et modernes*. L'Harmattan, Paris.
- Yves Bestgen et Sophie Piérard. 2006. *À la pêche aux marqueurs linguistiques de la structure du discours*. Journées internationales d'Analyse statistique des Données Textuelles - JADT, 2 : 749–757.

- Sandra Bond et John R. Hayes. 1984. [Cues people use to paragraph text](#). *Research in the Teaching of English*, 18(2) :147–167.
- Nathalie Casemajor Loustau. 2013. [Les topographies du pouvoir de mark lombardi : l'œuvre dans la carte](#). *Espace art actuel*, 103-104.
- Grégory Castéra, Yaël Kreplak, et Franck Leibovci. 2014. *Des récits ordinaires*. Les Presses du réel.
- Claire Colin, Thomas Conrad, et Aude Leblond. 2017. *Pratiques et poétiques du chapitre du XIXe au XXIe siècle*. Presses universitaires de Rennes, Rennes.
- Laurence Dahan-Gaida. 2023. *L'Art du diagramme : Sciences, littérature, arts*. Presses universitaires de Vincennes, Saint-Denis.
- Mark Z. Danielewski. 2002. *La Maison des feuilles : par Zampànò avec une introduction et des notes de Johnny Errand*. Denoël, Paris.
- Gilles Deleuze. 2002. *Francis Bacon : Logique de la sensation*. Éd. du Seuil, Paris.
- Paul Fraisse. 1974. *Psychologie du rythme*. Presses Universitaires de France, Paris.
- Shaojun Ji. 2008. What do paragraph divisions indicate in narrative texts? *Journal of Pragmatics*, 40 :1719–1730.
- Stephen King. 2002. *La Petite Fille qui aimait Tom Gordon*. Librairie générale française, Le Livre de Poche, Paris.
- Daniel Levitin, Parag Chordiyab, et Vinod Menon. 2012. [Musical rhythm spectra from bach to joplin obey a 1/f power law](#). *Proceedings of the National Academy of Sciences*, 109(10) :3716–3720.
- Thomas Philip Nehrlich et Fabienne Kilchör. 2014. [Interpunktion und textanfänge. stilmerkmale von kleists prosa in der datenvisualisierung](#). *Gedankenstriche - ein Journal des Kleist-Museums*, 3 :12–39.
- Julien Prévieux. 2011–2015. [Atelier de dessin - b.a.c. du 14e arrondissement de paris](#).
- Peter Richards, Susan Schwartzberg, Scott Snibbe, Tomas Apodaca, et Amy Balkin. 2005. [Cabspotting](#).
- Jean Rousset. 1962. *Forme et signification : Essai sur les structures littéraires de Corneille à Claudel*. José Corti, Paris.
- Michael Ruocco. 2013. [Movie barcode compresses entire movies into a single image](#). *Cartoon Brew*.
- Pius Servien. 1930. *Les Rythmes comme introduction physique à l'esthétique : Nouvelles méthodes d'analyse et leur application à la musique, aux rythmes du français et aux mètres doriens*. Boivin, [S.l.].
- Emmanuel Souchier. 1998. [L'image du texte pour une théorie de l'énonciation éditoriale](#). *Les cahiers de médiologie*, 6(2) :137–145.
- Heather A. Stark. 1988. [What do paragraph markings do?](#) *Discourse Processes*, 11(3) :275–303.
- Claude Patricia Tardif. 2018. [Un atlas des spectres de textes littéraires, avec une classification en fonction de la longueur des paragraphes](#).
- Claude Patricia Tardif. 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"](#).
- René Thom, Giulio Giorello, et Simona Morini. 1983. *Paraboles et catastrophes : Entretiens sur les mathématiques, la science et la philosophie réalisés par Giulio Giorello et Simona Morini*. Flammarion, Paris.
- Paul Valéry. 1919-1957. *Introduction à la méthode de Léonard de Vinci*. Éd. de la nouvelle revue française, Gallimard, Paris.
- Stephen Wilson. 2010. *Art + Sciences Now*. Thames & Hudson, London, Paris.
- George Kingsley Zipf. [1935] 1974. *La Psychologie du langage : Une introduction à la philologie dynamique*. Retz-C.E.P.L., Paris.