



## Manuel de présentation et d'utilisation du logiciel *Narra 2.0*

*Narra 2.0* est un logiciel applicatif développé afin de mesurer la longueur des paragraphes ou des chapitres de tout type de textes.

Dépôt auprès de l'APP  
sous le n° IDDN.FR.001.280013.000.S.P.2016.000.31235

Ce manuel de l'utilisateur offre une vue d'ensemble des caractéristiques du logiciel et fournit un pas à pas pour la réalisation de ses différentes fonctionnalités.

© Claude CLd, Patricia Claude Tardif, 2016-2023

Tous droits réservés.

<https://www.claude-cld.com>

## Remerciements

---

Je remercie Iheb Ben Amor, ingénieur et docteur en informatique de l'université Paris-Cité pour avoir développé avec moi *Narra 2.0* pendant plus d'un an en 2016.

Je remercie la Maison des Sciences de l'Homme (MSH) Paris Nord pour avoir soutenu ce projet. Grâce à ce soutien, l'informaticienne Flavie Tonon a réalisé la seconde version du logiciel en 2020.

## Table des matières

---

<b>REMERCIEMENTS</b> .....	2
<b>TABLE DES MATIÈRES</b> .....	3
<b>À PROPOS DU LOGICIEL <i>NARRA 2.0</i></b> .....	4
<i>Finalités du logiciel</i> .....	4
<i>Présentation du logiciel</i> .....	4
<b>UTILISATION DU LOGICIEL <i>NARRA 2.0</i></b> .....	6
<i>Les interfaces d'accueil utilisateur</i> .....	6
<i>Les interfaces de résultats</i> .....	8
<b>LES FONCTIONNALITES DU LOGICIEL <i>NARRA 2.0</i></b> .....	12
<i>Les données statistiques brutes</i> .....	12
<i>Les données statistiques agrégées</i> .....	13

## À propos du logiciel *Narra 2.0*

---

*Narra 2.0* a été développé dans le cadre de ma thèse de recherche-  
création en arts visuels (université de Paris 8 Vincennes-Saint-  
Denis), codirection en études littéraires (université du Québec à  
Montréal – UQÀM).

### ■ Finalités du logiciel

*Narra 2.0* est un outil logiciel permettant de mesurer la longueur des paragraphes alinéaires et des chapitres de tout type de textes, notamment de textes littéraires, de générer un histogramme en fonction de la longueur des séquences et des divisions textuelles, d'analyser et de fournir des données statistiques de base.

Il permet ainsi d'analyser la morphologie des textes façonnée par la longueur de leurs paragraphes et, dans une moindre mesure, de leurs chapitres.

### ■ Présentation du logiciel

Avant de lancer le logiciel, il convient de disposer de textes sous format ePub ou .doc. Ces textes peuvent être notamment des textes littéraires, des romans, des nouvelles ou des comtes comprenant des paragraphes et/ou des chapitres.

*Narra 2.0* génère :

- une visualisation de la longueur des paragraphes ou des chapitres par un histogramme en suivant l'ordre du texte sélectionné et par une courbe en fonction de la fréquence des paragraphes répartis selon leur longueur ;
- des données statistiques brutes, telles que le nombre de caractères, le nombre de paragraphes, la longueur du paragraphe le plus long, et agrégées, telles que la longueur moyenne, l'écart-type.

En entrée, le logiciel fonctionne à condition de choisir un texte sous format ePub ou doc.

En sortie, le logiciel génère un fichier CSV par texte contenant la longueur de tous ses paragraphes ou de ses chapitres dans l'ordre d'apparition du texte ainsi que les données statistiques susmentionnées.

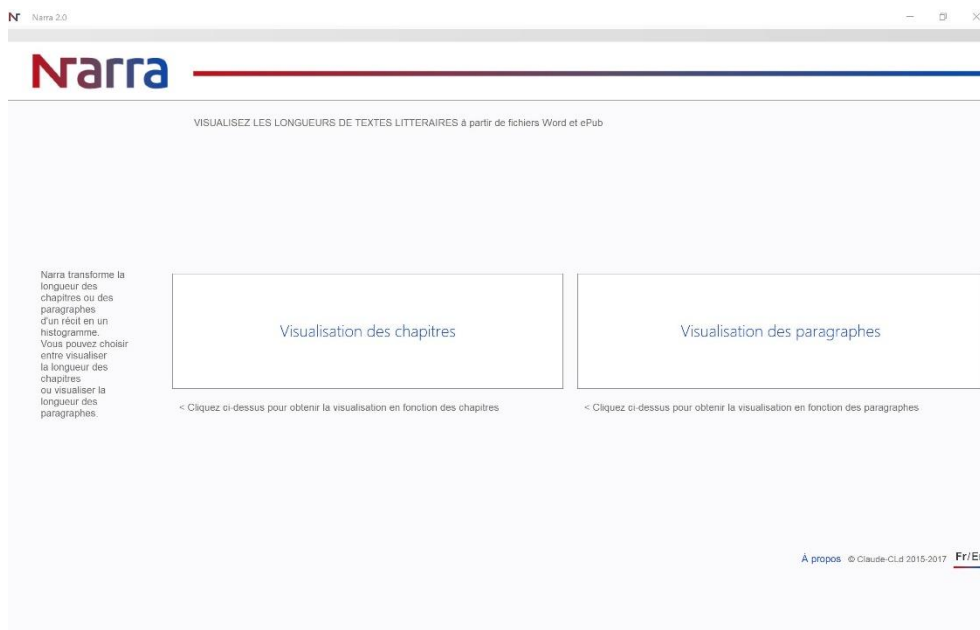
## Utilisation du logiciel *Narra 2.0*

Cette section présente un pas-à-pas d'utilisation des différentes fonctionnalités du logiciel.

Avant de lancer le logiciel, il convient de disposer de textes sous format ePub ou doc. Ces textes peuvent être notamment des textes littéraires mais pas uniquement.

### Les interfaces d'accueil utilisateur

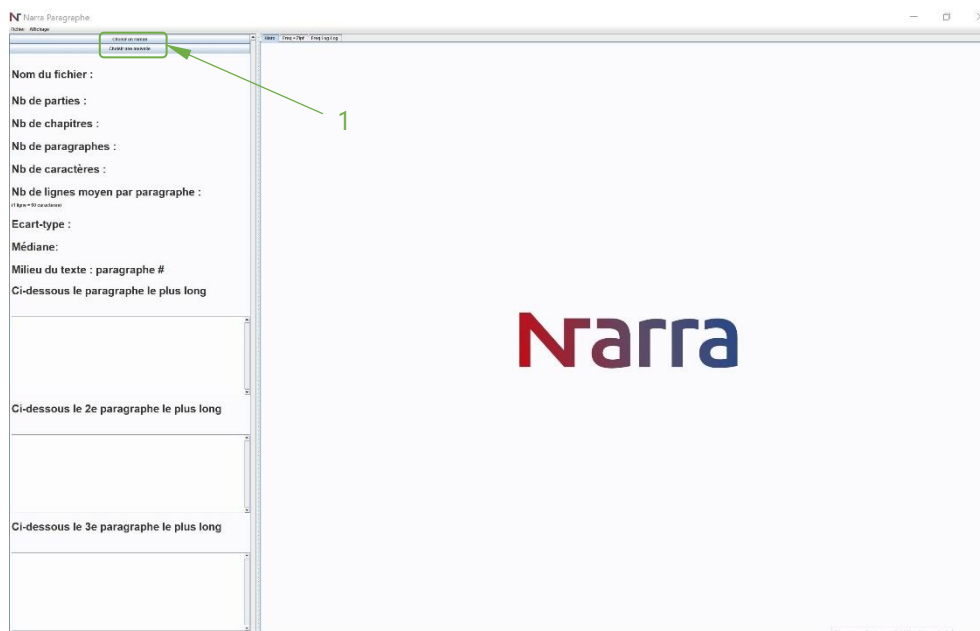
La **page de démarrage** permet à l'utilisateur de choisir s'il veut obtenir une analyse du texte en fonction de la longueur de ses paragraphes ou en fonction de la longueur de ses chapitres.



Pour changer l'interface linguistique et basculer vers l'anglais, cliquez sur le bouton « En » en bas à droite.

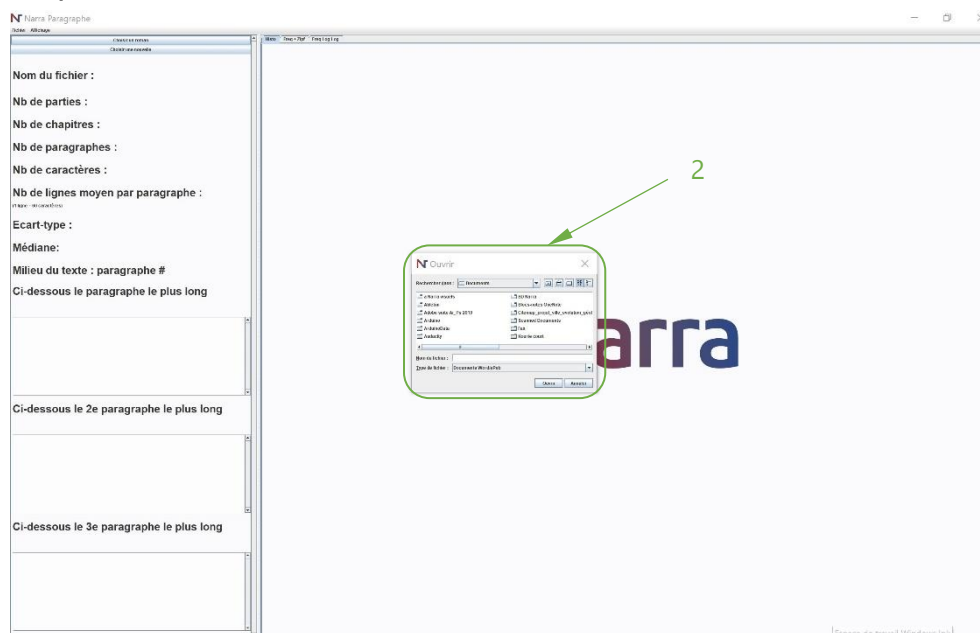
Que vous ayez choisi les paragraphes ou les chapitres, vous arrivez sur l'**interface suivante vous permettant de choisir un texte long ou court**.

Cliquez ainsi sur « Choisir un roman » pour un texte long ou sur « Choisir une nouvelle » pour un texte court (1).



Quel que soit le choix, le traitement est identique ; seule l'échelle du nombre des paragraphes pour l'histogramme change. La méthode qui consiste à mesurer les paragraphes est en revanche identique.

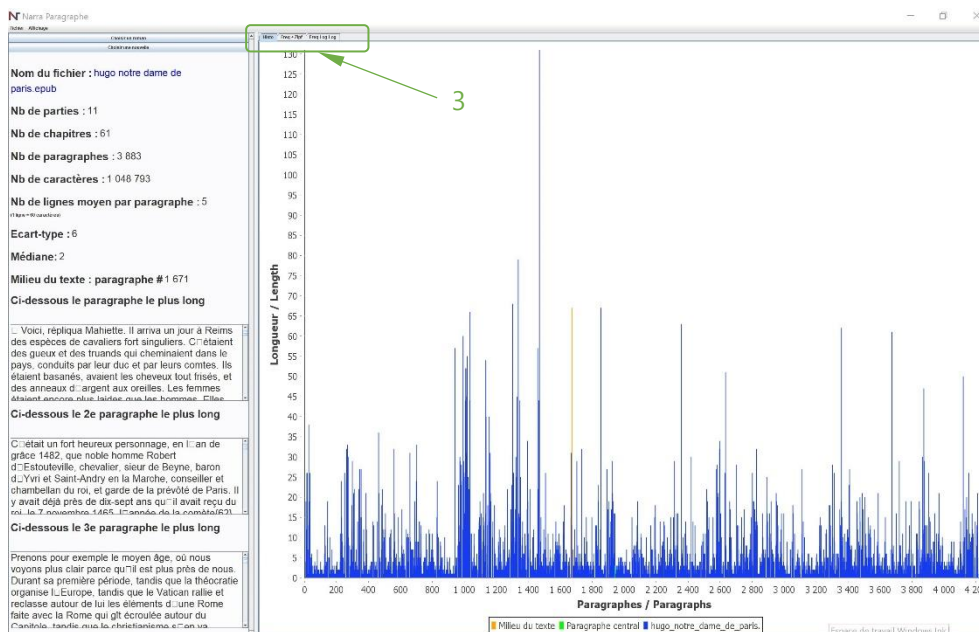
Après avoir choisi « roman » ou « nouvelle », une pop-up s'ouvre (2) et vous demande de choisir un fichier, sous format ePub ou doc. Cliquez ensuite sur « ouvrir ».



## Les interfaces de résultats

Après avoir choisi le fichier correspond au texte souhaité, le logiciel génère trois types de graphique (3) : un histogramme, une courbe des fréquences en valeurs et une courbe log-log des fréquences.

Après avoir cliqué sur « visualisation des paragraphes », une seconde interface remplace la précédente. Sur la gauche, y figurent le titre du fichier sélectionné, quelques données statistiques ainsi que le texte des trois premiers paragraphes les plus longs du texte. Sur la droite, un histogramme dont chaque ligne verticale correspond à un paragraphe, du premier paragraphe, à gauche, au dernier, à droite.



Dans cet exemple, le fichier choisi correspond à l'essai de Victor Hugo, *Notre-Dame des Paris. 1482*, publication 1832.

La hauteur de chacune des lignes verticales dépend de la longueur du paragraphe correspondant. Dans ce texte comprenant plus de 4 200 paragraphes alinéaires (paragraphes narratifs et dialogaux), le paragraphe le plus long mesure 130 lignes, soit 7 800 (130 x 60) signes espaces compris.

Notons que le logiciel indique que le roman de Victor Hugo comprend 11 parties (ou livres) et 59 chapitres, alors que le roman comprend 57 chapitres. L'écart de 2 chapitres s'explique par le fait que le

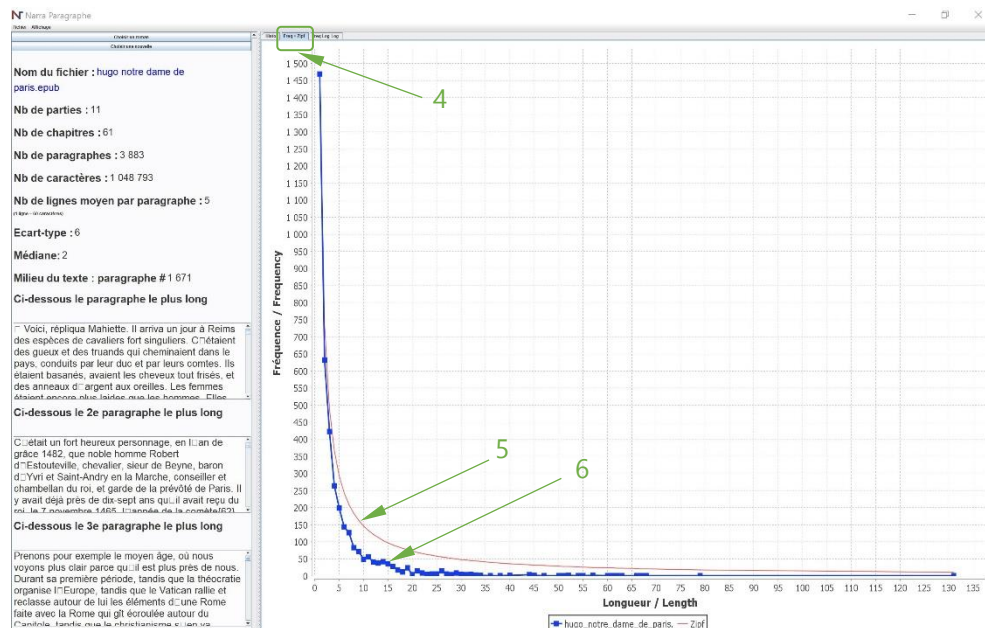


logiciel a considéré les deux notes (Note 1 et Note 2), à la fin du livre, comme des chapitres. Il s'agit d'une des limites du logiciel.

Pour les paragraphes, le logiciel génère aussi une courbe des fréquences en valeurs ou en log-log afin de réduire l'effet des valeurs extrêmes.

En cliquant sur l'onglet « Fréq + Zipf » (4), un histogramme est généré en valeurs cumulées avec la courbe de la loi de Zipf comme ci-dessous. Le panneau de gauche reste inchangé.

En rouge, est représenté la courbe de la loi de Zipf (5) ; en bleu, la courbe des fréquences en fonction de la taille des paragraphes du texte sélectionné, ici *Notre-Dame de Paris* de Victor Hugo (6).



Pourquoi la loi de distribution Zipf ? La question est de savoir si la distribution des paragraphes selon leur longueur approche une suite harmonique (la loi de Zipf), si la longueur des paragraphes est liée à la fréquence de leur usage ? Existerait-il une « éventuelle relation causale entre la longueur d'un mot et la fréquence de ses occurrences ».

George Kingsley Zipf (1902-1950) a commencé par mesurer la longueur des mots dans les textes littéraires (avant lui Jean-Baptiste Estoup dont il aurait repris la méthode). Il 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  $1/r$ . Les mots les plus courts sont les plus utilisés, inversement

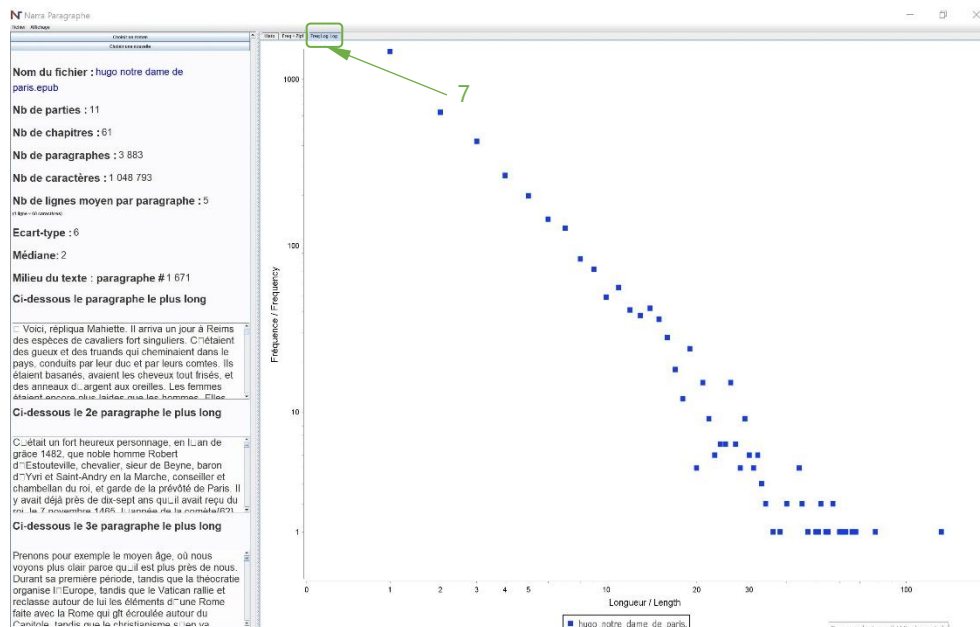
les plus longs sont les moins utilisés. Il trouva que le mot le plus fréquent dans un échantillon apparaît en moyenne une fois tous les dix mots, le deuxième mot le plus fréquent une fois tous les vingt mots..., le n-ième une fois tous les  $10 \times n$  mots.

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. Ainsi, le mot le plus fréquent (rang 1) apparaît 8 000 fois, le deuxième 4 000 fois ( $1/2$ ) et le troisième 2 660 fois ( $1/3$ ), etc.

La loi de Zipf se vérifie quels que soient la langue et le texte. Benoît Mandelbrot théorisa cette loi empirique en 1950 et démontra qu'elle était un cas particulier d'une loi de puissance (*power law*) plus générale, appelée loi de Mandelbrot, s'appliquant 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.

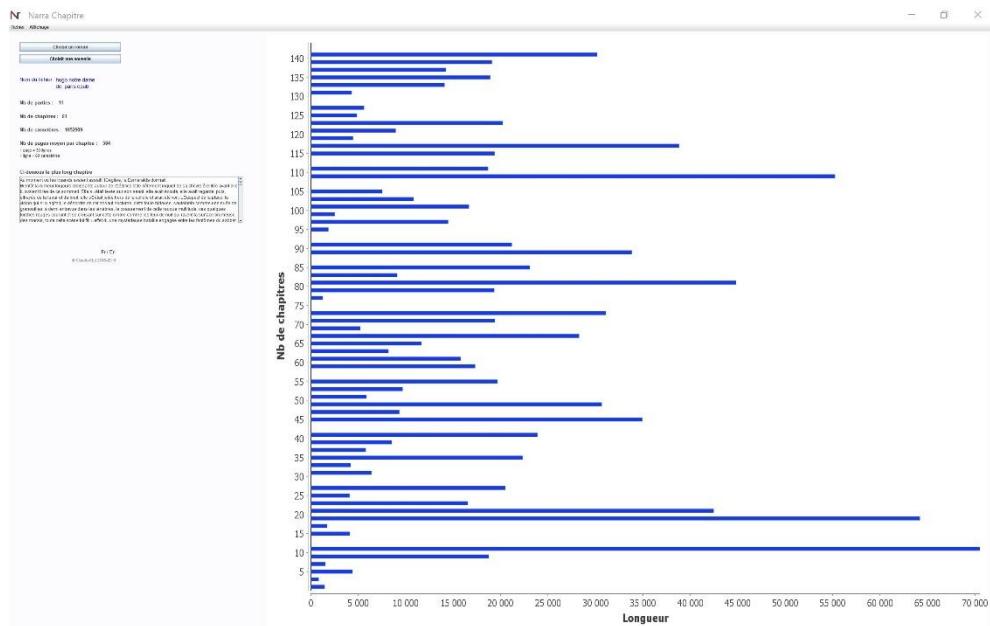
Aussi présumons-nous qu'une telle relation causale entre la longueur des mots et sa fréquence d'usage s'applique aux paragraphes, y compris dans les textes littéraires. Les graphiques ci-dessous confirment l'admissibilité de l'hypothèse, selon laquelle la longueur des paragraphes est inversement proportionnelle à leur fréquence (loi de Zipf).

En cliquant sur l'onglet « Fréq log-log » (7), la courbe de la fréquence en fonction de la taille des paragraphes du texte sélectionné figure à droite en repère log-log.



Après avoir cliqué sur « visualisation des chapitres », l'interface remplace celle de l'accueil. Sur la gauche, y figurent le titre du fichier choisi, quelques données statistiques ainsi que le texte des trois premiers paragraphes les plus longs du texte.

Vous obtenez l'histogramme suivant dont chaque trait horizontal correspond à un chapitre.



La première ligne en haut correspond au premier chapitre du texte, la dernière en bas, au dernier chapitre. Leur longueur dépend de la longueur du chapitre correspondant.

## Les fonctionnalités du logiciel *Narra 2.0*

---

Outre l'histogramme et la courbe des fréquences, le logiciel génère des données statistiques dont seulement quelques-unes sont visibles dans l'interface. Les autres le sont dans un fichier CSV généré par le logiciel.

Trois types de données statistiques sont produites :

1. des valeurs brutes, telles que la mesure de la taille des paragraphes et des chapitres, à leur nombre, etc. ;
2. des indicateurs de tendance centrale (ou de position de valeurs centrales), tels que la moyenne et la médiane de la longueur des paragraphes et des chapitres et le milieu du texte.
3. un indicateur de dispersion (l'écart-type), un indicateur de forme (la localisation du paragraphe correspondant au milieu du texte en nombre de caractères) et d'autres données.

Le logiciel mesure la longueur des paragraphes et des chapitres en nombre de caractères (ou signes espaces compris), traduit en nombre de lignes sur l'équivalence 60 signes pour 1 ligne afin de faciliter une saisie plus rapide de leur dimension.

### ■ Les données statistiques brutes

Ci-dessous les données statistiques brutes que le logiciel génère. Les données correspondent à ceux du texte pris en exemple *Notre-Dame de Paris*.

Nb de parties	11
Nb de chapitres	59
Nb de paragraphes	3 883
Nb de caractères	1 048 793
Taille 1 <sup>er</sup> para plus long	131
Taille 2 <sup>e</sup> para plus long	79
Taille 3 <sup>e</sup> para plus long	68
Taille 4 <sup>e</sup> para plus long	67

Taille 5 <sup>e</sup> para plus long	66
Num 1 <sup>er</sup> para plus long	1 311
Num 2 <sup>e</sup> para plus long	1 191
Num 3 <sup>e</sup> para plus long	1 162
Num 4 <sup>e</sup> para plus long	1 671
Num 5 <sup>e</sup> para plus long	943
Localisation 1 <sup>e</sup> para plus long sur nb tot para	34 %
Localisation 2 <sup>e</sup> para plus long sur nb tot para	31 %
Localisation 3 <sup>e</sup> para plus long sur nb tot para	30 %
Localisation 4 <sup>e</sup> para plus long sur nb tot para	43 %
Localisation 5 <sup>e</sup> para plus long sur nb tot para	24 %
Localisation 1 <sup>er</sup> para plus long sur nb tot para	42 %
Localisation 2 <sup>e</sup> para plus long sur nb tot para	37 %
Localisation 3 <sup>e</sup> para plus long sur nb tot para	34 %
Localisation 4 <sup>e</sup> para plus long sur nb tot para	50 %
Localisation 5 <sup>e</sup> para plus long sur nb tot para	26 %
1 <sup>er</sup> para le plus long	131
2 <sup>e</sup> para le plus long	79
3 <sup>e</sup> para le plus long	68
Taille 4 <sup>e</sup> para plus long	67
Taille 5 <sup>e</sup> para plus long	66

## ■ Les données statistiques agrégées

Ci-dessous les données statistiques agrégées que le logiciel génère.

### Les indicateurs de tendance centrale

Moyenne : nb de lignes moyen par paragraphe	5
Médiane : moitié des para sup et moitié inf	2

## Un indicateur de dispersion

Écart-type	6
------------	---

## Un indicateur de forme

Num para du milieu du texte (médiane des lignes)	1 671
Localisation para du milieu du texte (médiane des lignes)	43,03 %

La localisation du paragraphe correspondant à la médiane en nombre de lignes indique une asymétrie du texte ; la série est étalée à gauche, à savoir les paragraphes sont plus long au début.

## Autres données (1)

1 <sup>er</sup> para plus long	<p>– Voici, rÉpliqua Mahiette. Il arriva un jour É Reims des espÉces de cavaliers fort singuliers. CÉtaient des gueux et des truands qui cheminaient dans le pays, conduits par leur duc et par leurs comtes. Ils Étaient basanÉs, avaient les cheveux tout frisÉs, et des anneaux d'argent aux oreilles. Les femmes Étaient encore plus laides que les hommes. Elles avaient le visage plus noir et toujours dÉcouvert, un mÉchant roquet sur le corps, un vieux drap tissu de cordes liÉ sur l'Épaule, et la chevelure en queue de cheval. Les enfants qui se vautraient dans leurs jambes auraient fait peur É des singes. Une bande d'excommuniÉs. Tout cela venait en droite ligne de la basse Égypte É Reims par la Pologne. Le pape les avait confessÉs, É ce qu'on disait, et leur avait donnÉ pour pÉnitence d'aller sept ans de suite par le monde, sans coucher dans des lits. Aussi ils s'appelaient Penanciers et puaien. Il paraÉt qu'ils avaient ÉtÉ autrefois sarrasins, ce qui fait qu'ils croyaient É Jupiter, et qu'ils rÉclamaient dix livres tournois de tous archevÉques, ÉvÉques et abbÉs crossÉs et mitrÉs. C'est une bulle du pape qui leur valait cela. Ils venaient É Reims dire la bonne aventure au nom du roi d'Alger et de l'empereur d'Allemagne. Vous pensez bien qu'il n'en fallut pas davantage pour qu'on leur interdÉt l'entrÉe de la ville. Alors toute la bande campa de bonne grÉce prÉs de la porte de Braine, sur cette butte oÉ il y a un moulin, É cÉtÉ des trous des anciennes crayÉres. Et ce fut dans Reims É qui les irait voir. Ils vous regardaient dans la main et vous disaient des prophÉties merveilleuses. Ils Étaient de force É prÉdire É Judas qu'il serait pape. Il courait cependant sur eux de mÉchants bruits d'enfants volÉs et de bourses coupÉes et de chair humaine mangÉe. Les gens sages disaient</p>
--------------------------------	---

	<p>aux fous : N'y allez pas, et y allaient de leur cÅt en cachette. C'Åtait donc un emportement. Le fait est qu'ils disaient des choses Åtonner un cardinal. Les mÅres faisaient grand triomphe de leurs enfants depuis que les Ågyptiennes leur avaient lu dans la main toutes sortes de miracles Åcrits en paÅen et en turc. L'une avait un empereur, l'autre un pape, l'autre un capitaine. La pauvre Chantefleurie fut prise de curiositÅ. Elle voulut savoir ce qu'elle avait, et si sa jolie petite AgnÅs ne serait pas un jour impÅratrice d'ArmÅnie ou d'autre chose. Elle la porta donc aux Ågyptiens</p>
<p>2<sup>e</sup> para plus long</p>	<p>C'Åtait un fort heureux personnage, en l'an de grÅce 1482, que noble homme Robert d'Estouteville, chevalier, sieur de Beyne, baron d'Yvri et Saint-Andry en la Marche, conseiller et chambellan du roi, et garde de la prÅvÅtÅ de Paris. Il y avait d'ÅjÅ prÅs de dix-sept ans qu'il avait reÅsu du roi, le 7 novembre 1465, l'annÅe de la comÅte(62), cette belle charge de prÅvÅt de Paris, qui Åtait rÅputÅe plutÅt seigneurie qu'office, dignitas, dit Joannes LÅmnÅus, quÅ cum non exigua potestate politiam concernente, atque prÅrogativis multis et juribus conjuncta est(63). La chose Åtait merveilleuse en 82 qu'un gentilhomme ayant commission du roi et dont les lettres d'institution remontaient Å l'Åpoque du mariage de la fille naturelle de Louis XI avec monsieur le bÅttard de Bourbon. Le mÅme jour oÅ Robert d'Estouteville avait remplacÅ Jacques de Villiers dans la prÅvÅtÅ de Paris, maÅtre Jean Dauvet remplaÅait messire HÅlye de Thorrettes dans la premiÅre prÅsidence de la cour de parlement, Jean Jouvenel des Ursins supplantait Pierre de Morvilliers dans l'office de chancelier de France, Regnault des Dormans d'Åsappointait Pierre Puy de la charge de maÅtre des requÅtes ordinaires de l'hÅtel du roi. Or, sur combien de tÅtes la prÅsidence, la chancellerie et la maÅtrise s'Åtaient-elles promenÅes depuis que Robert d'Estouteville avait la prÅvÅtÅ de Paris ! Elle lui avait ÅtÅ baillÅe en garde, disaient les lettres patentes</p>
<p>3<sup>e</sup> para plus long</p>	<p>Prenons pour exemple le moyen Åge, oÅ nous voyons plus clair parce qu'il est plus prÅs de nous. Durant sa premiÅre pÅriode, tandis que la thÅocratie organise l'Europe, tandis que le Vatican rallie et reclasse autour de lui les ÅlÅments d'une Rome faite avec la Rome qui gÅt ÅcroulÅe autour du Capitole, tandis que le christianisme s'en va recherchant dans les d'Åcombres de la civilisation antÅrieure tous les Åtages de la sociÅtÅ et rebÅtit avec ces ruines un nouvel univers hiÅrarchique dont le sacerdoce est la clef de voÅte, on entend sourdre d'abord dans ce chaos, puis on voit peu Å peu sous le souffle du christianisme, sous la main des barbares, surgir des d'Åblais des architectures mortes, grecque et romaine, cette mystÅrieuse architecture romane, sÅur des maÅonneries thÅocratiques de l'Ågypte et de l'Inde, emblÅme inaltÅrable du catholicisme pur, immuable hiÅroglphe de l'unitÅ papale. Toute la pensÅe d'alors est Åcrite en effet dans ce sombre style roman. On y sent partout l'autoritÅ, l'unitÅ, l'impÅnÅtrable, l'absolu, GrÅgoire VII</p>

## Autres données (2)

Nb de para de plus de 100 lignes	1
Nb de para de plus de 50 lignes	17
Nb de para de plus de 20 lignes	126
Pourcentage de para de 1 et 2 lignes	54 %
Pourcentage de para de 1 ligne	38 %