Etude et caractérisation des disfluences scripturales dans les manuscrits de Stendhal

Nom : VIKHROVA
Prénom : Anne
UFR LLASIC

Mémoire de master 2 recherche – 30 crédits – Sciences du langage

Spécialité ou Parcours : spécialité Industries de la langue

Sous la direction de Thomas LEBARBÉ Année universitaire 2013-2014

Jury : **T.A. Lebarbé, G. Antoniadis, J.-M. Colletta**



Etude et caractérisation des disfluences scripturales dans les manuscrits de Stendhal

Nom : VIKHROVA
Prénom : Anne
UFR LLASIC

Mémoire de master 2 recherche - 30 crédits – Sciences du langage

Spécialité ou Parcours : spécialité Industries de la langue

Sous la direction de Thomas LEBARBÉ Année universitaire 2013-2014

Jury : **T.A. Lebarbé, G. Antoniadis, J.-M. Colletta**

Remerciements

Mes remerciements les plus profonds à l'université Stendhal et plus particulièrement aux enseignants et amis, qui m'ont encouragé pendant ces deux ans du cursus.

UNIVERSITE Stendhal GRENOBLE 3

Déclaration anti-plagiat

DECLARATION

1. Ce travail est le fruit d'un travail personnel et constitue un document original.

2. Je sais que prétendre être l'auteur d'un travail écrit par une autre personne est une pratique sévèrement sanctionnée par la loi.

3. Personne d'autre que moi n'a le droit de faire valoir ce travail, en totalité ou en partie, comme le sien.

4. Les propos repris mot à mot à d'autres figurent entre guillemets (citations).

5. Les écrits sur lesquels je m'appuie dans ce mémoire sont systématiquement référencés selon un système de renvoi bibliographique clair et précis.

Nom : Vikhrova Prénom : Anne

Date : 29 Mai 2014 Signature : 

Avertissement

Nous mettons les termes que nous souhaitons mettre en emphase en italique lorsque nous les introduisons pour la première fois dans le texte. S'il s'agît de noms de fonctions ou des éléments XML nous les mettons en italique à chaque référence. De même, nous mettons toujours en italique le nom du corpus *Journaux*. Nous utilisons aussi cette convention pour faire référence aux sections et sous-sections du mémoire.

Les références que nous présentons dans la bibliographie ne sont pas toutes citées dans le texte, nous les gardons tout de même, car ils nous ont servi de référence pour ce travail.

MOTS-CLÉS : disfluences scripturales, énonciation, linguistique, informatique, littérature, génétique de texte, manuscrits, Stendhal, humanités numériques.

RÉSUMÉ

L'objet du projet de recherche que nous présentons dans ce mémoire porte sur des phénomènes de l'écrit constatés sur les brouillons d'auteurs qui sont analogues aux disfluences de l'oral. Nous donnons le terme *disfluences scripturales* pour décrire ces phénomènes que nous étudions et caractérisons au sein du corpus des manuscrits de Stendhal. Notre étude porte spécifiquement sur le corpus *Journaux* de cette collection d'écrits. Nous nous appuyons sur les principes linguistiques et utilisons les méthodes informatiques pour étudier ce corpus.

Notre approche, d'un point de vue linguistique, s'appuie sur l'apport de la linguistique de l'énonciation à l'étude des disfluences scripturales. Cette analyse en corpus se fonde sur une démarche de traitement automatique des langues - au sens de l'outil et la modélisation informatique comme micro-macro-scope sur un corpus de matériau langagier numérisé. Nous proposons une caractérisation et une discussion sur la notion de disfluences scripturales, à l'aune des analyses plus présentes dans la littérature scientifiques des disfluences de l'oral.

KEYWORDS : scriptural disflencies, enunciation, linguistics, computer sciences, literature, literary criticism, manuscripts, Stendhal, digital humanities.

ABSTRACT

The research project that we present in this memoir concerns phenomena observed in written language and that we specifically observe in the rough drafts of an author's works. This language phenomena is analogous to what are known as *oral disfluencies*. We attribute the term *scriptural disfluencies* to the objects of our study, which we characterise within the corpus of Stendhal's manuscripts. Our study focuses specifically on the corpus *Journaux* (Journals) of this collection of writing. We rely on linguistic principles and use computational methods for this study.

Our approach from the linguistic perspective acknowledges the contribution of the theory of enunciation to the study of scriptural disfluencies. This corpus analysis is also founded in an approach for natural language processing – specifically in the adaptation of an instrument of observation designed for application on a corpus of digitised manuscripts. We propose a linguistic characterisation of the phenomenon along with a discussion around the notion of scriptural disfluencies in light of existing analyses of oral disfluencies in scientific literature.

Sommaire

REMERCIEMENTS 3
AVERTISSEMENT 5
SOMMAIRE 7
TABLE DE FIGURES 9
CONTEXTE HISTORIQUE 13
LE PROJET MANUSCRITS DE STENDHAL EN LIGNE 14
LES JOURNAUX DE STENDHAL - LIEU D'ENTRAINEMENT DE L'ECRIVAIN 14

1 INTRODUCTION 17
1.1 PRESENTATION DU PROJET 17
1.2 POSITIONNEMENT ET STRUCTURE DU MEMOIRE 18
1.3 METHODOLOGIE 19

2 PARTIE LINGUISTIQUE 21
2.1 DEFINITION DE LA (DIS|DYS)FLUENCE 21
2.2 LES DISFLUENCES DE L'ORAL 22
2.3 LES DISFLUENCES DE L'ECRIT OU LES DISFLUENCES SCRIPTURALES 23
2.4 LA LINGUISTIQUE DE L'ENONCIATION 23
2.5 LES BROUILLONS DE L'ECRIT 24
2.5.1 L'intérêt du linguiste pour des brouillons de l'écrit 26
2.5.2 La primauté de l'oral dans les structures de l'écrit 27
2.5.3 Le développement de l'énoncé et les strates d'écriture 29
2.5.4 Le fondement commun de l'énonciation à l'écrit et à l'oral – la pensée 29
2.6 LA CARACTERISATION DES DISFLUENCES SCRIPTURALES– DEFINITION DES CRITERES FORMELS 30
2.6.1 Le chunk 31
2.6.2 Exemples de chunk 32
2.6.3 Le segment 34
2.6.4 Les disfluence par rapport au segment 34
2.7 LE SYSTEME D'ANNOTATION 35
2.7.1 Part-of-speech (POS) Tags 36
2.7.2 Table de la catégorie ISTOK /ISCHUNK (annotation du statut ISTOK/ISCHUNK de la rature) 39
2.7.3 Attribution des éléments répétés du segment contenant la rature 39
2.7.4 Niveau de modification – la stratification du document 40
2.7.5 Table descriptive de positionnement de la rature par rapport au chunk 41

3 PARTIE INFORMATIQUE 45
3.1 METHODES DE TRAITEMENT AUTOMATIQUE DES DISFLUENCES DE L'ORAL 45
3.2 METHODES DE TRAITEMENT DES DISFLUENCES DE L'ECRIT – LE CAS DE TREETAGGER 46
3.3 LE POTENTIEL DESCRIPTIF DU LANGAGE XML 46
3.3.1 La structure XML 46
3.3.2 De la page manuscrite à la page XML – la grammaire de description CLELIA 48
3.3.3 Les éléments XML pertinents à l'étude des disfluences scripturales 49
3.3.4 Table explicative des attributs XML pertinents à l'étude 51
3.4 LA BASE DE DONNEES 51
3.4.1 La base de données manuscrits_stendhal 51
3.4.2 La base de données et le corpus dynamique 52
3.5 LE LANGAGE PHP ET L'AFFICHAGE DYNAMIQUE DES DONNEES 53
3.5.1 Le moteur de recherche 54
3.5.2 Extraction de données avec des requêtes SQL 55

3.5.3 Traitement de données XML avec des expressions régulières 56

4 PARTIE LINGUISTIQUE - INFORMATIQUE 60

4.1 INTERPRETATION LINGUISTIQUE DE L'ENCODAGE INFORMATIQUE 60

4.1.1 Lecture du contexte XML 60

4.1.2 Le contexte de paragraphe et le contexte immédiat de la disfluence 63

4.1.3 La stratification du texte du point de vue du contexte XML 63

4.2 CARACTERISATION LINGUISTIQUE DES DISFLUENCES SCRIPTURALES 63

4.2.1 Le regroupement selon le critère ISTOK/ISCHUNK 64

4.2.2 La distribution des catégories grammaticales selon le critère ISTOK/ISCHUNK 65

4.2.3 Les modifications de 1e niveau et de 2e niveau 65

4.2.4 Les unités répétées 67

4.2.5 L'étude des opérations de la prosodie et de la syntaxe et l'étude de la topologie du texte 70

4.3 FACTEURS INATTENDUS 70

5 CONCLUSION 72

ANNEXE A 73

ANNEXE B 132

ANNEXE C 135

LISTE DES CORPUS DES MANUSCRITS DE STENDHAL 135

INDEX 138

BIBLIOGRAPHIE 140

LIVRES ET ARTICLES EN-LIGNE 140

SITOGRAPHIE 142

RESSOURCES ET MANUELS INFORMATIQUES 142

Table de Figures

Figure 1 : Exemple de phrase avec un chunk nominal qui intègre un adjectif. 32
Figure 2 : Exemple de phrase avec un chunk pronominal constitué d'un pronom et un verbe auxiliaire. Le chunk pronominal est suivi d'un chunk verbal. 32
Figure 3 : Exemple de phrase avec trois chunks constitués autour la préposition *de* (en haut). Le deuxième exemple démontre l'intégration d'un chunk nominal qui porte aussi un adjectif (en bas). 33
Figure 4 : Exemple de phrase où plusieurs verbes constituent un chunk verbal après un chunk pronominal avec pronom et verbe auxiliaire. 33
Figure 5 : Exemple de phrase avec deux types de conjonctions enchainées : *et, qu' (l'élision de que)*. 33
Figure 6 : Exemple de phrase avec un chunk adjectival qui se produit à cause d'une conjonction. 33
Figure 7 : Exemple du cas rare d'un chunk adverbial composé de l'adverbe *uniquement* et le cas plus commun de l'intégration du même adverbe dans un chunk nominal. 34
Figure 8 : Table des annotations des *parts of speech* (POS) adaptées dans le cadre de l'étude 39
Figure 9 : Table explicative de la catégorie ISTOK/ISCHUNK 39
Figure 10 : Table explicative de l'annotation des éléments répétés 40
Figure 11 : Table explicative des niveaux de modification du document 40
Figure 12 : Table descriptive de positionnement de la rature par rapport au chunk 41
Figure 13 : Représentation graphique des relations entre les critères que nous définissons. 42
Figure 14 : Un exemple d'un document XML 47
Figure 15 : Exemple d'un fichier DTD associé au document XML de la figure 6. 48
Figure 16 : Schéma explicatif de la relation client-serveur et les scripts PHP/MySQL, Javascript, CSS. 54
Figure 17 : Schéma du moteur de recherche développé pour le corpus des manuscrits de Stendhal 55
Figure 18 : Requête SQL qui permet de sélectionner les pages du corpus contenant un élément prédéfini par l'utilisateur à partir d'un formulaire HTML. Cet exemple est le cœur du fonctionnement du moteur de recherche que nous utilisons. 56
Figure 19 : Requête SQL qui permet de trier nos annotations selon le critère ISTOK/ISCHUNK et le 1e niveau de modification. Cette requête renvoie les données regroupées par catégorie grammaticale (cf. Annexe B, Figure 2). 56
Figure 20 : Exemple de l'utilisation de la fonction preg_match_all() 57

Figure 21 : Exemple de l'utilisation de la fonction preg_match()....57
Figure 22 : Exemple de l'utilisation de la fonction str_replace() 57
Figure 23 : Graph des regroupements des disfluences scripturales selon le critère ISTOK/ISCHUNK....64
Figure 24 : Distribution des occurrences de catégories grammaticales pour chaque groupe du critère ISTOK/ISCHUNK (Nous avons regroupés les sous-catégories grammaticales sous leur catégories respectives pour une meilleure représentation de cette distribution. Ex : VERBE, PRON). 65
Figure 25 : Distribution des modifications du 1e et du 2e niveau pour les groupes du critère ISTOK/ISCHUNK....66
Figure 26 : Distribution des occurrences de catégories grammaticales pour chaque groupe du critère ISTOK/ISCHUNK et selon le niveau de modification (1/2)....67
Figure 27 : Table relevant des catégories grammaticales des unités répétées pour le 1e niveau de modification. 68
Figure 28 : Table relevant des catégories grammaticales des unités répétées pour le 2e niveau de modification. 69
Figure 29 : Table des annotations du corpus *Journaux*. 131
Figure 30 : Distribution des occurrences de catégories grammaticales pour chaque groupe du critère ISTOK/ISCHUNK 132
Figure 31 : Distribution des occurrences de catégories grammaticales pour chaque groupe du critère ISTOK/ISCHUNK et selon le niveau de modification (1/2)....133
Figure 32 : Distribution des catégories grammaticales pour les deux groupes ISTOK et ISCHUNK dans l'ordre décroissant. 134
Figure 33 : Distribution chronologique des termes utilisés pour décrire les disfluences (Cette représentation est non exhaustive.)....137

Préambule

Contexte historique

> *Dans son rapport au temps, Stendhal est pleinement un homme issu de la Révolution, en accord symbolique, dans un premier moment, avec le premier héritier historique de celle-ci, Napoléon, le « fils de la Révolution » : l'épopée napoléonienne, avec ses ombres et ses lumières, donne une unité historique à la période des* Journaux et Papiers de nos trois premiers tomes…*Stendhal ne s'interroge pas sur l'opportunité d' « épouser » son temps ou d'y résister : le scripteur de* Journaux et Papiers *de cette période* est *son temps, suit son tempo épique, prend acte de la rupture définitive introduite par la Révolution*[1] *et adhère au temps symboliquement homogène qui en est issu*[2].

La bibliothèque municipale de Grenoble possède une collection de manuscrits de Stendhal composée d'environ 20 000 feuillets [Faure *et al.,* 2009]. Elle est composé de divers documents de l'écrivain : des journaux personnels, des lettres, et des dossiers de travail. Si ces documents existent toujours c'est parce que ils ne'ont pas été publiés –la pratique des éditeurs de l'époque étant de détruire le manuscrit. Aujourd'hui, cette collection d'écrits est à la fois une attestation de l'époque historique à laquelle elle appartient et aux pratiques et processus de création de l'écrivain Stendhal. La collection permet de porter un regard sur ce moment précis de l'histoire [Meynard *et al.,* 2013]. L'écrivain Stendhal, comme le disent les éditrices des *Journaux et Papiers* (2013), ne « problématise pas vraiment le « temps historique » ni le moment historique auquel il appartient. Pourtant, le lecteur qui, plus de deux siècles plus tard porte le regard sur ces écrits le fait avec sa compréhension du contexte historique.

[1] Rupture qui correspond à l'émergence du nouveau sens du mot « révolution » : une histoire linéaire qui remplace le temps circulaire. Voir M. Ozouf, article « Révolution », *Dictionnaire critique de la Révolution française,* F. Furet et M. Ozouf (éd), Paris, Flammarion, 1988, p. 847-858.
[2] Dans *Journaux et Papiers* de [Meynard *et al.,* 2013], p. 8 - 9

Le projet Manuscrits de Stendhal en ligne

La collection de ces manuscrits contient des documents précieux et très fragiles. Par conséquence, la bibliothèque permet leur consultation pour des raisons justifiées et de manière ponctuelle [Faure *et al.,* 2009]. L'objectif est de mieux préserver ces documents historiques. Cet objectif de préservation est aussi une des motivations principales de créer la plateforme en ligne pour ce fonds. La plateforme des manuscrits de Stendhal rassemble la collection sous forme numérique. Elle permet au public d'avoir accès à l'ensemble de documents numérisés et de les consulter sur un site web. Le public visé est principalement constitué des trois groupes suivants :

- les chercheurs historiens, les chercheurs en lettres et les généticiens de texte
- les chercheurs en linguistique
- les enseignants et étudiants en lettres

La plateforme est accessible également aux transcripteurs des manuscrits qui sont les principaux contributeurs à la constitution de ce corpus. C'est grâce à ces contributions que les pages numérisées de script sont converties à un corpus de texte. Le corpus textuel permet d'atteindre une meilleure lisibilité des pages. Grâce aux transcriptions le public peut plus facilement accéder au sens du script et aux idées de l'écrivain. Les transcriptions permettent également de représenter tous les aspects du manuscrit qui ne sont pas conservés dans des éditions imprimées du texte. La transcription diplomatique du manuscrit permet de représenter la mise en page du script, la typographie et les éléments graphiques de la page. Elle permet aussi d'observer les modifications du script ou les phénomènes de réécriture [Faure *et al.,* 2009] auxquelles nous nous intéressons particulièrement. Pour ce projet nous nous intéressons aux processus et pratiques de l'écrivain à travers les disfluences scripturales dans les manuscrits.

Les journaux de Stendhal - lieu d'entrainement de l'écrivain

L'édition papier des *Journaux et Papiers* de Stendhal est, comme le corpus échantillonné que nous étudions pour ce projet de recherche, un ensemble d'écritures d'un écrivain qui n'est pas encore publié et n'est pas encore connu [Meynard *et al.*, 2013]. De plus, par leur caractère personnel les journaux et papiers de l'écrivain Stendhal peuvent être considérés comme des documents biographiques. Stendhal, comme beaucoup d'autres de son époque, s'inscrit dans la pratique de tenir un journal [Meynard *et al.,* 2013]. Nous pouvons interpréter son activité

d'écriture comme étant une activité qui lui permet de se comprendre ou de se connaître [Meynard *et al., 2013*]. Cependant, les journaux et papiers divers que Stendhal a laissés et qui font partie de la collection de ces manuscrits, sont aussi des documents qui attestent du fait que Stendhal s'entrainait à écrire. Certainement, les journaux et papiers se distinguent des documents de travail par leur caractère personnel, mais l'activité diariste de Stendhal n'était pas simplement pour travailler sa compréhension de lui-même, mais aussi de travailler l'écriture. Il utilise ses journaux pour s'entrainer à écrire et donc à se former au métier d'écrivain.

Son travail d'écrivain ne se trouve donc pas seulement dans ses dossiers de travail mais aussi dans ses journaux [Meynard *et al., 2013*]. Nous étudions donc les manuscrits de ses journaux et leur rapport avec les théories linguistiques d'énonciation. Le processus d'écriture chez l'écrivain met en évidence des processus divers et dynamiques qui se rassemblent sous la branche disciplinaire de la linguistique d'énonciation. Les relations plus étroites entre l'écrit et l'oral peuvent en suite s'établir lorsque nous entreprenons l'analyse fine de l'activité d'écriture. Enfin, nous voyons que même dans la représentation informatique du corpus les processus de modification tel que les *ajouts,* les *interlignes* et les *variantes en interligne* évoquent les différentes formes de reformulation qui se produisent lors de la construction énonciative.

1 Introduction

1 Introduction

1.1 Présentation du projet

Le projet de recherche que nous présentons dans ce mémoire entre dans une approche pluridisciplinaire d'étude de la langue et de Traitement Automatique des Langues (TAL). Nous nous situons à la fois entre l'informatique, la linguistique et la littérature. Nous sommes à l'intersection de plusieurs axes disciplinaires qui se retrouvent, de nos jours, de plus en plus souvent en relation. Ces axes fusionnent ensemble dans la filière disciplinaire intitulée les *humanités numériques.*

Le sujet de notre étude porte sur les disfluences scripturales dans les manuscrits de Stendhal. Dans ce cadre, nous disposons d'un corpus dont la constitution fait partie d'un projet de recherche en génétique de texte, en enseignement littéraire, en linguistique et permettant de mener des projets d'édition. Ce projet existe depuis plusieurs années et sa longévité atteste de sa valeur reconnue par la communauté scientifique et les fonds patrimoniaux de la région Rhône-Alpes. Nous avons alors la possibilité de faire des études linguistiques sur ce corpus, relevant de plusieurs scripteurs travaillant souvent en collaboration, dont un scripteur l'écrivain Stendhal lui-même. Du point de vue linguistique, nous observons plusieurs processus d'écriture :

- spontanée,
- par édition en plusieurs étapes,
- par transcription de l'oral à l'écrit par diction.

La constitution du corpus est réalisée à l'aide de techniques fondamentales en informatique. Notre étude linguistique et le traitement informatique du corpus repose donc sur l'adaptation des techniques informatiques appliquées. Dans le cadre de notre projet, le corpus est notamment structuré en XML et repose sur une base de données SQL.

Notre travail de recherche linguistique repose sur l'analyse de la variété et de la variabilité des disfluences scripturales, phénomène langagier que nous étudions. Ce phénomène est déjà connu par de nombreuses études de la langue orale effectuées sur des corpus oraux. De plus, de

nombreuses contributions théoriques de la filière de la linguistique d'énonciation sur des processus de reformulation et de paraphrase ont apporté des regards distincts sur ces processus de construction énonciative.

Dans le cadre de notre projet, nous analysons les variables de marques de disfluences écrites présentées dans le corpus avec l'objectif de les caractériser après une observation fine d'un échantillon, le corpus *Journaux,* sur lequel est fondée notre étude. Les résultats de cette observation nous permettront de mettre en rapport les disfluences scripturales avec les études linguistiques déjà réalisées autour des disfluences de l'oral.

1.2 Positionnement et structure du mémoire

La problématique de ce projet de recherche tourne autour des axes de la linguistique, des lettres et de la génétique de texte. Cependant, il touche également l'axe informatique car la représentation des données joue un rôle crucial à au moins deux niveaux :

- la récupération de données destinées à l'observation
- la caractérisation des données en vue de l'étude linguistique et de traitement automatique

Le mémoire est donc présenté en trois parties majeures :

Dans la section *Linguistique* nous parlons de l'apport de la linguistique de l'énonciation à l'étude des disfluences scripturales. Nous détaillons les critères linguistiques que nous utilisons dans l'étude du corpus.

Dans la section *Informatique* nous parlons du traitement automatique (en TAL) des disfluences. Nous expliquons les méthodes informatiques que nous utilisons pour l'étude de ce corpus.

Dans la troisième section, *Linguistique-Informatique,* nous présentons l'interprétation linguistique du corpus structuré en XML. Nous terminons par une présentation de nos observations du corpus.

1.3 Méthodologie

La méthodologie consiste à construire un outil d'observation qui permettra d'interagir avec le serveur et la base de données appelée *manuscrits_stendhal* et de récolter les données destinées à l'observation. Nous décomposons cette méthode en différentes activités menées en parallèle :

- Etude et compréhension de la valeur de la linguistique énonciative dans l'étude du discours et de la construction énonciative aussi bien à l'oral qu'à l'écrit.
- Précision des paramètres qui seront utilisés pour l'observation et qui serviront de critères de caractérisation.
- Réalisation d'un outil informatique pour consulter le corpus de manuscrits et étudier le phénomène de disfluence scripturale, en se focalisant sur la rature ou *biffe*[3].
- Observation du corpus et annotation des éléments repérés selon les paramètres définies.
- Caractérisation de l'échantillon selon ces paramètres.
- Discussion sur l'apport de notre caractérisation dans le contexte d'études sur l'oral et sur l'écrit en linguistique de l'énonciation comme en lettres.

[3] Nous utilisons les termes biffe et rature de manière interchangeable. Le terme *biffe* reflète celui de l'outil XML développé pour ce projet et nous le reprenons pour préserver cette cohérence.

2 Partie linguistique

Dans cette partie du mémoire nous introduisons et définissons le phénomène du langage que nous étudions, celui des disfluences scripturales. Nous explicitons le rapport de ce phénomène avec les disfluences de l'oral à travers la théorie linguistique de l'énonciation et leur fondement commun : la pensée. Nous expliquons aussi l'intérêt d'étudier les brouillons de l'écrit ou dans notre cas le manuscrit pour mieux comprendre ce phénomène. Nous introduisons et expliquons les méthodes linguistiques sur lesquelles nous nous fondons pour notre étude des disfluences scripturales.

2 Partie Linguistique

2.1 Définition de la (dis|dys)fluence

L'énonciation est un processus de coordination de la pensée, qui est multidimensionnelle [Robert, 1999]. La disfluence donc, en dépit de la connotation négative apportée par le préfix « dis » rajouté à la racine « fluence », s'agît d'un phénomène tout à fait commun. Ce phénomène est toutefois intéressant à étudier car il permet d'étudier les processus de cette coordination et la façon dont ils sont reflétés dans les unités et structures de la langue.

Une distinction existe dans la littérature scientifique entre les deux homophones les « disfluences » et les « dysfluences ». L'usage interchangeable de ces deux termes relèvent le plus souvent d'une traduction imprécise. Le fait que les phénomènes décrits par ces termes ont des réalisations similaires rajoute à la confusion. Nous souhaitons, donc, expliquer cette différence. Le terme « dysfluence » est le plus souvent utilisé en référence aux bégaiements, ou aux productions qui ont une dimension pathologique [Bove, 2008]. Les sources peuvent être à la fois physiologiques et psychologiques et les productions affectent de manière chronique l'efficacité de la communication. Les locuteurs ayant ces troubles chroniques ont souvent besoin d'intervention de la part de spécialistes du langage, des psychologues et des néorologues [Piérart, 2011]. Pourtant, il y a certainement des réalisations syntaxiques observées dans le bégaiement qui sont similaires aux disfluences « normatives ». Les répétitions de syntagmes, les phrases incomplètes, les révisions se produisent aussi régulièrement chez les locuteurs bégayants [Piérart, 2011]. En effet, ils font tous partie des effets observables de la restructuration syntaxique, qui est du même type que celle des disfluences non-pathologiques. La différence se trouve dans le degré, la fréquence et le contrôle volontaire par le locuteur de ces productions. Néanmoins, il se trouve souvent que le terme « dysfluence » soit traduit du français vers l'anglais comme « disfluency » et qu'en anglais la distinction soit faite entre « pathological disfluency » et « non-pathological disfluency », sans modification orthographique mais par modification adjectivale.

La disfluence dite « normale » peut avoir les mêmes types de réalisations (répétitions, hésitations, réparations inattendues) au niveau superficiel de la structure de l'énoncé sans impliquer les troubles chroniques anatomiques ou cognitifs de la « dysfluence » pathologique. Le phénomène est considéré comme étant lié au caractère improvisé ou spontané du discours oral en cours de production [Blanche-Benveniste, 1991]. Les disfluences sont des marques du processus

chez le locuteur de la construction de son énoncé. Nous utiliserons donc le terme « disfluence » pour faire référence à ce phénomène commun du langage.

2.2 Les disfluences de l'oral

Les disfluences de l'oral ont été profondément étudiées en linguistique. Nous présentons dans l'*Annexe C* (Figure 33) une distribution chronologique des contributions scientifiques fondamentales à l'étude de ces phénomènes et les termes résultants[4]. Les disfluences sont considérés comme étant lié au caractère improvisé ou spontané du discours oral en cours de production [Blanche-Benveniste, 1991]. Selon Blanche-Benveniste (1991), comme d'autres linguistes qui ont contribués profondément à l'étude de la relation entre l'oral et l'écrit, le bon usage du français repose largement sur des prescriptions venant de l'écrit. Ceci entraîne des prescriptions sur ce que c'est l'énoncé idéal ou l'énoncé normatif. Pourtant, à l'oral nous apercevons que la production spontanée entraîne toujours des marqueurs d'interruption, de reformulation et de modification de l'énoncé.

En Traitement Automatique des Langues (TAL) l'étude des disfluences à permis de créer des systèmes efficaces de traitement de la parole [Goulian *et al., 2002 ;* Shriberg *et al.*, 2006 ; Dister *et al.,* 2009 ; Hindle, 1983]. Ces traitements développés en TAL concernent le repérage des séquences disfluentes afin de réparer ces séquences en synthèse. Généralement un traitement automatique peut se présenter sous cette forme :

- repérage,
- filtrage,
- correction automatique ou nettoyage

L'objectif est de rendre des énoncés oraux aussi fluents que possible afin de privilégier la cohérence des énoncés [Goulian *et al., 2002 ;* Shriberg *et al.*, 2006 ; Dister *et al.,* 2009 ; Hindle, 1983]. Nous exposons en plus de détails les méthodes appliquées à l'oral et les corpus sur lesquels ces traitements ont été effectués dans la partie *Informatique* (cf. *Méthodes de traitement automatique des disfluences de l'oral*).

[4] Notre représentation graphique est inspirée de la liste constituée par R. Bove de [Bove, 2008].

2.3 Les disfluences de l'écrit ou les disfluences scripturales

Nous utiliserons le terme *disfluences scripturales* pour un phénomène langagier analogue qui se réalise à l'écrit et que nous observons dans les manuscrits de l'écrivain Stendhal. Nous constatons que les disfluences scripturales font partie du processus d'énonciation à l'écrit. Nous nous appuyons sur la théorie de la linguistique d'énonciation parce que les contributions de cette théorie peuvent aussi contribuer à notre caractérisation du phénomène.

2.4 La linguistique de l'énonciation

Nous ne faisons pas ici l'éventail exhaustif des contributions de la théorie de l'énonciation à la discipline de la linguistique, mais nos réflexions et réactions portent largement sur des travaux de Culioli [Culioli, 1999], de Blanche-Benveniste [Blanche-Benveniste, 1991 ; Blanche-Benveniste, 1993], Fuchs [Fuchs, 1994] et dans une fonction interdisciplinaire sur la contribution des idées de De Biasi [Biasi, 2007 ; Biasi, 2013] à l'étude du manuscrit et généralement à la génétique textuelle.

Pour Antoine Culioli [Culioli, 1999], la linguistique de l'énonciation est l'étude des pratiques discursives et rhétoriques parmi l'ensemble des pratiques langagières. La linguistique de l'énonciation s'intègre dans l'ensemble des filières de la linguistique parce qu'elle concerne aussi bien les unités lexicales, que les structures syntaxiques et prosodiques. Cependant, par son regard spécifique sur les *pratiques,* cette branche de la discipline est beaucoup plus concernée par la dimension pragmatique du langage. Si les études sur la syntaxe et la grammaire concernent les unités linguistiques et les relations entre eux, l'effet des unités et des structures sur l'énonciation a toujours une dimension qui est de l'ordre pragmatique. Il y a un processus et donc une *activité* de signification dans lequel s'inscrit l'énonciateur. Ceci est la raison pour laquelle Culioli voit tout le processus de la production langagière comme signification, y compris les reformulations et échecs de l'énoncé. Contrairement à la pragmatique traditionnelle, selon laquelle les reformulations ne signifient pas puisqu'elles produisent des énoncés déformés ou trop écartés de la norme [Fuchs, 1994]. L'énonciation est donc un processus progressif qui se déroule dans le temps et qui peut porter des traces de sa production [Culioli, 1999]. Ces traces sont aussi signifiantes que les énoncés plus fluents. À l'écrit, les traces ou marques de la production sont les

disfluences scripturales que nous étudions. L'intérêt linguistique que nous portons sur ce corpus de manuscrits de l'écrivain Stendhal se fonde sur deux aspects principaux :

- la relation entre l'oral et l'écrit dans la production énonciative
- la multidimensionalité de la pensée et le processus de coordination et de développement de celui-ci lors de l'écriture.

Le fait que les marques de production sont évidentes dans les brouillons de l'écrit nous permet de les repérer et de les caractériser.

2.5 Les brouillons de l'écrit

[…] l'oral improvisé est analogue […] [en termes de marques d'improvisation[5]] aux brouillons de l'écrit. Une comparaison légitime ne devrait se faire qu'entre l'oral improvisé et l'écrit à l'état de brouillon[6]

En littérature et en génétique des textes, un écrit à l'état de brouillon est un « avant-texte ». Le terme fait référence à la fois aux états provisoires et inachevés d'un texte et à son processus de production. Il est généralement appliqué aux brouillons de textes publiés. Il est donc important de noter qu'un avant-texte n'est pas simplement un brouillon, mais plus précisément un brouillon qui précède un texte final.

L'avant-texte est aussi l'espace observable des étapes de modification, ponctuées par des suppressions et des ajouts, d'un texte en voie de rédaction et en métamorphose vers un état final et l'énoncé idéal. Le manuscrit est le lieu de l'activité de la création écrite où le processus de rédaction et de relecture peut se décomposer en multiples étapes. L'étude donc des manuscrits permet une observation minutieuse des processus synchrones et asynchrones ainsi que des étapes successives qui contribuent au développement de l'énoncé.

Dans la collection des Manuscrits de Stendhal, il y a des pages de brouillons qui sont proprement des « avant-textes » parce qu'ils précèdent des publications, mais la collection contient principalement des journaux personnels et des dossiers de travail qui n'ont jamais été publiés et donc « ne deviendront jamais texte » [Meynard, 2014].

[5] Blanche-Benveniste compte parmi les marques d'improvisation les répétitions, ratures, recherches de mots, hésitations, « euh », inachèvements (1991, p. 56).
[6] [Blanche-Benveniste, 1991].

Le manuscrit présente un cas unique pour l'étude linguistique de l'adaptation de l'énoncé à l'écriture. Nous observons sur la page la conversion de la pensée multidimensionnelle [Robert, 1999] vers la forme linéarisée de l'écriture. De la même façon, l'adaptation de l'énoncé à l'écrit entraîne une linéarisation des structures prosodiques [Martin, 2013]. Toutes les modifications que nous observons sur la page font preuve du modèle paradigmatique qui fait le lien entre la cognition et le langage [Blanche-Benveniste, 1991 ; Fuchs, 1994 ; Dufaye et Gournay, 2013, Robert, 1999]. De plus, observer un document de brouillon nous amène à considérer l'activité d'énonciation sous trois aspects spécifiques. Nous lions ensuite ces aspects aux critères que nous utilisons pour la caractérisation linguistique du corpus de *Journaux.*

Le premier aspect est celui des opérations qui font partie du processus d'énonciation (de reformulation, d'expansion, ou de réduction de l'énoncé). Le deuxième aspect apporte une dimension de temporalité en décomposant le processus en étapes. Nous pouvons ainsi distinguer les modifications faites lors de l'écriture de celles apportées aux moments ultérieurs. Enfin, par le troisième aspect il est possible de décrire la page de brouillon en termes d'intervenants et ainsi de distinguer les modifications apportées par le même scripteur de celles apporté par d'autres.

Le premier aspect des opérations est très rigoureusement décrit par la grammaire de description informatique, CLELIA, adopté pour le projet. Ainsi, les diverses opérations de modification de l'énoncé que nous observons dans la page manuscrite sont représentés dans le corpus. Nous détaillons plus précisément les termes associés, les *ajouts,* les *biffes,* les *interlignes,* les *variantes en interligne*—pour ne mentionner que les principaux—dans la partie *Informatique.*

Le deuxième aspect permet de lier les disfluences scripturales avec l'oral spontané. Si selon Blanche-Benveniste les productions de l'oral spontané sont comparables aux brouillons de l'écrit, il est intéressant d'observer et de caractériser les types de productions qui viennent particulièrement des modifications en cours d'écriture. Les modifications qui résultent de l'activité de relecture étant d'un autre ordre. Elles attestent aussi à la capacité du langage d'encourager le scripteur ou le relecteur à la paraphrase ou à la reformulation. Dans les cas des manuscrits de Stendhal, nous voyons souvent l'expansion ou le développement du texte [Meynard, 2014]. Pourtant, le point de vue du scripteur à la relecture du document diffère de son point de vue lors du premier jet. Les détails du texte sur lesquelles il focalise ne sont pas les mêmes. C'est à dire que relire un énoncé déjà produit n'entraîne pas les mêmes processus de coordination de la pensée.

Nous considérons le troisième aspect qui concerne les interventions d'autres scripteurs ou de relecteurs. Lorsque notre corpus de *Journaux,* sur lequel nous commençons cette étude de

caractérisation ne contient pas beaucoup de modifications de ce type, le critère nous servira ultérieurement pour l'étude d'autres corpus dans lesquels il est plus présent.
Nous sommes dans la capacité d'observer les disfluences scripturales dans le corpus des *Journaux* sous les trois aspects que nous venons de mentionner. Ils nous serviront pour la caractérisation linguistique de ces phénomènes.

2.5.1 L'intérêt du linguiste pour des brouillons de l'écrit

Jean Véronis [Véronis, 2004] transmet parfaitement le type de relation que nous observons entre l'oral est l'écrit. Pour une langue telle que le français, le statut de l'écrit est maintenu comme étant de la plus haute importance. Véronis cite également Claire Blanche-Benveniste dans son explication de cette relation.

> *Deux raisons expliquent sans doute cet état de fait. Tout d'abord, l'idée que l'écrit, qui représente la norme, est seul digne d'étude est ancré de façon très profonde dans la pensée occidentale : Denys de Thrace affirmait déjà au Ier siècle que la grammaire est la connaissance de la langue des poètes et des prosateurs. Divers auteurs ont dénoncé les préjugés dont l'oral était victime : il est encore très largement considéré comme une langue fautive, brouillonne et inorganisée (cf.Blanche-Benveniste, 1997). Cette situation est très largement celle de la recherche en traitement automatique des langues : les phénomènes propres à l'oral sont souvent considérés comme des erreurs et dysfonctionnements par rapport à la « vraie » grammaire, celle de l'écrit policé.*

La grammaire policée de l'écrit étant quelque chose qui n'appartient qu'aux écrivains ou prosateurs fait simplement partie d'un contrôle hégémonique de la langue. Le statut accordé à l'écrit pour le français, comme pour d'autres langues du monde, fait partie des moyens de contrôle souvent mis en place et maintenu par des structures de pouvoir. Une discussion plus approfondie sur cet aspect se fonde proprement dans la branche de la sociolinguistique.

Ceci étant dit, étudier la langue orale permet de répertorier les phénomènes de l'énonciation qui ne s'intègrent pas dans la grammaire décrite pour le français. Ceci inclus les unités aussi bien que les structures grammaticales. Cependant, il ne faut pas croire que l'oral et l'écrit sont des langues différentes ou totalement opposées. Comme le dit Blanche-Benveniste, ces deux modes ont plus de similarités que de différences [Blanche-Benveniste, 1994]. Ces

similarités et distinctions ont besoin d'être décrites. L'étude des brouillons de l'écrit permet de le faire. Les similarités que nous y trouvons permettent de se rapprocher des fondements cognitifs de l'énonciation. Les distinctions permettent d'apercevoir l'existence des règles propres à l'écrit, la grammaire, que nous venons de mentionner. Aussi, elles amènent peut-être à préciser les règles et comment celles-ci interviennent dans des situations d'énonciation spécifiques.

Etudier les manuscrits permet d'observer le même processus d'énonciation traduit vers la forme écrite. L'activité est la même, c'est-à-dire celle de l'énonciation. La source est aussi la même, celle de la pensée. Les processus de coordination de la pensée entraînent les mêmes types d'opérations (de reformulation, de paraphrase, de réduction et d'expansion de texte). Les disfluences scripturales qui interviennent lors du processus d'écriture sont corrigés par le scripteur au moyen des séquences de reformulation.

Entreprendre l'étude du manuscrit, en se fondant sur les similarités entre les deux modes de l'écrit et de l'oral, demande une sensibilisation profonde aux conventions des deux. Ensuite, décrire les unités, les structures et les processus de ces phénomènes permet de mener l'étude profonde de la langue, comme l'ont déjà fait tant de linguistes auparavant. Les observations tirées de l'étude peuvent servir pour mieux décrire les distinctions ou inversement éliminer des critères discriminants mal-fondés.

2.5.2 La primauté de l'oral dans les structures de l'écrit

Il est bien connu pour les linguistes qui travaillent sur la relation entre l'écrit et l'oral que la prosodie affecte la syntaxe. De plus, la prosodie, comme la syntaxe, peut être décrite en termes de structure. Toutes les deux font partie d'un ensemble de structurations qui agissent sur l'énoncé et qui ont été identifiés par les linguistes [Martin, 2013b]. Nous énumérons ces structurations ici pour les expliciter :

- prosodiques,
- morphologiques,
- syntaxiques,
- informationnelles

Ces quatre structurations opèrent sur la structure de l'énoncé. La question est comment fonctionne le processus de coordination de ces quatre facteurs lors de la construction de l'énoncé ? Lorsque nous nous intéressons à la relation entre l'oral et l'écrit nous considérons les

effets que la prosodie peut avoir sur l'énoncé lorsqu'elle guide la structure syntaxique de celle-ci. Nous nommons deux effets qui sont importants dans la construction de l'énoncé :

- la désambiguïsation des unités potentiellement ambigües [Martin, 2013a].
- La ponctuation ou la mise en relief des unités (dans un développement discursif).

Lorsque nous parlons des facteurs qui interviennent lors de la construction de l'énoncé, nous parlons du processus qui s'appelle l'encodage. Dans le processus de la transmission de l'énoncé le processus de l'encodage est effectué par l'énonciateur. Le processus qui est effectué par le récepteur de l'énoncé s'appelle le décodage. Il implique une analyse des facteurs prosodiques, syntaxiques, morphologiques et informationnels. Est-ce que, comme le dit Martin ces facteurs sont analysées simultanément ou l'une après l'autre [Martin, 2013b] ? Comment est-ce que le récepteur découpe l'énoncé lors du décodage ?

La lecture de l'énoncé écrit est aussi un processus de décodage de manière similaire à l'activité d'audition. Cependant, c'est un processus qui implique la perception visuelle des signes codifiés du langage. L'écrit est non seulement caractérisé par une certaine linéarité [Martin, 2013a], mais aussi par l'encodage de l'énoncé par des symboles graphiques. Ainsi, après la transmission de l'énoncé sur papier, le lecteur décode l'énoncé et met en pratique ses connaissances de la relation entre graphème et phonème. Il décode aussi les autres marques de l'énoncé écrit, notamment celles de la ponctuation, la coordination et la subordination.

La linéarisation à l'écrit est également observable du point de vue de la syllabe. Lors de la production de l'énoncé l'enchainement de syllabes produit la structure prosodique de l'énoncé, qui est aussi modulé par l'accentuation et la durée des syllabes [Martin, 2013a]. Selon Martin, la structure est donc non linéaire, elle est plutôt hiérarchique. Elle tend à s'aplatir lors de l'adaptation de l'énoncé aux conventions de l'écrit. L'orthographe ne représente pas la structure prosodique. Cette structure aussi doit être décodée par le lecteur.

Ce qui rend intéressant l'étude de l'écrit, du point de vue de la prosodie et des aspects intervenants de l'oral, est qu'évidemment les aspects qui sont transmis à l'écrit sont décodés par les lecteurs, mais pas toujours de la même façon. Le contact du récepteur avec l'énoncé met en relation les compétences de décodage du récepteur avec les compétences de l'énonciateur qui a encodé l'énoncé au départ. Il peut aussi mettre en question le système d'encodage, celui de la grammaire de l'écrit, et les possibilités et limites qu'elle définie pour l'expression énonciative. C'est un fait auquel l'écrivain est profondément sensible. De plus, nous pensons que le processus d'encodage dans lequel s'inscrit l'écrivain cible un décodage spécifique par le récepteur. De son

côté, le récepteur de l'énoncé effectue un décodage à partir des marques de l'écrit qui permettent de reconstruire la forme orale de l'énoncé.

2.5.3 Le développement de l'énoncé et les strates d'écriture

Le fait que le processus de rédaction implique de multiples étapes de modification suggère que le choix des insertions est modulé par la dynamique entre les structures sous-jacentes de la syntaxe et la prosodie. Lors des étapes de transformation du texte ces structures peuvent ressortir ou être plus évidentes alors qu'à l'état final, d'un texte il est plus difficile de retrouver les chevauchements entre ces structures car ils sont absorbés par la fluence. La progression du document vers son état final peut sans doute être caractérisée par un développement stratifié.
La décomposition des processus peut donc se faire ainsi :

- une première phrase spontanée ou quasi-spontanée qui correspond au premier jet et qui met en évidence les piétinements d'expression et les recherches de mots produits par la coordination du processus cognitif chez l'écrivain.
- des phases successives de relecture et de modification par l'auteur.
- des phases successives de relecture et de modification par ses relecteurs.
- des phases de modification par l'auteur suite aux relectures antérieures.

Le manuscrit devient le lieu d'évènements de relectures multiples, un espace qui permet des reformulations, des paraphrases et sans doute de multiples hésitations. Ainsi se construit un discours alimenté par de multiples acteurs qui agissent successivement pour transformer la page en lieu de collaboration et d'interaction langagière.

2.5.4 Le fondement commun de l'énonciation à l'écrit et à l'oral – la pensée

La relation entre l'écrit et l'oral est donc que la source commune est pensée ou la cognition. Les modes de l'oral et de l'écrit sont simplement deux moyens de transmettre la pensée. L'oral se réalise acoustiquement et l'écrit graphiquement.

L'étude des manuscrits permet d'observer les disfluences scripturales aux différents niveaux. Ceci permet de décomposer le processus afin de l'étudier :

- au niveau de structures de phrase
- au niveau des unités lexicales
- au niveau des composants des unités lexicales (les morphèmes-phonèmes).

Nous tenons à dire donc que les étapes entre l'encodage et le décodage de l'énoncé entrent pleinement en jeu même au premier aborde de la transmission de la pensée vers l'écrit. De plus, ils peuvent être décrites en termes de ces structures, unités et composants. Nous pouvons ensuite, à travers leur étude, remonter aux processus cognitifs. Enfin, à travers les opérations de coordination de la pensée, nous accédons à une meilleure compréhension de l'énonciation.

Ce que nous pouvons apprendre de l'étude des manuscrits concernera plus précisement :

- les pratiques du scripteur ou des scripteurs (dans notre cas, Stendhal).

- les opérations d'expansion ou de réduction du texte

- la coordination de la pensée vis-à-vis l'énonciation (nous ne faisons pas dans ce cas la distinction entre l'oral et l'écrit car la cognition précède les deux modes.

2.6 La caractérisation des disfluences scripturales– définition des critères formels

« Languages make use of different tools, but do so within the framework of a common process of meaning construction, much of which has yet to be elucidated »
[Robert, 1999, p. 22].

Dans l'objectif de caractériser les données que nous avons extraites de la base des *manuscrits_stendhal* nous avons adapté un système d'annotation que nous décrivons plus tard (cf. 2.7 *Système d'annotation*). Dans cette adaptation nous nous fondons sur le point de vue catégoriel qui se trouve non seulement en linguistique mais dans les sciences en général. Définir des critères formels nous permettra de relever les caractéristiques des phénomènes traités. Nous allons détailler le raisonnement, spécificités et les influences du système dans les sections suivants. Plus particulièrement, nous expliquerons les influences et aussi nos choix d'adapter le système d'annotation en se fondant sur des théories de la linguistique formelle et des méthodes informatiques actuelles en traitement automatique des langues.

L'approche que nous adoptons pour caractériser les disfluences scripturales du corpus des manuscrits repose sur ces principales notions :

- la notion de *chunk* introduite par Abney [Abney, 1994]

- la notion du *segment* de T. A. Lebarbé [Lebarbé, 2002].

- la catégorie grammaticale

- le positionnement de l'unité par rapport au chunk

- la repérage de la répétition de l'unité

- le critère de temporalité de la modification de l'unité

Les deux approches de la segmentation de phrases, le *chunk* et le *segment,* nous ont apporté une méthode de délimitation du texte que nous étudions, de situation de morceaux d'interruption et donc de description formelle de la variation et la variabilité du phénomène de cet étude.

Avant d'expliciter le système d'annotation que nous proposons pour ces éléments nous définissons brièvement les notions de *chunk* ainsi que du *segment* par rapport à la phrase.

2.6.1 Le *chunk*

La notion de *chunk* introduit par Abney (1994) permet de segmenter des phrases en unités dans un traitement de surface du texte. Les unités sont découpées en suivant des principes à la fois proprement grammaticaux mais aussi ceux qu'Abney identifie comme étant prosodiques. Ainsi, délimiter une phrase en chunks permet plus facilement de relever des structures grammaticales récurrentes. En principe, le chunk permet aussi de relever des motifs prosodiques.

Selon Abney la segmentation de la phrase en *chunks* est ancrée dans un processus naturel de lecture. Il décrit le processus de *clustering*[7] d'unités lexicales pour expliquer comment ces unités se regroupent autour des unités lexicales qui constituent la *tête*[8] du groupe [Abney, 1994]. Les unités de tête des chunks entretiennent des relations hierarchiques à l'intérieur de la phrase.

Les segmentations en chunks que nous utilisons pour découper et situer les unités des disfluences suivent donc les règles suivantes :

- *Des chunks nominaux :* l'ouverture d'un chunk nominal est souvent marquée par la présence d'un déterminant. Le chunk lui même contient un substantif ou un nom propre et peux être précédé ou suivi d'un adjectif.

- *Des chunks pronominaux :* l'ouverture d'un chunk pronominal est marqué par un pronom suivi d'un verbe. Dans les cas où le verbe est constitué d'un verbe auxiliaire suivi d'un verbe conjugué ou infinitif le verbe auxiliaire est intégré dans le chunk pronominal et le deuxième verbe

[7] Venant de l'anglais et utilisé pour parler des regroupements des unités
[8] traduction de l'anglais du terme *head* utilisé par Abney.

constitue un chunk à part. Sinon, le verbe qui s'intègre dans le chunk pronominal peut aussi intégrer des adverbes.

- *Des chunks prépositionnels :* l'ouverture d'un chunk prépositionnel est marqué par une préposition, laquelle peut être suivie soit de verbe –conjugué ou non – soit d'un motif du *chunk nominal* qui est parfaitement assimilé dans le *chunk prépositionnel.*

- *Des chunks verbaux :* un verbe conjugué ou à l'infinitif peut apparaître en tant qu'un chunk indépendant s'il est précédé d'un chunk complet, tel qu'un *chunk pronominal*, et suivi d'un chunk complet d'un autre type. Un chunk verbal peut être constitué de plusieurs verbes. Des déterminants démonstratifs après un chunk verbal et avant un chunk nominal ou pronominal s'intègrent dans le chunk verbal précédent.

- *Des chunks de conjonction :* des conjonctions telles que « *que, et, ou, comme* » etcetera constituent aussi leur propres chunks. Ceci permet de plus facilement repérer ces jonctures de phrases. Il est aussi intéressant d'observer la distribution de la disfluence écrite par rapport à ces unités.

- *Des chunks adjectivaux et des chunks adverbiaux :* généralement, le chunk nominal absorbe une séquence adjectivale. Pourtant, dans certaines conditions, comme dans la présence d'une conjonction, un adjectif ou adverbe peut être son propre chunk.

2.6.2 Exemples de chunk

Nous donnons ici quelques exemples pour enlever les ambiguïtés lors de notre démarche de segmentation en chunk.

Les chunks nominaux :

IDPage : 738 R. 302 Rés., volume 1, tome 3, feuillet 399, verso	[où] [la pe**é**tite Phélipe] [y sera][.]

Figure 1 : Exemple de phrase avec un chunk nominal qui intègre un adjectif.

Les chunks pronominaux :

IDPage : 385 R. 5896 Rés., volume 25, feuillet 120, verso	[au 53.e `mile`] [nous **avons** sommes] [descendus] [dans une valée] [assez agréable] [et] [surtout trés fertile][.]

Figure 2 : Exemple de phrase avec un chunk pronominal constitué d'un pronom et un verbe auxiliaire. Le chunk pronominal est suivi d'un chunk verbal.

Les chunks prépositionnels :

IDPage : 124 R. 5896 Rés., volume 22, feuillet 72, verso	[La Lecture] [des Mémoires] [**de la vie**] [de Marmontel]
IDPage : 2662 R. 5896 Rés., volume 26, feuillet 129, recto	[**chez** dinant] [chez M.lle R.] [il ne parla] [que] [de la mauvaise qualité] [du vin] [de Bor][.]

Figure 3 : Exemple de phrase avec trois chunks constitués autour la préposition *de* (en haut). Le deuxième exemple démontre l'intégration d'un chunk nominal qui porte aussi un adjectif (en bas).

Les chunks verbaux :

IDPage : 736 R. 302 Rés., volume 1, tome 3, feuillet 398, verso	[Mante] [et] [moi] [nous nous sommes] [allé promener] [**aux** à la Terrasse] [des feuillans] [de 2 h.] [à 4 ¼][.]

Figure 4 : Exemple de phrase où plusieurs verbes constituent un chunk verbal après un chunk pronominal avec pronom et verbe auxiliaire.

Les chunks de conjonction :

IDPage : 697 R. 302 Rés., volume 1, tome 3, feuillet 380, verso	[et] [qu'][il ne l'a pas] [eue][.]

Figure 5 : Exemple de phrase avec deux types de conjonctions enchainées : *et, qu' (l'élision de que).*

Les chunks adjectivaux :

IDPage : 2340 R. 302 Rés., volume 1, tome 3, feuillet 406, verso	[et] [dit ces paroles] [et] [d'autres] [du même sens] [avec les intonations les plus **vraies** *naturelles*] [et] [les plus larges][.]

Figure 6 : Exemple de phrase avec un chunk adjectival qui se produit à cause d'une conjonction.

Les chunks adverbiaux :

IDPage : 3618 R. 9975 Rés., volume , feuillet	[uniquement] [par l'influence] [du général] [dont] [il devait] [la protection **uniquement**] [à

15, verso	M.me de Berulle][.]

Figure 7 : Exemple du cas rare d'un chunk adverbial composé de l'adverbe *uniquement* et le cas plus commun de l'intégration du même adverbe dans un chunk nominal.

2.6.3 Le *segment*

Nous nous appuyions aussi sur la notion de *segment* présentée dans les travaux doctoraux de T. A. Lebarbé, qui développait la notion d'une unité de traitement intermédiaire entre la phrase et le *chunk*. La définition du segment nous a servi pour ce projet de caractérisation linguistique puisque nous nous avons appliqué l'approche de segmentation. La définition et délimitation du segment à l'intérieur de la phrase repose sur le repérage de marqueurs récurrents tels les conjonctions de subordination et de coordination, les pronoms relatifs et la ponctuation [Lebarbé, 2002]. Selon la définition, le segment peut être constitué de un à plusieurs chunks consécutifs. Cette définition nous a permis de procéder à une analyse du contexte des unités biffées, c'est-à-dire des points ou séquences d'interruption du texte. Repérer des segments à l'intérieur des phrases nous a permis de :

- délimiter la partie de la phrase concernée par la modification
- situer les parties biffées et modifiées par rapport aux chunks inclusives du segment
- récupérer suffisamment de contexte pour repérer des répétitions si ils se présentent dans l'énoncé.

Notre segmentation manuelle en chunks et segments du contexte des disfluences que nous étudions du corpus des *Journaux* est surtout illustrative. Il est possible d'avoir le segment s'arrêter sur un chunk prépositionnel. Pourtant, dans le souci de description de l'échantillon nous tenons de commencer un segment juste après une ponctuation terminal ou non-terminal ou juste après une conjonction. Nous arrêtons le segment juste avant une ponctuation terminal ou non-terminal ou juste avant une conjonction.

2.6.4 Les disfluence par rapport au segment

La disfluence peut constituer une unité, plusieurs unités, un chunk entier ou même s'étendre sur plusieurs chunks. Cette diversité de réalisation peut aussi être mis en relation avec

l'unité de segment. Ceci est particulièrement vrai lorsque la disfluence se produit sur la borne d'un chunk qui se trouve aussi sur la borne d'un segment.

Dans ces cas spécifiques le segment devient aussi utile pour la description des relations syntaxiques et prosodiques des unités de disfluence par rapport au segment, à la phrase et même par rapport au paragraphe. Ainsi nous constatons que les relations que les unités de disfluences entretiennent avec le contexte des autres unités et structures sont de l'ordre hiérarchique. Il devient donc évident que la compréhension et description des processus de reformulation au-delà de la caractérisation nécessitera une description plus fine de ces relations hiérarchiques.

Nous présentons les phrases segmentées et comportant des chunk*s* du corpus étudiés dans l' *Annexe A*.

2.7 Le système d'annotation

Notre caractérisation des disfluences scripturales demande de définir des critères ou paramètres formels de description. Nous pouvons ensuite nous appuyer sur ces critères dans l'observation non seulement de la variation mais aussi des similarités des productions de ce corpus. Nous proposons donc d'observer le corpus en nous fondant sur les critères formels suivants :

- Un système existant et relativement indiscutable de catégories grammaticales de Part-of-speech (POS) qui permet d'identifier et repérer des unités lexicales discrètes.
- Un système de classification gradué entre le *token* et le *chunk* qui permet d'identifier le type de disfluence du point de vue syntaxique et éventuellement du point de vue prosodique aussi.
- Un système de classification qui permet de qualifier le système de *token/chunk* en précisant si une disfluence porte sur tout le chunk, une partie du chunk, plusieurs chunks, ou sinon à quelle position est-ce que la disfluence se situe dans le chunk identifié.
- Un système de classification qui permet de repérer des éléments répétés s'ils se produisent dans la phrase et dans quelle position par rapport au chunk.
- Un classement de *niveau* ou de *strate* de la modification que nous établissons en étudiant le contexte XML immédiat de chaque unité de disfluence. Ce critère est celui qui nous permet de décrire à quelle strate de réécriture appartient une disfluence et éventuellement combien de strates de réécriture et de modification y a t-il sur une page donnée.

L'ensemble de ces systèmes nous aidera d'identifier les aspects formels qui caractérisent les objets de notre étude. Nous expliquons en plus de détails les critères de chaque système ou rubrique dans les sections qui suivent. Nous présentons un diagramme des relations des critères dans la Figure

2.7.1 Part-of-speech (POS) Tags

Nous reprenons certaines catégories grammaticales décrites par le Part-of-speech tagger du *French Treebank*. Nous fonder sur ce système existant permet de nous repérer dans le travail d'annotation des catégories grammaticales des unités biffées. Nous expliquons brièvement les quelques adaptations que nous faisons pour notre traitement manuel du corpus.

Le système de *POS tags* pour le Français que nous avons consulté pour le traitement de notre cas existe sous deux variants :

- Le French Treebank POS Tagset
- Le CC Tagset

Le POS Tagset du *French Treebank* ou *FTB* est utilisé par le Stanford POS Tagger, Treetagger et Morfette [SW1]. Il identifie quatorze (14) catégories et trente-quatre (34) sous-catégories de la langue française. Sa variante *CC Tagset* est une version optimisé du FTB avec autant d'applications—notamment *LGTagger* et *MElt*—et quelques différences à noter que nous citons par la suite.

- Il y a entre les deux *Tagsets* des différences qui concernent les noms d'étiquettes.
- Le *CC Tagset* a généralement moins de sous-catégories pour les catégories de verbes, de pronoms et autres.
- Le *CC Tagset* a une étiquette ET pour les mots étrangers.
- Le *CC Tagset* a un tag pour des formes interrogatives des adjectives, des adverbes, des pronoms et des déterminants.

Nous adaptons les catégories existantes pour servir notre analyse linguistique et adaptons également les tags au niveau superficiel. Nous utilisons v : à la place de VER par exemple [SW1].

Les choix de nos tags visent à refléter les POS que nous observons dans les biffes (ratures) en faisant clairement les distinctions dans les cas des modes et aspects de verbe, sans rajouter trop de sous-catégories. Nous rajoutons pourtant la sous-catégorie :amorce pour les cas de verbe (v) et de substantifs (SUBS) qui se présentent toujours incomplets sur la page manuscrite. Nous rencontrons ces cas assez fréquemment dans l'échantillon du corpus traité pour les prendre en compte dans l'annotation POS.

En principe, un schéma d'annotation grammatical est représentatif des catégories existantes de la langue représentée. Ceci implique la capabilité du système d'être adapté aux données linguistiques traitées. Notre travail d'annotation POS se fonde principalement sur les catégories grammaticales existantes et nous adaptons un système d'étiquettes qui reflètent majoritairement les *tagsets* adaptés du French Treebank. Lorsque nous sommes toujours confrontés aux catégories grammaticales du français ce système tient toujours. Les différences notoires que nous adoptons concernent principalement les noms d'étiquettes. Nous sommes amenés aussi à rajouter quelques catégories qui ne sont pas membres des *tagsets* du French Treebank, mais qui reflètent pourtant nos données. Le principe d'annotation reste toujours le même et il est fondamental à cette partie du projet.

La table suivante présente notre adaptation des *tagsets* à partir des catégories POS déjà existants pour le français et elle contient deux catégories que nous ajoutons pour le cas spécifique du corpus. Nous avons quatorze (14) catégories et vingt-cinq (25) sous-catégories.

TAG	CATEGORIE
ADJ	Adjectif
ADJ : amorce	Adjectif amorcé (ou incomplet)
ADV	Adverbe
ADV : neg	Adverbe de négation (ne, pas)
ADV : amorce	Adverbe amorcé (ou incomplet)
DET	Déterminant
DET : art	Déterminant article
DET : pos	Déterminant pronom possessif
DET : dem	Déterminant démonstratif
CON	Conjonction

CON : cc	Conjonction de coordination (mais, or, et, ou, car, donc, ni, etc.)
CON : sub	Conjonction de subordination (que, quand, lorsque, comme, si, puisque, quoique, etc.)
SUBS	Substantif (nom commun)
SUBS : amorce	Substantif amorcé (ou incomplet)
NPP	Nom propre
PRON	Pronom
PRON : dem	Pronom démonstratif (ce, celui, celle, ceux)
PRON : ind	Pronom indéfini (autre, certaine, chacun, etc.)
PRON : pers	Pronom personnel
PRON : pos	Pronom possessif (mien, tien, etc.)
PRON : rel	Pronom relatif (laquelle, que (COD), qu', où, dont).
PRON : dir	Pronom direct (me, te, lui, leur)
PREP	Préposition
PREP :det	Préposition + déterminant amalgame
VERBE	Verbe
v : fs	Verbe futur simple
v : ind	Verbe indicatif
v : cond	Verbe conditionnel
v : imp	Verbe impératif
v : inf	Verbe infinitif
v : ps	Verbe passé simple
v : sub	Verbe forme subjonctif
v : pc	Verbe passé composé
v : amorce	Verbe amorcé ou incomplet
PONCT	Ponctuation
SYM	Symbol : Nous comptons parmi les symboles : les caractères grecques ou autres que Stendhal utilise s'ils ne font pas partie d'une phrase traduite ou transcrite de la langue.
UKN	Les cas auxquels il n'est pas possible

	actuellement d'attribuer une catégorie.
interjection	Les exclamations dans le texte telles que : « Ah », etcetera.
ET	Les mots en langue étrangère

Figure 8 : Table des annotations des *parts of speech* (POS) adaptées dans le cadre de l'étude

2.7.2 Table de la catégorie ISTOK /ISCHUNK[9] (annotation du statut ISTOK/ISCHUNK de la rature)

La propriété ISTOK/ISCHUNK permet de regrouper les unités de disfluence en fonction de cinq grandes catégories que nous explicitons brièvement dans la table suivante :

ISTOK/ ISCHUNK	DEFINITION
ISTOK	La rature est focalisée sur un seul mot à l'intérieur d'un chunk.
ISTOK+	La rature est focalisée sur plus qu'un mot à l'intérieur d'un chunk.
ISTOK-	La rature est focalisée sur un ou plusieurs caractères.
ISCHUNK	La rature est un chunk complet.
ISCHUNK+	La rature est plus qu'un chunk ou s'étend sur plusieurs chunks.

Figure 9 : Table explicative de la catégorie ISTOK/ISCHUNK

2.7.3 Attribution des éléments répétés du segment contenant la rature

Le Penn Treebank utilise un système d'annotation pour distinguer entre les bornes (B) et l'intérieur d'un chunk (I) [SW3]. Nous reprenons le système d'annotation qui permet de décrire la relation du mot par rapport au chunk pour un mot répété. Toutefois, nous tenons à notre définition de ce qui peut constituer un chunk, (cf. Section 2.5.1), que nous fondons sur celle d'Abney [Abney, 1994]. La table suivante décrit l'unité répété en termes de sa position par rapport à un chunk.

ELEMENT REPETE	DEFINITION
0	Le segment ne contient pas de mot répété[10]

[9] Cf. a définition de chunk que nous donnons et sur lequel nous fondons le chunking que nous faisons manuellement pour l'analyse de l'échantillon du corpus.

\RB	Le segment contient un mot répété sur la borne d'un chunk.
\RI	Le segment contient un mot répété à l'intérieur d'un chunk.

Figure 10 : Table explicative de l'annotation des éléments répétés

2.7.4 Niveau de modification – la stratification du document

Le contexte du paragraphe (Contexte De Paragraphe) et celui de la ligne (Contexte Droit et Contexte Gauche) permet d'identifier l'étape de modification selon le principe de stratification décrite par [Biasi, 2013] et qui est de l'intérêt aux généticiens de texte et aux éditeurs. Nous décrivons cette stratification par le paramètre *Niveau,* que nous expliquons aussi dans la table suivante :

NIVEAU	DEFINITION
1	La rature et modification se produite dans le flux d'écriture
2	La rature et modification se produit à une étape ultérieure de relecture. Dans le contexte immédiat la biffe (rature) est accompagnée d'ajout, interligne, ou variante en interligne. Ces derniers peuvent être soit de la main de Stendhal lui même soit d'un de ses nombreux relecteurs.
3*	La rature et modification se produit sur une étape encore successive. *Le niveau successif de relecture peut s'incrémenter. Il est propre au document concerné et la *profondeur* [Biasi, 2013] des strates peut se varier en fonction de nombre de relectures auxquelles un texte a été soumis.

Figure 11 : Table explicative des niveaux de modification du document

[10] Nous supposons qu'un syntagme contient plusieurs chunks. Notre observation du corpus nous permet donc de repérer des mots répétés dans le syntagme et non pas dans le chunk de la rature même.

2.7.5 Table descriptive de positionnement de la rature par rapport au chunk

Ce paramètre permet de situer l'unité de disfluence par rapport au chunk ou chunks que nous identifions à l'intérieur des segments :

POSITION DANS LE CHUNK	DEFINITION
Début	La rature est un ou plusieurs mots en début d'un chunk.
Fin	La rature est un ou plusieurs mots en fin d'un chunk.
Mil	La rature est un ou plusieurs mots au milieu d'un chunk.
Tout	La rature est un chunk complet.

Figure 12 : Table descriptive de positionnement de la rature par rapport au chunk

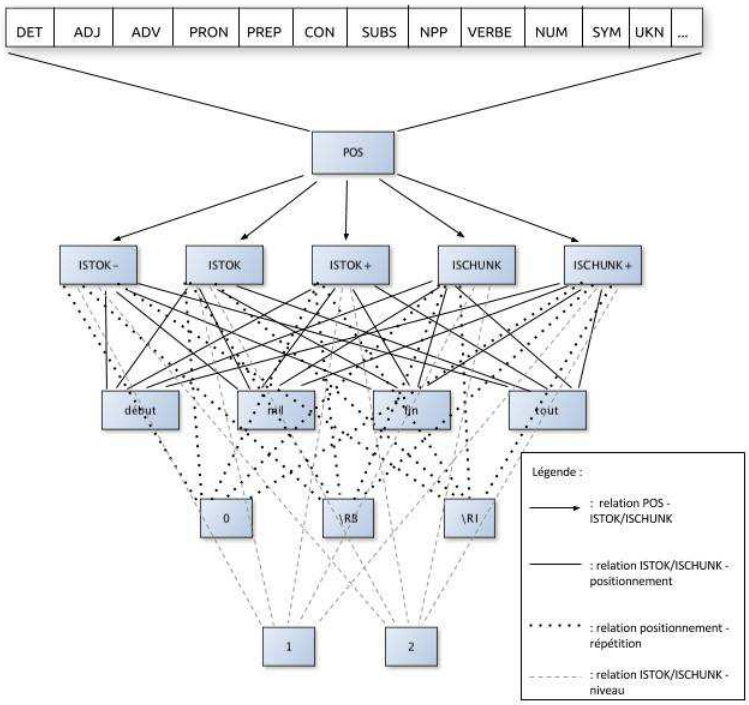


Figure 13 : Représentation graphique des relations entre les critères que nous définissons.

Les notions théoriques de la linguistique que nous avons présenté dans cette partie du mémoire sont intégrales à l'étude linguistique de l'énonciation. Elles sont aussi fondamentales à l'étude des disfluences scripturales qui sont intégrales aux processus de l'énonciation à l'écrit. Les critères formels que nous avons détaillés ont permis de procéder à l'étude de ce phénomène de l'écrit à l'aide des outils informatiques dont nous parlerons dans la partie suivante.

3 Partie informatique

Dans cette section du mémoire nous décrivons les méthodes informatiques que nous utilisons pour le traitement de ce corpus. Nous parlons aussi des méthodes développées en TAL pour le traitement des disfluences de l'oral, des méthodes pour le traitement de l'écrit. Nous expliquons pourquoi ces méthodes ne sont pas suffisantes pour le traitement de notre corpus. Ensuite, nous expliquons les méthodes que nous utilisons. Ces méthodes brièvement :

- de description de données : via un corpus structuré en XML,
- de gestion de données : via une base SQL,
- de requête dans les données : via le langage PHP

3 Partie Informatique

3.1 Méthodes de traitement automatique des disfluences de l'oral

La plupart des travaux de modélisation des disfluences relèvent du traitement de l'oral, où les corpus oraux servent à entraîner des systèmes de reconnaissance ou d'analyse de la parole. En traitement de l'oral, les méthodes qui permettent de modéliser efficacement des séquences de mots qui constituent ou qui contiennent les disfluences. Une méthode est d'élaborer des règles à partir des données linguistiques d'un corpus oral. Quelques exemples que nous citons ici relèvent des travaux d'André Valli et Jean Véronis [Valli et Véronis, 1999], ainsi que de Mathieu Constant et Anne Dister [Dister et Constant, 2012].

Jean Véronis et André Valli ont étudié le fonctionnement d'un étiqueteur morpho-syntaxique, celui de *Cordial,* sur le corpus oral de G.A.R.S. (*Groupe Aixois de Recherche en syntaxe – Université de Provence)*. La méthode consistait à créer des règles adaptées aux conventions de transcription du corpus et qui peuvent donc relever des disfluences transcrites dans le corpus. De façon similaire, le traitement du corpus du corpus *VALIBEL (Variété Linguistiques du Français en Belgique)* par les travaux de Constant et Dister [Constant et Dister, 2012] consistait d'appliquer des règles spécifiquement conçues pour les transcriptions du corpus.

D'autres méthodes élaborées pour le repérage automatique de disfluences de l'oral utilisent les méthodes d'analyse de surface par segmentation de l'énoncé en chunks, la description des relations entre eux et l'ajout des informations sémantiques ou pragmatiques [Goulian *et al.*, 2002]. Les algorithmes qui repèrent automatiquement les disfluences de l'oral recherchent des séquences de répétition ou des points d'interruption dans des énoncés et suppriment les séquences gênantes à la fluence [Bove, 2008]. Ces méthodes visent à pousser plus loin l'intelligence des systèmes élaborés.

Des méthodes appliquées en TAL permettent de nettoyer les énoncés des disfluences et rendre le texte fluent et cohérent. Pourtant, à partir des modélisations de ces systèmes et des corpus transcrits il est toujours possible d'observer la non-linéarité des énoncés. Ceci caractérise l'oral spontané, mais aussi l'écrit à l'état de brouillon.

3.2 Méthodes de traitement des disfluences de l'écrit – le cas de Treetagger

Treetagger est un outil qui permet d'annoter des textes linguistiquement. Cet outil est donc très présent dans le domaine de traitement des corpus écrits. De plus, il peut être utilisé pour plusieurs langues et a été utilisé pour traiter l'allemand, l'anglais, le français, l'italien, le néerlandais, l'espagnol, le bulgare, le russe, le portugais, le galicien, le chinois, le swahili, le slovak, le latin, l'estonien et des textes de l'ancien français [SW5].

Ce système de traitement de texte n'est pourtant pas actuellement adapté au traitement du corpus des manuscrits de Stendhal. La raison pour ceci est que Treetagger nettoie le texte des balises XML structurantes. Le problème avec l'approche est que le corpus XML du corpus des manuscrits a une structure sémantique spécifique pour la mise en forme du texte, les éléments typographiques et notamment les disfluences scripturales. Le traitement par *Treetagger*[11] n'est donc pas adapté pour le traitement des disfluences scripturales encodées dans le XML du corpus. De manière similaire au traitement de l'oral nous observons avec Treetagger le nettoyage du corpus des séquences disfluentes. Lorsque nous nous intéressons spécifiquement à ces séquences d'interruption et de modification ce traitement ne permet pas d'étudier le texte du point de vue évolutif ou permettant de prendre en compte les reformulations. Nous entreprenons donc une autre méthode pour le repérage et la caractérisation ces phénomènes spécifiques.

3.3 Le potentiel descriptif du langage XML

Le potentiel descriptif de la structure XML est directement provenant de la capacité d'imbrication et d'héritage d'éléments qui permet la structure arborescente [Meynard et Lebarbé, 2011]. Cette capacité permet aussi de créer un corpus dynamique est donc très adaptable à l'exploitation de données pour l'affichage sur le web [Faure *et al.,* 2009].

3.3.1 La structure XML

Les principaux éléments composants d'un document XML sont les suivants, selon l'explication donnée en introduction dans le manuel informatique écrit par Brochard [RM3] :

[11] Ce traitement a été effectué grâce au projet *Anatext* qui est une plateforme développée par l'université Stendhal. *Anatext* permet de constituer un corpus de textes littéraires pour leur étude linguistique et statistique.

- *l'entête* : la déclaration du document du type <?xml version="1.0" encoding="UTF-8" ?>
- *la racine :* l'élément racine de l'arbre qui contient l'ensemble d'éléments du document. Les éléments ont entre eux une relation hiérarchique.
- *l'arbre :* est l'ensemble de ces mêmes éléments organisés hiérarchiquement à partir de la *racine.*

Pour être exploitable par les différents environnements informatiques la structure XML de données doit conformer à des normes spécifiques définies par le W3C. Premièrement, un document XML doit être bien formé selon les principes décrits auparavant. C'est-à-dire, qu'il doit impérativement avoir la déclaration du document XML en *entête* ainsi qu'un élément *racine.* De plus, les éléments de l'arbre doivent être délimités correctement, sans l'imbrication de balises. Un document XML bien formé doit aussi être validé par un document DTD associé à ce premier. Le DTD contient une déclaration de tous les éléments et attributs qui sont permis dans la structure de l'arbre XML. Une déclaration du DTD apparaît également dans l'entête du document XML avec le nom du fichier DTD et l'extension .dtd . De plus, il faut préciser dans la déclaration le chemin vers le fichier DTD si celui-ci se trouve dans un autre répertoire que le fichier XML. La figure donne un exemple d'un document XML avec une déclaration de DTD pour référence.

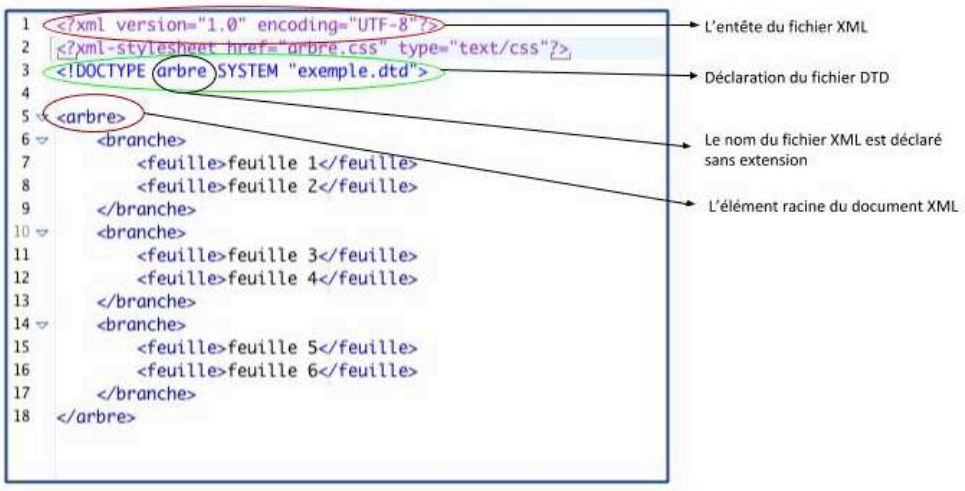


Figure 14 : Un exemple d'un document XML

Le fichier DTD pour le même document ressemblera à la figure suivante.

```
1
2
3  <!--     EXEMPLE DTD    -->
4  <!--      Mai 2014      -->
5  <!--      Arbre XML     -->
6
7  <!ELEMENT arbre (branche*)>
8
9  <!ELEMENT branche (feuille*)>
10     <!ELEMENT feuille (#PCDATA)>
```

Figure 15 : Exemple d'un fichier DTD associé au document XML de la figure 6.

3.3.2 De la page manuscrite à la page XML – la grammaire de description CLELIA

Les objectifs de représentation du manuscrit demandent un travail considérable de transcription pour convertir le contenu textuel autographe et allographe en texte numérique. Ce texte peut ensuite être interprété informatiquement. Dans un deuxième temps, puisque le projet de manuscrits de Stendhal vise à représenter les plus finement possible toutes les spécificités représentatives du document (cf. *Section 3.3.3*) une méthode pour convertir le corpus de pages numérisées a été conçue pour répondre à ces objectifs de description.

Ces besoins ont donc amené l'équipe *Manuscrits Stendhal* de créer une grammaire de description structurée en XML. Cette grammaire permet de décrire les spécificités de ce corpus par rapport à chaque individuelle page manuscrite. Le nom, CLELIA, est l'acronyme qui par coïncidence et par créativité linguistique correspond au nom du personnage Clélia du roman *La Chartreuse de Parme.* Les motivations pour créer CLELIA étaient nombreuses. Parmi elles, le besoin de décrire et d'adapter à l'informatique le corpus des manuscrits, comme nous venons de dire. Dans un deuxième temps, la norme a été motivée par la volonté d'adapter l'outil informatique à une équipe de chercheurs et contributeurs ayant des compétences informatiques de différents degrés. En général, les personnes venant de différentes disciplines n'ont pas les mêmes niveaux de compétences ni les mêmes pratiques en informatique. Cependant, tous ceux qui travaillent sur le projet ont tout de même besoin d'accéder à l'environnement de travail et au corpus de manuscrits [Meynard et Lerbarbé, 2011]. Il est important que les membres de l'équipe prennent conscience de la valeur sémantique des éléments de la structure XML sur lequel repose le travail de constitution et d'annotation du corpus. C'est ainsi qu'établir une grammaire de description avec la valeur sémantique des constituents clairement définies fait toujours partie des principes fondamentaux du projet. Une équipe largement francophone travaillant sur un auteur français est un facteur principal de l'encodage adapté à la langue française.

Nous pouvons représenter tous les aspects de la page avec la structure arborescente du XML et les éléments déclarés dans le .dtd de CLELIA [Lebarbé et Meynard, 2013]. La structure du XML reste adaptée aux objectifs de représentation des corpus écrits et oraux grâce à cette structure, mais aussi à la plasticité de description hiérarchique. Un schéma de la structure arborescente de CLELIA peut expliciter à la fois la structure XML et le potentiel de description d'un document manuscrit :

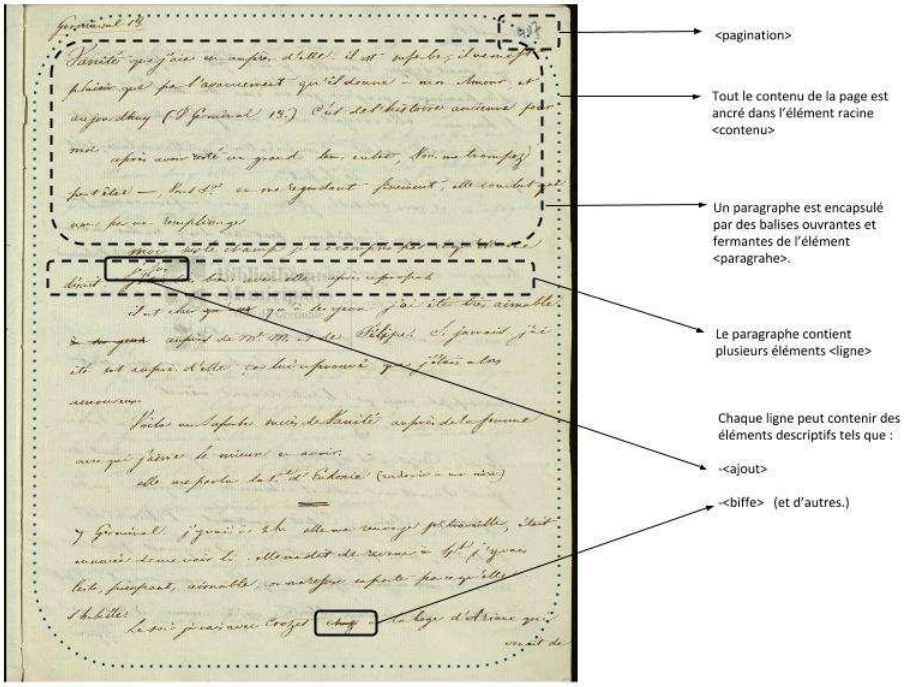


Figure : description de la page manuscrite vers une structure de description en XML. Exemple de page est : R. 302 Rés., volume 1, tome 3, feuillet 407, recto.

3.3.3 Les éléments XML pertinents à l'étude des disfluences scripturales

Parmi les éléments du XML qui permettent de représenter informatiquement la page manuscrite et la convertir en corpus dynamique [Faure *et al., 2013*] sont des éléments structuraux que nous utilisons pour nous repérer à l'intérieur du texte. Parmi les éléments structuraux nous comptons par exemple les <paragraphes/> et <lignes/>. Ensuite, à l'intérieur de la structure du texte nous procédons au repérage d'éléments qui permettent de représenter de façon détaillée les particularités scripturales de ces manuscrits. Ces éléments décrivent à la

fois des aspects typographiques, mais qui peuvent, selon le cas servir à l'analyse de la disfluence scripturale. Le manuel du projet *Manuscrits de Stendhal*[12] décrit en détail les éléments qui comprennent la structure XML et nous nous appuyons sur ce manuel ici pour la liste descriptive de ces éléments. Comprendre leur valeur sémantique et leur fonction à l'intérieur de la structure XML est essentielle pour ce projet linguistique.

- *Biffe* → le texte est biffé ou raturé par le scripteur.
- *Ajout* → le texte en ajout se distingue de l'interligne dans le sens où il est inséré dans la ligne d'écriture selon le manuel [SW4].
- *Interligne* → le texte est une modification insérée dans la ligne d'écriture [SW4].
- *Variante_en_interligne* → le texte est un choix que l'écrivain se laisse par rapport à une formulation ou un mot.
- *Surcharge* → une modification directement au dessus d'un mot déjà écrit
- *Bloc_en_interligne* → le texte en interligne de la page est un bloc entier.
- *Exposant* → le texte est une abréviation .
- *Marginal* → le texte est un ajout au texte mis dans en marge de la page.
- *Souligne* → le texte est souligné.
- *Code* → Le texte est un mot ou une expression utilisé pour un autre.
- *Traduction* → Le texte est dans une autre langue.
- *Blanc* → La place a été gardé par le scripteur pour remplir plus tard.
- *Contre_reclame* → une insertion dans le texte par le scripteur qui peut être une réaction ou une contradiction au texte.

Suivant les définitions de cette liste explicative d'éléments, des ambiguïtés peuvent se poser entre les éléments d'*ajout*, d'*interligne* et de *variante en interligne*. Ceci peut être le cas pour les transcripteurs aussi. La décision d'annoter un mot en *ajout* à la place de l'*interligne* peut constituer un choix interprétatif. Les transcriptions peuvent donc varier selon les transcripteurs. Pour les objectifs de notre projet les éléments *ajout, interligne* et *variante en interligne* constituent une modification à un moment ultérieur de la production initiale du

[12] Cf. la *Bibliographie*, dans la section *Sites-web et Manuels Informatiques* pour la référence complète et le lien vers le site du projet.

texte. Nous tenons compte du fait que nous ne pouvons pas dire avec précision à quel moment la modification est apportée.

3.3.4 Table explicative des attributs XML pertinents à l'étude

De plus qu'avoir les éléments pour décrire et délimiter [SW4] les différents aspects du script dans des manuscrits, la structure XML permet aussi d'apporter des informations supplémentaires à l'aide des attributs des éléments. Il y a différents types d'attributs pour les différents types d'éléments. Ils peuvent apporter de l'information concernant le scripteur, le type d'outil utilisé pour modifier le script, la date de modification si cette information est disponible et d'autres types d'informations qui servent à enrichir le texte. Pour reprendre un exemple d'un attribut qui peut être intéressant pour l'analyse des disfluences scripturales nous donnons en exemple les attributs associés à l'élément *biffe :*

- scripteur → donne le nom du scripteur s'il diffère du scripteur du texte [SW4].
- outil d'écriture → donne le nom de l'outil d'écriture (de biffe) s'il diffère de l'outil d'écriture du texte (ex: un texte à l'encre noire biffé au crayon)[SW4].

La liste complète de ces attributs se trouve également sur le site du manuel du projet[13].

3.4 La base de données

La base de données est un système de gestion de données qui permet de structurer et maintenir les données à l'aide d'une structure de tables relationnelles. Le potentiel pour la gestion de données langagières est clairement présent et très répandu dans le domaine de traitement de langues.

3.4.1 La base de données *manuscrits_stendhal*

Le projet des *Manuscrits de Stendhal* a parmi ses objectifs la représentation d'une collection littéraire et un corpus linguistique [Lebarbé, 2009]. Dans le contexte du projet les membres de l'équipe se confrontent très souvent à la richesse et à la complexité de cette

[13] Cf. *Bibliographie,* section *Sites-web,* référence bibliographique [SW4].

collection. En conséquence, ils adoptent les méthodes de la discipline scientifique dans laquelle se situe le projet. Dans le cas du projet *Manuscrits de Stendhal*, il était nécessaire aussi d'adopter des principes de la génétique textuelle afin de « rendre interprétables génétiquement des documents » à l'ensemble de la production [Biasi, 2013]. Nous énumérons ces principes :

1. organiser,
2. contextualiser,
3. étudier et interpréter.

La gestion de la collection avec une base de données SQL trouve une solution adaptée. La base de données SQL permet d'élaborer une approche de structuration appropriée au caractère des données ainsi que des besoins des chercheurs impliqués. Autrement dit, la base de données accorde une structure qui est toujours fondamentale à la représentation des connaissances. La problématique d'organisation des ensembles de données de même type ne change pas beaucoup entre disciplines qu'elles soient linguistiques, littéraires ou venant d'autres courants des sciences humaines. Pourtant, chaque collection, même à l'intérieur de la même discipline aura ses propres besoins de représentation et d'organisation. Comprendre ces besoins de description et d'organisation est essentiel pour leur représentation [Gray, *et al.,* 2013].

La base de données SQL fonctionne sous le principe que les entités documentées et répertoriées de la collection peuvent être regroupées grâce à des propriétés communes. Les identifiants, ou clés permettent de relier les tables de la base et regrouper des ensembles ayant un identifiant commun. Ce principe de regroupement par moyen de clé est un principe fondamental des bases de données SQL. De telle manière aussi peuvent être reliées des pages manuscrites d'un corpus ou domaine de notre corpus d'étude.

3.4.2 La base de données et le corpus dynamique

Pour comprendre la façon dont il est possible d'exploiter le corpus organisé dans une base de données il faut apprehender le rôle de la structure XML accordée aux documents. Le XML permet de représenter les différents aspects de mise en forme, de typographie, de description génétique ou linguistique. Ainsi, les pages manuscrites sont identifiées et peuvent être traitées en tant qu'objets informatiques [Faure *et al.,* 2009]. Ensuite, requêter la base de données des manuscrits permet de générer dynamiquement des objets textuels du corpus, de les trier, regrouper et afficher en fonction des besoins spécifiques du chercheur [Faure *et al.,* 2009].

3.5 Le langage PHP et l'affichage dynamique des données

Nous utilisons le langage PHP pour dialoguer avec la base *manuscrits_stendhal* et adapter les contenus de ce corpus à nos besoins. Notre méthode pour ce faire est relativement simple et suit les principes fondamentaux de la relation « client-serveur ». Nos scripts sont hébergés dans un répertoire sur le serveur qui contient aussi la base de données du corpus. Les relations entre les scripts associés sont déclarées dans les fichiers selon les conventions précises :

- soit dans l'entête HTML du fichier
- soit à l'intérieur du PHP avec la fonction *include()* ou *require()* selon les cas.

Un des avantages notoires du PHP est la facilité d'utilisation du langage [RM5]. De plus, il répond particulièrement bien aux besoins d'affiche dynamique du HTML et du XML. Enfin, il est également compatible avec d'autres langages Web tels que le Javascript [RM5].

Un schéma de la relation client-server, et l'ensemble des scripts PHP/MySql, Javascript et CSS, clarifiera la dynamique qui caractérise l'interaction de ces langages distincts de programmation :

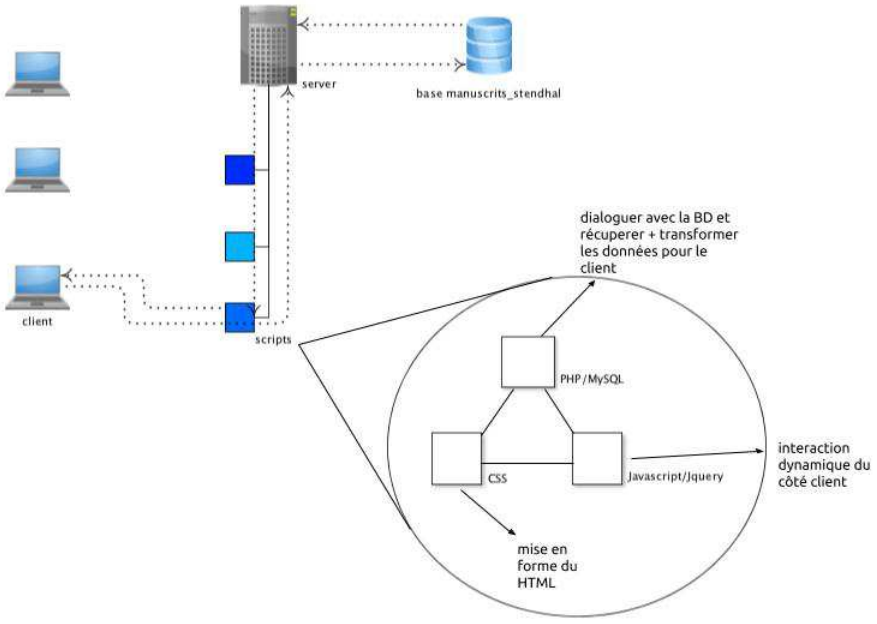


Figure 16 : Schéma explicatif de la relation client-serveur et les scripts PHP/MySQL, Javascript, CSS.

3.5.1 Le moteur de recherche[14]

Afficher les données, que nous récoltons de la base de données, dans le navigateur implique la réalisation d'un moteur de recherche simple. Ce moteur de recherche est essentiellement un formulaire HTML qui permet :

1. de sélectionner un phénomène d'énonciation écrite qui correspond à la fois à un élément défini de la structure XML de CLELIA et
2. de trier l'ensemble de la collection des manuscrits selon les corpus[15] décrits par les chercheurs sur Stendhal.

Il nous est possible donc de cibler la recherche du phénomène sur un corpus spécifique de la base de données. La réalisation de cet outil du point de vue informatique repose sur les

[14] Le moteur de recherche et l'outil d'observation se trouvent à cette adresse : http://stendhal.msh-alpes.fr/tests_anne/new/form.php
[15] Pour la liste complète des corpus cf. *Annexe C*

capacités des langages PHP et MySQL. Ils permettent de requêter la base de données selon les critères prédéfinis et présentés à l'utilisateur sous forme de menu. Ensuite, nous procédons à l'affichage des résultats de la sélection pour l'utilisateur. Comme l'indique la figure d'illustration de ce processus de sélection, l'élément de sélection et le corpus sélectionné agissent ensemble pour déterminer le résultat de la requête de l'utilisateur.

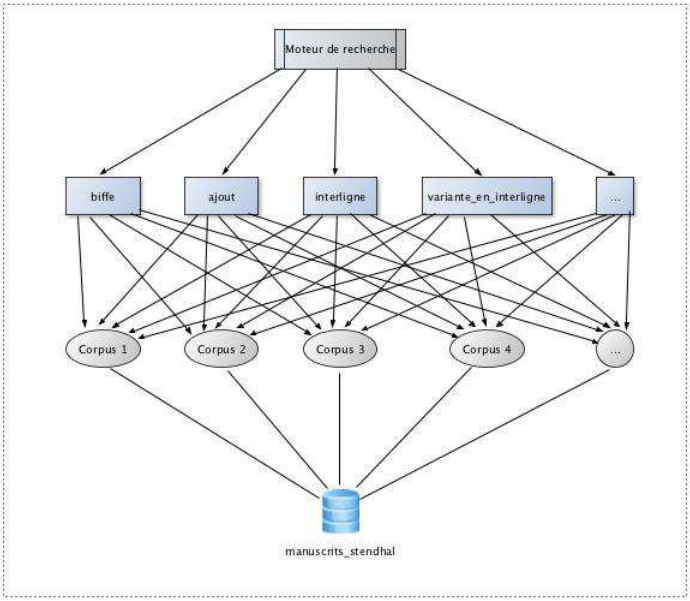


Figure 17 : Schéma du moteur de recherche développé pour le corpus des manuscrits de Stendhal

Du point de vue de l'évolution de l'outil, le menu de sélection d'éléments peut être ajusté pour correspondre aux besoins spécifiques de la recherche sur le corpus.

3.5.2 Extraction de données avec des requêtes SQL

Le langage PHP permet de se servir du SQL pour requêter des données contenues dans des bases de données. La base que nous avons à notre disposition est celle de *manuscrits_stendhal*. Nous avons deux moyens de faire des requêtes dans cette base :

- à travers l'interface de phpMyAdmin

- à travers des scripts écrits en PHP

Un exemple de requête SQL écrite dans un script PHP :

```
216
217 // requête SELECT dans la bd manuscrits_stendhal
218 $rq = "SELECT    CLELIA_Page.IDPage,
219                  CLELIA_Page.Contenu ,  //nous recherchons dans la table CLELIA_Page le champ Contenu
220                  CLELIA_CorpusDePage.IDCorpus ,
221                  CLELIA_Corpus.Corpus,
222                  CLELIA_Page.Cote,
223                  CLELIA_Page.Volume,
224                  CLELIA_Page.NumeroPage,
225                  CLELIA_Page.RectoVerso
226        FROM      CLELIA_Page,
227                  CLELIA_CorpusDePage ,
228                  CLELIA_Corpus
229        WHERE     CLELIA_Page.Valide
230        LIKE      \"O\"              //nous recherchons parmi les pages validées
231        and       CLELIA_Page.Contenu
232        LIKE      \"%$element%\"     //nous précisons l'élément de recherche
233        AND       CLELIA_CorpusDePage.IDPage = CLELIA_Page.IDPage
234        AND       CLELIA_Corpus.IDCorpus = CLELIA_CorpusDePage.IDCorpus
235        AND       CLELIA_Corpus.IDCorpus = \"$corpus\"";      //nous précisons le corpus
236
237
```

Figure 18 : Requête SQL qui permet de sélectionner les pages du corpus contenant un élément prédéfini par l'utilisateur à partir d'un formulaire HTML. Cet exemple est le cœur du fonctionnement du moteur de recherche que nous utilisons.

Un exemple de requête SQL à travers l'interface *PhpMyAdmin* :

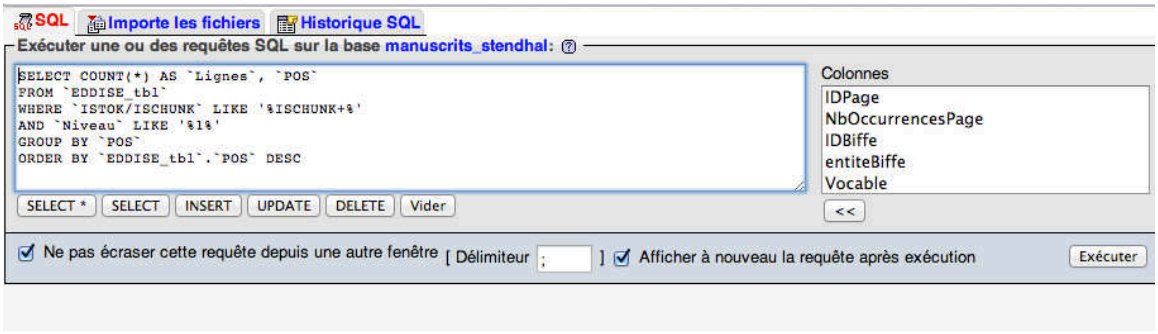


Figure 19 : Requête SQL qui permet de trier nos annotations selon le critère ISTOK/ISCHUNK et le 1e niveau de modification. Cette requête renvoie les données regroupées par catégorie grammaticale (cf. Annexe B, Figure 2).

3.5.3 Traitement de données XML avec des expressions régulières

A l'aide des expressions régulières le langage PHP permet de rechercher les balises XML qui ont un intérêt pour nous. Il existe des fonctions PHP spécifiquement conçues à ces fins. Les fonctions *preg_match(), preg_match_all()* et *str_replace()* permettent de rechercher des séquences spécifiques dans des chaînes de caractères. Dans chaque cas, nous définissons dans les paramètres de la fonction 1) le motif recherché, 2) le contexte de la recherche, 3) la variable qui récupère les résultats, 4) l'ordonnancement des résultats (optionnel). Avec les résultats obtenus de la fonction *preg_match()* ou *preg_match_all()* il est possible ensuite de manipuler les résultats.

Nous pouvons utiliser la fonction *str_replace()* pour réinsérer le motif dans un autre contexte ou dans le même en mettant en valeur ce motif avec le CSS. Nous donnons en exemple les trois types de fonctions que nous avons utilisé :

```
$pat1 = "/<$element(.*?)<\/$element>/s";
preg_match_all($pat1, $lg->Contenu, $lesbiffes, PREG_SET_ORDER);
```

Figure 20 : Exemple de l'utilisation de la fonction preg_match_all()

```
preg_match($pat2, $leparagraphe, $lecontexte) ;
```

Figure 21 : Exemple de l'utilisation de la fonction preg_match()

```
$pat2 = "/(.*?)".str_replace("/", "\/",$labiffe[0])."(.*?)<\/paragraphe>/s";
```

Figure 22 : Exemple de l'utilisation de la fonction str_replace()

Avec l'exemple de la Figure 19 nous définissons le motif avec la variable $pat1, recherchons dans le contexte défini par $lg->Contenu et récupérons les résultats sous forme de tableau dans la variable $lesbiffes. Figure 20 donne une variante de la fonction avec le motif défini par $pat2, la recherche dans $leparagraphe et les résultats de la recherche dans $lecontexte.

La fonction *str_replace()* (Figure 21) permet de retrouver de nouveau chaque motif du tableau rendu $lesbiffes, défini par la variable $labiffe[0], et de réinsérer le motif dans le texte (toujours entouré de son contexte XML).

Dans cette section nous avons décrit les méthodes informatiques qui nous ont permis d'interroger le corpus linguistique avec lequel nous travaillons. Ces méthodes nous ont servies dans la création de l'outil informatique d'observation du corpus. Elles sont fondamentales à la constitution de ce corpus et les objectifs des projets autour des manuscrits. Elles sont également essentielles pour le travail linguistique que nous menons. Les observations que nous présentons dans la section suivante relèvent directement de ces méthodes.

4 Partie linguistique- informatique

Nous avons utilisé des principes linguistiques et des méthodes informatiques pour réaliser l'étude que nous présentons dans ce mémoire. Ceci nous a permis de caractériser les disfluences scripturales du corpus *Journaux* des manuscrits de Stendhal. Dans cette troisième partie du mémoire nous présentons nos observations, relevées de requêtes dans une nouvelle table d'annotations que nous avons créé dans la base *manuscrits_stendhal.* La caractérisation linguistique des disfluences scripturales relève de ces observations du corpus *Journaux*. Nous expliquons d'abord la sémantique de l'encodage XML et comment nous l'utilisons pour interpréter les disfluences du manuscrit.

4 Partie linguistique - informatique

4.1 Interprétation linguistique de l'encodage informatique

Nous observons le corpus linguistique grâce à la structure XML de CLELIA que nous mentionnons auparavant. Le langage XML a une fonction décidément descriptive et essentielle pour la constitution et manipulation de ce corpus. En essence, les éléments XML constituants de la structure de chaque page de manuscrits permettent de déclarer et de décrire des éléments caractéristiques du texte de la page. La description est suffisamment fine pour permettre la représentation *pseudo-diplomatique*[16] du texte visé par l'équipe éditoriale. De plus, la description typographique du texte aux phénomènes proprement linguistiques que nous étudions dans le cadre de ce projet. Il est nécessaire de se repérer à l'intérieur de cette structure XML et de bien comprendre la valeur sémantique et descriptive de l'encodage des éléments XML ainsi que la valeur sémantique de leur combinaison.

4.1.1 Lecture du contexte XML

Chaque cas de disfluence étant transcrit en XML nous pouvons alors observer les différents cas qui se produisent dans le texte et de les distinguer :

- Les disfluences qui se produisent en cours ou dans le flux d'écriture.
- Les disfluences qui résultent d'une relecture et sont accompagnées d'une ou plusieurs modifications (ajout, interligne, variante en interligne).

Nous pouvons également distinguer entre les disfluences qui résultent des corrections par la main de l'auteur, des corrections et des remarques de ses nombreux relecteurs :

- Un *ajout*, une *interligne*, une *variante en interligne*, une *marginale* de la main de Stendhal
- Un *ajout*, une *interligne*, une *variante en interligne*, une *marginale* de la main autre que Stendhal

[16] La transcription pseudo-diplomatique respect la forme et les spécificités de la page et du script, mais ne vise pas de reproduire l'image de la page manuscrite.

Les étapes de relecture et commentaires peuvent être très complexes sur le plan de l'empilement ou de la stratification du document [Biasi, 2013]. Travailler sur des manuscrits et préparer les transcriptions permetten de se rendre compte de la complexité de la tâche de représenter les « mouvements d'écriture » et reformulations à l'étape de brouillon. Il y a une distinction importante faite dans le corpus entre les « biffes accompagnées d'ajouts» et les « biffes accompagnées d'interlignes ». Ceci permet de différencier les processus de reformulation. Le processus d'écriture chez Stendhal étant plutôt du cadre d'augmentation et de développement du texte, ses manuscrits présentent beaucoup plus souvent des ajouts que des suppressions, par exemple [Meynard, 2014]. Il est donc nécessaire et intéressant de distinguer les différents façons dont il insère les modifications dans la page.

Le contexte XML révèle les cooccurrences des disfluences avec des abréviations (*exposant),* des *traductions*, et d'autres phénomènes de productions transcrites dans la structure XML [Faure *et al.,* 2009]. Ces cas peuvent engendrer des études autour des productions de disfluences avec d'autres types de phénomènes scripturales. Nous notons ces possibilités, bien qu'elles ne fassent pas partie de notre projet actuel de recherche. De même, l'échantillon du corpus des *Journaux* que nous observons pour cette première caractérisation ne contient pas tous les cas de variation décrits auparavant. En effet, il y a peu de cas de relectures et donc de modifications autres que ceux de Stendhal. Nous énumérons cependant les différents cas que nous avons observé dans le corpus. Ainsi faisant une distinction entre ceux qui signalent une première ou une deuxième strate de modification, ou bien des strates successives, auxquelles nous attribuons un *niveau* :

1. *Biffe sans cooccurrence avec d'autres éléments sur la ligne* → Une partie biffée de la phrase en combinaison avec d'autres éléments semble au départ le plus simple des cas. Plus précisément, le contexte XML signale une modification qui a été fait dans le flux d'écriture et donc ce que peut bien être considérer d'appartenir au premier jet du scripteur. Nous attribuons le *niveau 1* à ce contexte.

2. *Biffe en cooccurrence avec un ajout* → Le contexte XML d'un élément <biffe> en cooccurrence d'un élément <ajout> signale une modification ultérieure du flux d'écriture. Un ajout peut signaler soit une relecture du document entier ou une partie du document. Nous ne pouvons pas dire avec précision le point d'arrêt du scripteur mais nous constatons cependant un retour sur la phrase. Dans ce cas nous sommes au *niveau 2* de modification. Le contexte XML annoté nous permet également de dire si l'ajout est de la part du scripteur lui même ou de la main d'un relecteur.

3. *Biffe en cooccurrence avec une interligne* → Similairement au contexte d'ajout une biffe en cooccurrence avec une interligne signale un processus de modification lors d'une relecture et donc un *niveau 2*.

4. *Biffe en cooccurrence avec une variante en interligne* → Une biffe en cooccurrence avec une variante en interligne signale une modification de *niveau 2*.

5. *Biffe en cooccurrence avec une surcharge* → Ce type de production est particulièrement intéressant pour l'analyse des manuscrits. Le contexte indique un mot réécrit directement sur un autre. De plus, les surcharges sont souvent plus difficiles à déchiffrer à cause de la juxtaposition d'écriture. Nous considérons que ces productions sont du *niveau 1* et donc qu'ils sont quasi-spontanées. Pourtant, elles présentent des cas intéressants de discussion : une réécriture sur un mot vient-elle plus probablement d'une réécriture lors du premier jet du manuscrit ou peut elle être le produit d'une relecture ? La première option présente un cas de vraisemblance aux hésitations qui peuvent être apparentées à celles de l'oral. La deuxième est un rappel que l'étude fine des manuscrits est toujours fondamentale pour pouvoir poser des hypothèses concrètes sur ce type de cooccurrence.

6. *Biffe en cooccurrence avec un exposant* → les éléments *exposant* signalent une abréviation adoptée par Stendhal afin de raccourcir certains mots (Mlle pour *Mademoiselle*, St. pour *Saint*, p.r. pour *pour*). Les biffes dans le contexte d'un de ces éléments sont aussi des disfluences qui se produisent lors de la transformation de la pensée sous forme écrite.

7. *Biffe en cooccurrence avec un élément code* → les éléments de code signalent une diversion de l'orthographe contemporaine du français des conventions spécifiques adoptées par l'écrivain ou des abréviations. Une étude qualitative ou quantitative de leur cooccurrence avec la biffe peut présenter des analyses intéressantes sur le processus énonciatif chez l'écrivain Stendhal, mais ne fais pas sujet de notre étude actuelle.

8. *Biffe en cooccurrence avec* une *traduction* → Comme le cas précédent de cooccurrence avec un élément *code* étudier les disfluences scripturales en cooccurrence avec des phrases traduites d'autres langues peut apporter encore un point de vu intéressant à l'étude des productions de l'écrivain Stendhal.

4.1.2 Le contexte de paragraphe et le contexte immédiat de la disfluence

L'unité qui sert de contexte dans notre corpus est le paragraphe. Il permet d'observer suffisamment le contexte des éléments analysés. Néanmoins, nous constatons en consultant des descriptions de modes de transcriptions utilisés dans les études sur l'oral que le paragraphe n'est pas la seule option de découpage [Bouvet et Morel, 2002]. Ceci peut être vu comme un choix interprétatif, plus ou moins fondé, bien évidemment en fonction des cas, mais un choix toute de même. Ainsi, l'unité de découpage peut être plus grand que le paragraphe – citons le cas de l'*hyperparagraphe*[17]. De même, les ratures dans les manuscrits peuvent s'étendre sur plusieurs paragraphes et il peut donc y avoir des ajustements sur cette unité de découpage selon les cas.

4.1.3 La stratification du texte du point de vue du contexte XML

La situation de l'unité de disfluence dans le contexte XML que nous avons décrit auparavant permet de décrire le processus énonciatif en termes de temporalité et ainsi de distinguer l'une modification produit spontanément de celle produit ultérieurement. C'est largement pour cette raison que nous avons annoté les données de *Niveau.* Ceci permet de transmettre la description stratifiée réalisée en XML grâce aux éléments <ajout/>, <interligne/>, <variante_en_interligne/> etc. Nous décrivons ces strates de façon incrémentale en utilisant :

- 1 pour le premier jet, l'écriture et modification dans le flux et donc sans cooccurrence d'ajouts, interlignes ou notes en marge,
- 2 en occurrence des éléments qui désignent les ajouts, interlignes, variantes en interligne, notes en marge, etc.
- 3* où à partir du contexte XML nous observons des étapes successives de modification.

4.2 Caractérisation linguistique des disfluences scripturales

Nos objectifs de caractérisation nous ont amené à définir un ensemble de critères sur lesquels nous nous fondons pour nos annotations du corpus *Journaux.* Ces critères portent sur des aspects formels des unités de disfluences et le contexte de phrases dans lesquelles ils apparaissent. Ils nous ont permis d'observer non seulement la variation dans les cas des disfluences scripturales du corpus de ces journaux mais aussi de repérer des similarités parmi la

[17] Cf. Bouvet et Morel (2002), *Le ballet et la musique de la parole.*

variation que nous observons, même à l'intérieur de ce petit échantillon de la production écrite de l'écrivain Stendhal. En nous fondant sur ces critères formels pouvons regrouper les données, ce qui nous permettra de mieux les comprendre. Nous pouvons donc les regrouper selon les groupes que nous avons définis, selon les catégories grammaticales, et selon la strate ou *niveau* de modification. Nous pouvons également observer quelles unités sont répétées et à quelles catégories grammaticales elles appartiennent. Les annotations pour chaque unité du corpus, avec la segmentation en *chunks* et en *segments* est aussi présenté dans *l'Annexe A*.

Nous avons transféré les annotations du corpus dans une nouvelle table que nous avons rajouté à la base *manuscrits_stendhal.* Ainsi, nous pouvons consulter les annotations pour chaque unité à partir de l'interface de l'outil d'observation. Nous pouvons également requêter ces informations directement à partir de l'interface *PhpMyAdmin.* Dans les sous-sections suivantes nous présentons nos observations à l'aide des graphiques qui permettent de mieux illustrer les ensembles que nous pouvons observer. Les ensembles qui permettent cette caractérisation peignent une image complexe et intéressante du phénomène.

4.2.1 Le regroupement selon le critère ISTOK/ISCHUNK

Voici ce que nous relevons de nos observations de la distribution du groupe ISTOK/ISCHUNK sur l'ensemble de 542 unités étudiées :

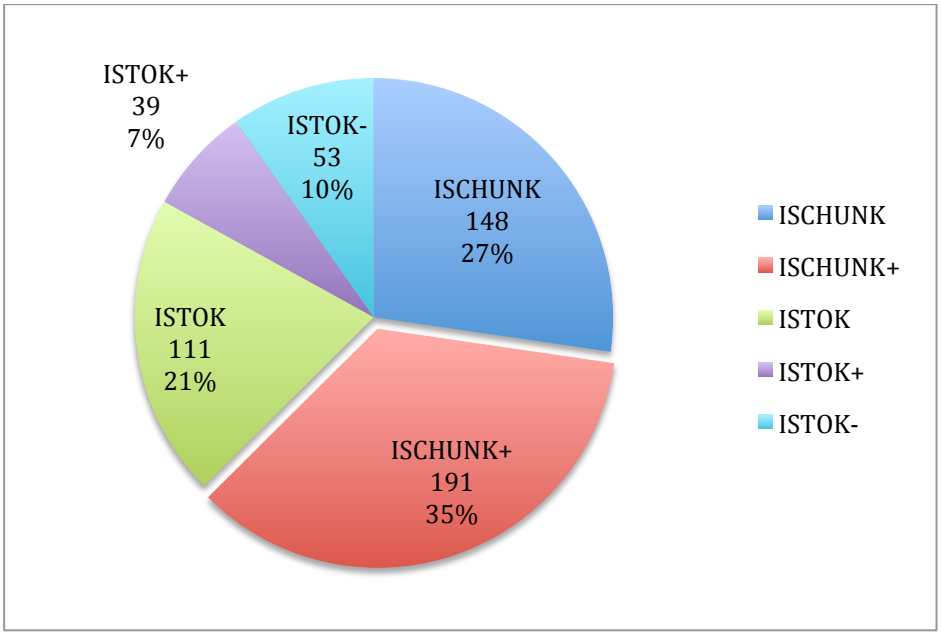


Figure 23 : Graph des regroupements des disfluences scripturales selon le critère ISTOK/ISCHUNK

Nous pouvons ensuite révéler pour chaque groupe la distribution quantitative de catégories grammaticales ou encore de montrer la distribution des éléments répétés des segments de réparation en fonction du groupe ou de la catégorie. Le potentiel de description est notable et les chiffres peuvent devenir beaucoup plus parlants si l'étude est menée sur tout le corpus des manuscrits.

4.2.2 La distribution des catégories grammaticales selon le critère ISTOK/ISCHUNK

Cette représentation d'un diagramme en bâton permet aussi de voir que cette distribution peut être décrite en termes de catégorie grammaticale pour chaque groupe ISTOK/ISCHUNK. Nous présentons ce graphe aussi dans l'*Annexe B* pour plus de facilité de consultation des détails.

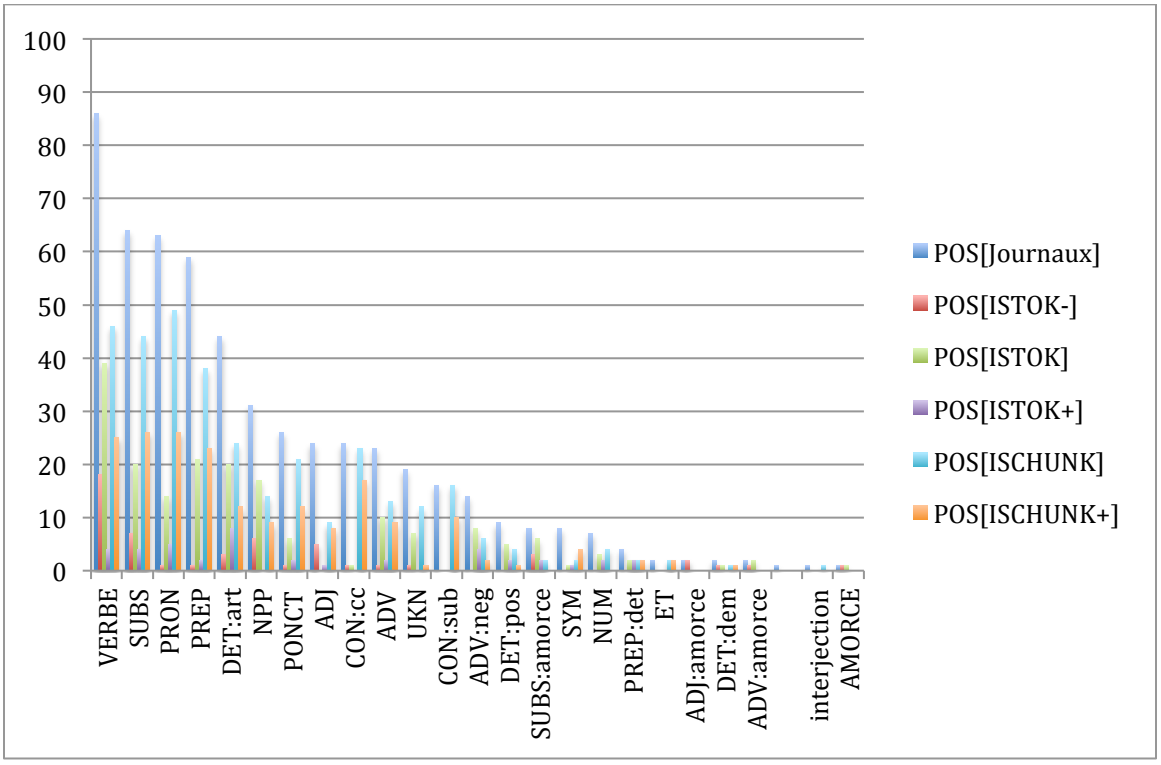


Figure 24 : Distribution des occurrences de catégories grammaticales pour chaque groupe du critère ISTOK/ISCHUNK (Nous avons regroupés les sous-catégories grammaticales sous leur catégories respectives pour une meilleure représentation de cette distribution. Ex : VERBE, PRON).

4.2.3 Les modifications de 1e niveau et de 2e niveau

Nous avons établi un critère de niveau pour distinguer la modification de l'énoncé en cours de l'écriture de la modification sur la relecture. Ce critère nous permet maintenant de révéler les unités et structures de l'écrit spontané.

De la même manière que les modifications du premier niveau, nous pouvons également démontrer les unités et structures venant des modifications sur la relecture du document. Nous présentons par la suite une représentation graphique des disfluences scripturales venant du premier jet (1e niveau) comparées à celles venant des modifications postérieures (2e niveau).

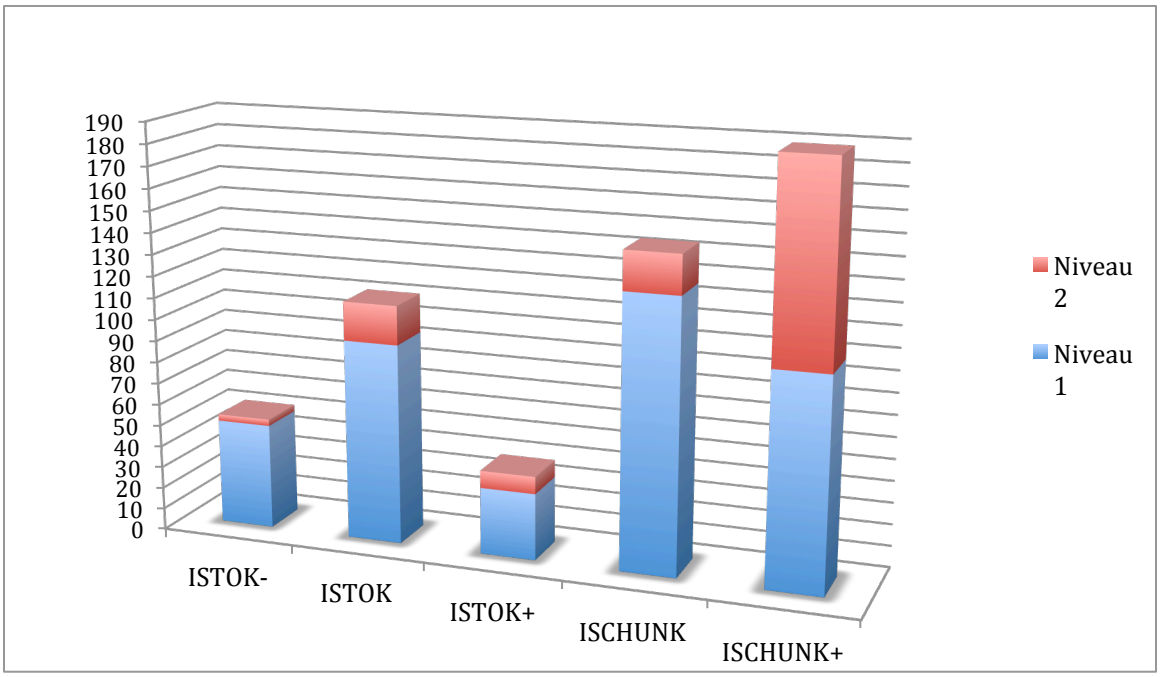


Figure 25 : Distribution des modifications du 1e et du 2e niveau pour les groupes du critère ISTOK/ISCHUNK

La même distinction de niveau peut aussi être faite pour chaque catégorie grammaticale. Pour un aperçu plus détaillé nous présentons le graphe suivant (Figure 26) également dans l'*Annexe B* (Figure 30).

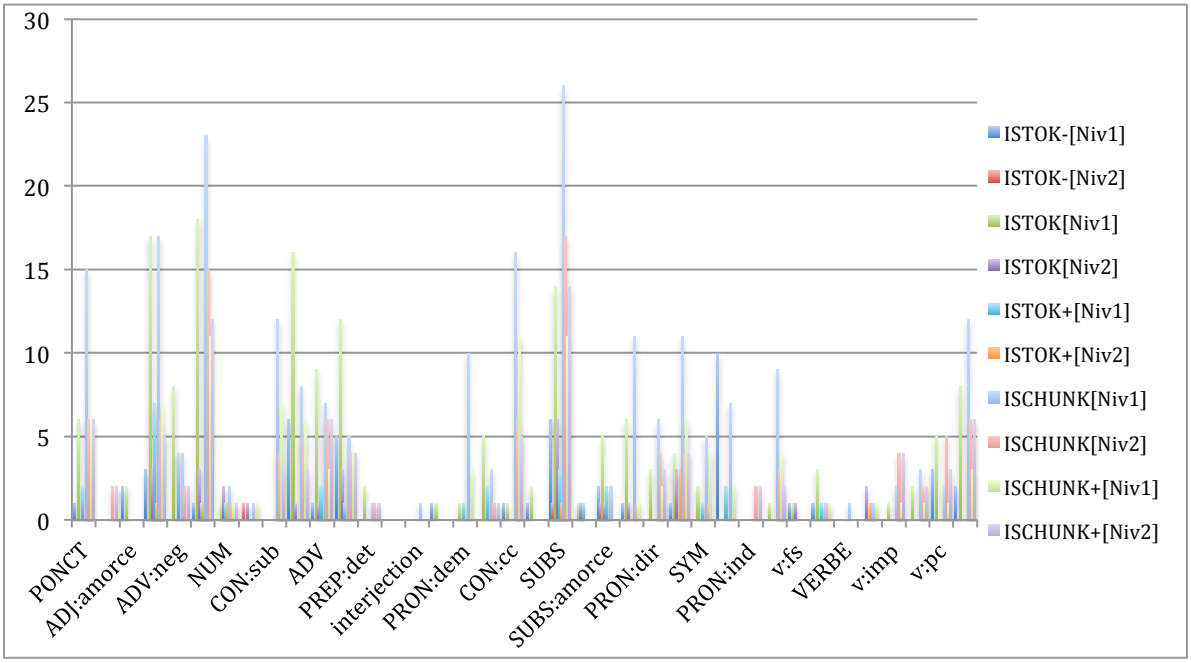


Figure 26 : Distribution des occurrences de catégories grammaticales pour chaque groupe du critère ISTOK/ISCHUNK et selon le niveau de modification (1/2).

Les modifications produites lors du premier jet (1e niveau) se regroupent selon l'appartenance au critère ISTOK/ISCHUNK et aussi par catégorie grammaticale. Nous observons la prévalence du groupe ISCHUNK pour les modifications du premier niveau.

4.2.4 Les unités répétées

L'annotation des unités répétées permet d'observer quelles unités du corpus sont répétées, à quelles catégories grammaticales elles appartiennent et si la répétition apparaît sur la borne ou à l'intérieur d'un *chunk.* Nous présentons ces résultats dans les figures qui suivent (Figure 27 et Figure 28).

Occurrence	POS	RR
15	PREP	\RB
11	DET:art	\RB
8	CON:sub	\RB
7	ADV:neg	\RB
7	CON:cc	\RB
6	v:amorce	\RB

6	v:ind	\RB
6	SUBS	\RB
6	NPP	\RB
5	PRON:pers	\RB
5	PONCT	\RB
5	ADV	\RI
4	SUBS:amorce	\RB
4	PRON:rel	\RB
3	PRON:dem	\RB
2	v:inf	\RB
2	PRON:dir	\RB
2	ADJ:amorce	\RB
2	ADJ	\RB
1	v:ps	\RI
1	v:imp	\RB
1	v:cond	\RB
1	UKN	\RB
1	PRON:pos	\RI
1	interjection	\RB
1	ADV:amorce	\RB
1		\RB

Figure 27 : Table relevant des catégories grammaticales des unités répétées pour le 1ᵉ niveau de modification.

Occurrence	POS	RR
4	SUBS	\RB
3	v:amorce	\RI
3	PRON:pers	\RB
3	DET:art	\RI
3	ADV	\RI
2	v:ps	\RB
2	v:ind	\RI
2	CON:cc	\RB
1	v:pc	\RB
1	v:imp	\RB
1	SUBS:amorce	\RI

1	PRON:rel	\RB
1	PRON:pos	\RI
1	PRON:dir	\RB
1	PRON:dem	\RI
1	ADV:neg	\RI
1	ADJ	\RB

Figure 28 : Table relevant des catégories grammaticales des unités répétées pour le 2ᵉ niveau de modification.

4.2.5 L'étude des opérations de la prosodie et de la syntaxe et l'étude de la topologie du texte

Nous observons ainsi le croisement entre les disciplines dans la pratique d'interprétation et de représentation de ce corpus de manuscrits. Par exemple, si l'objectif est de comprendre les effets des structurations de la prosodie et de la syntaxe lors de l'encodage de l'énoncé par le scripteur, la comparaison des unités et structures qui proviennent des modifications du premier jet avec celles des relectures successives ont de l'intérêt pour le linguiste. De la même façon cet intérêt pour le généticien de texte porte sur des étapes progressives d'écriture et de relecture auxquelles un texte a été soumis. Du point de vue linguistique nous observons le développement de l'énoncé, du point de vue de la génétique de texte le développement du document. Comme nous avons constaté lors de l'étude, c'est grâce aux moyens informatiques que l'objectif de représentation de ce corpus est atteinte. Pour notre part, la caractérisation linguistique des disfluences scripturales, repose également sur les méthodes informatiques et les principes linguistiques afin de mener l'étude de ce corpus. Dans l'ensemble, les objectifs sont validées quand nous arrivons à décrire le développement non-linéaire de la pensée [Fuchs, 1994 ; Robert, 1999] ou le développement topographique d'un texte ou un ensemble de textes [Biasi, 2013].

4.3 Facteurs inattendus

Le projet de transcription et de numérisation de ces manuscrits dépend des contributions des transcripteurs, annotateurs et relecteurs. La transcription d'un manuscrit est un travail minutieux. Pourtant, il peut y avoir des erreurs commises par les transcripteurs qui n'ont pas été saisis par les relecteurs. Ces erreurs sont souvent mineures, mais ils ont tout de même un effet sur les données que nous utilisons pour l'analyse des disfluences scripturales. Pour ces raisons, il est souhaitable de revenir directement au manuscrit afin de résoudre les cas d'ambiguïté rencontrés dans les données. C'est une des raisons pour laquelle nous avons un lien permettant d'accéder directement à la page manuscrite et sa transcription pour chaque extrait récupéré par notre moteur de recherche.

De même, l'annotation que nous avons effectuée pour la caractérisation du corpus a été réalisée manuellement. Les erreurs sur les données peuvent provenir de nos annotations et relèvent de l'erreur humaine. Notre travail a été fait avec attention particulière afin d'être aussi précis que possible dans les annotations. Pourtant, si nous avons commis des erreurs pour une ou plusieurs catégories ils affecteront les résultats que nous présentons ici. Tout de même, pour ce petit échantillon les erreurs sont faciles à révéler et corriger dans l'objectif d'être aussi précis que possible pour la représentation des données.

Dans cette section nous avons présenté la caractérisation que nous avons réalisée sur le corpus *Journaux* des manuscrits de Stendhal. Cette caractérisation représente un échantillon de l'ensemble des productions écrites de l'écrivain qui est actuellement à la disposition des chercheurs et du public.

Comme nous l'avons évoqué l'étude a été réalisée sur des transcriptions validées afin de minimiser les erreurs provenant des données. Néanmoins le facteur de l'erreur humaine dans le cas des transcriptions et aussi de l'analyse linguistique largement réalisée à la main est toutefois très réelle.

Notre caractérisation n'est pas représentative de tout le corpus. Cependant, elle permet de remarquer les unités et les structures et relever les pratiques spécifiques de Stendhal dans ce corpus *Journaux.*

5 Conclusion

Etudier ce corpus du point de vue linguistique (et en s'appuyant sur des outils informatiques) peut vraiment apporter aux études génétiques des manuscrits, car nous apportons des critères formels dans un champ d'étude qui a toujours été largement interprétatif. L'outil informatique joue un rôle important dans le développement des méthodes rigoureuses d'observation scientifique. Les résultats de l'étude peuvent apporter des révélations importantes et concrètes aux objectifs de la discipline de la génétique textuelle et au domaine de la littérature. Il est intéressant aussi pour nous du point de vue de la linguistique d'observer le processus de création et de l'énonciation chez l'écrivain Stendhal. En observant les types d'hésitations et la façon dont l'écrivain formule et reformule son énoncé tout au long du processus de rédaction permet de mieux comprendre ce processus de l'adaptation de la pensée et les choix les plus minutieux de la plume de l'écrivain. Les étapes multiples qui constituent le mouvement de l'avant-texte vers un texte final deviennent plus évidentes lors de l'observation fine du corpus. Sans doute, l'intervention des relecteurs multiples rend encore plus intriguant cette observation du processus de l'énonciation sous forme écrite. À travers l'étude des journaux, des pensées intimes, des conversations ainsi que des graines des idées qui sèmeront certainement l'ensemble de l'œuvre de l'écrivain au long de sa vie notre compréhension de l'écrivain Stendhal à travers ses journaux est plus enrichie.

Annexe A

Table représentative des annotations effectuées sur le corpus *Journaux*

Légende :

- **les mots en gras :** les unités biffées du texte.
- *Les mots en italique :* les ajouts
- */En italique et entourés de barres obliques/ :* les interlignes
- *//En italique et entourés de double barres obliques// :* les variantes en interligne
- [Entre crochets] : les mots appartenant à un chunk complet

IDPage	Référence Feuillet	Vocable	unité biffée	Segment de transformation	POS	ISTOK/ISCHUNK	RR	Positionnement	Type de Réparation	Niveau
2	R. 5896 Rés., volume 22, feuillet 15, verso	il	~~il n'~~	[cependant] [il n'en fut pas moins grand] [au moment même] [**il n'** la Somme]	PRON:dem	ISTOK+	0	debut	auto	1
2	R. 5896 Rés., volume 22, feuillet 15, verso	n'	~~il n'~~	[cependant] [il n'en fut pas moins grand] [au moment même] [**il n'** la Somme]	ADV:neg	ISTOK+	0	fin	auto	1
3	R. 5896 Rés., volume 22, feuillet 16, recto	B.	~~B.~~	[Ennui profond] [que] [D.] [et] [M.] [donnent] [à B. Tencin][.]	NPP	ISTOK	0	mil	auto	1
3	R. 5896 Rés., volume 22, feuillet 16, recto		~~'~~	[qu'elle diffèrence] [!]	PONCT	ISTOK	0	mil	auto	1
5	R. 5896 Rés., volume 22, feuillet 17, recto	Franceleu	~~Francaleu~~	[que] [Damis reconnaissant][**Francaleu** Baliveau]	NPP	ISTOK	0	debut	auto	1
7	R. 5896 Rés., volume 22, feuillet 18, recto	déroul	~~déroul~~	[nous **déroul** parlons] [du projet] [de M.e Cardxx] [de le marier] [à M.lle Pauline]	v:amorce	v:amorce	0	mil	auto	1
7	R. 5896 Rés., volume 22, feuillet 18, recto	pre	~~pre~~	[Je vois] [qu'][il **pre** aime mieux] [prendre] [une bonne petite fille] [qui] [lui promette] [les plaisirs] [du cœur] [à toutes] [les grandeurs possibles][.]	v:amorce	v:amorce	0	mil	auto	1
	R. 5896 Rés., v									

	20, verso									
19	5896 Rés., volume 22, feuillet 24, recto	la	~~la~~	[J'ai vu hier] [le bon genre] [de **la** plaisanterie] [non pas sublimé] [mais] [bien indiqué] [par M.r de Possai][.]	DET:art	ISTOK	0	mil	auto	1
33	R. 5896 Rés., volume 16, feuillet 3, recto	à	~~à mon~~	[mais] [cette habitude] [tient = elle] [**à mon** au système] [de passions] [qui] [doit toujours] [dominer chez] [moi]	DET:pos	ISTOK+	0	debut	auto	1
33	R. 5896 Rés., volume 16, feuillet 3, recto	mon	~~à mon~~	[mais] [cette habitude] [tient = elle] [**à mon** au système] [de passions] [qui] [doit toujours] [dominer chez] [moi]	DET:pos	ISTOK+	0	mil	auto	1
1	R. 5896 Rés., volume 22, feuillet 15, recto	comp	~~comp~~	[que] [j'aie] [**comp** trouvés] [de ma vie]	v:amorce	ISCHUNK	0	debut	auto	1
37	R. 5896 Rés., volume 16, feuillet 5, verso	Fraire	~~Fraire~~	[je suivis] [de l'oeil][le pretre] [qui] [etait] [**Fraire** là] [en étole]	NPP	ISTOK	0	debut	auto	1
37	R. 5896 Rés., volume 16, feuillet 5, verso	dans	~~dans~~	[la femme] [portait toujours] [la petite figure] [**dans** sur ses bras]	PREP	ISTOK	0	debut	auto	1
529	R. copie Arbelet Rés., volume cahier n° 50, feuillet 20, recto	Rebuffet	~~Rebuffet~~	[Ce père] [me semble] [avoir beaucoup] [du caractère] [α] [de l'esprit] [de mon cousin] [**Rebuffet**] [et] [être] [comme] [lui] [peu apprécié][.]	NPP	ISCHUNK	0	tout	auto	2
529	R. copie Arbelet Rés., volume cahier n° 50, feuillet 20, recto	Madame	~~Madame~~	[J'ai] [trouvé] [**Madame** Livia] [libre] [α] [plongée] [dans l'ennui][.]	NPP	ISCHUNK	0	debut	auto	?
529	R. copie Arbelet Rés., v	P	~~p~~	[et] [qu'][un moyen presque sur] [d'éviter] [ce goufre affreux] [est] [de se livrer]	NPP	ISTOK	0	fin	auto	1

	olume cahier n° 50, feuillet 20, recto			[comme] [Lady **P**]						
439	R. 5896 Rés., volume 22, feuillet 93, recto	e	~~e~~	[,] [et] [cette Représentation] [celle] [qui] [a jamais fait**e**] [l'impression la plus forte] [sur moi][.]	v:pc	ISTOK-	0	fin	auto	1
79	R. 5896 Rés., volume 22, feuillet 27, recto	à	~~à~~	[Edouard M.] [m'annonce] [qu'][il ser**à**] [à Paris] [dans les 1.rs jours] [de Thermidor][.]	v:fs	ISTOK-	0	fin	auto	1
87	R. 5896 Rés., volume 22, feuillet 55, recto'	m'	~~m'~~	[si] [son ame] [ressemble] [à ses écrits] [il ne doit pas] [goûter du tout] [la joie acre] [de celui = ci][.] [**m'** l'optimiste]	DET:pos	ISTOK	0	debut	auto	1
91	R. 5896 Rés., volume 22, feuillet 57, recto	M	~~M~~	[**Me** Goldoni] [pense] [comme] [moi] [sur la plupart] [de ces comédies] [en vers] [que] [l'on donnait] [en France] [vers 1750][.]	NPP	ISCHUNK	0	tout	auto	1
92	R. 5896 Rés., volume 22, feuillet 57, verso	pe	~~pe~~	[-] [**pe** manque] [d'expression]	v:amorce	ISCHUNK	0	tout	auto	1
92	R. 5896 Rés., volume 22, feuillet 57, verso	peintre	~~peintre~~	[Goldoni] [est] [peut étre] [le **peintre** poète] [le plus naturel] [qui] [existe][.]	SUBS	ISTOK	0	fin	auto	1
93	R. 5896 Rés., volume 22, feuillet 58, recto	cela	~~cela~~	[**cela**] [q] [et] [la première][.]	PRON:dem	ISCHUNK	0	tout	auto	1
93	R. 5896 Rés., volume 22, feuillet 58, recto	q	~~q~~	[cela] [**q**] [et] [la première][.]	CON:sub	ISCHUNK	0	tout	auto	1
93	R. 5896 Rés., volume 22, feuillet 58, recto	être	~~être~~	[pour **être** faire] [des ouvrages sublimes] [il faut ne vivre] [que] [pour son gènie][.]	v:inf	ISTOK	0	mil	auto	1
94	R. 5896 Rés., volume 22, feuillet 58, verso	me	~~me de~~	[je lis ensuite] [La finta Amalatta] [de Goldoni] [qui] [**me de** m'engage] [à mettre tout de suite] [à exécution] [un projet] [formé] [le Dimanche] [jour]	PRON:dir	ISCHUNK	\RB	debut	auto	1

94	R. 5896 Rés., volume 22, feuillet 58, verso	de	~~me de~~	[je lis ensuite] [La finta Amalatta] [de Goldoni] [qui] [**me de** m'engage] [à mettre tout de suite] [à exécution] [un projet] [formé] [le Dimanche] [jour]	v:amorce	ISTOK+	\RB	fin	auto	1
98	R. 5896 Rés., volume 22, feuillet 60, verso	te	~~te~~	[quel parti] [peut=on tirer] [de la Vanité**te**] [?]	SUBS:amorce	ISTOK-	0	fin	auto	1
98	R. 5896 Rés., volume 22, feuillet 60, verso	Lund	~~Lund~~	[12][.] [**Lund** Dim]	SUBS	ISCHUNK	0	tout	auto	1
98	R. 5896 Rés., volume 22, feuillet 60, verso	figure	~~figure~~	[je remarque] [la **figure** *phisionomie* basse] [et] [quelquefois méchante] [des prêtres][.]	SUBS	ISTOK	0	mil	auto	2
98	R. 5896 Rés., volume 22, feuillet 60, verso	de	~~de la~~	[**de la**] [on va] [à l'église] [après les prieres] [on accomp][.] [jusqu'][à la derniére demeure][.]	PREP	ISCHUNK	0	debut	auto	1
98	R. 5896 Rés., volume 22, feuillet 60, verso	la	~~de la~~	[**de la**] [on va] [à l'église] [après les prieres] [on accomp][.] [jusqu'][à la derniére demeure][.]	ADV	ISCHUNK	0	fin	auto	1
102	R. 5896 Rés., volume 22, feuillet 62, verso	*...*	*...*	[la Vanité][,] [*...* **le pr** le Cavalière][est même dur] [p.r la dame] [lorsqu'][il lui annonce brusquement] [la mort] [de son mari][.]	UKN	ISCHUNK	0	debut	auto	1
102	R. 5896 Rés., volume 22, feuillet 62, verso	le	~~le pr~~	[**le pr** Le Cavaliere][est même dur] [p.r la dame] [lorsqu'][il lui annonce brusquement] [la mort] [de son mari][.]	DET:art	ISCHUNK	0	mil	auto	1
102	R. 5896 Rés., volume 22, feuillet 62, verso	pr	~~le pr~~	[**le pr** Le Cavaliere][est même dur] [p.r la dame] [lorsqu'][il lui annonce brusquement] [la mort] [de son mari][.]	SUBS	ISCHUNK	0	fin	auto	1

103	R. 5896 Rés., volume 22, feuillet 63, recto	les	~~Les acteurs~~	[**Les acteurs** Tartuffe] [fait toujours] [la faute] [de ne pas retenir] [Orgon] [lorsqu'][il s'emporte] [contre son fils] [et] [qu'][il dit] [ne me retenez pas][.]	DET:art	ISCHUNK	0	debut	auto	1
103	R. 5896 Rés., volume 22, feuillet 63, recto	acteurs	~~Les acteurs~~	[**Les acteurs** Tartuffe] [fait toujours] [la faute] [de ne pas retenir] [Orgon] [lorsqu'][il s'emporte] [contre son fils] [et] [qu'][il dit] [ne me retenez pas][.]	SUBS	ISCHUNK	0	fin	auto	1
104	R. 5896 Rés., volume 22, feuillet 63, verso	pri	~~pri~~	[il a **pri** mis] [dans le role] [de Laf.] [beaucoup] [de vers] [de ce grand hom]	v:amorce	ISTOK-	0	fin	auto	1
105	R. 5896 Rés., volume 22, feuillet 64, recto	ù	~~ù~~	[Je lis] [de la Vérité] [par Brissot=Varville] [o**ù**u] [plutôt] [je le parcours] [cet ouvrage] [pourra] [m'être trés utile][.]	CON:cc	ISTOK-	0	fin	auto	1
105	R. 5896 Rés., volume 22, feuillet 64, recto	qui	~~qui~~	[je lis] [sa méthode] [de Conduire] [la raison] [**qui** dont] [ce] [qui |m'|] [m'interesse] [peut tenir] [en 3 phrases][.]	PRON:rel	ISCHUNK	\RB	tout	auto	1
110	R. 5896 Rés., volume 22, feuillet 65, verso	é	~~é~~	[,] [outre cela] [il s'échauf**é**e] [une petite fois] [devant moi][,]	v:pc	ISTOK-	0	fin	auto	1
112	R. 5896 Rés., volume 22, feuillet 66, verso	la	~~la Correspondance~~	[je lis] [avec beaucoup] [de plaisir] [à la Bibliothèque] [n.le] [**la Correspondance** *les lettres* autographes] [de Voltaire] [à Maupertuis][.]	DET:art	ISCHUNK	\RB	debut	auto	1
112	R. 5896 Rés., volume 22, feuillet 66, verso	Correspondance	~~la Correspondance~~	[je lis] [avec beaucoup] [de plaisir] [à la Bibliothèque] [n.le] [**la Correspondance** *les lettres* autographes] [de Voltaire] [à Maupertuis][.]	SUBS	ISCHUNK	\RB	mil	auto	1
120	R. 5896 Rés., volume 22, feuillet 70, verso	lettre	~~lettre en~~	[et] [au même moment] [tirant] [vanité] [d'une **lettre**] [**en** réponse insignifiante] [p.r la Vanité]	PREP	ISCHUNK+	0	fin	auto	1
120	R. 5896 Rés., volume 22,	en	~~lettre en~~	[et] [au même moment] [tirant] [vanité] [d'**une lettre**] [**en** réponse insignifiante]	SUBS	ISCHUNK+	0	debut	auto	1

	feuillet 70, verso			[p.r la Vanité]						
124	R. 5896 Rés., volume 22, feuillet 72, verso	de	~~de la vie~~	[La Lecture] [des Mémoires] [**de la vie**] [de Marmontel]	PREP	ISCHUNK	\RB	debut	auto	1
124	R. 5896 Rés., volume 22, feuillet 72, verso	la	~~de la vie~~	[La Lecture] [des Mémoires] [**de la vie**] [de Marmontel]	DET:art	ISCHUNK	\RB	mil	auto	1
124	R. 5896 Rés., volume 22, feuillet 72, verso	vie	~~de la vie~~	[La Lecture] [des Mémoires] [**de la vie**] [de Marmontel]	SUBS	ISCHUNK	\RB	fin	auto	1
86	R. 5896 Rés., volume 22, feuillet 54, verso	*...*	*...*	[v.] [les ***…*** Pensèes diverses][.]	UKN	ISCHUNK			auto	1
99	R. 5896 Rés., volume 22, feuillet 61, recto	sa	~~sa femme~~	[à 8 ¾] [j'entre] [chez mr Carara] [j'y trouve] [**sa femme** Madame][,] [Adèle] [et] [M.r Davrange] [Insp.r] [aux Revües] [je crois][.]	DET:pos	ISCHUNK		debut	auto	1
99	R. 5896 Rés., volume 22, feuillet 61, recto	femme	~~sa femme~~	[à 8 ¾] [j'entre] [chez mr Carara] [j'y trouve] [**sa femme** Madame][,] [Adèle] [et] [M.r Davrange] [Insp.r] [aux Revües] [je crois][.]	SUBS	ISCHUNK		fin	auto	1
99	R. 5896 Rés., volume 22, feuillet 61, recto	e	~~e~~	[Probablement M. e Car.] [leur a parlé] [après ma sortie] [de mon prétendu bonheur] [avec Is. P **e**][.]	NPP	ISTOK-	0	fin	auto	1
99	R. 5896 Rés., volume 22, feuillet 61, recto		'	[j'ai m'al fait] [mes affaires][.]	PONCT	ISTOK	0	mil	auto	1
100	R. 5896 Rés., volume 22, feuillet 61, verso			[]	UKN	ISTOK	0	tout	auto	1
100	R. 5896 Rés., volume 22, feuillet 61, verso	qu'	~~qu'Adèle veut se marier~~	[Ces dames] [me disent] [**qu'][Adèle veut] [se marier**][qu'un homme][qui tient] [de très près] [à la Cour] [fait la Cour] [à Adèle]	CON:sub	ISCHUNK+	\RB	tout	auto	1
100	R. 5896 Rés., volume 22, feuillet 61, verso	Adèle	~~qu'Adèle veut se marier~~	[Ces dames] [me disent] [**qu'][Adèle veut] [se marier**][qu'un homme][qui tient] [de très près] [à la Cour] [fait la Cour] [à Adèle]	NPP	ISCHUNK+	\RB	debut	auto	1
100	R. 5896 Rés., volume 22, feuillet 61,	veut	~~qu'Adèle veut se marier~~	[Ces dames] [me disent] [**qu'][Adèle veut] [se marier**][qu'un homme][qui tient] [de très près] [à la Cour] [fait la Cour] [à Adèle]	v:ind	ISCHUNK+	\RB	fin	auto	1

	verso									
100	R. 5896 Rés., volume 22, feuillet 61, verso	se	~~qu'Adèle veut se marier~~	[Ces dames] [me disent] [**qu'][Adèle veut] [se marier**][qu'un homme][qui tient] [de très près] [à la Cour] [fait la Cour] [à Adèle]	PRON:dir	ISCHUNK+	\RB	debut	auto	1
100	R. 5896 Rés., volume 22, feuillet 61, verso	marier	~~qu'Adèle veut se marier~~	[Ces dames] [me disent] [**qu'][Adèle veut] [se marier**][qu'un homme][qui tient] [de très près] [à la Cour] [fait la Cour] [à Adèle]	v:inf	ISCHUNK+	\RB	fin	auto	1
100	R. 5896 Rés., volume 22, feuillet 61, verso	elle	~~elle~~	[**elle** en voilà][assez]	PRON:rel	ISTOK	0	debut	auto	1
126	R. 5896 Rés., volume 22, feuillet 73, verso	é	~~é~~	[Pendant peu de Semaines] [de ma vie] [j'ai été témoin] [d'événemens aussi intér**é**essans] [pour moi][.]	ADJ	ISTOK-	0	fin	auto	1
129	R. 5896 Rés., volume 22, feuillet 75, recto	p.r.	~~p.r~~	[il semble] [que] [ce dernier feu follet] [d'amour] [**p.r** p.r Gate]		ISTOK	\RB	debut	auto	1
129	R. 5896 Rés., volume 22, feuillet 75, recto	celui	~~celui~~	[ou] [est] [ce] [**celui**] [l'effet] [des plaidoyers] [de l'Avocat]	PRON:dem	ISCHUNK	0	tout	auto	1
427	R. 5896 Rés., volume 25, feuillet 141, verso	trois	~~trois~~	[Distinguer] [**trois** *quatre* genres] [de conversation][.]	ADJ	ISTOK	0	debut	auto	2
427	R. 5896 Rés., volume 25, feuillet 141, verso	de	~~de~~	[**de** la gaité]	PREP	ISTOK	0	debut	auto	1
427	R. 5896 Rés., volume 25, feuillet 141, verso	3.e	~~3.~~e où on est gai et curieux	**[3.e] [où] [on est gai] [et] [curieux]**	NUM	ISCHUNK+	0	tout	auto	1
427	R. 5896 Rés., volume 25, feuillet 141, verso	où	~~3.~~e où on est gai et curieux	**[3.e] [où] [on est gai] [et] [curieux]**	ADV	ISCHUNK+	0	tout	auto	1
427	R. 5896 Rés., volume 25, feuillet 141, verso	on	~~3.~~e où on est gai et curieux	**[3.e] [où] [on est gai] [et] [curieux]**	PRON:pers	ISCHUNK+	0	debut	auto	1

427	R. 5896 Rés., volume 25, feuillet 141, verso	est	~~3.~~e où on est gai et curieux	[**3.e] [où] [on est gai] [et] [curieux]**	v:ind	ISCHUNK+	0	mil	auto	1
427	R. 5896 Rés., volume 25, feuillet 141, verso	gai	~~3.~~e où on est gai et curieux	[**3.e] [où] [on est gai] [et] [curieux]**	ADJ	ISCHUNK+	0	fin	auto	1
427	R. 5896 Rés., volume 25, feuillet 141, verso	et	~~3.~~e où on est gai et curieux	[**3.e] [où] [on est gai] [et] [curieux]**	CON:cc	ISCHUNK+	0	tout	auto	1
427	R. 5896 Rés., volume 25, feuillet 141, verso	curieux	~~3.~~e où on est gai et curieux	[**3.e] [où] [on est gai] [et] [curieux]**	ADJ	ISCHUNK+	0	tout	auto	1
128	R. 5896 Rés., volume 22, feuillet 74, verso	avec	~~avec~~	[nous l'avons vu] [passer] [**avec** mon oncle][,] [M.e R.]	PREP	ISTOK	0	debut	auto	1
131	R. 5896 Rés., volume 22, feuillet 76, recto	aux	~~aux~~	[Je lui ai] [trouvé] [la figure] [jusqu'][**aux** *à* la bouche très bien.]	PREP	ISTOK	\RB	debut	auto	1
131	R. 5896 Rés., volume 22, feuillet 76, recto	à	~~à~~	[en général] [sa conduite] [**à** a été très bonne]	v:pc	ISTOK	0	debut	auto	1
132	R. 5896 Rés., volume 22, feuillet 76, verso	rait	~~rait~~	[que] [la Phisionomie] [me le permet**rait**]	v:cond	ISTOK-	0	fin	auto	1
132	R. 5896 Rés., volume 22, feuillet 76, verso	comb	~~comb~~	[Mais je sens] [**comb** par moi-même][combien tous les Signes][que][donnent] [les gens passionnés] [peuvent être] [trompeur][.]	ADV:amorce	ISTOK-	\RB	debut	auto	1
140	R. 5896 Rés., volume 22, feuillet 80, verso	che	~~che~~	[**che** à la porte] [de M.e Anet]	v:amorce	ISTOK-	0	debut	auto	1
141	R. 5896 Rés., volume 22, feuillet 81, recto	n'	~~n'~~	[il **n'**en parlerait quelquefois]	ADV:neg	ISTOK	0	mil	auto	1

142	R. 5896 Rés., volume 22, feuillet 81, verso	2	~~2 heures~~	[Cet intervalle] [de **2 heures** midi] [à Cinq] [fut charmant] [pour moi]	NUM	ISTOK+	0	mil	auto	2
142	R. 5896 Rés., volume 22, feuillet 81, verso	heures	~~2 heures~~	[Cet intervalle] [de **2 heures** midi] [à Cinq] [fut charmant] [pour moi]	SUBS	ISTOK+	0	fin	auto	2
406	R. 5896 Rés., volume 25, feuillet 131, recto	va	~~va à~~	[nous distingons] [la cote] [qui] [**va] [à** court] [vers Cherbourg] [à 25 lieues] [de nous][.]	v:ind	ISCHUNK+	0	tout	auto	2
406	R. 5896 Rés., volume 25, feuillet 131, recto	à	~~va à~~	[nous distingons] [la cote] [qui] [**va] [à** court] [vers Cherbourg] [à 25 lieues] [de nous][.]	PREP	ISCHUNK+	0	debut	auto	2
407	R. 5896 Rés., volume 25, feuillet 131, verso	pour	~~pour~~	[Quitté] **[pour** le hâvre] [à 4.h] [du matin_] [Crozet a eu] [à la poste] [la même espèce] [de desapointement] [à peu près] [que] [moi] [en sortant] [de Strasbourg]	PREP	ISTOK	0	debut	auto	1
149	R. 5896 Rés., volume 22, feuillet 85, recto	expier	~~expier~~	[qui] [fait **expier** punir][des crimes] [par d'autres Crimes]	v:inf	ISTOK	0	mil	auto	1
149	R. 5896 Rés., volume 22, feuillet 85, recto	voir	~~voir et~~	[le génie comique] [est] [celui] [qui] [nous donne] [une bien bonne] [et] [sure raison] [de **voir**] [**et**] [reconnaître] [notre excellence][.]	CON:cc	ISCHUNK+	0	fin	auto	1
149	R. 5896 Rés., volume 22, feuillet 85, recto	et	~~voir et~~	[le génie comique] [est] [celui] [qui] [nous donne] [une bien bonne] [et] [sure raison] [de **voir**] [**et**] [reconnaître] [notre excellence][.]	CON:cc	ISCHUNK+	0	debut	auto	1
149	R. 5896 Rés., volume 22, feuillet 85, recto	à	~~à~~	[est] [**à** au véritable hôte] [de S.t Michel] [en Savoye][.]	PREP	ISTOK	\RB	debut	auto	1
149	R. 5896 Rés., volume 22, feuillet 85, recto	est	~~est plus~~	[La plaisanterie] [**est plus** saisit plus vîte] [que le Comique]	v:ind	ISCHUNK	\RI	debut	auto	1

149	R. 5896 Rés., volume 22, feuillet 85, recto	plus	~~est plus~~	[La plaisanterie] [**est plus** saisit plus vîte] [que le Comique]	ADV	ISCHUNK	\RI	mil	auto	1
150	R. 5896 Rés., volume 22, feuillet 85, verso	je	~~je suis~~	[il suit] [de là] [que] [**je suis** j'ai] [p.r le Vaniteux] [douze scènes] [au moins] [qui] [me fourniront] [nombre] [de traits Comiques]	pron:rel	ISCHUNK	\RB	debut	auto	1
150	R. 5896 Rés., volume 22, feuillet 85, verso	suis	~~je suis~~	[il suit] [de là] [que] [**je suis** j'ai] [p.r le Vaniteux] [douze scènes] [au moins] [qui] [me fourniront] [nombre] [de traits Comiques]	v:ind	ISCHUNK	\RB	fin	auto	1
36	R. 5896 Rés., volume 16, feuillet 5, recto	les	~~les bras~~	[Petite fille morte] [**les bras** Ses petites mains jointes]	DET:art	ISCHUNK	0	debut	auto	1
36	R. 5896 Rés., volume 16, feuillet 5, recto	bras	~~les bras~~	[Petite fille morte] [**les bras** Ses petites mains jointes]	SUBS	ISCHUNK	0	fin	auto	1
554	R. 5896 Rés., volume 16, feuillet 4, recto	est	~~est~~	[Ce second mouvement] [de raison] [**est** vient] [de *Tracy* Y.]	v:ind	ISTOK	0	debut	auto	1
528	R. copie Arbelet Rés., volume cahier n° 50, feuillet 19, verso	Wi	~~Wi~~	[and] [for the **Wi** parts][.]	AMORCE	ISTOK-	0	mil	auto	1
526	R. copie Arbelet Rés., volume cahier n° 50, feuillet 18, verso	i	~~i~~	[dit] [M. r de Ta**i**lleyrand][.]	NPP	ISTOK-	0	mil	auto	1
521	R. copie Arbelet Rés., volume cahier n° 50, feuillet 15, verso	e	~~e~~	[Le Gouvernement] [α] [les circonstances] [ne sont pas arrangé**e**s] [de manière] [à ce] [qu'][elles puissent être amusantes][.]	v:pc	ISTOK-	0	fin	auto	1
519	R. copie Arbelet Rés., volume cahier n° 50, feuillet 14, verso	un	~~un infini~~	[tout y est] [traité] [avec **un** art **infini**][.]	DET:art	ISTOK	0	debut	auto	1

519	R. copie Arbelet Rés., volume cahier n° 50, feuillet 14, verso	infini	~~un infini~~	[tout y est] [traité] [avec **un** art **infini**][.]	ADJ	ISTOK	0	fin	auto	1
518	R. copie Arbelet Rés., volume cahier n° 50, feuillet 14, recto	qui	~~qui étaient~~	[Les frais] [de cette fête][**qui] [étaient] [**assez considérables][étaient supportés]	PRON:rel	ISCHUNK+	\RB	tout	auto	2
518	R. copie Arbelet Rés., volume cahier n° 50, feuillet 14, recto	étaient	~~qui étaient~~	[Les frais] [de cette fête][**qui] [étaient] [**assez considérables][étaient supportés]	v:imp	ISCHUNK+	\RB	tout	auto	2
517	R. copie Arbelet Rés., volume cahier n° 50, feuillet 13, verso	α	~~α~~	[**α**] [de Castagnettes]	SYM	ISCHUNK	0	tout	auto	1
517	R. copie Arbelet Rés., volume cahier n° 50, feuillet 13, verso	*...*	*...*	[α] [d'autres instrumens] [qu'] [on dit] [d'***…*** origine Grecque][.]	UKN	ISTOK	0	mil	auto	1
507	R. copie Arbelet Rés., volume cahier n° 50, feuillet 7, recto	Dam	~~Dam~~	[**Dam** Bacch][a particulièrement réussi] [à bien exprimer] [l'ironie][.]	NPP	ISTOK-	0	debut	auto	1
505	R. copie Arbelet Rés., volume cahier n° 50, feuillet 6, recto	è	~~è~~	[Ce qui] [le fit appeler] [Pergole**è**se][.]	NPP	ISTOK-	0	fin	auto	1
505	R. copie Arbelet Rés., volume cahier n° 50, feuillet 6, recto	-1	~~-1~~	[**(1)**]	SYM	ISCHUNK	0	tout	indeterminé	2
506	R. copie Arbelet Rés., volume cahier n° 50, feuillet 6, verso	et	~~et~~	[David Perez] [né] [à Naples] [à composé] [un Credo] [qui] [se chante encore] [dans l'église] [des Pères] [de l'Oratoire] [à certaines solemnités] [**et** où] [l'on va][l'entendre]	CON:cc	ISCHUNK	0	tout	auto	2
506	R. copie Arbelet Rés., v	comme	comme un original. On le	**[comme] [un original][.] [On le traite] [ainsi] [comme] [Leo][,]**	CON:sub	ISCHUNK+	\RB	tout	auto	1

	olume cahier n° 50, feuillet 6, verso		traite ainsi comme Leo, Scarlatti, Porpora α Durante.	**[Scarlatti][,] [Porpora] [α] [Durante.]**						
506	R. copie Arbelet Rés., volume cahier n° 50, feuillet 6, verso	un	comme un original. On le traite ainsi comme Leo, Scarlatti, Porpora α Durante.	**[comme] [un original][.] [On le traite] [ainsi] [comme] [Leo][,] [Scarlatti][,] [Porpora] [α] [Durante.]**	DET:art	ISCHUNK+	\RB	debut	auto	1
506	R. copie Arbelet Rés., volume cahier n° 50, feuillet 6, verso	original	comme un original. On le traite ainsi comme Leo, Scarlatti, Porpora α Durante.	**[comme] [un original][.] [On le traite] [ainsi] [comme] [Leo][,] [Scarlatti][,] [Porpora] [α] [Durante.]**	SUBS	ISCHUNK+	\RB	fin	auto	1
506	R. copie Arbelet Rés., volume cahier n° 50, feuillet 6, verso	.	comme un original. On le traite ainsi comme Leo, Scarlatti, Porpora α Durante.	**[comme] [un original][.] [On le traite] [ainsi] [comme] [Leo][,] [Scarlatti][,] [Porpora] [α] [Durante.]**	PONCT	ISCHUNK+	\RB	tout	auto	1
506	R. copie Arbelet Rés., volume cahier n° 50, feuillet 6, verso	On	comme un original. On le traite ainsi comme Leo, Scarlatti, Porpora α Durante.	**[comme] [un original][.] [On le traite] [ainsi] [comme] [Leo][,] [Scarlatti][,] [Porpora] [α] [Durante.]**	PRON:pers	ISCHUNK+	\RB	debut	auto	1
506	R. copie Arbelet Rés., volume cahier n° 50, feuillet 6, verso	le	comme un original. On le traite ainsi comme Leo, Scarlatti, Porpora α Durante.	**[comme] [un original][.] [On le traite] [ainsi] [comme] [Leo][,] [Scarlatti][,] [Porpora] [α] [Durante.]**	DET:art	ISCHUNK+	\RB	mil	auto	1
506	R. copie Arbelet Rés., volume cahier n° 50, feuillet 6, verso	traite	comme un original. On le traite ainsi comme Leo, Scarlatti, Porpora α Durante.	**[comme] [un original][.] [On le traite] [ainsi] [comme] [Leo][,] [Scarlatti][,] [Porpora] [α] [Durante.]**	v:ind	ISCHUNK+	\RB	fin	auto	1

506	R. copie Arbelet Rés., volume cahier n° 50, feuillet 6, verso	ainsi	comme un original. On le traite ainsi comme Leo, Scarlatti, Porpora α Durante.	**[comme] [un original][.] [On le traite] [ainsi] [comme] [Leo][,] [Scarlatti][,] [Porpora] [α] [Durante.]**	ADV	ISCHUNK+	\RB	tout	auto	1
506	R. copie Arbelet Rés., volume cahier n° 50, feuillet 6, verso	comme	comme un original. On le traite ainsi comme Leo, Scarlatti, Porpora α Durante.	**[comme] [un original][.] [On le traite] [ainsi] [comme] [Leo][,] [Scarlatti][,] [Porpora] [α] [Durante.]**	CON:sub	ISCHUNK+	\RB	tout	auto	1
506	R. copie Arbelet Rés., volume cahier n° 50, feuillet 6, verso	Leo	comme un original. On le traite ainsi comme Leo, Scarlatti, Porpora α Durante.	**[comme] [un original][.] [On le traite] [ainsi] [comme] [Leo][,] [Scarlatti][,] [Porpora] [α] [Durante.]**	NPP	ISCHUNK+	\RB	tout	auto	1
506	R. copie Arbelet Rés., volume cahier n° 50, feuillet 6, verso	,	comme un original. On le traite ainsi comme Leo, Scarlatti, Porpora α Durante.	**[comme] [un original][.] [On le traite] [ainsi] [comme] [Leo][,] [Scarlatti][,] [Porpora] [α] [Durante.]**	PONCT	ISCHUNK+	\RB	tout	auto	1
506	R. copie Arbelet Rés., volume cahier n° 50, feuillet 6, verso	Scarlatti	comme un original. On le traite ainsi comme Leo, Scarlatti, Porpora α Durante.	**[comme] [un original][.] [On le traite] [ainsi] [comme] [Leo][,] [Scarlatti][,] [Porpora] [α] [Durante.]**	NPP	ISCHUNK+	\RB	tout	auto	1
506	R. copie Arbelet Rés., volume cahier n° 50, feuillet 6, verso	,	comme un original. On le traite ainsi comme Leo, Scarlatti, Porpora α Durante.	**[comme] [un original][.] [On le traite] [ainsi] [comme] [Leo][,] [Scarlatti][,] [Porpora] [α] [Durante.]**	PONCT	ISCHUNK+	\RB	tout	auto	1
506	R. copie Arbelet Rés., volume cahier n° 50, feuillet 6, verso	Porpora	comme un original. On le traite ainsi comme Leo, Scarlatti,	**[comme] [un original][.] [On le traite] [ainsi] [comme] [Leo][,] [Scarlatti][,] [Porpora] [α] [Durante.]**	NPP	ISCHUNK+	\RB	tout	auto	1

			Porpora α Durante.							
506	R. copie Arbelet Rés., volume cahier n° 50, feuillet 6, verso	α	comme un original. On le traite ainsi comme Leo, Scarlatti, Porpora α Durante.	**[comme] [un original][.] [On le traite] [ainsi] [comme] [Leo][,] [Scarlatti][,] [Porpora] [α] [Durante.]**	SYM	ISCHUNK+	\RB	tout	auto	1
506	R. copie Arbelet Rés., volume cahier n° 50, feuillet 6, verso	Durante	comme un original. On le traite ainsi comme Leo, Scarlatti, Porpora α Durante.	**[comme] [un original][.] [On le traite] [ainsi] [comme] [Leo][,] [Scarlatti][,] [Porpora] [α] [Durante.]**	NPP	ISCHUNK+	\RB	tout	auto	1
506	R. copie Arbelet Rés., volume cahier n° 50, feuillet 6, verso	.	comme un original. On le traite ainsi comme Leo, Scarlatti, Porpora α Durante.	**[comme] [un original][.] [On le traite] [ainsi] [comme] [Leo][,] [Scarlatti][,] [Porpora] [α] [Durante.]**	PONCT	ISCHUNK+	\RB	tout	auto	1
503	R. copie Arbelet Rés., volume cahier n° 50, feuillet 5, recto	*...*	*...*	[Naples] [7**…*** bre] [10 8 bre 1803]	SUBS	ISTOK-	0	mil	auto	2
503	R. copie Arbelet Rés., volume cahier n° 50, feuillet 5, recto	mais	mais en 1803 Il n'y en avait plus que trois	[**mais en 1803] [Il n'y en avait plus] [que] [trois**]	CON:cc	ISCHUNK+	tout		auto	2
503	R. copie Arbelet Rés., volume cahier n° 50, feuillet 5, recto	en	mais en 1803 Il n'y en avait plus que trois	[**mais en 1803] [Il n'y en avait plus] [que] [trois**]	PREP	ISCHUNK+	debut		auto	2
503	R. copie Arbelet Rés., volume cahier n° 50, feuillet 5, recto	1803	mais en 1803 Il n'y en avait plus que trois	[**mais en 1803] [Il n'y en avait plus] [que] [trois**]	NUM	ISCHUNK+	fin		auto	2
503	R. copie Arbelet Rés., volume cahier n° 50, feuillet	il	mais en 1803 Il n'y en avait plus que trois	[**mais en 1803] [Il n'y en avait plus] [que] [trois**]	PRON:pers	ISCHUNK+	debut		auto	2

	5, recto									
503	R. copie Arbelet Rés., volume cahier n° 50, feuillet 5, recto	n'	mais en 1803 Il n'y en avait plus que trois	[**mais en 1803] [Il n'y en avait plus] [que] [trois**]	ADV:neg	ISCHUNK+	mil		auto	2
503	R. copie Arbelet Rés., volume cahier n° 50, feuillet 5, recto	y	mais en 1803 Il n'y en avait plus que trois	[**mais en 1803] [Il n'y en avait plus] [que] [trois**]	ADV	ISCHUNK+	mil		auto	2
503	R. copie Arbelet Rés., volume cahier n° 50, feuillet 5, recto	en	mais en 1803 Il n'y en avait plus que trois	[**mais en 1803] [Il n'y en avait plus] [que] [trois**]	PREP	ISCHUNK+	mil		auto	2
503	R. copie Arbelet Rés., volume cahier n° 50, feuillet 5, recto	avait	mais en 1803 Il n'y en avait plus que trois	[**mais en 1803] [Il n'y en avait plus] [que] [trois**]	v:imp	ISCHUNK+	mil		auto	2
503	R. copie Arbelet Rés., volume cahier n° 50, feuillet 5, recto	plus	mais en 1803 Il n'y en avait plus que trois	[**mais en 1803] [Il n'y en avait plus] [que] [trois**]	ADV	ISCHUNK+	fin		auto	2
503	R. copie Arbelet Rés., volume cahier n° 50, feuillet 5, recto	que	mais en 1803 Il n'y en avait plus que trois	[**mais en 1803] [Il n'y en avait plus] [que] [trois**]	CON:sub	ISCHUNK+	tout		auto	2
503	R. copie Arbelet Rés., volume cahier n° 50, feuillet 5, recto	trois	mais en 1803 Il n'y en avait plus que trois	[**mais en 1803] [Il n'y en avait plus] [que] [trois**]	ADJ	ISCHUNK+	tout		auto	2
503	R. copie Arbelet Rés., volume cahier n° 50, feuillet 5, recto	50	~~50~~	[Il y a plus] [de véritable] [amour] [pour cet art] [dans **50** *10* lazzaroni] [que] [dans tout *un* **le** public] [**qui**] [**s'extasie**] [**un Dimanche**] [**au conservatoire**] [**de la rue Bergère**][**.**]	NUM	ISTOK	0	mil	auto	2

503	R. copie Arbelet Rés., volume cahier n° 50, feuillet 5, recto	le	~~le~~	[Il y a plus] [de véritable] [amour] [pour cet art] [dans **50** *10* lazzaroni] [que] [dans tout *un* **le** public] [**qui**] [**s'extasie**] [**un Dimanche**] [**au conservatoire**] [**de la rue Bergère**][**.**]	DET:art	ISTOK	0	mil	auto	2
503	R. copie Arbelet Rés., volume cahier n° 50, feuillet 5, recto	qui	~~qui s'extasie un Dimanche au conservatoire de la rue Bergère~~	[Il y a plus] [de véritable] [amour] [pour cet art] [dans **50** *10* lazzaroni] [que] [dans tout *un* **le** public] [**qui**] [**s'extasie**] [**un Dimanche**] [**au conservatoire**] [**de la rue Bergère**][**.**]	PRON:rel	ISCHUNK+	0	tout	auto	2
503	R. copie Arbelet Rés., volume cahier n° 50, feuillet 5, recto	s'	~~qui s'extasie un Dimanche au conservatoire de la rue Bergère~~	[Il y a plus] [de véritable] [amour] [pour cet art] [dans **50** *10* lazzaroni] [que] [dans tout *un* **le** public] [**qui**] [**s'extasie**] [**un Dimanche**] [**au conservatoire**] [**de la rue Bergère**][**.**]	PRON:dir	ISCHUNK+	0	debut	auto	2
503	R. copie Arbelet Rés., volume cahier n° 50, feuillet 5, recto	extasie	~~qui s'extasie un Dimanche au conservatoire de la rue Bergère~~	[Il y a plus] [de véritable] [amour] [pour cet art] [dans **50** *10* lazzaroni] [que] [dans tout *un* **le** public] [**qui**] [**s'extasie**] [**un Dimanche**] [**au conservatoire**] [**de la rue Bergère**][**.**]	v:ind	ISCHUNK+	0	fin	auto	2
503	R. copie Arbelet Rés., volume cahier n° 50, feuillet 5, recto	un	~~qui s'extasie un Dimanche au conservatoire de la rue Bergère~~	[Il y a plus] [de véritable] [amour] [pour cet art] [dans **50** *10* lazzaroni] [que] [dans tout *un* **le** public] [**qui**] [**s'extasie**] [**un Dimanche**] [**au conservatoire**] [**de la rue Bergère**][**.**]	DET:art	ISCHUNK+	0	debut	auto	2
503	R. copie Arbelet Rés., volume cahier n° 50, feuillet 5, recto	Dimanche	~~qui s'extasie un Dimanche au conservatoire de la rue Bergère~~	[Il y a plus] [de véritable] [amour] [pour cet art] [dans **50** *10* lazzaroni] [que] [dans tout *un* **le** public] [**qui**] [**s'extasie**] [**un Dimanche**] [**au conservatoire**] [**de la rue Bergère**][**.**]	SUBS	ISCHUNK+	0	fin	auto	2

503	R. copie Arbelet Rés., volume cahier n° 50, feuillet 5, recto	au	~~qui s'extasie un Dimanche au conservatoire de la rue Bergère~~	[Il y a plus] [de véritable] [amour] [pour cet art] [dans **50** *10* lazzaroni] [que] [dans tout *un* **le** public] [**qui**] [**s'extasie**] [**un Dimanche**] [**au conservatoire**] [**de la rue Bergère**][**.**]	PREP:det	ISCHUNK+	0	debut	auto	2
503	R. copie Arbelet Rés., volume cahier n° 50, feuillet 5, recto	conservatoire	~~qui s'extasie un Dimanche au conservatoire de la rue Bergère~~	[Il y a plus] [de véritable] [amour] [pour cet art] [dans **50** *10* lazzaroni] [que] [dans tout *un* **le** public] [**qui**] [**s'extasie**] [**un Dimanche**] [**au conservatoire**] [**de la rue Bergère**][**.**]	SUBS	ISCHUNK+	0	fin	auto	2
503	R. copie Arbelet Rés., volume cahier n° 50, feuillet 5, recto	de	~~qui s'extasie un Dimanche au conservatoire de la rue Bergère~~	[Il y a plus] [de véritable] [amour] [pour cet art] [dans **50** *10* lazzaroni] [que] [dans tout *un* **le** public] [**qui**] [**s'extasie**] [**un Dimanche**] [**au conservatoire**] [**de la rue Bergère**][**.**]	PREP	ISCHUNK+	0	debut	auto	2
503	R. copie Arbelet Rés., volume cahier n° 50, feuillet 5, recto	la	~~qui s'extasie un Dimanche au conservatoire de la rue Bergère~~	[Il y a plus] [de véritable] [amour] [pour cet art] [dans **50** *10* lazzaroni] [que] [dans tout *un* **le** public] [**qui**] [**s'extasie**] [**un Dimanche**] [**au conservatoire**] [**de la rue Bergère**][**.**]	DET:art	ISCHUNK+	0	mil	auto	2
503	R. copie Arbelet Rés., volume cahier n° 50, feuillet 5, recto	rue	~~qui s'extasie un Dimanche au conservatoire de la rue Bergère~~	[Il y a plus] [de véritable] [amour] [pour cet art] [dans **50** *10* lazzaroni] [que] [dans tout *un* **le** public] [**qui**] [**s'extasie**] [**un Dimanche**] [**au conservatoire**] [**de la rue Bergère**][**.**]	SUBS	ISCHUNK+	0	mil	auto	2
503	R. copie Arbelet Rés., volume cahier n° 50, feuillet 5, recto	Bergère	~~qui s'extasie un Dimanche au conservatoire de la rue Bergère~~	[Il y a plus] [de véritable] [amour] [pour cet art] [dans **50** *10* lazzaroni] [que] [dans tout *un* **le** public] [**qui**] [**s'extasie**] [**un Dimanche**] [**au conservatoire**] [**de la rue Bergère**][**.**]	NPP	ISCHUNK+	0	fin	auto	2

503	R. copie Arbelet Rés., volume cahier n° 50, feuillet 5, recto	tems	tems où les mœurs étaient si gaies Paris sous le régent.	[Les grands artistes] [que] [Naples] [a produits] [vécurent] [vers 1726] **[tems] [où] [les mœurs étaient si gaies] [Paris sous le régent.]**	SUBS	ISCHUNK+	0	debut	auto	2
503	R. copie Arbelet Rés., volume cahier n° 50, feuillet 5, recto	où	tems où les mœurs étaient si gaies Paris sous le régent.	[Les grands artistes] [que] [Naples] [a produits] [vécurent] [vers 1726] **[tems] [où] [les mœurs étaient si gaies] [Paris sous le régent.]**	ADV	ISCHUNK+	0	tout	auto	2
503	R. copie Arbelet Rés., volume cahier n° 50, feuillet 5, recto	les	tems où les mœurs étaient si gaies Paris sous le régent.	[Les grands artistes] [que] [Naples] [a produits] [vécurent] [vers 1726] **[tems] [où] [les mœurs étaient si gaies] [Paris sous le régent.]**	DET:art	ISCHUNK+	0	debut	auto	2
503	R. copie Arbelet Rés., volume cahier n° 50, feuillet 5, recto	mœurs	tems où les mœurs étaient si gaies Paris sous le régent.	[Les grands artistes] [que] [Naples] [a produits] [vécurent] [vers 1726] **[tems] [où] [les mœurs étaient si gaies] [Paris sous le régent.]**	SUBS	ISCHUNK+	0	fin	auto	2
503	R. copie Arbelet Rés., volume cahier n° 50, feuillet 5, recto	étaient	tems où les mœurs étaient si gaies Paris sous le régent.	[Les grands artistes] [que] [Naples] [a produits] [vécurent] [vers 1726] **[tems] [où] [les mœurs étaient si gaies] [Paris sous le régent.]**	v:imp	ISCHUNK+	0	tout	auto	2
503	R. copie Arbelet Rés., volume cahier n° 50, feuillet 5, recto	si	tems où les mœurs étaient si gaies Paris sous le régent.	[Les grands artistes] [que] [Naples] [a produits] [vécurent] [vers 1726] **[tems] [où] [les mœurs étaient si gaies] [Paris sous le régent.]**	CON:cc	ISCHUNK+	0	tout	auto	2

503	R. copie Arbelet Rés., volume cahier n° 50, feuillet 5, recto	gaies	tems où les mœurs étaient si gaies Paris sous le régent.	[Les grands artistes] [que] [Naples] [a produits] [vécurent] [vers 1726] **[tems] [où] [les mœurs étaient si gaies] [Paris sous le régent.]**	ADJ	ISCHUNK+	0	tout	auto	2
503	R. copie Arbelet Rés., volume cahier n° 50, feuillet 5, recto	Paris	tems où les mœurs étaient si gaies Paris sous le régent.	[Les grands artistes] [que] [Naples] [a produits] [vécurent] [vers 1726] **[tems] [où] [les mœurs étaient si gaies] [Paris sous le régent.]**	NPP	ISCHUNK+	0	tout	auto	2
503	R. copie Arbelet Rés., volume cahier n° 50, feuillet 5, recto	sous	tems où les mœurs étaient si gaies Paris sous le régent.	[Les grands artistes] [que] [Naples] [a produits] [vécurent] [vers 1726] **[tems] [où] [les mœurs étaient si gaies] [Paris sous le régent.]**	PREP	ISCHUNK+	0	debut	auto	2
503	R. copie Arbelet Rés., volume cahier n° 50, feuillet 5, recto	le	tems où les mœurs étaient si gaies Paris sous le régent.	[Les grands artistes] [que] [Naples] [a produits] [vécurent] [vers 1726] **[tems] [où] [les mœurs étaient si gaies] [Paris sous le régent.]**	DET:art	ISCHUNK+	0	mil	auto	2
503	R. copie Arbelet Rés., volume cahier n° 50, feuillet 5, recto	régent	tems où les mœurs étaient si gaies Paris sous le régent.	[Les grands artistes] [que] [Naples] [a produits] [vécurent] [vers 1726] **[tems] [où] [les mœurs étaient si gaies] [Paris sous le régent.]**	SUBS	ISCHUNK+	0	fin	auto	2
503	R. copie Arbelet Rés., volume cahier n° 50, feuillet 5, recto	.	tems où les mœurs étaient si gaies Paris sous le régent.	[Les grands artistes] [que] [Naples] [a produits] [vécurent] [vers 1726] **[tems] [où] [les mœurs étaient si gaies] [Paris sous le régent.]**	PONCT	ISCHUNK+	0	tout	auto	2
501	R. copie Arbelet Rés., volume cahier n° 50, feuillet	Collo	~~Collo~~	[et] [**Collo** *Capo] [*di Monti]	NPP	ISCHUNK	0	tout	auto	2

	4, recto									
501	R. copie Arbelet Rés., volume cahier n° 50, feuillet 4, recto	Lou	~~Lou~~	[Volupté] [du Roi **Loui** Joseph]	NPP	ISTOK	0	mil	auto	1
500	R. copie Arbelet Rés., volume cahier n° 50, feuillet 3, verso	eut	~~eut~~	[dans chaque bouteille] [duquel] [on **eut** *aurait* fait fondre] [2 livres] [de sucre][.]	v:ps	ISTOK	0	mil	auto	2
553	R. copie Arbelet Rés., volume cahier n° 50, feuillet 2, verso	a	~~a~~	[C'est] [je crois] [Frédérick][=][=][**a** Gasse]	PREP	ISTOK	0	debut	auto	1
553	R. copie Arbelet Rés., volume cahier n° 50, feuillet 2, verso	ni	~~ni~~	[Raoul] [de Créqui] [de Fioravanti**ni**]	NPP	ISTOK-	0	fin	auto	1
499	R. copie Arbelet Rés., volume cahier n° 50, feuillet 3, recto	E	~~E~~	[La plus belle vue] [du monde] [probablement **E** est] [celle] [dont] [on jouit] [de la maison] [de l' hermite] [.]	v:ind	ISTOK-	0	mil	auto	1
499	R. copie Arbelet Rés., volume cahier n° 50, feuillet 3, recto	Scheger	~~Scheger~~	[les noms] [de M.e] [de Staël] [et][de **Scheger** Selegel]	NPP	ISTOK	0	fin	auto	1
499	R. copie Arbelet Rés., volume cahier n° 50, feuillet 3, recto	Selegel	~~Sclegel~~	[de **Selegel**]**[.]** [Shleghel.]	NPP	ISTOK	0	fin	auto	1
497	R. copie Arbelet Rés., volume cahier n° 50, feuillet 2, recto	-1	~~-1~~	[**(1)**]	SYM	ISCHUNK	0	tout	auto	1
266	R. 5896 Rés., volume 22, feuillet 41, recto	me	~~me dit~~	[Vu] [Dalban] [et] [Lavaudou] [rue jaque] [n.° 139.] [D.] [**me dit**] [juge F. mallein] [et] [fr.] [Faure] [comme moi][.]	PRON:dir	ISCHUNK	0	debut	auto	1

266	R. 5896 Rés., volume 22, feuillet 41, recto	dit	~~me dit~~	[Vu] [Dalban] [et] [Lavaudou] [rue jaque] [n.° 139.] [D.] [**me dit**] [juge F. mallein] [et] [fr.] [Faure] [comme moi][.]	v:ind	ISCHUNK	0	fin	auto	1
275	R. 5896 Rés., volume 22, feuillet 45, verso	donna	~~donna~~	[la **donna** vedova scaltra]	NPP	ISTOK	0	mil	auto	1
276	R. 5896 Rés., volume 22, feuillet 46, recto	serai	~~serai~~	[à l'avenir] [lorsque] [je **serai** *devrai être* présenté] [à quelcun]	v:fs	ISTOK	0	mil	auto	2
278	R. 5896 Rés., volume 22, feuillet 47, recto	par	~~par~~	[il aurait] [pu] [**par** encore donner] [des explications]	PREP	ISTOK	0	debut	auto	1
280	R. 5896 Rés., volume 22, feuillet 48, recto	était	~~était~~	[il **était** avait] [un habit bleu]	v:imp	ISTOK	0	fin	auto	1
283	R. 5896 Rés., volume 22, feuillet 49, verso	m'	~~m'~~	[M.lle D.] [**m'**avait] [témoigné] [1.h auparavant] [dans sa loge]	DET:pos	ISTOK	0	debut	auto	1
284	R. 5896 Rés., volume 22, feuillet 50, recto		'	[a letter] [to my great father] [upon neuilly house][’] [intrigue]	PONCT	ISCHUNK	0	tout	auto	1
323	R. 5896 Rés., volume 16, feuillet 4, verso	4h	~~4h~~	[nous sortons] [de Gemenos] [à **4h** 5h][.] [route fraiche] [et] [assez agréable] [au fonds] [d’un petit vallon]	SYM	ISTOK	0	fin	auto	1
323	R. 5896 Rés., volume 16, feuillet 4, verso	grands	~~grands~~	[ruisseau bordé] [de **grands** peupliers] [d’Italie] [dans toute leur fraicheur][.]	ADJ	ISTOK	0	mil	auto	1
323	R. 5896 Rés., volume 16, feuillet 4, verso	arride	~~arride et~~	[descente pleine **arride] [et]** [/humide/ froide] [*/et] [sans arbres/*] [de 1.h]	ADJ	ISCHUNK+	\RB	fin	auto	2
323	R. 5896 Rés., volume 16, feuillet 4, verso	et	~~arride et~~	[descente pleine **arride] [et]** [/humide/ froide] [*/et] [sans arbres/*] [de 1.h]...	CON:cc	ISCHUNK+	\RB	tout	auto	2
344	R. 5896 Rés., volume 22,	mon	~~mon~~	[Si mes loteries] [de **mon** octavien=arrigo //henri //=][fair Monfort] [et] [du	PRON:pos	ISTOK	0	mil	auto	2

	feuillet 26, verso			Pervertisseur] [reussissent] [Je puis avoir] [de 2 000 f.] [à 6 000 f.] [à manger] [cet hiver]						
345	R. 5896 Rés., volume 22, feuillet 27, verso	à	~~à~~	[et] [l'avidité] [qu'] [**à** a un hom][.] [qui] [n'est pas] [habitué] [à la considération] [de rapeler] [sans cesse] [à soi] [et] [aux autres] [celle] [qu'] [il a instantanément][.]	v:ind	ISTOK	0	debut	auto	1
349	R. 5896 Rés., volume 22, feuillet 29, verso	détresse	~~détresse~~	[=] [il nous parle] [de la pénurie **détresse** escroquante] [du Marquis] [de L'Angle][.]	SUBS	ISTOK	0	mil	auto	2
351	R. 5896 Rés., volume 22, feuillet 30, verso	de	~~de ce grand~~	[Je retrouve] [dans les écrits] [**de ce grand** di quel grande] [plusieurs] [des Pensées] [que] [j'avais déjà] [eues][.]	PREP	ISCHUNK	\RB	debut	auto	1
351	R. 5896 Rés., volume 22, feuillet 30, verso	ce	~~de ce grand~~	[Je retrouve] [dans les écrits] [**de ce grand** di quel grande] [plusieurs] [des Pensées] [que] [j'avais déjà] [eues][.]	PRON:dem	ISCHUNK	\RB	mil	auto	1
351	R. 5896 Rés., volume 22, feuillet 30, verso	grand	~~de ce grand~~	[Je retrouve] [dans les écrits] [**de ce grand** di quel grande] [plusieurs] [des Pensées] [que] [j'avais déjà] [eues][.]	ADJ	ISCHUNK	\RB	fin	auto	1
353	R. 5896 Rés., volume 22, feuillet 31, verso		~~le~~	[S'il les écartait] [il produirait] [sur l'ame] [**le** un Sentiment délicieux] [du même genre] [que] [celui qu'y fait] [naitre] [une églogue] [de Virgile][.]	DET:art	ISTOK	0	debut	auto	2
353	R. 5896 Rés., volume 22, feuillet 31, verso		~~te~~	[Il a produit quelquefois] [cet**te** effet] [sur moi] [dans son Charmant role] [de Zéphire][.]	DET:dem	ISTOK-	0	debut	auto	2
353	R. 5896 Rés., volume 22, feuillet 31, verso	à	~~à la~~	[Psiché] [**à la** m'a charmé]	PREP	ISTOK+	0	debut	auto	1
353	R. 5896 Rés., volume 22, feuillet 31, verso	la	~~à la~~	[Psiché] [**à la** m'a charmé]	DET:art	ISTOK+	0	mil	auto	1
354	R. 5896 Rés., volume 22, feuillet 32, recto	une	~~une s~~	[Tristesse] [redouble] [dans **une s** l'ame sensible] [d'Isabelle]	DET:art	ISTOK+	0	mil	auto	1

354	R. 5896 Rés., volume 22, feuillet 32, recto	s	~~une s~~	[Tristesse] [redouble] [dans **une s** l'ame sensible] [d'Isabelle]	SUBS:amorce	ISTOK+	0	mil	auto	1
356	R. 5896 Rés., volume 22, feuillet 33, recto	comme	~~comme~~	[**comme**] [()][valent][=][elles] [les belles scênes] [de Shakespeare] [?][)] [mais] [il y a] [de grands défauts] [de Scenegiatura]	CON:sub	ISCHUNK	0	tout	auto	1
357	R. 5896 Rés., volume 22, feuillet 33, verso	il	~~il s'agit~~	[au lieu] [que] [dans R. diable] **[il s'agit]** [cela vous touche][.]	PRON:pers	ISCHUNK	0	debut	auto	1
357	R. 5896 Rés., volume 22, feuillet 33, verso	s'	~~il s'agit~~	[au lieu] [que] [dans R. diable] **[il s'agit]** [cela vous touche][.]	PRON:dir	ISCHUNK	0	mil	auto	1
357	R. 5896 Rés., volume 22, feuillet 33, verso	s'agit	~~il s'agit~~	[au lieu] [que] [dans R. diable] **[il s'agit]** [cela vous touche][.]	v:ind	ISCHUNK	0	fin	auto	1
357	R. 5896 Rés., volume 22, feuillet 33, verso	,	~~,~~	[C'est][,] [lui] [,] [réponds=je][,]	PONCT	ISCHUNK	0	tout	indeterminé	?
358	R. 5896 Rés., volume 22, feuillet 34, recto	et	~~et~~	[il semble] [qu'][il n'ait manqué] [à ce Shakespeare] [si] [naturel] [**et**] [si] [passionné][,] [et] [si] [fort][,]	CON:cc	ISCHUNK	\RB	tout	auto	1
359	R. 5896 Rés., volume 22, feuillet 34, verso	des	~~des~~	[ceux] [de Voltaire] [**des** suposent presque tous] [le Caractére] [du roi][de Prusse]	PREP:det	ISTOK	0	debut	auto	1
360	R. 5896 Rés., volume 22, feuillet 35, recto	par	~~par ex~~	[une peinture] [de Couvent] [**par ex** cette piéce] [se raproche] [du sistême] [d'Alfieri][.]	PREP	ISCHUNK	0	debut	auto	1
360	R. 5896 Rés., volume 22, feuillet 35, recto	ex	~~par ex~~	[une peinture] [de Couvent] [**par ex** cette piéce] [se raproche] [du sistême] [d'Alfieri][.]	SUBS	ISCHUNK	0	fin	auto	1
361	R. 5896 Rés., volume 22, feuillet 35, verso	Paroles	~~Paroles et~~	**[Paroles] [et]** [musique sans nul mérite][,] [Paroles] [du dernier bête]	CON:cc	ISCHUNK+	\RB	tout	indeterminé	?
361	R. 5896 Rés., v	et	~~Paroles et~~	**[Paroles] [et]** [musique sans nul mérite][,]	SUBS	ISCHUNK+	\RB	tout	indeterminé	?

	olume 22, feuillet 35, verso			[Paroles] [du dernier bête]						
362	R. 5896 Rés., volume 22, feuillet 36, recto	un	~~un Concer~~	[qui] [m'a apris] [qu'] [à Perpignan] [les habitans] [avaient] [donné] [**un Concer** une Sérénade] [à M.e Moreau][.]	DET:art	ISCHUNK	\RB	debut	auto	1
362	R. 5896 Rés., volume 22, feuillet 36, recto	Concer	~~un Concer~~	[qui] [m'a apris] [qu'] [à Perpignan] [les habitans] [avaient] [donné] [**un Concer** une Sérénade] [à M.e Moreau][.]	SUBS:amorce	ISCHUNK	\RB	fin	auto	1
363	R. 5896 Rés., volume 22, feuillet 36, verso	t	~~t~~	[Il semble] [que] [dans le poème] [on **t** ait évité exprès] [ce qui] [pouvait être bon][.]	UKN	ISTOK-	0	mil	auto	1
363	R. 5896 Rés., volume 22, feuillet 36, verso	mais	~~mais~~	[**mais** mais] [non charmantes] [on voit] [que] [ce] [n'est pas] [a comprehensive Soul] [qui] [les a] [faites][.]	CON:cc	ISCHUNK	\RB	tout	auto	1
364	R. 5896 Rés., volume 22, feuillet 37, recto	préférant	~~préférant son intérêt~~	[**préférant son intérêt]** [nullement amoureux][de la gloire]	VERBE	ISCHUNK	0	debut	auto	1
364	R. 5896 Rés., volume 22, feuillet 37, recto	son	~~préférant son intérêt~~	[**préférant son intérêt]** [nullement amoureux][de la gloire]	DET:pos	ISCHUNK	0	mil	auto	1
364	R. 5896 Rés., volume 22, feuillet 37, recto	intérêt	~~préférant son intérêt~~	[**préférant son intérêt]** [nullement amoureux][de la gloire]	SUBS	ISCHUNK	0	fin	auto	1
365	R. 5896 Rés., volume 22, feuillet 37, verso	e	~~e~~	[,] [suivi **e**] [de l'Entrevue platitude] [de Vigée][.]	v:ind	ISTOK-	0	fin	auto	1
366	R. 5896 Rés., volume 22, feuillet 38, recto	n'	~~n'auraie~~	[qui] [**n'aurai**e ne seraient pas échappées] [à Picard]	ADV:neg	ISTOK+	\RB	debut	auto	1
366	R. 5896 Rés., volume 22, feuillet 38, recto	auraie	~~n'auraie~~	[qui] [**n'aurai**e ne seraient pas échappées] [à Picard]	v:cond	ISTOK+	\RB	mil	auto	1

366	R. 5896 Rés., volume 22, feuillet 38, recto	*...*	*...*	[***…*** maladie] [du gout] [,] [sous la Monarchie][.]	UKN	ISTOK	0	debut	auto	1
3609	R. 5896 Rés., volume 25, feuillet 120, recto	;	~~;~~	[**;**] [notre passion] [en parlant ensemble] [est] [la Vanité][,]	PONCT	ISTOK-	0	tout	auto	1
386	R. 5896 Rés., volume 25, feuillet 121, recto	par	~~par malheur~~	[Mais] **[par malheur** *dans son monde-ci*] [l'homme] [est] [un peu machine]	PREP	ISCHUNK	0	debut	auto	2
386	R. 5896 Rés., volume 25, feuillet 121, recto	malheur	~~par malheur~~	[Mais] **[par malheur** *dans son monde-ci*] [l'homme] [est] [un peu machine]	SUBS	ISCHUNK	0	fin	auto	2
389	R. 5896 Rés., volume 25, feuillet 122, verso	d'	~~d'étonner~~	[le Pont de bateaux] [ne fait] [d'autre effet] [que] [**d'étonner**] [de faire] [desirer] [un Pont fixe][.]	PREP	ISCHUNK	\RB	debut	auto	1
389	R. 5896 Rés., volume 25, feuillet 122, verso	étonner	~~d'étonner~~	[le Pont de bateaux] [ne fait] [d'autre effet] [que] [**d'étonner**] [de faire] [desirer] [un Pont fixe][.]	v:inf	ISCHUNK	\RB	fin	auto	1
394	R. 5896 Rés., volume 25, feuillet 125, recto	cass	~~cass~~	[il nous a] [montré] [des vitres] [de sa lanterne **cass**][qui] [ont été **cas br** félées][par des graviers] [que] [le vent] [soulevait] [et] [portait] [jusqu'][à cette lanterne][.]	v:amorce	ISTOK-	\RI	fin	auto	1
394	R. 5896 Rés., volume 25, feuillet 125, recto	cas	~~cas~~	[il nous a] [montré] [des vitres] [de sa lanterne **cass**][qui] [ont été **cas br** félées][par des graviers] [que] [le vent] [soulevait] [et] [portait] [jusqu'][à cette lanterne][.]	v:amorce	ISTOK-	\RI	fin	auto	1
394	R. 5896 Rés., volume 25, feuillet 125, recto	br	~~br~~	[il nous a] [montré] [des vitres] [de sa lanterne **cass**][qui] [ont été **cas br** félées][par des graviers] [que] [le vent] [soulevait] [et] [portait] [jusqu'][à cette lanterne][.]	v:amorce	ISTOK-	0	fin	auto	1

390	R. 5896 Rés., volume 25, feuillet 123, recto	taint	~~taint~~	[mais] [au milieu] [des **taint** teinturiers] [de Rouen]	SUBS:amorce	ISTOK-	\RB	mil	auto	1
390	R. 5896 Rés., volume 25, feuillet 123, recto	L	~~L~~	[**L** Ce matin] [30] [à 5 heures] [**du matin**] [nous sommes] [partis] [de Rouen] [pour le Hâvre] [et] [nous avons] [vû] [un beau quartier] [le long] [de la Seine]	DET:art	ISTOK-	0	debut	auto	1
390	R. 5896 Rés., volume 25, feuillet 123, recto	du	~~du matin~~	[**L** Ce matin] [30] [à 5 heures] [**du matin**] [nous sommes] [partis] [de Rouen] [pour le Hâvre] [et] [nous avons] [vû] [un beau quartier] [le long] [de la Seine]	PREP:det	ISCHUNK	0	fin	auto	1
390	R. 5896 Rés., volume 25, feuillet 123, recto	matin	~~du matin~~	[**L** Ce matin] [30] [à 5 heures] [**du matin**] [nous sommes] [partis] [de Rouen] [pour le Hâvre] [et] [nous avons] [vû] [un beau quartier] [le long] [de la Seine]	SUBS	ISCHUNK	0	fin	auto	1
391	R. 5896 Rés., volume 25, feuillet 123, verso	et	~~et Honfleur~~	[,] [**et] [Honfleur** le hâvre] [et] [Honfleur] [qui] [est] [de l'autre coté]	CON:cc	ISCHUNK+	\RB	tout	auto	1
391	R. 5896 Rés., volume 25, feuillet 123, verso	Honfleur	~~et Honfleur~~	[,] [**et] [Honfleur** le hâvre] [et] [Honfleur] [qui] [est] [de l'autre coté]	NPP	ISCHUNK+	\RB	tout	auto	1
391	R. 5896 Rés., volume 25, feuillet 123, verso	situées	~~situées sur un coteau qui~~	[cette suite] [de maisons][se termine] [par une grande quantité] [de maisons] [du même genre] [**situées**] [**sur un coteau**] [**qui**] [s'élevant] [en amphithéatre] [sur le côteau] [qui] [s'éléve] [lui même] [en cet endroit] [et] [qui] [est juste] [à l'extrêmité] [de la grande rue] [du Hâvre][.]	v:pc	ISCHUNK+		tout	auto	1
391	R. 5896 Rés., volume 25, feuillet 123, verso	sur	~~situées sur un coteau qui~~	[cette suite] [de maisons][se termine] [par une grande quantité] [de maisons] [du même genre] [**situées**] [**sur un coteau**] [**qui**] [s'élevant] [en amphithéatre] [sur le côteau] [qui] [s'éléve] [lui même] [en cet endroit] [et] [qui] [est juste] [à l'extrêmité] [de la grande rue] [du Hâvre][.]	PREP	ISCHUNK+		debut	auto	1
391	R. 5896 Rés., volume 25, feuillet 123, verso	un	~~situées sur un coteau qui~~	[cette suite] [de maisons][se termine] [par une grande quantité] [de maisons] [du même genre] [**situées**] [**sur un coteau**] [**qui**] [s'élevant] [en amphithéatre] [sur le	DET:art	ISCHUNK+		mil	auto	1

				côteau] [qui] [s'éléve] [lui même] [en cet endroit] [et] [qui] [est juste] [à l'extrêmité] [de la grande rue] [du Hâvre][.]						
391	R. 5896 Rés., volume 25, feuillet 123, verso	coteau	~~situées sur un coteau qui~~	[cette suite] [de maisons][se termine] [par une grande quantité] [de maisons] [du même genre] [**situées**] [**sur un coteau**] [**qui**] [s'élevant] [en amphithéatre] [sur le côteau] [qui] [s'éléve] [lui même] [en cet endroit] [et] [qui] [est juste] [à l'extrêmité] [de la grande rue] [du Hâvre][.]	SUBS	ISCHUNK+		fin	auto	1
391	R. 5896 Rés., volume 25, feuillet 123, verso	qui	~~situées sur un coteau qui~~	[cette suite] [de maisons][se termine] [par une grande quantité] [de maisons] [du même genre] [**situées**] [**sur un coteau**] [**qui**] [s'élevant] [en amphithéatre] [sur le côteau] [qui] [s'éléve] [lui même] [en cet endroit] [et] [qui] [est juste] [à l'extrêmité] [de la grande rue] [du Hâvre][.]	PRON:rel	ISCHUNK+		tout	auto	1
392	R. 5896 Rés., volume 25, feuillet 124, recto	l	~~l~~	[l' extrémité] [de cette rue] [opposée] [au coteau] [est **l** un] [des bassins][.]	DET:art	ISTOK-	0	mil	auto	1
392	R. 5896 Rés., volume 25, feuillet 124, recto	beaucoup	~~beaucoup~~	[Nous avons] [vû] [sur la mer] [**beaucoup**] [**de** /*quelques*/ bateaux pêcheurs] [(][40] [ou] [50][)]	ADV	ISCHUNK+	0	tout	auto	2
392	R. 5896 Rés., volume 25, feuillet 124, recto	de	~~de~~	[**beaucoup de** //*quelques*// bateaux pêcheurs]	PREP	ISCHUNK+		debut	auto	2
392	R. 5896 Rés., volume 25, feuillet 124, recto	ext	~~ext~~	[le bruit] [des vagues] [quoique] [la mer] [fut calme] [**ext** excite vivement] [l'attention][.]	v:amorce	ISTOK-	0	debut	auto	1
393	R. 5896 Rés., volume 25, feuillet 124, verso	à	~~à~~	[Nous avons] [monté] [à la lanterne] [d'un] [des phares] [et] [de là] [nous avons] [vû] [la côte] [de Cherbourg] [à environ 30 lieues] [de nous] [et] [**à** devant nous] [toute la plaine liquide][.]	PREP	ISTOK	0	debut	auto	1
393	R. 5896 Rés., volume 25, feuillet 124, verso	que	~~que~~	[**que**] [nous avons] [entendu] [quelques coups] [de canon] [qui] [nous ont] [paru] [venir] [de fort loin]	CON:sub	ISCHUNK	0	tout	auto	1
395	R. 5896 Rés., volume 25, feuillet 125, verso	que	~~que~~	[les rues] [du havre] [étaient inondées] [et] [on défendit] [d'avoir] [du feu] [et] [de la lumiére] [de peur] [**que** qu'][une maison] [s'écroulant]	CON:sub	ISCHUNK	\RB	tout	auto	1

397	R. 5896 Rés., volume 25, feuillet 126, verso	*...*	*...*	[chemin] [faisant] [nous avons passé] [sur un coteau] [qui] [***…* la mer**] [d'un côté] [nous faisait voir] [la mer]	UKN	ISCHUNK	0	debut	auto	1
397	R. 5896 Rés., volume 25, feuillet 126, verso	la	~~la mer~~	[chemin] [faisant] [nous avons passé] [sur un coteau] [qui] [***…* la mer**] [d'un côté] [nous faisait voir] [la mer]	DET:art	ISCHUNK	0	mil	auto	1
397	R. 5896 Rés., volume 25, feuillet 126, verso	mer	~~la mer~~	[chemin] [faisant] [nous avons passé] [sur un coteau] [qui] [***…* la mer**] [d'un côté] [nous faisait voir] [la mer]	SUBS	ISCHUNK	0	fin	auto	1
402	R. 5896 Rés., volume 25, feuillet 129, recto	matelots	~~matelots~~	[reflexions] [sur l'influence ordinaire] [de la mer] [sur les **matelots** /*matelots*/ **marins**] [qui] [ont] [plus] [d'esprit] [α] [de caractère] [que] [des gens] [d'un état correspondant] [sur terre][.]	SUBS	ISTOK	\RB	mil	auto	2
402	R. 5896 Rés., volume 25, feuillet 129, recto	marins	~~marins~~	[reflexions] [sur l'influence ordinaire] [de la mer] [sur les **matelots** /*matelots*/ **marins**] [qui] [ont] [plus] [d'esprit] [α] [de caractère] [que] [des gens] [d'un état correspondant] [sur terre][.]	SUBS	ISTOK	\RB	fin	auto	2
402	R. 5896 Rés., volume 25, feuillet 129, recto	mis	~~mis~~	[la tête] [s'est embarrassée] [matelots] [enfin étant] [**mis** descendu] [dans la chambre] [où] [il s est] [assis]	v:amorce	ISCHUNK	0	tout	auto	1
405	R. 5896 Rés., volume 25, feuillet 130, verso	,	~~,~~	[nous avons] [vu] [le Phare] [,] [montré] [par un jeune homme intelligent][.]	PONCT	ISCHUNK	0	tout	auto	1
410	R. 5896 Rés., volume 25, feuillet 133, recto	à	~~à~~	[Ce coteau] [est] [vis à vis] [le château] [de Montmorency] [**à** */séparés par/* 3 lieues] [de plaine][.]	PREP	ISTOK	0	debut	auto	2
415	R. 5896 Rés., volume 25, feuillet 135, verso	–	~~–~~	[]	UKN	ISCHUNK	0	tout	auto	1
416	R. 5896 Rés., volume 25, feuillet 136, recto	la	~~la mer nous~~	[enfin] [est] [venu] [le moment] [où] [**la mer nous** la côte] [est disparue][à nos yeux][.]	DET:art	ISCHUNK+	0	debut	auto	1
416	R. 5896 Rés., v	mer	~~la mer nous~~	[enfin] [est] [venu] [le moment] [où] [**la**	SUBS	ISCHUNK+	0	fin	auto	1

	olume 25, feuillet 136, recto			**mer nous** la côte] [est disparue][à nos yeux][.]						
416	R. 5896 Rés., volume 25, feuillet 136, recto	nous	~~la mer nous~~	[enfin] [est] [venu] [le moment] [où] [**la mer nous** la côte] [est disparue][à nos yeux][.]	PRON:pers	ISCHUNK+	0	debut	auto	1
416	R. 5896 Rés., volume 25, feuillet 136, recto	ve	~~ve~~	[mais] [un nuage] [qui] [est] [**ve** survenu] [et] [qui semblait] [nous annoncer] [une pluie durable] [nous a dérobé] [cette vue] [et] [en augmentant] [par la pensée] [le vent]	v:amorce	ISCHUNK	0	tout	auto	1
417	R. 5896 Rés., volume 25, feuillet 136, verso	m'	~~m'ont~~	[a 9 h. ½] [le besoin] [d'être] [assis] [et] [de me mettre] [à l'abri] [de la pluie] [**m'ont** /*m'a*/] [fait descendre] [dans la chambre] [où] [au bout] [d'½ heure]	PRON:dir	ISCHUNK	\RB	debut	auto	2
417	R. 5896 Rés., volume 25, feuillet 136, verso	ont	~~m'ont~~	[a 9 h. ½] [le besoin] [d'être] [assis] [et] [de me mettre] [à l'abri] [de la pluie] [**m'ont** /*m'a*/] [fait descendre] [dans la chambre] [où] [au bout] [d'½ heure]	v:pc	ISCHUNK	\RB	fin	auto	2
418	R. 5896 Rés., volume 25, feuillet 137, recto	le	~~le~~	[mon bonnet] [de coton] [sur la tête] [et] [**le** comme] [c'était] [le moment] [du plus grand vent]	DET:art	ISTOK	0	debut	auto	1
418	R. 5896 Rés., volume 25, feuillet 137, recto	je	~~je~~	[**je** mon grand balancement] [et] [surtout] [celui] [de mon bonnet] [de coton] [ont beaucoup] [fait] [rire] [Henri] [et] [Félix][.]	PRON:pers	ISTOK	0	debut	auto	1
419	R. 5896 Rés., volume 25, feuillet 137, verso	ne	~~ne~~	[peut être] [son vaisseau] [battu] [par la tempête] [est] [pour lui] [un péril moindre] [que] [**ne** ne l'est] [pour un cocher] [un caprice] [de ses chevaux][#]	ADV:neg	ISTOK	\RB	debut	auto	1
419	R. 5896 Rés., volume 25, feuillet 137, verso	*...*	*...*	[***...*] [**,] [dans l'état actuel] [de notre nos idées] [sur la mer]	UKN	ISCHUNK+	0	tout	auto	1
419	R. 5896 Rés., volume 25, feuillet 137, verso	notre	~~notre~~	[*...*] [,] [dans l'état actuel] [de **notre** nos idées][sur la mer]	ADJ	ISTOK	0	mil	auto	1
419	R. 5896 Rés., volume 25, feuillet 137, verso	paraîtrait	~~paraîtrait~~	[une manœuvre] [au fort] [de l'ouragan] [**paraîtrait** *serait* il] [lâche] [à l'abordage][et] [à la vue] [de l'épée][.]	v:cond	ISCHUNK	0	debut	auto	2
421	R. 5896 Rés., volume 25,	–	~~–~~	[]	PONCT	ISCHUNK	0	tout	auto	1

	feuillet 138, verso									
423	R. 5896 Rés., volume 25, feuillet 138, verso	et	~~et englo~~	[**et] [englo]** [il m'a semblé][que] [je supporterais] [cela très bien] [et] [ce] [qui] [est essentiel]	CON:cc	ISCHUNK+	0	tout	auto	1
423	R. 5896 Rés., volume 25, feuillet 138, verso	englo	~~et englo~~	[**et] [englo]** [il m'a semblé][que] [je supporterais] [cela très bien] [et] [ce] [qui] [est essentiel]	v:amorce	ISCHUNK+	0	debut	auto	1
423	R. 5896 Rés., volume 25, feuillet 138, verso	la	~~la~~	[il m'a semblé] [que] [me voyant **la**] [sur une planche] [et] [après tous mes efforts] [désespérant] [tout à fait][,]	ADV	ISTOK	0	fin	auto	1
423	R. 5896 Rés., volume 25, feuillet 138, verso	aller	~~aller sur~~	[mais] [il était] [trop tard] [pour **aller**] [**sur** */entrer dans/* le chantier] [et] [pour examiner] [les détails][.]	v:inf	ISCHUNK+	0	fn	auto	2
423	R. 5896 Rés., volume 25, feuillet 138, verso	sur	~~aller sur~~	[mais] [il était] [trop tard] [pour **aller**] [**sur** */entrer dans/* le chantier] [et] [pour examiner] [les détails][.]	PREP	ISCHUNK+	0	debut	auto	2
425	R. 5896 Rés., volume 25, feuillet 140, verso	s	~~s~~	[je n'ai rien] [observé] [de nouveau**s**] [:]	ADJ	ISTOK-	0	fin	auto	1
425	R. 5896 Rés., volume 25, feuillet 140, verso	le	~~le~~	[**le** l'honneur] [d'avoir] [vû] [naitre] [Corneille] [a] [donné] [à la ville] [de Rouen] [une terrible sévérité] [en matière] [de spectacle][.]	DET:art	ISTOK	\RB	debut	auto	1
425	R. 5896 Rés., volume 25, feuillet 140, verso	pour	~~pour~~	[on ne daigne pas] [applaudir] [le débutant] [**pour** dès son entrée] [pour l'encourage][.]	PREP	ISCHUNK	\RB	debut	auto	1
426	R. 5896 Rés., volume 25, feuillet 141, recto	*...*	*...*	[ma principale remarque] **[*...*** est] [qu'] [on] [ne fait point usage][de l'accent *aigu*] [ainsi] [été se prononce] [aitai]	UKN	ISCHUNK	0	tout	auto	1
88	R. 5896 Rés., volume 22, feuillet 55, verso	le	~~le su~~	[Si je suivais] [ce projet] [mes pièces] [n'auraient absolument rien] [de commun] [avec les siennes] [que] **[le su** l'objet.]	DET:art	ISCHUNK	\RB	debut	auto	1

88	R. 5896 Rés., volume 22, feuillet 55, verso	su	~~le su~~	[Si je suivais] [ce projet] [mes pièces] [n'auraient absolument rien] [de commun] [avec les siennes] [que] **[le su** l'objet.]	SUBS:amorce	ISCHUNK	\RB	fin	auto	1
308	R. 302 Rés., volume 1, tome 3, feuillet 372, verso	femmes	~~femmes~~	[Nous pourrions] [baiser] [deux **femmes** comtesses] [qui] [logent prés] [de chez nous]	SUBS	ISTOK		mil	auto	1
567	R. 5896 Rés., volume 22, feuillet 87, recto	qu'	~~qu'en de~~	[**qu'**] [**en de**] [que] [par la passion] [de |d'| Orgon][.]	CON:cc	ISCHUNK+	\RB	debut	auto	1
567	R. 5896 Rés., volume 22, feuillet 87, recto	en	~~qu'en de~~	[**qu'**] [**en de**] [que] [par la passion] [de |d'| Orgon][.]	PREP	ISCHUNK+	\RB	mil	auto	1
567	R. 5896 Rés., volume 22, feuillet 87, recto	de	~~qu'en de~~	[**qu'**] [**en de**] [que] [par la passion] [de |d'| Orgon][.]	v:amorce	ISCHUNK+	\RB	fin	auto	1
569	R. 5896 Rés., volume 22, feuillet 88, recto	de	~~de l'~~	[puisqu'] [il voit] [**de l'** *une* hypocrisie **naturelle** *continuelle*] [dans toute ma conduite,]	PREP	ISTOK+	0	debut	auto	2
569	R. 5896 Rés., volume 22, feuillet 88, recto	l'	~~de l'~~	[puisqu'] [il voit] [**de l'** *une* hypocrisie **naturelle** *continuelle*] [dans toute ma conduite,]	DET:art	ISTOK+	0	mil	auto	2
569	R. 5896 Rés., volume 22, feuillet 88, recto	naturelle	~~naturelle~~	[puisqu'] [il voit] [**de l'** *une* hypocrisie **naturelle** *continuelle*] [dans toute ma conduite,]	ADJ	ISTOK	0	fin	auto	2
571	R. 5896 Rés., volume 22, feuillet 89, recto	é	~~é~~	[Prévôt] [de g **é** enêve]	NPP	ISTOK-	0	mil	auto	1
659	R. 5896 Rés., volume 22, feuillet 154, recto	de	~~de~~	[que] [j'ai] [**de** trop] [de ce] [**que** *dont*] [les autres] [n'ont pas assez][,]	PREP	ISTOK	0	debut	auto	1
659	R. 5896 Rés., volume 22, feuillet 154, recto	que	~~que~~	[que] [j'ai] [**de** trop] [de ce] [**que** *dont*] [les autres] [n'ont pas assez][,]	PRON:rel	ISCHUNK	0	tout	auto	2

660	R. 5896 Rés., volume 22, feuillet 154, verso	il	~~il y~~	[**il y** Je lui fais tout]	DET:art	ISTOK+	0	debut	auto	1
660	R. 5896 Rés., volume 22, feuillet 154, verso	y	~~il y~~	[**il y** Je lui fais tout]	ADV	ISTOK+	0	mil	auto	1
664	R. 5896 Rés., volume 22, feuillet 156, verso	,	~~,~~	[Son fils cadet][**,**] [et] [sa fille]	PONCT	ISTOK	0	tout	auto	1
666	R. 5896 Rés., volume 22, feuillet 157, verso	très	~~trés~~	[je l'ai regardée] [**trés** tout le tems]	ADJ	ISTOK	0	debut	auto	1
667	R. 5896 Rés., volume 22, feuillet 158, recto	que	~~que~~	[**que**] [on a sonné]	PRON:rel	ISCHUNK	0	tout	auto	1
668	R. 5896 Rés., volume 22, feuillet 158, verso	e	~~e~~	[J'étais extrememement poli**e]**[,]	ADJ	ISTOK-	0	fin	auto	1
669	R. 5896 Rés., volume 22, feuillet 159, recto	me	~~me~~	[qu'][elle **me** le prendrait ensuite][.]	PRON:dir	ISTOK	0	mil	auto	1
653	R. 5896 Rés., volume 22, feuillet 151, recto	29	~~29~~	[teneva le mani nelle mie il] [**29** 30][.]	NUM	ISCHUNK	0	tout	auto	1
662	R. 5896 Rés., volume 22, feuillet 155, verso	ne	~~ne m~~	[je **ne m** me serais sauvé][,]	ADV:neg	ISTOK+	\RI	mil	auto	1
662	R. 5896 Rés., volume 22, feuillet 155, verso	m	~~ne m~~	[je **ne m** me serais sauvé][,]	PRON:pos	ISTOK+	\RI	mil	auto	1
671	R. 5896 Rés., volume 22, feuillet 160, recto	sans	~~sans qu'il~~	[**sans**] [**qu'**] [**il** ***---***] [sans affectation]	PREP	ISCHUNK+	\RB	tout	auto	1
671	R. 5896 Rés., volume 22, feuillet 160,	qu'	~~sans qu'il~~	[**sans**] [**qu'**] [**il** ***---***] [sans affectation]	CON:cc	ISCHUNK+	\RB	debut	auto	1

	recto									
671	R. 5896 Rés., volume 22, feuillet 160, recto	il	~~sans qu'il~~	[**sans**] **[qu'**] [**il** *---*] [sans affectation]	PRON:rel	ISCHUNK+	\RB	debut	auto	1
671	R. 5896 Rés., volume 22, feuillet 160, recto	Répo	~~Répo~~	[aux **Répo**] [on dit] [du monde][.]	UKN	ISTOK	0	fin	auto	1
683	R. 5896 Rés., volume 27, feuillet 1, recto	qu'	~~qu'il~~	[de son regiment] [**qu'**][**il** qui lui dit] [qu'il lui scavait] [une ressource]	PRON:rel	ISCHUNK+	\RB	tout	auto	1
683	R. 5896 Rés., volume 27, feuillet 1, recto	il	~~qu'il~~	[de son regiment] [**qu'**][**il** qui lui dit] [qu'il lui scavait] [une ressource]	PRON:pers	ISCHUNK+	\RB	debut	auto	1
683	R. 5896 Rés., volume 27, feuillet 1, recto	ce	~~cen~~	[a auxerre] [mais] [**cen** mais] [ce n'était plus] [la douceur] [de l' abesse] [de x]	PRON:dem	ISCHUNK	\RB	tout	auto	1
683	R. 5896 Rés., volume 27, feuillet 1, recto	n	~~cen~~	[a auxerre] [mais] [**cen** mais] [ce n'était plus] [la douceur] [de l' abesse] [de x]	ADV:neg	ISCHUNK	\RB	tout	auto	1
684	R. 5896 Rés., volume 27, feuillet 1, verso	h	~~h~~	[qu'][elle ressentait][.] [**h** fiere] [et] [hautaine]	ADJ:amorce	ISTOK-	\RB	debut	auto	1
684	R. 5896 Rés., volume 27, feuillet 1, verso	qui	~~qui devait~~	[Comme] [un enfant] [**qui**] [**devait**] [dont] [elle devait faire] [son negre]	PRON:rel	ISCHUNK	0	tout	auto	1
684	R. 5896 Rés., volume 27, feuillet 1, verso	devait	~~qui devait~~	[Comme] [un enfant] [**qui**] [**devait**] [dont] [elle devait faire] [son negre]	v:imp	ISCHUNK	\RB	tout	auto	1
684	R. 5896 Rés., volume 27, feuillet 1, verso	quel	~~quel~~	[Melle Dauphin] [**quel** quelque temps apprès]	ADJ:amorce	ISTOK-	\RB	debut	auto	1
684	R. 5896 Rés., volume 27, feuillet 1, verso	ch	~~ch~~	[de son ordre] [**ch** jouissait] [de 25,000f]	v:amorce	ISTOK-	0	debut	auto	1
684	R. 5896 Rés., v	n'	~~n'éta~~	[mais] [**n'éta** cadet] [de famille] [n'étant	ADV:neg	ISTOK+	\RB	debut	auto	1

	olume 27, feuillet 1, verso			devenu riche]						
684	R. 5896 Rés., volume 27, feuillet 1, verso	éta	~~n'éta~~	[mais] [**n'éta** cadet] [de famille] [n'étant devenu riche]	v:amorce	ISTOK+	\RB	debut	auto	1
685	R. 5896 Rés., volume 27, feuillet 2, recto	elle	~~elle donna~~	[Le soir] [**elle donna** cependant] [elle donna] [la lettre] [a lire] [a Mr dubour]	PRON:pers	ISTOK+	\RB	debut	auto	2
		donna	~~elle donna~~	[*Le soir*] [**elle donna** cependant] [elle donna] [la lettre] [a lire] [a Mr dubour]	v:ps	ISTOK+	\RB	mil	auto	2
686	R. 5896 Rés., volume 27, feuillet 2, verso	et	~~et fit~~	[elle se mit] [au lit] [**et**] [**fit**] [car] [sa fureur] [-] [l'agitait singulierement] [elle fit beaucoup] [de reproches] [a M.r Dubour]	CON:cc	ISCHUNK+	\RI	tout	auto	1
686	R. 5896 Rés., volume 27, feuillet 2, verso	fit	~~et fit~~	[elle se mit] [au lit] [**et**] [**fit**] [car] [sa fureur] [-] [l'agitait singulierement] [elle fit beaucoup] [de reproches] [a M.r Dubour]	v:ps	ISCHUNK+	\RI	tout	auto	1
686	R. 5896 Rés., volume 27, feuillet 2, verso	qu'	~~qu'il~~	[et] [déclara] [**qu'**][**il**] [que] [dès qu'][il pourrait marcher] [il épouserait] [Melle Dauphin]	CON:sub	ISCHUNK+	\RB	tout	auto	1
686	R. 5896 Rés., volume 27, feuillet 2, verso	il	~~qu'il~~	[et] [déclara] [**qu'**][**il**] [que] [dès qu'][il pourrait marcher] [il épouserait] [Melle Dauphin]	PRON:pers	ISCHUNK+	\RB	tout	auto	1
279	R. 5896 Rés., volume 22, feuillet 47, verso	outre	~~outre~~	[qui] [lui a apporté] [300 mile francs] **[outre** sans compter] [les espérances.]	PREP	ISTOK	0	debut	auto	1
385	R. 5896 Rés., volume 25, feuillet 120, verso	avons	~~avons~~	[au 53.e mile] [nous **avons** sommes] [descendus] [dans une valée] [assez agréable] [et] [surtout trés fertile][.]	v:ind	ISTOK	0	debut	auto	1

691	R. 302 Rés., volume 1, tome 3, feuillet 377, recto	à	~~à~~	[qu'][il a reçu] [du ministre Berthier] [sur une prétendue Conspiration] [a ordonné] [**à** a |à| un chef d'Escadron]	PREP	ISTOK	\RB	debut	auto	1
691	R. 302 Rés., volume 1, tome 3, feuillet 377, recto	dans	~~dans le tems~~	[Cette lettre] [de Paris] [arriva] [au G.al Brune] [**dans le tems** */le même jour/*] [que] [le G.al]	PREP	ISCHUNK	0	debut	auto	2
691	R. 302 Rés., volume 1, tome 3, feuillet 377, recto	le	~~dans le tems~~	[Cette lettre] [de Paris] [arriva] [au G.al Brune] [**dans le tems** */le même jour/*] [que] [le G.al]	DET:art	ISCHUNK	0	mil	auto	2
691	R. 302 Rés., volume 1, tome 3, feuillet 377, recto	tems	~~dans le tems~~	[Cette lettre] [de Paris] [arriva] [au G.al Brune] [**dans le tems** */le même jour/*] [que] [le G.al]	SUBS	ISCHUNK	0	fin	auto	2
693	R. 302 Rés., volume 1, tome 3, feuillet 378, recto	*…*	~~*…*~~	[tous ces noms sont extraits] [des ***…***] [Discours] [de Dalbon][,] [ouvrage trés médiocre][.]	SUBS	ISTOK+	0	fin	auto	1
694	R. 302 Rés., volume 1, tome 3, feuillet 379, recto	l'	~~l'exa~~	[elle fait appeler] [le philosophe] [**l'exa** déguisé] [en Grec]	DET:art	ISCHUNK		debut		
694	R. 302 Rés., volume 1, tome 3, feuillet 379, recto	exa	~~l'exa~~	[elle fait appeler] [le philosophe] [**l'exa** déguisé] [en Grec]	PRON:dir	ISCHUNK	0	fin	auto	1
700	R. 302 Rés., volume 1, tome 3, feuillet 382, recto	à	~~à~~	[la ville] [qui] [est] [à peu près ronde][,] [**à** a 600. t] [de Diamêtre][.]	PREP	ISTOK	\RB	debut	auto	1
701	R. 302 Rés., volume 1, tome 3, feuillet 382, verso	é	~~é~~	[Je me tromperai presque toujours] [lorsque] [je croirai] [un homme totalement] [d'un Caractè**é**re][,]	SUBS	ISTOK-	0	fin	auto	1

706	R. 302 Rés., volume 1, tome 3, feuillet 385, recto	é	~~é~~	[c'est] [le chef lieu] [de ce Dép.t] [que] [le G.al Spital] [ancien chef d'état major] [de l'aile gauche commandée][,]	SUBS	ISTOK-	0	fin	auto	1
707	R. 302 Rés., volume 1, tome 3, feuillet 385, verso	m'	~~m'~~	[on me **m'** met] [10 sangsues][.]	v:amorce	ISTOK-	\RI	mil		
713	R. 302 Rés., volume 1, tome 3, feuillet 387, verso	y	~~y~~	[les leurs aidées] [d'une fine soubrette] [mettaient tout] [en usage] [pour **y** découvrir] [ce qu'][ils y fesaient] [α][c][.]	ADV	ISTOK	\RI	mil	auto	1
713	R. 302 Rés., volume 1, tome 3, feuillet 387, verso	c'	~~c'est~~	[**c'est**] [Connaître] [à fond] [les hommes][,]	PRON:dem	ISCHUNK	0	debut	auto	1
713	R. 302 Rés., volume 1, tome 3, feuillet 387, verso	est	~~c'est~~	[**c'est**] [Connaître] [à fond] [les hommes][,]	v:ind	ISCHUNK	0	fin	auto	1
713	R. 302 Rés., volume 1, tome 3, feuillet 387, verso	é	~~é~~	[Beéaucoup de travaux] [et] [jamais de solitude] [me guériront][.]	ADV	ISTOK-	0	debut	auto	1
714	R. 302 Rés., volume 1, tome 3, feuillet 388, recto	jo	~~jo~~	[à 7 heures] [du soir [elle s'exerçait] [à **jo** repeter] [une simphonie] [d'Haidn][,]	v:amorce	ISTOK-	0	mil	auto	1
710	R. 302 Rés., volume 1, tome 3, feuillet 388, verso	le	~~le Courrrier~~	[je pars] [p.r L.] [le] [dans **le Courrrier** la diligence]	DET:art	ISTOK+	\RI	mil	auto	1
710	R. 302 Rés., volume 1, tome 3, feuillet 388, verso	Courrier	~~le Courrrier~~	[je pars] [p.r L.] [le] [dans **le Courrrier**] [la diligence]	SUBS	ISTOK+	\RI	fin	auto	1

710	R. 302 Rés., volume 1, tome 3, feuillet 388, verso	;	~~;~~	[elle me donne] [de ses cheveux][**;**.]	PONCT	ISCHUNK	0	tout	auto	1
718	R. 302 Rés., volume 1, tome 3, feuillet 390, verso	la	~~la Cav~~	[Inspecteur g.al] [de **la Cav** l'infanterie][.]	DET:art	ISTOK+	\RI	mil	auto	1
718	R. 302 Rés., volume 1, tome 3, feuillet 390, verso	Cav	~~la Cav~~	[Inspecteur g.al] [de **la Cav** l'infanterie][.]	SUBS:amorce	ISTOK+	\RI	fin	auto	1
718	R. 302 Rés., volume 1, tome 3, feuillet 390, verso	,	~~,~~	[Jobert a] [obtenu] [un Sabre] [d'honneur][**,**] [qu'][il méritait si bien][.]	PONCT	ISCHUNK	0	tout	auto	1
718	R. 302 Rés., volume 1, tome 3, feuillet 390, verso	t	~~t~~	[Guerin a] [exposé] [son**t** Superbe Tableau] [de Phedre] [et] [hyppolite] [le 4 Brumaire 11][.]	DET:pos	ISTOK	0	debut	auto	1
690	R. 302 Rés., volume 1, tome 3, feuillet 376, verso	une	~~une~~	[nous étions] [à Table] [à **une** minuit] [et] [demi][,]	DET:art	ISTOK	0	mil	auto	1
687	R. 302 Rés., volume 1, tome 3, feuillet 378, verso	g	~~g~~	[la Donna] [contraria] [al Consi**g**lio][.]	SUBS	ISTOK-	0	fin	auto	1
731	R. 302 Rés., volume 1, tome 3, feuillet 397, recto	et	~~et~~	[elle y ètait] [avec M.e Mortier][,] [**et**] [la petite Félipe][,] [et] [Wagner][.]	CON:cc	ISCHUNK	0	tout	auto	1
731	R. 302 Rés., volume 1, tome 3, feuillet 397, recto	Nous	~~Nous sommes~~	[**Nous sommes** W. a] [accompagné] [Félipe][,]	PRON:pers	ISCHUNK	0	debut	auto	1

731	R. 302 Rés., volume 1, tome 3, feuillet 397, recto	sommes	~~Nous sommes~~	[**Nous sommes** W. a] [accompagné] [Félipe][,]	v:ind	ISCHUNK	0	fin	auto	1
731	R. 302 Rés., volume 1, tome 3, feuillet 397, recto	avons	~~avons~~	[nous y **avons** sommes] [restés] [¾ d'heures][.]	v:ind	ISTOK	0	mil	auto	1
731	R. 302 Rés., volume 1, tome 3, feuillet 397, recto	exp	~~exp~~	[et] [qu'][il faudrait] [2] [ou] [3 heures] [pour **exp** y plier] [les termes] [de la langue][.]	v:amorce	ISTOK	0	mil	auto	1
732	R. 302 Rés., volume 1, tome 3, feuillet 397, verso	é	~~é~~	[ou] [du moins] [elle veut] [que] [je le croie] [car] [elle m'a dit] [qu'][elle avait bien compris] [ma démarche][d**é**e Lundi][,]	PREP	ISTOK-	0	debut	auto	1
732	R. 302 Rés., volume 1, tome 3, feuillet 397, verso		~~'~~	[cependant] [n'etait ce] [que] [les yeux] [d'u tempérament éveillé][,]	PREP:det	ISTOK	0	debut	auto	1
697	R. 302 Rés., volume 1, tome 3, feuillet 380, verso		~~'~~	[et] [qu'][il ne l'a pas] [eue][.]	PRON:pers	ISTOK-	0	mil	auto	1
736	R. 302 Rés., volume 1, tome 3, feuillet 398, verso	aux	~~aux~~	[Mante] [et] [moi] [nous nous sommes] [allé promener] [**aux** à la Terrasse] [des feuillans] [de 2 h.] [à 4 ¼][.]	PREP	ISTOK	\RB	debut	auto	1
738	R. 302 Rés., volume 1, tome 3, feuillet 399, verso	é	~~é~~	[où] [la pe**é**tite Phélipe] [y sera][.]	ADJ	ISTOK-	0	mil	auto	1
738	R. 302 Rés., volume 1, tome 3, feuillet 399, verso	n'	~~n'ela~~	[et] [point de grace] [**n'ela** ni de touchant] [ne disant pas tout] [bonnement la premiere chose] [qui] [me vient][.]	ADV:neg	ISCHUNK	\RB	debut	auto	1

738	R. 302 Rés., volume 1, tome 3, feuillet 399, verso	ela	~~n'ela~~	[et] [point de grace] [**n'ela** ni de touchant] [ne disant pas tout] [bonnement la premiere chose] [qui] [me vient][.]	UKN	ICHUNK	\RB	fin	auto	1
715	R. 302 Rés., volume 1, tome 3, feuillet 389, recto	travai	~~travai~~	[je pourrais] [**travai** me loger prés] [des Tuilleries]	v:amorce	ISTOK-	0	debut	auto	1
715	R. 302 Rés., volume 1, tome 3, feuillet 389, recto	é	~~é~~	[lorsque] [je ne voudrai pas] [aller] [au sp**é**ectacle] [je pourrai encore] [disposer] [de mon tems] [dépuis 5 h.] [jusqu'][à 6] [pour la promenade]	SUBS	ISTOK-	0	mil	auto	1
717	R. 302 Rés., volume 1, tome 3, feuillet 390, recto	depuis	~~depuis~~	[M.r Daru] [qui] [s'est] [marié] [**depuis** /il y a/ 3 mois] [est ici][,]	PREP	ISTOK	0	debut	auto	2
733	R. 302 Rés., volume 1, tome 3, feuillet 398, recto	mais	~~mais cela~~	[C'est la gloire][,] [**mais**] [**cela**] [et] [une Ame ardente] [alors] [elle vaut peut être mieux] [que] [la jeunesse][.]	CON:cc	ISCHUNK+	0	tout	auto	1
733	R. 302 Rés., volume 1, tome 3, feuillet 398, recto	cela	~~mais cela~~	[C'est la gloire][,] [**mais**] [**cela**] [et] [une Ame ardente] [alors] [elle vaut peut être mieux] [que] [la jeunesse][.]	PRON:dem	ISCHUNK+	0	tout	auto	1
733	R. 302 Rés., volume 1, tome 3, feuillet 398, recto	décr	~~décr~~	[la fille] [me traitait] [avec l'intimité] [de la gaîté] [et] [de la jeunesse][,] [**décr** naturelle] [aux Provinciales]	UKN	ISCHUNK	0	tout	auto	1
733	R. 302 Rés., volume 1, tome 3, feuillet 398, recto	par	~~par que~~	[à celle] [de mon oncle] [qui] [à 46 ans] [tombe déjà] [dans la faiblesse morale] [et] [**par**] [**que**] [par suite Phisique] [de la Vieillesse][.]	PREP	ISCHUNK+	\RB	tout	auto	1

733	R. 302 Rés., volume 1, tome 3, feuillet 398, recto	que	~~par que~~	[à celle] [de mon oncle] [qui] [à 46 ans] [tombe déjà] [dans la faiblesse morale] [et] [**par**] [**que**] [par suite Phisique] [de la Vieillesse][.]	CON:sub	ISCHUNK+	0	tout	auto	1
909	R. 302 Rés., volume 1, tome 1, feuillet 196, recto	ou	~~ou les~~	[**ou**] [**les** MM.es Marionet] [de Viénnois]	CON:cc	ISCHUNK	0	tout	auto	1
909	R. 302 Rés., volume 1, tome 1, feuillet 196, recto	les	~~ou les~~	[**ou**] [**les** MM.es Marionet] [de Viénnois]	DET:art	ISTOK	0	debut	auto	1
943	R. 5896 Rés., volume 25, feuillet 67, verso	Chez	~~Chez~~	[**Chez** écouter] [et] [suivre davantage [le naturel] [dans ma conduite] [et] [mon stile][.]	PREP	ISCHUNK	0	debut	auto	1
737	R. 302 Rés., volume 1, tome 3, feuillet 399, recto	s	~~s~~	[elle m'inspire tant de dégoût**s** que je n'ai rien trouvé à lui dire]	SUBS	ISTOK-	0	fin	auto	1
737	R. 302 Rés., volume 1, tome 3, feuillet 399, recto	scéance	~~scéance~~	[le reste] [de la **|scéance|** leçon]	SUBS	ISTOK	0	fin	auto	1
737	R. 302 Rés., volume 1, tome 3, feuillet 399, recto	encor	~~encor~~	[Ce] [qui] [nous faisait] [**encor** tenir] [les côtés][.]	SUBS	ISTOK	0	fin	auto	1
737	R. 302 Rés., volume 1, tome 3, feuillet 399, recto	Ce	~~Ce~~	[et] [lui avait dit] [:] [**Ce**] [(][de moi][)] ["][Ce jeune homme] [est] [bien né]	PRON:dem	ISCHUNK	0	tout	auto	1

737	R. 302 Rés., volume 1, tome 3, feuillet 399, recto	n	~~n~~	[je **n**'ai] [une place] [au B.] [au] [de la Guerre] [qui] [me rapporte] [1 500][.]	ADV:neg	ISTOK	0	mil	auto	1
90	R. 5896 Rés., volume 22, feuillet 56, verso	de	~~de la Comédie~~	[Voila le pouvoir] [**de la Comédie**] [du spectacle]	PREP	ISCHUNK	0	debut	auto	1
90	R. 5896 Rés., volume 22, feuillet 56, verso	la	~~de la Comédie~~	[Voila le pouvoir] [**de la Comédie**] [du spectacle]	DET:art	ISCHUNK	0	mil	auto	1
90	R. 5896 Rés., volume 22, feuillet 56, verso	Comédie	~~de la Comédie~~	[Voila le pouvoir] [**de la Comédie**] [du spectacle]	SUBS	ISCHUNK	0	fin	auto	1
90	R. 5896 Rés., volume 22, feuillet 56, verso	*...*	~~*...*~~	[je lis] [un j][.] [de Pal][.] [le 28] [peut etre] [***...***] [machiavel][.]	UKN	ISCHUNK	0	tout	auto	1
1504	R. 302 Rés., volume 1, tome 1, feuillet 185, recto	4	~~4~~	[**4** 195]	NUM	ISCHUNK	0	tout	auto	
1506	R. 302 Rés., volume 1, tome 1, feuillet 186, recto	dans	~~dans Virgile.~~	[.][.][.][.][.] [ruit] [in vetitum] [;] [nefas] [!] [**Dans Virgil**]	PREP	ISCHUNK	0	debut	auto	1
1506	R. 302 Rés., volume 1, tome 1, feuillet 186, recto	Virgile	~~dans Virgile.~~	[.][.][.][.][.] [ruit] [in vetitum] [;] [nefas] [!] [**Dans Virgil**]	NPP	ISCHUNK	0	fin	auto	1
1814	R. 302 Rés., volume 1, tome 3, feuillet 418, recto	dab	~~dab~~	[au lieu] [d'odieuse] [qu'][elle serait] [**dab** dans le premier instant] [à mes yeux][.]	UKN	ISTOK	0	debut	hetero	2
1813	R. 302 Rés., volume 1,	le	~~le tableau de~~	[que[[leurs idées basses] [auraient entiérement] [sali] [**le tableau**] [**de** mon	DET:art	ISCHUNK+	0	debut	auto	1

	tome 3, feuillet 417, verso			avenir] [pour un jour ou deux]						
1813	R. 302 Rés., volume 1, tome 3, feuillet 417, verso	tableau	~~le tableau de~~	[que[[leurs idées basses] [auraient entiérement] [sali] [**le tableau**] [**de** mon avenir] [pour un jour ou deux]	SUBS	ISCHUNK+	0	mil	auto	1
1813	R. 302 Rés., volume 1, tome 3, feuillet 417, verso	de	~~le tableau de~~	[que[[leurs idées basses] [auraient entiérement] [sali] [**le tableau**] [**de** mon avenir] [pour un jour ou deux]	PREP	ISCHUNK+	0	debut	auto	1
1815	R. 302 Rés., volume 1, tome 3, feuillet 418, verso	je	~~je~~	[et] [**je *...*** eux passent] [dans la ville] [pour riches]	PRON:pers	ISTOK	0	debut	auto	1
1815	R. 302 Rés., volume 1, tome 3, feuillet 418, verso	*...*	~~*...*~~	[et] [**je *...*** eux passent] [dans la ville] [pour riches]	UKN	ISTOK	0	mil	auto	1
1877	R. copie Arbelet Rés., volume cahier n° 50, feuillet 22, recto	J'	~~J'ai trouvé chez tous mes amis d'Italie~~	**[J'ai trouvé] [chez tous mes amis] [d'Italie] [moins] [d'esprit] [que] [je ne m'y attendais][.]**	PRON:pers	ISCHUNK+	0	debut	hetero	2
1877	R. copie Arbelet Rés., volume cahier n° 50, feuillet 22, recto	ai	~~J'ai trouvé chez tous mes amis d'Italie~~	**[J'ai trouvé] [chez tous mes amis] [d'Italie] [moins] [d'esprit] [que] [je ne m'y attendais][.]**	v:ind	ISCHUNK+	0	fin	hetero	2
1877	R. copie Arbelet Rés., volume cahier n° 50, feuillet 22, recto	trouvé	~~J'ai trouvé chez tous mes amis d'Italie~~	**[J'ai trouvé] [chez tous mes amis] [d'Italie] [moins] [d'esprit] [que] [je ne m'y attendais][.]**	v:pc	ISCHUNK+	0	tout	hetero	2
1877	R. copie Arbelet Rés., volume cahier n° 50, feuillet 22, recto	chez	~~J'ai trouvé chez tous mes amis d'Italie~~	**[J'ai trouvé] [chez tous mes amis] [d'Italie] [moins] [d'esprit] [que] [je ne m'y attendais][.]**	PREP	ISCHUNK+	0	debut	hetero	2
1877	R. copie Arbelet Rés., volume cahier n° 50, feuillet	tous	~~J'ai trouvé chez tous mes amis d'Italie~~	**[J'ai trouvé] [chez tous mes amis] [d'Italie] [moins] [d'esprit] [que] [je ne m'y attendais][.]**	PRON:ind	ISCHUNK+	0	mil	hetero	2

	22, recto									
1877	R. copie Arbelet Rés., volume cahier n° 50, feuillet 22, recto	mes	~~J'ai trouvé chez tous mes amis d'Italie~~	**[J'ai trouvé] [chez tous mes amis] [d'Italie] [moins] [d'esprit] [que] [je ne m'y attendais][.]**	DET:pos	ISCHUNK+	0	mil	hetero	2
1877	R. copie Arbelet Rés., volume cahier n° 50, feuillet 22, recto	amis	~~J'ai trouvé chez tous mes amis d'Italie~~	**[J'ai trouvé] [chez tous mes amis] [d'Italie] [moins] [d'esprit] [que] [je ne m'y attendais][.]**	SUBS	ISCHUNK+	0	fin	hetero	2
1877	R. copie Arbelet Rés., volume cahier n° 50, feuillet 22, recto	d'	~~J'ai trouvé chez tous mes amis d'Italie~~	**[J'ai trouvé] [chez tous mes amis] [d'Italie] [moins] [d'esprit] [que] [je ne m'y attendais][.]**	PREP	ISCHUNK+	0	debut	hetero	2
1877	R. copie Arbelet Rés., volume cahier n° 50, feuillet 22, recto	Italie	~~J'ai trouvé chez tous mes amis d'Italie~~	**[J'ai trouvé] [chez tous mes amis] [d'Italie] [moins] [d'esprit] [que] [je ne m'y attendais][.]**	NPP	ISCHUNK+	0	fin	hetero	2
1877	R. copie Arbelet Rés., volume cahier n° 50, feuillet 22, recto	moins	~~moins d'esprit que je~~ ne ~~m'y attendais~~	**[J'ai trouvé] [chez tous mes amis] [d'Italie] [moins] [d'esprit] [que] [je ne m'y attendais][.]**	ADV	ISCHUNK+	0	tout	hetero	2
1877	R. copie Arbelet Rés., volume cahier n° 50, feuillet 22, recto	d'	~~moins d'esprit que je~~ ne ~~m'y attendais~~	**[J'ai trouvé] [chez tous mes amis] [d'Italie] [moins] [d'esprit] [que] [je ne m'y attendais][.]**	PREP	ISCHUNK+	0	debut	hetero	2
1877	R. copie Arbelet Rés., volume cahier n° 50, feuillet 22, recto	esprit	~~moins d'esprit que je~~ ne ~~m'y attendais~~	**[J'ai trouvé] [chez tous mes amis] [d'Italie] [moins] [d'esprit] [que] [je ne m'y attendais][.]**	SUBS	ISCHUNK+	0	fin	hetero	2
1877	R. copie Arbelet Rés., volume cahier n° 50, feuillet 22, recto	que	~~moins d'esprit que je~~ ne ~~m'y attendais~~	**[J'ai trouvé] [chez tous mes amis] [d'Italie] [moins] [d'esprit] [que] [je ne m'y attendais][.]**	CON:sub	ISCHUNK+	0	tout	hetero	2
1877	R. copie Arbelet Rés., volume cahier n° 50, feuillet 22, recto	je	~~moins d'esprit que je~~ ne ~~m'y attendais~~	**[J'ai trouvé] [chez tous mes amis] [d'Italie] [moins] [d'esprit] [que] [je ne m'y attendais][.]**	PRON:pers	ISCHUNK+	0	debut	hetero	2

1877	R. copie Arbelet Rés., volume cahier n° 50, feuillet 22, recto	ne	~~moins d'esprit que je~~ ne ~~m'y attendais~~	**[J'ai trouvé] [chez tous mes amis] [d'Italie] [moins] [d'esprit] [que] [je ne m'y attendais][.]**	ADV:neg	ISCHUNK+	0	mil	hetero	2
1877	R. copie Arbelet Rés., volume cahier n° 50, feuillet 22, recto	m'	~~moins d'esprit que je~~ ne ~~m'y attendais~~	**[J'ai trouvé] [chez tous mes amis] [d'Italie] [moins] [d'esprit] [que] [je ne m'y attendais][.]**	PRON:dir	ISCHUNK+	0	mil	hetero	2
1877	R. copie Arbelet Rés., volume cahier n° 50, feuillet 22, recto	y	~~moins d'esprit que je~~ ne ~~m'y attendais~~	**[J'ai trouvé] [chez tous mes amis] [d'Italie] [moins] [d'esprit] [que] [je ne m'y attendais][.]**	ADV	ISCHUNK+	0	mil	hetero	2
1877	R. copie Arbelet Rés., volume cahier n° 50, feuillet 22, recto	attendais	~~moins d'esprit que je~~ ne ~~m'y attendais~~	**[J'ai trouvé] [chez tous mes amis] [d'Italie] [moins] [d'esprit] [que] [je ne m'y attendais][.]**	v:imp	ISCHUNK+	0	fin	hetero	2
1878	R. copie Arbelet Rés., volume cahier n° 50, feuillet 23, verso	paraît	~~parait que j'ai fait quelques lieues, sur le Fleuve of Knowing.~~	**[paraît] [que] [j'ai] [fait] [quelques lieues][,] [sur le Fleuve] [of Knowing][.]**[/I have made/] [/some Miles/] [/on the River/] [/of Knowing/][/./]	v:ind	ISCHUNK+	0	tout	hetero	2
1878	R. copie Arbelet Rés., volume cahier n° 50, feuillet 23, verso	que	~~parait que j'ai fait quelques lieues, sur le Fleuve of Knowing.~~	**[paraît] [que] [j'ai] [fait] [quelques lieues][,] [sur le Fleuve] [of Knowing][.]**[/I have made/] [/some Miles/] [/on the River/] [/of Knowing/][/./]	CON:sub	ISCHUNK+	0	tout	hetero	2
1878	R. copie Arbelet Rés., volume cahier n° 50, feuillet 23, verso	j'	~~parait que j'ai fait quelques lieues, sur le Fleuve of Knowing.~~	**[paraît] [que] [j'ai] [fait] [quelques lieues][,] [sur le Fleuve] [of Knowing][.]**[/I have made/] [/some Miles/] [/on the River/] [/of Knowing/][/./]	PRON:pers	ISCHUNK+	0	debut	hetero	2

1878	R. copie Arbelet Rés., volume cahier n° 50, feuillet 23, verso	ai	~~parait que j'ai fait quelques lieues, sur le Fleuve of Knowing.~~	**[paraît] [que] [j'ai] [fait] [quelques lieues][,] [sur le Fleuve] [of Knowing][.]**[/I have made/] [/some Miles/] [/on the River/] [/of Knowing/][/./]	v:ind	ISCHUNK+	0	fin	hetero	2
1878	R. copie Arbelet Rés., volume cahier n° 50, feuillet 23, verso	fait	~~parait que j'ai fait quelques lieues, sur le Fleuve of Knowing.~~	**[paraît] [que] [j'ai] [fait] [quelques lieues][,] [sur le Fleuve] [of Knowing][.]**[/I have made/] [/some Miles/] [/on the River/] [/of Knowing/][/./]	v:pc	ISCHUNK+	0	tout	hetero	2
1878	R. copie Arbelet Rés., volume cahier n° 50, feuillet 23, verso	quelques	~~parait que j'ai fait quelques lieues, sur le Fleuve of Knowing.~~	**[paraît] [que] [j'ai] [fait] [quelques lieues][,] [sur le Fleuve] [of Knowing][.]**[/I have made/] [/some Miles/] [/on the River/] [/of Knowing/][/./]	ADJ	ISCHUNK+	0	debut	hetero	2
1878	R. copie Arbelet Rés., volume cahier n° 50, feuillet 23, verso	lieues	~~parait que j'ai fait quelques lieues, sur le Fleuve of Knowing.~~	**[paraît] [que] [j'ai] [fait] [quelques lieues][,] [sur le Fleuve] [of Knowing][.]**[/I have made/] [/some Miles/] [/on the River/] [/of Knowing/][/./]	SUBS	ISCHUNK+	0	fin	hetero	2
1878	R. copie Arbelet Rés., volume cahier n° 50, feuillet 23, verso	,	~~parait que j'ai fait quelques lieues, sur le Fleuve of Knowing.~~	**[paraît] [que] [j'ai] [fait] [quelques lieues][,] [sur le Fleuve] [of Knowing][.]**[/I have made/] [/some Miles/] [/on the River/] [/of Knowing/][/./]	PONCT	ISCHUNK+	0	tout	hetero	2
1878	R. copie Arbelet Rés., volume cahier n° 50, feuillet 23, verso	sur	~~parait que j'ai fait quelques lieues, sur le Fleuve of Knowing.~~	**[paraît] [que] [j'ai] [fait] [quelques lieues][,] [sur le Fleuve] [of Knowing][.]**[/I have made/] [/some Miles/] [/on the River/] [/of Knowing/][/./]	PREP	ISCHUNK+	0	debut	hetero	2
1878	R. copie Arbelet Rés., volume cahier n° 50, feuillet 23, verso	le	~~parait que j'ai fait quelques lieues, sur le Fleuve of Knowing.~~	**[paraît] [que] [j'ai] [fait] [quelques lieues][,] [sur le Fleuve] [of Knowing][.]**[/I have made/] [/some Miles/] [/on the River/] [/of Knowing/][/./]	DET:art	ISCHUNK+	0	mil	hetero	2

1878	R. copie Arbelet Rés., volume cahier n° 50, feuillet 23, verso	Fleuve	~~parait que j'ai fait quelques lieues, sur le Fleuve of Knowing.~~	**[paraît] [que] [j'ai] [fait] [quelques lieues][,] [sur le Fleuve] [of Knowing][.]**[/I have made/] [/some Miles/] [/on the River/] [/of Knowing/][/./]	SUBS	ISCHUNK+	0	fin	hetero	2
1878	R. copie Arbelet Rés., volume cahier n° 50, feuillet 23, verso	of	~~parait que j'ai fait quelques lieues, sur le Fleuve of Knowing.~~	**[paraît] [que] [j'ai] [fait] [quelques lieues][,] [sur le Fleuve] [of Knowing][.]**[/I have made/] [/some Miles/] [/on the River/] [/of Knowing/][/./]	ET	ISCHUNK+	0	debut	hetero	2
1878	R. copie Arbelet Rés., volume cahier n° 50, feuillet 23, verso	Knowing	~~parait que j'ai fait quelques lieues, sur le Fleuve of Knowing.~~	**[paraît] [que] [j'ai] [fait] [quelques lieues][,] [sur le Fleuve] [of Knowing][.]**[/I have made/] [/some Miles/] [/on the River/] [/of Knowing/][/./]	ET	ISCHUNK+	0	fin	hetero	2
1878	R. copie Arbelet Rés., volume cahier n° 50, feuillet 23, verso	.	~~parait que j'ai fait quelques lieues, sur le Fleuve of Knowing.~~	**[paraît] [que] [j'ai] [fait] [quelques lieues][,] [sur le Fleuve] [of Knowing][.]**[/I have made/] [/some Miles/] [/on the River/] [/of Knowing/][/./]	PONCT	ISCHUNK+	0	tout	hetero	2
1878	R. copie Arbelet Rés., volume cahier n° 50, feuillet 23, verso	Bar	Bar.α Lamb.	**[Bar][.] [α] [Lamb][.]**	SUBS	ISCHUNK+	0	tout	hetero	2
1878	R. copie Arbelet Rés., volume cahier n° 50, feuillet 23, verso	.	Bar.α Lamb.	**[Bar][.] [α] [Lamb][.]**	PONCT	ISCHUNK+	0	tout	hetero	2
1878	R. copie Arbelet Rés., volume cahier n° 50, feuillet 23, verso	α	Bar.α Lamb.	**[Bar][.] [α] [Lamb][.]**	CON:cc	ISCHUNK+	0	tout	hetero	2
1878	R. copie Arbelet Rés., volume cahier n° 50, feuillet 23, verso	Lamb	Bar.α Lamb.	**[Bar][.] [α] [Lamb][.]**	SUBS	ISCHUNK+	0	tout	hetero	2
1878	R. copie	.	Bar.α Lamb.	**[Bar][.] [α] [Lamb][.]**	PONCT	ISCHUNK+	0	tout	hetero	2

	Arbelet Rés., volume cahier n° 50, feuillet 23, verso									
1878	R. copie Arbelet Rés., volume cahier n° 50, feuillet 23, verso	m'	~~m'ont paru manquer d'esprit.~~	**[m'ont] [paru] [manquer] [d'esprit][.][**/*sauter*/**][**/*cette ligne*/**]**	PRON:dir	ISCHUNK+	0	debut	hetero	2
1878	R. copie Arbelet Rés., volume cahier n° 50, feuillet 23, verso	ont	~~m'ont paru manquer d'esprit.~~	**[m'ont] [paru] [manquer] [d'esprit][.][**/*sauter*/**][**/*cette ligne*/**]**	v:ind	ISCHUNK+	0	fin	hetero	2
1878	R. copie Arbelet Rés., volume cahier n° 50, feuillet 23, verso	paru	~~m'ont paru manquer d'esprit.~~	**[m'ont] [paru] [manquer] [d'esprit][.][**/*sauter*/**][**/*cette ligne*/**]**	v:pc	ISCHUNK+	0	tout	hetero	2
1878	R. copie Arbelet Rés., volume cahier n° 50, feuillet 23, verso	manquer	~~m'ont paru manquer d'esprit.~~	**[m'ont] [paru] [manquer] [d'esprit][.][**/*sauter*/**][**/*cette ligne*/**]**	v:inf	ISCHUNK+	0	tout	hetero	2
1878	R. copie Arbelet Rés., volume cahier n° 50, feuillet 23, verso	d'	~~m'ont paru manquer d'esprit.~~	**[m'ont] [paru] [manquer] [d'esprit][.][**/*sauter*/**][**/*cette ligne*/**]**	PREP	ISCHUNK+	0	debut	hetero	2
1878	R. copie Arbelet Rés., volume cahier n° 50, feuillet 23, verso	esprit	~~m'ont paru manquer d'esprit.~~	**[m'ont] [paru] [manquer] [d'esprit][.][**/*sauter*/**][**/*cette ligne*/**]**	SUBS	ISCHUNK+	0	fin	hetero	2
1878	R. copie Arbelet Rés., volume cahier n° 50, feuillet 23, verso	.	~~m'ont paru manquer d'esprit.~~	**[m'ont] [paru] [manquer] [d'esprit][.][**/*sauter*/**][**/*cette ligne*/**]**	PONCT	ISCHUNK+	0	tout	hetero	2
1878	R. copie Arbelet Rés., volume cahier n° 50, feuillet 23, verso	ournal	~~ournal~~	[j'ai lu] [le J**ournal** *uvénal*] [de Cesarotti][.]	SUBS:amorce	ISTOK-	0	mil	hetero	2

1812	R. 302 Rés., volume 1, tome 3, feuillet 417, recto	influent	~~influent~~	[et] [en général] [des têtes stupides] **[influent**] [ou] [des cœurs froids] [/*ou*/] [/*tous les deux ensemble*/][/,/] [influent] [sur mon bonheur][,]	v:ind	ISCHUNK	\RB	tout	auto	1
1375	R. 302 Rés., volume 1, tome 1, feuillet 100, recto	m'	~~m'~~	[Lemazurier] [peut] [**m'**être utile] [pour cela][.]	PRON:dir	ISTOK	0	debut	auto	1
1375	R. 302 Rés., volume 1, tome 1, feuillet 100, recto	y	~~y~~	[/*sans compter*/] [/*la loge Cr./*] [/*y allait*/] [/*chez elle*/] [/*2 fois*/] [/*par semaine.*/]	ADV	ISTOK	0	debut	auto	2
2330	R. 302 Rés., volume 1, tome 1, feuillet 89, recto	qu'	~~qu'elle~~	[m'expliqua] [qu'][elle] [que] [ses frères] [et] [elles avaient] [une Domestique] [de confiance] [à l'Albe] [qui] [avait fait] [5] [quintaux] [60 liv][.] [de coton][,] [α][.] [a] [α][.][a][.]	CON:sub	ISCHUNK+	\RB	tout	auto	1
2330	R. 302 Rés., volume 1, tome 1, feuillet 89, recto	elle	~~qu'elle~~	[m'expliqua] [qu'][elle] [que] [ses frères] [et] [elles avaient] [une Domestique] [de confiance] [à l'Albe] [qui] [avait fait] [5] [quintaux] [60 liv][.] [de coton][,] [α][.] [a] [α][.][a][.]	SUBS	ISCHUNK+	0	tout	auto	1
2339	R. 302 Rés., volume 1, tome 3, feuillet 406, recto	s	~~s~~	[à amuser] [p.r être] [trouvés aimable**s**][.]	ADJ	ISTOK-	0	fin	auto	1
2340	R. 302 Rés., volume 1, tome 3, feuillet 406, verso	vraies	~~vraies~~	[et] [dit ces paroles] [et] [d'autres] [du même sens] [avec les intonations les plus **vraies** *naturelles*] [et] [les plus larges][.]	ADJ	ISTOK	0	fin	auto	2
2345	R. 302 Rés., volume 1, tome 3, feuillet 408, verso	Rue	~~Rue~~	[en passant] [devant une marchande] [de mode] [au coin] [de la **Rue** Place] [des Victoires]	SUBS	ISTOK	0	fin	auto	1
2345	R. 302 Rés., volume 1, tome 3, feuillet 408, verso	achet	~~achet~~	[elle entre] [dans un magazin] [p.r **achet** essayer]	v:amorce	ISTOK	0	fin	auto	1
2341	R. 302 Rés., v	j'	~~j'étais~~	[**j'étais** *je fus* trés bien] [avec elle] [après ce	PRON:per	ISTOK+	\RB	debut	auto	2

	olume 1, tome 3, feuillet 407, recto			propos][.]	s					
2341	R. 302 Rés., volume 1, tome 3, feuillet 407, recto	étais	~~j'étais~~	[**j'étais** *je fus* trés bien] [avec elle] [après ce propos][.]	PRON:pers	ISTOK+	\RB	debut	auto	2
2341	R. 302 Rés., volume 1, tome 3, feuillet 407, recto	qu'	~~qu'ass~~	[il est clair] [**qu'**][**ass**] [qu'][à ses yeux] [j'ai été] [très aimable] [**à ses yeux**] [auprès de Me][.] [M][.] [et] [de Félipe][.]	CON:sub	ISCHUNK	\RB	tout	auto	1
2341	R. 302 Rés., volume 1, tome 3, feuillet 407, recto	ass	~~qu'ass~~	[il est clair] [**qu' ass** qu'][à ses yeux] [j'ai été] [très aimable] [**à ses yeux**] [auprès de Me][.] [M][.] [et] [de Félipe][.]	PREP	ISTOK	0	debut	auto	1
2341	R. 302 Rés., volume 1, tome 3, feuillet 407, recto	à	~~à ses yeux~~	[il est clair] [**qu' ass** qu'][à ses yeux] [j'ai été] [très aimable] [**à ses yeux**] [auprès de Me][.] [M][.] [et] [de Félipe][.]	PREP	ISCHUNK	0	debut	auto	1
2341	R. 302 Rés., volume 1, tome 3, feuillet 407, recto	ses	~~à ses yeux~~	[il est clair] [**qu' ass** qu'][à ses yeux] [j'ai été] [très aimable] [**à ses yeux**] [auprès de Me][.] [M][.] [et] [de Félipe][.]	DET:pos	ISCHUNK	0	mil	auto	1
2341	R. 302 Rés., volume 1, tome 3, feuillet 407, recto	yeux	~~à ses yeux~~	[il est clair] [**qu' ass** qu'][à ses yeux] [j'ai été] [très aimable] [**à ses yeux**] [auprès de Me][.] [M][.] [et] [de Félipe][.]	SUBS	ISCHUNK	0	fin	auto	1
2341	R. 302 Rés., volume 1, tome 3, feuillet 407, recto	chez	~~chez~~	[Le soir] [je vais] [avec Crozet] [**chez** à la loge] [d'Ariane]	ADV	ISTOK	0	debut	auto	1

2350	R. 302 Rés., volume 1, tome 3, feuillet 411, verso	se	~~se~~	[mais] [qui] [vient] [**se** principalement] [de cette mauvaise habitude][.]	PRON:dir	ISTOK	0	debut	auto	1
2350	R. 302 Rés., volume 1, tome 3, feuillet 411, verso	,	~~,~~	[Le Cours] [de la Conversation][**,**] [a amené] [ce] [qui] [suit]	PONCT	ISCHUNK	0	tout	auto	1
2354	R. 302 Rés., volume 1, tome 3, feuillet 413, verso	l'	~~l'idee~~	[Il entre] [dans **l'idee** l'impression] [du beau]	DET:art	ISTOK+	\RI	mil	auto	1
2354	R. 302 Rés., volume 1, tome 3, feuillet 413, verso	idee	~~l'idee~~	[Il entre] [dans **l'idee** l'impression] [du beau]	DET:art	ISTOK+	0	mil	auto	1
2354	R. 302 Rés., volume 1, tome 3, feuillet 413, verso	Je	~~Je n'ai~~	[**Je n'ai**] [Ma conduite] [n'a rien] [de cette Phisionomieaux] [yeux] [de Louäson][.] [Si] [je puis] [m'en faire] [aimer][,] [ce n'est] [que] [par l'extrême confiance][,] [et] [en l'aidant] [à voir] [dans le Témugin] [d'aujourdhuy] [le Gengis] [de 1820][.]	PRON:pers	ISCHUNK	0	tout	auto	1
2354	R. 302 Rés., volume 1, tome 3, feuillet 413, verso	n'	~~Je n'ai~~	[**Je n'ai**] [Ma conduite] [n'a rien] [de cette Phisionomieaux] [yeux] [de Louäson][.] [Si] [je puis] [m'en faire] [aimer][,] [ce n'est] [que] [par l'extrême confiance][,] [et] [en l'aidant] [à voir] [dans le Témugin] [d'aujourdhuy] [le Gengis] [de 1820][.]	ADV:neg	ISCHUNK	\RB	tout	auto	1
2354	R. 302 Rés., volume 1, tome 3, feuillet 413, verso	ai	~~Je n'ai~~	[**Je n'ai**] [Ma conduite] [n'a rien] [de cette Phisionomieaux] [yeux] [de Louäson][.] [Si] [je puis] [m'en faire] [aimer][,] [ce n'est] [que] [par l'extrême confiance][,] [et] [en l'aidant] [à voir] [dans le Témugin] [d'aujourdhuy] [le Gengis] [de 1820][.]	v:ind	ISCHUNK	0	tout	auto	1
2354	R. 302 Rés., volume 1, tome 3, feuillet 413, verso	é	~~é~~	[Apprendre] [à me borner] [en écrivant tondre] [mon Stile autrement] [les accéessoires] [me font] [oublier] [le principal.]	SUBS	ISTOK-	0	mil	auto	1

2356	R. 302 Rés., volume 1, tome 3, feuillet 414, verso	c'et	~~c'et~~	[J'ai senti] [en l'entendant esquisser] [que] [**c'et** le genre Comique] [était] [mes premieres Amours][.]	PRON:dem	ISCHUNK	0	tout	auto	1
2356	R. 302 Rés., volume 1, tome 3, feuillet 414, verso	c'et	~~c'et~~	[J'ai senti] [en l'entendant esquisser] [que] [**c'et** le genre Comique] [était] [mes premieres Amours][.]	v:ind	ISCHUNK	0	tout	auto	1
2357	R. 302 Rés., volume 1, tome 3, feuillet 415, recto	saurais	~~saurais~~	[je **saurais** suivrais tous] [les événemens] [qui] [arrivent] [à un hom] [qui] [veut faire] [jouer] [une piéce][.]	v:cond	ISTOK	0	mil	auto	1
2358	R. 302 Rés., volume 1, tome 3, feuillet 415, verso	à	~~à~~	[D.] [nous **à** a conté] [l'histoire] [de mon frère] [le major][,]	PREP	ISTOK	\RB	mil	auto	1
2360	R. 302 Rés., volume 1, tome 3, feuillet 416, verso	cela	~~cela et~~	[Digeon] [m'a offert] [son Mouchoir] [sans rien dire] [**cela] [et**] [et] [avec grace][.]	PRON:dem	ISCHUNK+	0	tout	auto	1
2360	R. 302 Rés., volume 1, tome 3, feuillet 416, verso	et	~~cela et~~	[Digeon] [m'a offert] [son Mouchoir] [sans rien dire] [**cela et**] [et] [avec grace][.]	CON:cc	ISCHUNK+	\RB	tout	auto	1
2597	R. 5896 Rés., volume 26, feuillet 126, recto	nous	~~nous~~	[**nous** ils parlent] [anglais] [avec P.]	PRON:pers	ISTOK	0	debut	auto	1
2654	R. 5896 Rés., volume 26, feuillet 127, verso	nous	~~nous y trouv~~	[nous partons] [à 10 1/2] [p.r aller] [à la fabrique] [d'indiennes] [de Petit] [et] [**nous y trouv** nous sortons] [par la porte] [de Rives]	PRON:pers	ISCHUNK	\RB	debut	auto	1
2654	R. 5896 Rés., volume 26, feuillet 127, verso	y	~~nous y trouv~~	[nous partons] [à 10 1/2] [p.r aller] [à la fabrique] [d'indiennes] [de Petit] [et] [**nous y trouv** nous sortons] [par la porte] [de Rives]	ADV	ISCHUNK	0	mil	auto	1
2654	R. 5896 Rés., volume 26, feuillet 127,	trouv	~~nous y trouv~~	[nous partons] [à 10 1/2] [p.r aller] [à la fabrique] [d'indiennes] [de Petit] [et] [**nous y trouv** nous sortons] [par la porte] [de	v:amorce	ISCHUNK	0	fin	auto	1

	verso			Rives]						
2662	R. 5896 Rés., volume 26, feuillet 129, recto	chez	~~chez~~	[**chez** dinant] [chez M.lle R.] [il ne parla] [que] [de la mauvaise qualité] [du vin] [de Bor][.]	ADV	ISCHUNK	\RB	debut	auto	1
2662	R. 5896 Rés., volume 26, feuillet 129, recto	en	~~en~~	[il **en** offrit] [du sien][,]	ADV	ISTOK	0	mil	auto	1
2663	R. 5896 Rés., volume 26, feuillet 129, verso	car	~~car~~	[**car**] [nous avons tous] [plus ou moins] [la manie]	CON:cc	ISCHUNK	0	tout	auto	1
2666	R. 5896 Rés., volume 26, feuillet 130, recto	nyon	~~nyon~~	[genêve] [à gauche] [en Amphithéâtre] [en face] [le coté] [de **nyon**] [à 3/4] [de lieue] [à droite] [le lac] [jusqu'][à Rolle]	NPP	ISTOK	0	fin	auto	1
2670	R. 5896 Rés., volume 26, feuillet 130, verso	hom	~~hom~~	[la mort] [de Socrate] [est] [d'un **hom** sage]	SUBS:amorce	ISTOK	0	mil	auto	1
2685	R. 302 Rés., volume 1, tome 3, feuillet 419, verso	grand	~~grand~~	[C] [**grand** petite Riviere][,]	ADJ	ISTOK	0	debut	auto	1
2685	R. 302 Rés., volume 1, tome 3, feuillet 419, verso	très	~~trés gra~~	[tout couverts] [de **trés gra** Peupliers trés grands]	ADV	ISTOK+	\RI	mil	auto	1
2685	R. 302 Rés., volume 1, tome 3, feuillet 419, verso	grand	~~trés gra~~	[tout couverts] [de **trés gra** Peupliers très grands]	ADJ	ISTOK+	\RB	fin	auto	1
2838	R. 5896 Rés., volume 20, feuillet 74, recto	à	~~à Asso~~	["][Journal] [du Voyage] [dans la Brianza] [**à Asso**] [[][Aout] [1818][]][.]["]	PREP	ISCHUNK	0	debut	auto	1
2838	R. 5896 Rés., volume 20, feuillet 74, recto	Asso	~~à Asso~~	["][Journal] [du Voyage] [dans la Brianza] [**à Asso**] [[][Aout] [1818][]][.]["]	UKN	ISCHUNK	0	fin	auto	1

2844	R. 5896 Rés., volume 22, feuillet 102, verso	;	~~;~~	[je regardais] [l'empereur] [**;** .]	PONCT	ISCHUNK	0	tout	auto	1
2845	R. 5896 Rés., volume 22, feuillet 103, recto		~~lorsque~~	[Si] [l'état] [où] [nous sommes][**lorsque** pendant] [que] [l'on décide] [de notre sort] [est] [d'un bon augure]	CON:sub	ISCHUNK	0	tout	auto	2
2845	R. 5896 Rés., volume 22, feuillet 103, recto		~~deviné~~	[j'en ai] [**deviné** prévu] [une] [aujourdhui] [longtems] [avant] [qu'][elle la fit][.]	v:pc	ISCHUNK	0	tout	auto	2
1817	R. 5896 Rés., volume 15, feuillet 154, recto	je	~~je n'ai cherché~~	[Trop souvent] [jusqu'][ici] [**je n'ai cherché**] [en écrivant] [sur les événemens] [qui] [m'arrivaient] [je n'ai] [que] [débondé] [mes passions]	PRON:pers	ISCHUNK	0	debut	auto	1
1817	R. 5896 Rés., volume 15, feuillet 154, recto	n'	~~je n'ai cherché~~	[Trop souvent] [jusqu'][ici] [**je n'ai cherché**] [en écrivant] [sur les événemens] [qui] [m'arrivaient] [je n'ai] [que] [débondé] [mes passions]	ADV:neg	ISCHUNK	0	mil	auto	1
1817	R. 5896 Rés., volume 15, feuillet 154, recto	ai	~~je n'ai cherché~~	[Trop souvent] [jusqu'][ici] [**je n'ai cherché**] [en écrivant] [sur les événemens] [qui] [m'arrivaient] [je n'ai] [que] [débondé] [mes passions]	v:in	ISCHUNK	0	mil	auto	1
1817	R. 5896 Rés., volume 15, feuillet 154, recto	cherché	~~je n'ai cherché~~	[Trop souvent] [jusqu'][ici] [**je n'ai cherché**] [en écrivant] [sur les événemens] [qui] [m'arrivaient] [je n'ai] [que] [débondé] [mes passions]	v:pc	ISCHUNK	0	fin	auto	1
1818	R. 5896 Rés., volume 15, feuillet 154, verso	;	~~;~~	[et] [se figure] [des faussetés][**;** .]	PONCT	ISCHUNK	0	tout	auto	1
1819	R. 5896 Rés., volume 15, feuillet 155, recto	formes	~~formes.~~	[au lieu] [de l'avoir] [sur le modèle idéal] [que] [je me **formes.** Suis] [formé][.]	SUBS	ISTOK+	0	mil	auto	1
1819	R. 5896 Rés., volume 15, feuillet 155, recto	.	~~formes.~~	[au lieu] [de l'avoir] [sur le modèle idéal] [que] [je me **formes.** Suis] [formé][.]	PONCT	ISTOK+	0	mil	auto	1

3085	R. 5896 Rés., volume 22, feuillet 99, recto	jour	~~jour~~	[mon oncle] [arrive] [le 12 frimaire] [**jour** lendemain] [du couronnement] [à 2 heures] [du matin][.]	SUBS	ISCHUNK	0	tout	auto	1
3617	R. 9975 Rés., volume , feuillet 16, recto	l	~~l~~	[La réponse] [à cette question] [résoudrait **l** enfin] [le problême][.]	DET:art	ISTOK-	\RB	mil	auto	1
3617	R. 9975 Rés., volume , feuillet 16, recto	de	~~de la nudité~~	[**de la nudité** vû] [le sombre nuage] [qui] [couvre] [toute la société] [de Banti]	PREP	ISCHUNK	0	debut	auto	1
3617	R. 9975 Rés., volume , feuillet 16, recto	la	~~de la nudité~~	[**de la nudité** vû] [le sombre nuage] [qui] [couvre] [toute la société] [de Banti]	DET:art	ISCHUNK	0	mil	auto	1
3617	R. 9975 Rés., volume , feuillet 16, recto	nudité	~~de la nudité~~	[**de la nudité** vû] [le sombre nuage] [qui] [couvre] [toute la société] [de Banti]	SUBS	ISCHUNK	0	fin	auto	1
3617	R. 9975 Rés., volume , feuillet 16, recto	ou	~~ou plusieurs femmes~~	[M.me] [de B.] [doit] [penser] [que] [Banti a] [une maitresse] [**ou**] [**plusieurs femmes**] [et] [**ne** trouve] [celles] [de sa société] [ou] [trop tristes][,] [ou] [trop assujetissantes][.]	CON:cc	ISCHUNK+	0	tout	auto	2
3617	R. 9975 Rés., volume , feuillet 16, recto	plusieurs	~~ou plusieurs femmes~~	[M.me] [de B.] [doit] [penser] [que] [Banti a] [une maitresse] [**ou**] [**plusieurs femmes**] [et] [**ne** trouve] [celles] [de sa société] [ou] [trop tristes][,] [ou] [trop assujetissantes][.]	PRON:ind	ISCHUNK+	0	debut	auto	2
3617	R. 9975 Rés., volume , feuillet 16, recto	femmes	~~ou plusieurs femmes~~	[M.me] [de B.] [doit] [penser] [que] [Banti a] [une maitresse] [**ou**] [**plusieurs femmes**] [et] [**ne** trouve] [celles] [de sa société] [ou] [trop tristes][,] [ou] [trop assujetissantes][.]	SUBS	ISCHUNK+	0	fin	auto	2
3617	R. 9975 Rés., volume , feuillet 16, recto	ne	~~ne~~	[M.me] [de B.] [doit] [penser] [que] [Banti a] [une maitresse] [**ou**] [**plusieurs femmes**] [et] [**ne** trouve] [celles] [de sa société] [ou] [trop tristes][,] [ou] [trop assujetissantes][.]	ADV:neg	ISTOK	0	debut	auto	1
3618	R. 9975 Rés., volume , feuillet 15, verso	uniquement	~~uniquement~~	[uniquement] [par l'influence] [du général] [dont] [il devait] [la protection **uniquement**] [à M.me de Berulle][.]	ADV	ISTOK	0	fin	auto	1
3618	R. 9975 Rés., volume , feuillet 15,	marquée	~~marquée~~	[La physionomie] [de la conduite] [ci dessus] [s'est **marquée** produite] [dans le voyage] [de V][.]	v:pc	ISTOK	0	fin	auto	1

	verso									
3618	R. 9975 Rés., volume , feuillet 15, verso	maintenant	~~maintenant~~	[Banti] [a souvent] [répété][cette faute] [plus] [ou] [moins] [:] [leurs tête] [à tête] [sont **maintenant**] [silencieux] [et] [froids][.]	ADV	ISTOK	0	fin	auto	1
3637	R. 90729 Rés., volume 2, feuillet 29, recto	le	~~le~~	[**le** huit ans] [après]	DET:art	ISTOK	0	debut	auto	1
3768	R. 5896 Rés., volume 16, feuillet 70, verso	de	~~de jalousie~~	[un air] [**de jalousie**] [d'envie] [et] [de desir][.]	PREP	ISCHUNK	0	debut	auto	2
3768	R. 5896 Rés., volume 16, feuillet 70, verso	jalousie	~~de jalousie~~	[un air] [**de jalousie**] [d'envie] [et] [de desir][.]	SUBS	ISCHUNK	0	fin	auto	2
3768	R. 5896 Rés., volume 16, feuillet 70, verso	voit	~~voit~~	[il **voit** cherche toujours] [une épigramme] [ou] [quelque chose] [d'aimable] [pour lui]	v:ind	ISTOK	0	mil	auto	1
3768	R. 5896 Rés., volume 16, feuillet 70, verso	j'	~~j'aurais grand besoin et~~	[M. A. monsieur] [**j'aurais**] [**grand besoin**] [**et**] [/*je desirerais*/] [/*me faire*/] [/*quelque*/]	PRON:pers	ISCHUNK+	\RB	debut	auto	1
3768	R. 5896 Rés., volume 16, feuillet 70, verso	aurais	~~j'aurais grand besoin et~~	[M. A. monsieur] [**j'aurais**] [**grand besoin**] [**et**] [/*je desirerais*/] [/*me faire*/] [/*quelque*/]	v:cond	ISCHUNK+	0	fin	auto	1
3768	R. 5896 Rés., volume 16, feuillet 70, verso	grand	~~j'aurais grand besoin et~~	[M. A. monsieur] [**j'aurais**] [**grand besoin**] [**et**] [/*je desirerais*/] [/*me faire*/] [/*quelque*/]	ADJ	ISCHUNK+	0	debut	auto	1
3768	R. 5896 Rés., volume 16, feuillet 70, verso	besoin	~~j'aurais grand besoin et~~	[M. A. monsieur] [**j'aurais**] [**grand besoin**] [**et**] [/*je desirerais*/] [/*me faire*/] [/*quelque*/]	SUBS	ISCHUNK+	0	fin	auto	1
3768	R. 5896 Rés., volume 16, feuillet 70, verso	et	~~j'aurais grand besoin et~~	[M. A. monsieur] [**j'aurais**] [**grand besoin**] [**et**] [/*je desirerais*/] [/*me faire*/] [/*quelque*/]	CON:cc	ISCHUNK+	0	tout	auto	1

3769	R. 5896 Rés., volume 16, feuillet 71, recto	dont	~~dont~~	[idée] [des mœurs] [de la nation] [**dont**] [de chez la quelle] [vous venez][.]	PRON:rel	ISCHUNK	0	tout	auto	1
3769	R. 5896 Rés., volume 16, feuillet 71, recto	Ah	~~Ah~~	[M.r A] [**Ah**] [ricanant][.] [Ah] [!] [je vous remercie][,]	interjection	ISCHUNK	\RB	tout	auto	1
3769	R. 5896 Rés., volume 16, feuillet 71, recto	bonne	~~bonne~~	[à mesure] [qu'][on avance] [dans la **bonne** société riche]	ADJ	ISTOK	0	mil	auto	1
4515	R. 5896 Rés., volume 22, feuillet 13, recto	*…*	~~*…*~~	[Tout sentiment] [s'évanouit] [()][quoique][***…***] [la perception] [reste invariable][)] [par la familiarité] [des mêmes objets][.]	UKN	ISCHUNK	0	tout	auto	2
4516	R. 5896 Rés., volume 22, feuillet 13, verso	Cette	~~Cette fable de~~	[**Cette fable**] [**de** les 15 premiers vers] [de cette fable] [me donneraient] [beaucoup] [plus] [d'émotion] [que toute la Tragédie[.]	PRON:dem	ISCHUNK+	\RI	debut	auto	1
4516	R. 5896 Rés., volume 22, feuillet 13, verso	fable	~~Cette fable de~~	[**Cette fable**] [**de** les 15 premiers vers] [de cette fable] [me donneraient] [beaucoup] [plus] [d'émotion] [que toute la Tragédie[.]	SUBS	ISCHUNK+	\RB	fin	auto	1
4516	R. 5896 Rés., volume 22, feuillet 13, verso	de	~~Cette fable de~~	[**Cette fable**] [**de** les 15 premiers vers] [de cette fable] [me donneraient] [beaucoup] [plus] [d'émotion] [que toute la Tragédie[.]	PREP	ISCHUNK+	\RB	debut	auto	1
4518	R. 5896 Rés., volume 22, feuillet 14, verso	rend	~~rend~~	[ce] [qui] [**rend** nous rend avantageux]	v:ind	ISTOK	\RI	debut	auto	1
4672	R. 5896 Rés., volume 04, feuillet 173, verso	comm	~~comm~~	[Je me demandais] [en dinant] [vis à vis] [de cette femme] [**comm** par quel méchanisme] [des yeux grands] [et] [même beaux] [peuvent ils] [être] [bêtes][?]	ADV:amorce	ISTOK	0	debut	auto	1

4672	R. 5896 Rés., volume 04, feuillet 173, verso	M.r	~~M.~~r	[Je me suis] [rapelé] [les yeux] [de **M.r** l'amiral M][.]	NPP	ISTOK	0	mil	auto	1
4672	R. 5896 Rés., volume 04, feuillet 173, verso	2	~~2.°~~	[**2.°** J'ai observé] [que] [la Courbe] [formée] [par la paupiére supérieure] [de A] [en B] [était inflexible]	NUM	ISTOK+	0	debut	auto	1
4672	R. 5896 Rés., volume 04, feuillet 173, verso	.	~~2.°~~	[**2.°** J'ai observé] [que] [la Courbe] [formée] [par la paupiére supérieure] [de A] [en B] [était inflexible]	PONCT	ISTOK+	0	debut	auto	1
4672	R. 5896 Rés., volume 04, feuillet 173, verso	°	~~2.°~~	[**2.°** J'ai observé] [que] [la Courbe] [formée] [par la paupiére supérieure] [de A] [en B] [était inflexible]	SYM	ISTOK+	0	debut	auto	1
4672	R. 5896 Rés., volume 04, feuillet 173, verso	de	~~de A en B~~	[**2.°** J'ai observé] [que] [la Courbe] [formée] [par la paupiére supérieure] [**de A**] [**en B**] [était inflexible]	PREP	ISCHUNK+	0	debut	auto	1
4672	R. 5896 Rés., volume 04, feuillet 173, verso	A	~~de A en B~~	[**2.°** J'ai observé] [que] [la Courbe] [formée] [par la paupiére supérieure] [**de A**] [**en B**] [était inflexible]	SYM	ISCHUNK+	0	fin	auto	1
4672	R. 5896 Rés., volume 04, feuillet 173, verso	en	~~de A en B~~	[**2.°** J'ai observé] [que] [la Courbe] [formée] [par la paupiére supérieure] [**de A**] [**en B**] [était inflexible]	PREP	ISCHUNK+	0	debut	auto	1
4672	R. 5896 Rés., volume 04, feuillet 173, verso	B	~~de A en B~~	[**2.°** J'ai observé] [que] [la Courbe] [formée] [par la paupiére supérieure] [**de A**] [**en B**] [était inflexible]	SYM	ISCHUNK+	0	fin	auto	1
4672	R. 5896 Rés., volume 04, feuillet 173, verso	que	~~que peut être l'inflexion C lui manquait~~	[**que**] [**peut-être**] [**l'inflexion**] [**C**] [**lui manquait**] [qu'][elle] [se rapprochait] [du demi=cercle parfait.]	CON:sub	ISCHUNK+	\RB	tout	auto	1

4672	R. 5896 Rés., volume 04, feuillet 173, verso	peut-être	~~que peut-être l'inflexion C lui manquait~~	[**que**] [**peut-être**] [**l'inflexion**] [**C**] [**lui manquait**] [qu'][elle] [se rapprochait] [du demi=cercle parfait.]	ADV	ISCHUNK+	0	tout	auto	1
4672	R. 5896 Rés., volume 04, feuillet 173, verso	l'	~~que peut-être l'inflexion C lui manquait~~	[**que**] [**peut-être**] [**l'inflexion**] [**C**] [**lui manquait**] [qu'][elle] [se rapprochait] [du demi=cercle parfait.]	DET:art	ISCHUNK+	0	tout	auto	1
4672	R. 5896 Rés., volume 04, feuillet 173, verso	inflexion	~~que peut-être l'inflexion C lui manquait~~	[**que**] [**peut-être**] [**l'inflexion**] [**C**] [**lui manquait**] [qu'][elle] [se rapprochait] [du demi=cercle parfait.]	SUBS	ISCHUNK+	0	tout	auto	1
4672	R. 5896 Rés., volume 04, feuillet 173, verso	C	~~que peut-être l'inflexion C lui manquait~~	[**que**] [**peut-être**] [**l'inflexion**] [**C**] [**lui manquait**] [qu'][elle] [se rapprochait] [du demi=cercle parfait.]	SYM	ISCHUNK+	0	tout	auto	1
4672	R. 5896 Rés., volume 04, feuillet 173, verso	lui	~~que peut-être l'inflexion C lui manquait~~	[**que**] [**peut-être**] [**l'inflexion**] [**C**] [**lui manquait**] [qu'][elle] [se rapprochait] [du demi=cercle parfait.]	PRON:dir	ISCHUNK+	0	debut	auto	1
4672	R. 5896 Rés., volume 04, feuillet 173, verso	manquait	~~que peut-être l'inflexion C lui manquait~~	[**que**] [**peut-être**] [**l'inflexion**] [**C**] [**lui manquait**] [qu'][elle] [se rapprochait] [du demi=cercle parfait.]	v:imp	ISCHUNK+	0	fin	auto	1
4672	R. 5896 Rés., volume 04, feuillet 173, verso	les	~~les yeux de~~ V	[J'avais] [cette idée] [d'après] [**les yeux] [de V**]	DET:art	ISCHUNK+	0	debut	auto	1
4672	R. 5896 Rés., volume 04, feuillet 173, verso	yeux	~~les yeux de~~ V	[J'avais] [cette idée] [d'après] [**les yeux] [de V**]	SUBS	ISCHUNK+	\RI	fin	auto	1
4672	R. 5896 Rés., volume 04, feuillet 173, verso	de	~~les yeux de~~ V	[J'avais] [cette idée] [d'après] [**les yeux] [de V**]	PREP	ISCHUNK+	\RB	debut	auto	1

4672	R. 5896 Rés., volume 04, feuillet 173, verso	V	~~les yeux de~~ V	[J'avais] [cette idée] [d'après] **[les yeux] [de V]**	SYM	ISCHUNK+	0	fin	auto	1
4672	R. 5896 Rés., volume 04, feuillet 173, verso	D	~~D cet homme sensible, dont~~	[J'avais] [cette idée] [d'après] **[les yeux]** [**de V**] [**D**] [**cet homme sensible**][**,**] [**dont**] [mes yeux]	PREP	ISCHUNK+	0	debut	auto	1
4672	R. 5896 Rés., volume 04, feuillet 173, verso	cet	~~D cet homme sensible, dont~~	[J'avais] [cette idée] [d'après] **[les yeux]** [**de V**] [**D**] [**cet homme sensible**][**,**] [**dont**] [mes yeux]	DET:dem	ISCHUNK+	0	mil	auto	1
4672	R. 5896 Rés., volume 04, feuillet 173, verso	homme	~~D cet homme sensible, dont~~	[J'avais] [cette idée] [d'après] **[les yeux]** [**de V**] [**D**] [**cet homme sensible**][**,**] [**dont**] [mes yeux]	SUBS	ISCHUNK+	0	mil	auto	1
4672	R. 5896 Rés., volume 04, feuillet 173, verso	sensible	~~D cet homme sensible, dont~~	[J'avais] [cette idée] [d'après] **[les yeux]** [**de V**] [**D**] [**cet homme sensible**][**,**] [**dont**] [mes yeux]	ADJ	ISCHUNK+	0	fin	auto	1
4672	R. 5896 Rés., volume 04, feuillet 173, verso	,	~~D cet homme sensible, dont~~	[J'avais] [cette idée] [d'après] **[les yeux]** [**de V**] [**D**] [**cet homme sensible**][**,**] [**dont**] [mes yeux]	PONCT	ISCHUNK+	0	tout	auto	1
4672	R. 5896 Rés., volume 04, feuillet 173, verso	dont	~~D cet homme sensible, dont~~	[J'avais] [cette idée] [d'après] **[les yeux]** [**de V**] [**D**] [**cet homme sensible**][**,**] [**dont**] [mes yeux]	PRON:rel	ISCHUNK+	0	tout	auto	1

Figure 29 : Table des annotations du corpus *Journaux.*

Annexe B

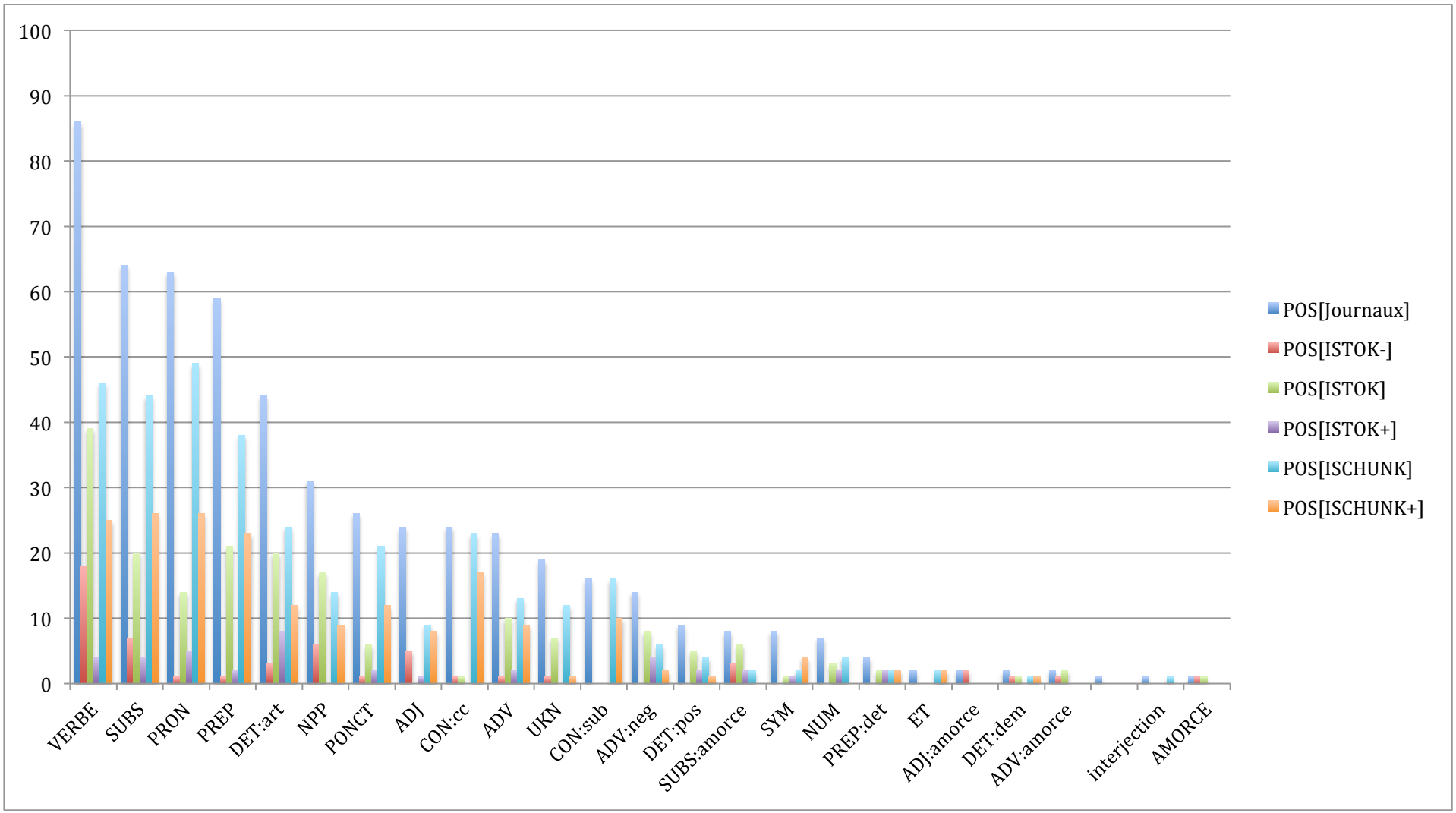


Figure 30 : Distribution des occurrences de catégories grammaticales pour chaque groupe du critère ISTOK/ISCHUNK

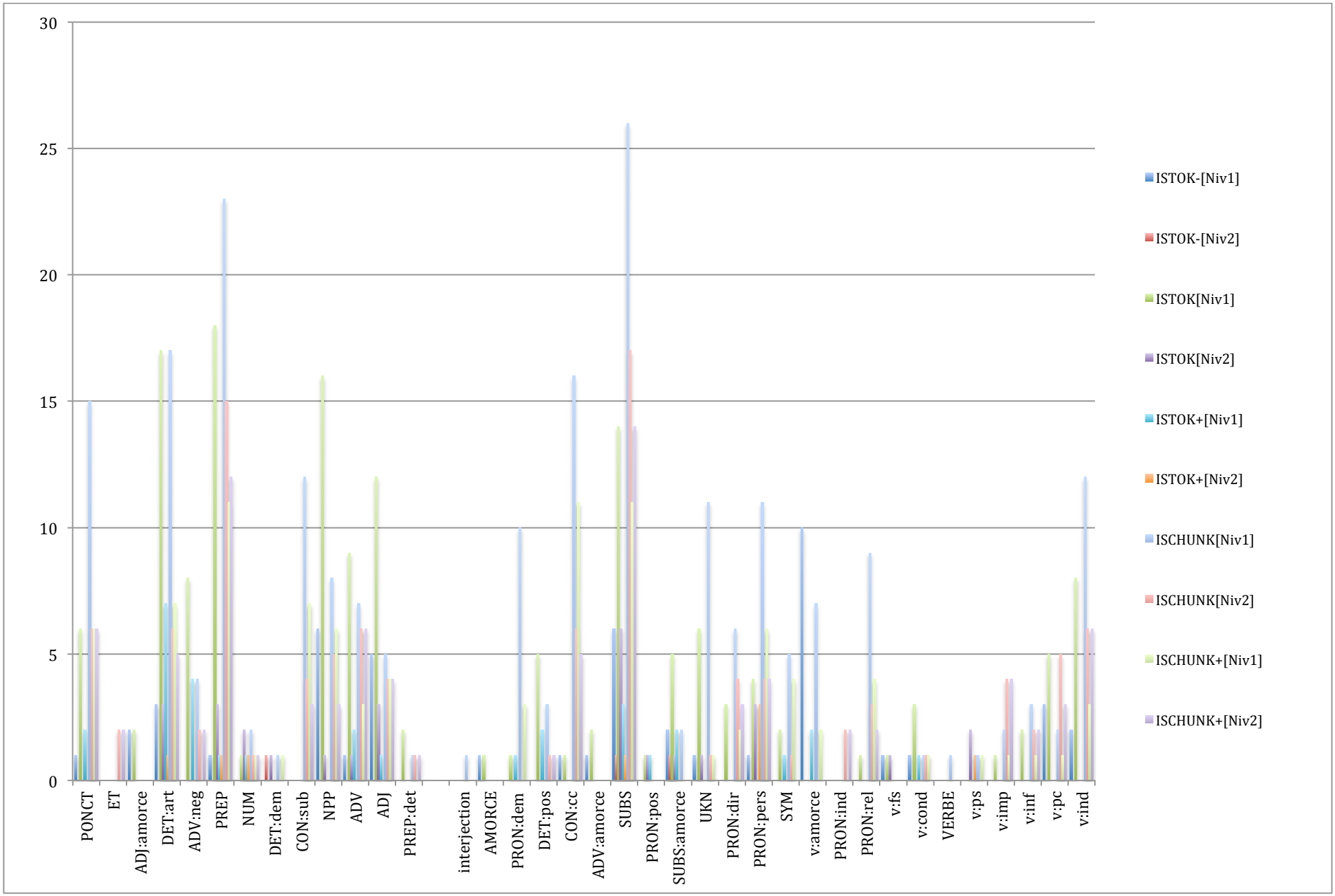


Figure 31 : Distribution des occurrences de catégories grammaticales pour chaque groupe du critère ISTOK/ISCHUNK et selon le niveau de modification (1/2).

POS	ISTOK[Niv1]	POS	ISCHUNK[Niv1]
PREP	18	SUBS	26
DET:art	17	PREP	23
NPP	16	DET:art	17
SUBS	14	CON:cc	16
ADJ	12	PONCT	15
ADV	9	CON:sub	12
ADV:neg	8	v:ind	12
v:ind	8	UKN	11
PONCT	6	PRON:pers	11
UKN	6	PRON:dem	10
DET:pos	5	PRON:rel	9
SUBS:amorce	5	NPP	8
v:pc	5	ADV	7
PRON:pers	4	v:amorce	7
PRON:dir	3	PRON:dir	6
v:cond	3	ADJ	5
ADJ:amorce	2	SYM	5
ADV:amorce	2	ADV:neg	4
PREP:det	2	DET:pos	3
v:inf	2	v:inf	3
SYM	2	NUM	2
NUM	1	SUBS:amorce	2
AMORCE	1	v:imp	2
PRON:dem	1	v:pc	2
CON:cc	1	PREP:det	1
PRON:pos	1	DET:dem	1
PRON:rel	1	interjection	1
v:fs	1	v:cond	1
v:imp	1	VERBE	1
v:amorce	0	v:ps	1
PRON:ind	0	ADJ:amorce	0
ET	0	ET	0
DET:dem	0		0
CON:sub	0	AMORCE	0
interjection	0	ADV:amorce	0
VERBE	0	PRON:pos	0
v:ps	0	PRON:ind	0
		v:fs	0

Figure 32 : Distribution des catégories grammaticales pour les deux groupes ISTOK et ISCHUNK dans l'ordre décroissant.

Annexe C

Liste des Corpus des manuscrits de Stendhal

- A Tour through Italy
- Caractères
- Correspondance personnelle
- Cours de l'Ecole Centrale
- Des Périls de la Langue Italienne
- Ebauches de pièces de théâtre
- Féder
- Filosofia nova et textes afférents
- Hamlet
- Histoire_de_la_Peinture_en_Italie*
- Journal de 1801
- Journal de 1802
- Journal de 1803
- Journal de 1804
- Journal de 1805
- Journal de 1806
- Journal de 1810
- Journal de 1811
- Journal de 1812
- Journal de 1813
- Journal de mon 3e Voyage à Paris (1804-1805)
- Journal de sa Vie depuis le 15 avril 1806 jusqu'au 3 Mai 1810
- Journaux
- Journaux de voyage
- Lamiel*
- Les deux hommes
- Letellier
- Le_Rose_et_le_Vert*
- Lucien_Leuwen*
- L_Italie_en_1818*
- Papiers
- Papiers de Colomb
- Pensées

- Perrino
- Pharsale
- Poésies
- Projets de pièces de théâtre
- Reliure
- Rome_Naples_et_Florence_en_1817*
- Souvenirs_d_egotisme*
- Style
- Textes historiques
- Textes politiques
- Textes sur la littérature
- Textes sur la musique
- Textes sur l'économie
- Traductions
- Traduction de l'Eneïde
- Traité de l'art d'écrire des comédies
- Travaux sur Hobbes
- Vie_de_Henry_Brulard*
- Voyages_en_France*
- Zelinde et Lindor

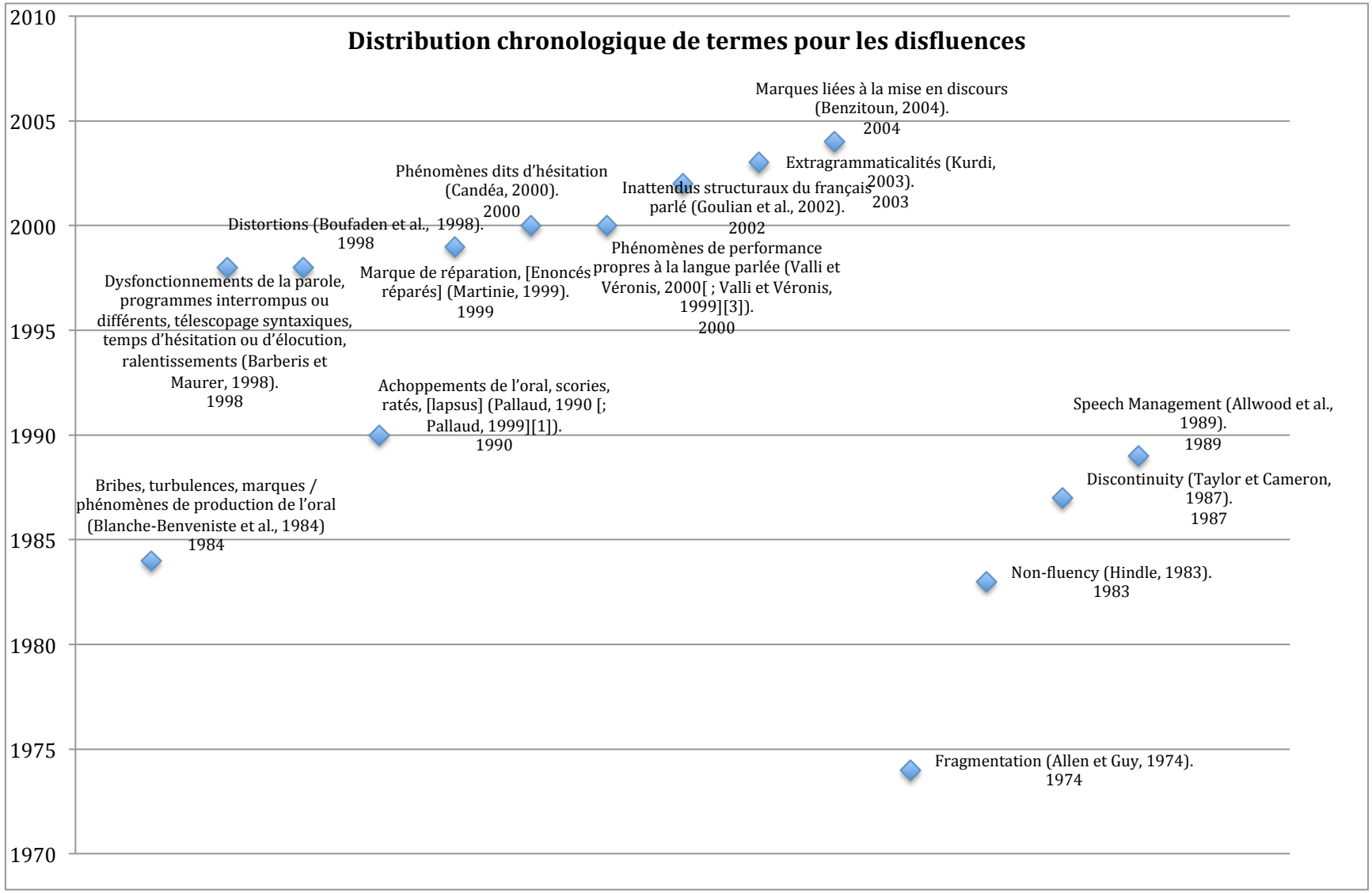


Figure 33 : Distribution chronologique des termes utilisés pour décrire les disfluences (Cette représentation est non exhaustive.)

Index

A

ajout · 37, 40, 45, 55, 56, 57, 58
ajouts · 13, 21, 22, 56, 58
avant-texte · 21, 67

B

biffe · 17, 37, 46, 57, 58
biffes · 22, 33, 56, 57
Blanche-Benveniste · 19, 20, 21, 22, 23, 133
Bove · 40, 134, 135
brouillons · 5, 6, 18, 21, 22, 23, 24

C

catégorie grammaticale · 8, 27, 51, 60, 62, 63
chunk · 6, 7, 8, 27, 28, 29, 30, 31, 32, 36, 37, 38, 63, 68
CLELIA · 6, 22, 43, 49, 55
cognition · 22, 26, 27
Constant · 40, 133
coordination · 19, 21, 22, 24, 25, 26, 27, 31, 34
corpus dynamique · 6, 41, 44, 47
Culioli · 20, 134

D

De Biasi · 20, 37
disfluences · 1, 2, 5, 6, 7, 8, 12, 15, 16, 18, 19, 20, 21, 22, 23, 24, 26, 27, 28, 31, 32, 38, 39, 40, 41, 44, 46, 54, 55, 56, 57, 58, 59, 60, 61, 65, 133, 134, 135
disfluences de l'oral · 5, 6, 18, 39, 40
disfluences scripturales · 1, 2, 5, 6, 7, 8, 12, 15, 16, 18, 20, 21, 22, 23, 24, 26, 27, 32, 38, 41, 44, 46, 54, 58, 59, 60, 61, 65
Dister · 19, 40, 133

E

énonciation · 5, 6, 12, 13, 16, 17, 18, 19, 20, 22, 23, 24, 26, 27, 38, 49, 67, 134
expressions régulières · 7, 51

F

Fuchs · 20, 22, 65, 134, 135

G

génétique de texte · 5, 15, 65
génétique textuelle · 20, 46, 67

H

hésitations · 21, 26, 57, 67
humanités numériques · 5, 15

I

interligne · 13, 22, 37, 45, 55, 56, 57, 58
interlignes · 13, 22, 56, 58

J

Journaux · 5, 11, 12, 16, 22, 23, 31, 54, 56, 59, 66, 130, 134

L

Lebarbé · 27, 31, 41, 43, 46, 134

M

manuscrit · 11, 12, 18, 20, 21, 22, 24, 26, 43, 54, 57, 65
manuscrits · 8, 11, 12, 13, 15, 17, 20, 21, 22, 24, 26, 27, 40, 43, 44, 46, 47, 48, 49, 50, 53, 54, 55, 56, 57, 58, 59, 60, 65, 66, 67, 130
Martin · 19, 22, 24, 25
Meynard · 11, 12, 13, 21, 22, 41, 43, 56, 134
moteur · 6, 8, 49, 50, 51, 65

N

niveau · 7, 8, 9, 26, 32, 33, 37, 51, 56, 57, 59, 61, 62, 63, 64, 128
numérique · 11, 15, 43

P

paraphrases · 26
pensée · 6, 13, 18, 19, 21, 22, 23, 24, 26, 27, 57, 65, 67, 96
PHP · 6, 8, 39, 47, 48, 49, 50, 51, 135, 136
POS · 6, 8, 32, 33, 34, 35, 63, 64, 68, 129, 135
pratiques discursives · 20
prosodie · 7, 24, 25, 26, 65

R

reformulations · 20, 26, 56
Robert · 19, 22, 27, 65, 135

S

segment · 6, 27, 28, 31, 36
Shriberg · 19, 135
SQL · 6, 8, 15, 39, 47, 50, 51
syntaxe · 7, 20, 24, 26, 40, 65

T

TAL · 15, 16, 19, 39, 40
token · 32
transcription · 12, 15, 40, 43, 55, 65
Treetagger · 6, 33, 40

U

unité · 11, 27, 28, 31, 32, 36, 37, 58, 59

V

variante en interligne · 37, 45, 55, 56, 57
Véronis · 23, 40, 135

X

XML · 6, 7, 8, 15, 16, 17, 32, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 51, 52, 54, 55, 56, 57, 58, 135, 136

Bibliographie

Livres et articles en-ligne

[Abney, 1994]. Abney, S. (1994). Parsing by chunks. *Bell Communications Research.* Récupéré [le 30 avril, 2014]. de http://www.ling.helsinki.fi/kit/2011s/clt350/docs/Abney-Chunks91.pdf

[Biasi, 2007] Biasi, P.-M. de. (n.d.). (2007). Édition horizontale, édition verticale. Pour une typologie des éditions génétiques (le domaine français 1980-1995).http://www.item.ens.fr. text. Récupéré le [12 mai, 2014] de http://www.item.ens.fr/index.php?id=13346

[Biasi, 2013] Biasi, P.-M. de. (2013). Pour une politique d'enrichissement du patrimoine écrit. *Trésors de l'écrit. 10 ans d'enrichissement du patrimoine écrit, Réunion des Musées Nationaux*, Paris, 1991, (p. 10-31). Récupéré le [8 Mars 2014] de http://www.item.ens.fr/index.php?id=13573.

[Blanche-Benveniste, 1993] Blanche-Benveniste, C. (1993). Le français parlé - études grammaticales. Paris, Éditions du CNRS, Collection « Sciences du langage », 1990. *L'Information Grammaticale*, *57*(1), 58–59. Récupéré [le 28 avril 2014] de http://www.persee.fr/web/revues/home/prescript/search/

[Blanche-Benveniste, 1991] Blanche-Benveniste, C. (1991). Les études sur l'oral et le travail d'écriture de certains poètes contemporains. *Langue française 89 : L'oral dans l'écrit*, pp. 52-71. Récupéré [le 17 déc. 2013] de http://www.persee.fr/articleAsPDF/lfr_0023-8368_1991_num_89_1_5763/article_lfr_0023-8368_1991_num_89_1_5763.pdf

[Bouvet et Morel, 2002] Bouvet, B., Morel, Mary-Annick. (2002). L'iconicité des indices gestuels et intonatifs. Dans *Le ballet et la musique de la parole* (9-16). Paris: Ophrys.

[Bove, 2008] Bove, R. (2008). *Analyse Syntaxique Automatique de l'oral : étude des disfluences* (Thèse de Doctorat). Université d'Aix-Marseilles, France.
Extrait de TEL : Serveur des Thèses Multi-disciplinaire. Récupéré [le 15 nov. 2013] de http://tel.archives-ouvertes.fr/docs/00/64/79/00/PDF/These_RB.pdf

[Constant et Dister, 2012] Constant, M., Dister A. (2012). Mots composés et disfluences. *Lexicometrica JADT*, 269-280. Récupéré [le 15 nov. 2013] de http://lexicometrica.univ-

paris3.fr/jadt/jadt2012/Communications/Constant,%20Matthieu%20et%20al.%20-%20Mots%20composes%20et%20disfluences.pdf.

[Culioli, 2013] Culioli, A. (2013). Pour une linguistique de l'énonciation : formalisation et opérations de repérage (Tome 2). Paris : Ophrys.

[Dufaye et Gournaye, 2013] Dufaye, L., Gournaye, L. (Eds.). (2013). Benveniste après un demi-siècle. Regards sur l'énonciation aujourd'hui. Paris : Ophrys.

[Faure *et al.,* 2009] Faure, A., Lebarbé, T., Meynard, C., & Touati, A. (2009). Manuscrits de Stendhal : Du patrimoine papier au document électronique. In *12e Conférence Internationale sur le Document Electronique* Montréal, Québec.

[Fuchs, 1994] Fuchs, C. (1994). La reformulation en discours : une pratique langagière. Dans *Paraphrase et Enonciation.* Paris : Ophrys. P. 3-33.

[Goulian *et al.,* 2002] Goulian, J., Antoine, J. Y . & Poirier, F. (2002). Compréhension automatique de la parole et tal : une approche syntaxico-sémantique pour le traitement des inattendus structuraux du français parlé. *In Actes de la Vème conférence Rencontre des Etudiants Chercheurs en Informatique pour le Traitement Automatique des Langues,* Nancy (France). Récupéré [le 07 déc. 2013] de http://www.loria.fr/projets/JEP-TALN/actes/TALN/posters/Poster12.pdf

[Gray *et al.,* 2013] Gray, J., Bounegru L., Chambers, L. (2013). Comprendre les données. Dans *Guide du Datajournalisme*. Paris : Groupe Eyrolles. P. 145-164.

[Lebarbé, 2013] Lebarbé, T. Réunion de travail. 18 avril 2013.

[Lebarbé, 2002] Lebarbé, T. (2002). *Hiérarchie Inclusive des Unités Linguistiques en Analyse Syntaxique Coopérative* (Thèse de Doctorat). Université de Caen, Caen, France.

[Meynard, 2014] Meynard, C. (2014). Entretien individuel, 16,17 janvier, 2014.

[Meynard et Lebarbé, 2011] Meynard, C. Lebarbé, T. (2011). Au croisement des lettres, de la linguistique et de l'informatique : Les Manuscrits de Stendhal en ligne. Fabula-LhT, n° 8, « Le partage des disciplines », Récupéré [le 21 mars 2014] de http://www.fabula.org/lht/8/lebarbe.html.

[Meynard *et al.,* 2013] Meynard, C., de Jacquelot, H., Corredor, M. R. (2013). *Journaux et Papiers Stendhal Volume I – 1797-1804*. ELLUG Université Stendhal : Grenoble.

[Piérart, 2011] Piérart, B. (2011). Introduction. Dans B. Piérart (Ed.), *Les bégaiements de l'adulte*. Editions Mardaga : Collines de Wavre, Belgique. Récupéré [le 29 mai 2014] de http://www.cairn.info/les-begaiements-de-l-adulte--9782804700737-page-13.htm

[Piu et Bove, 2007] Piu, M., Bove, R. (2007). *Annotation des disfluences dans les corpus oraux.* RÉCITAL, Toulouse, Récupéré [le 20 avril 2014] de http://sites.univ-provence.fr/delic/perso/bove/publis/RECITAL_2007_PIU_BOVE.pdf

[Robert, 1999] Robert, S. (1999). Cognitive Invariant and Linguistic Variability : From units to utterance. Dans C. Fuchs and S. Robert (Eds.), *Language Diversity and Cognitive Representations.* (p. 21- 32). Amsterdam/Philadelphia : John Benjamins Publishing Company.

[Shriberg *et al.,* 2006] Shriberg, E., Liu, Y., Stolke, A., Hillard, D., Ostendorf, M., Harper M. (2006). Enriching Speech Recognition with Automatic Detection of Sentence Boundaries and Disfluencies , *IEEE TRANSACTIONS ON SPEECH & AUDIO PROCESSING*. Récupéré [le 07 déc 2013] de http://www.speech.sri.com/cgi-bin/run-distill?papers/ieee-aslp2006-mde.ps.gz

[Valli et Vérnis, 1999] Valli, A., Véronis, J. (1999). Etiquetage grammatical de corpus oraux : problèmes et perspectives. *Revue français de linguistique appliquée,* 4(2) :113-133. Récupéré [le 07 déc. 2013] de http://sites.univ-provence.fr/veronis/pdf/1999rfla.pdf

Sitographie

[SW1]« French Treebank POS Tagset » Récupéré [le 30 avril 2014] De http://french-postaggers.tiddlyspot.com/#[[French%20TreeBank%20POS%20Tags]]

[SW2] French Treebank French Treebank II. *« French Tree-tagger Part of Speech Tags ».* Récupéré le [21/04/2014] de http://www.ims.uni-stuttgart.de/institut/mitarbeiter/schmid/french-tagset.html..

[SW3] Penn Treebank II. Récupéré [le 21 avril 2014] de http://www.clips.ua.ac.be/pages/mbsp-tags.

[SW4] Manuel de Transcription XML des Manuscrits de Stendhal. Récupéré [le 14 mai 2014] de http://stendhal.msh-alpes.fr/index.php?n=XML.Manuel.

Ressources et manuels informatiques

[RM1] http://www.manuelphp.com/mysql/mysqldump.php

[RM2] Nebra, M. (2013). Concevez votre site web avec PHP et MYSQL : Le développement d'un site dynamique enfin à votre portée ! 2e édition. Simple IT.

[RM3] http://www.nicolashoening.de/?twocents&nr=8

[RM4] Brochard, J. Introduction. Dans *XML Concepts et mise en œuvre.* Nantes : Editions ENI. P. 7- 15.

[RM5] Mariel, Stéphane. Introduction lapidaire à PHP. Dans *PHP 5* (2004). Paris : Editions Eyrolles. De la série : Les Cahiers du Programmeur. P. 5-19.

[RM6] Mariel, Stéphane. Echanges et contenus XML avec DOM. Dans *PHP 5* (2004). Paris : Editions Eyrolles. De la série : Les Cahiers du Programmeur. P. 121-145.

MOTS-CLÉS : disfluences scripturales, énonciation, linguistique, informatique, littérature, génétique de texte, manuscrits, Stendhal, humanités numériques.

RÉSUMÉ

L'objet du projet de recherche que nous présentons dans ce mémoire porte sur des phénomènes de l'écrit constatés sur les brouillons d'auteurs qui sont analogues aux disfluences de l'oral. Nous donnons le terme *disfluences scripturales* pour décrire ces phénomènes que nous étudions et caractérisons au sein du corpus des manuscrits de Stendhal. Notre étude porte spécifiquement sur le corpus *Journaux* de cette collection d'écrits. Nous nous appuyons sur les principes linguistiques et utilisons les méthodes informatiques pour étudier ce corpus.

Notre approche, d'un point de vue linguistique, s'appuie sur l'apport de la linguistique de l'énonciation à l'étude des disfluences scripturales. Cette analyse en corpus se fonde sur une démarche de traitement automatique des langues - au sens de l'outil et la modélisation informatique comme micro-macro-scope sur un corpus de matériau langagier numérisé. Nous proposons une caractérisation et une discussion sur la notion de disfluences scripturales, à l'aune des analyses plus présentes dans la littérature scientifiques des disfluences de l'oral.

KEYWORDS : scriptural disflencies, enunciation, linguistics, computer sciences, literature, literary criticism, manuscripts, Stendhal, digital humanities.

ABSTRACT

The research project that we present in this memoir concerns phenomena observed in written language and that we specifically observe in the rough drafts of an author's works. This language phenomena is analogous to what are known as *oral disfluencies*. We attribute the term *scriptural disfluencies* to the objects of our study, which we characterise within the corpus of Stendhal's manuscripts. Our study focuses specifically on the corpus *Journaux* (Journals) of this collection of writing. We rely on linguistic principles and use computational methods for this study.

Our approach from the linguistic perspective acknowledges the contribution of the theory of enunciation to the study of scriptural disfluencies. This corpus analysis is also founded in an approach for natural language processing – specifically in the adaptation of an instrument of observation designed for application on a corpus of digitised manuscripts. We propose a linguistic characterisation of the phenomenon along with a discussion around the notion of scriptural disfluencies in light of existing analyses of oral disfluencies in scientific literature.