

---

## TABLE DES MATIÈRES

REMERCIEMENTS.....	7
POUR COMPRENDRE CE QUI VA SUIVRE.....	9
COMMENT LIRE CE LIVRE.....	17
CHAPITRE I	
COMPTER LES MOTS ? PAS SI SIMPLE.....	19
Quelques idées reçues.....	20
Savoir ce que l'on compte : un cas pratique.....	20
Quelles données ? Quels outils ? Pour quoi faire ?.....	28
Compter avec des <i>nuages de mots</i> ?.....	29
Mais où sont passés les mots ? Le cas de Google Ngram Viewer.....	36
CHAPITRE II	
CONSTITUER UN CORPUS EN ANALYSE DU DISCOURS, UN MOMENT CRUCIAL.....	41
Le corpus en analyse du discours.....	41
Principes de base.....	41
<i>Le corpus est construit</i> .....	41
<i>Les corpus est contextualisé</i> .....	42
<i>Le corpus est contrastif/comparatif</i> .....	43
Le corpus en AD : de la clôture à l'ouverture, de l'homogénéité à l'hétérogénéité.....	44
<i>La notion de « moments de corpus » chez les historiens du discours</i> .....	45
<i>La notion de « corpus réflexif »</i> .....	47
<i>Les « corpus à géométrie variable »</i> .....	48
Quelques exemples de structuration du corpus.....	51
La structuration en genres de discours.....	51
Structuration selon les sources énonciatives.....	52

Structuration selon les espaces ou les sphères d'activité .....	53
Structuration selon les moments discursifs .....	54
Le nécessaire croisement des critères .....	55
<b>Le cas des corpus recueillis sur le web .....</b>	<b>58</b>

CHAPITRE III

**CONSTITUER UN CORPUS EN TROIS SCÉNARIOS .....** 63

**Scénario 1 – Un corpus médiatique autour d'une expression, d'un énoncé, d'une construction ou d'un mot-notion .....** 64

Étape 1 – Quel est mon objet d'étude, quel est mon corpus? .....

*Choix des observables* .....

*Délimitation du corpus : de la sphère médiatique au corpus de presse* .....

Étape 2 – Recueil du corpus .....

*Où recueillir les données?* .....

*Comment recueillir les données? Requêtes et mots-clés* .....

Étape 3 – Formatage et nettoyage des données .....

*Extensions des fichiers prises en compte par le logiciel* .....

*Codage des caractères* .....

*Nettoyage des scories* .....

Étape 4 – Quelques choix à effectuer dans la préparation des données .....

*Que faire des majuscules?* .....

*Que faire des mots composés?* .....

*Lemmatiser ou ne pas lemmatiser mon corpus?* .....

*Que peut m'apporter une annotation de mon corpus?* .....

Étape 5 – Concaténation des fichiers, structuration des données, balisage .....

*Concaténation* .....

*Structuration* .....

*Balisage* .....

**Scénario 2 – Un corpus autour d'un thème .....** 94

Étape 1 – Quel est mon objet d'étude, quel est mon corpus? .....

*Exemple 1* .....

*Exemple 2* .....

*Exemple 3* .....

Étape 2 – Recueil du corpus .....

*Exemple 1* .....

*Exemple 2* .....

*Exemple 3* .....

Étapes 3 et 5 .....

Étape 4 .....

**Scénario 3 – Un corpus de discours politiques .....** 97

Étape 1 – Quel est mon objet d'étude, quel est mon corpus? .....

Étape 2 – Recueil du corpus .....

*Où recueillir les données?* .....

*Comment recueillir les données?* .....

Étape 3 – Formatage et nettoyage des données .....

Étape 4 – Des choix à effectuer .....

Étape 5 – Structuration du corpus .....

## CHAPITRE IV

<b>COMPTER DANS LES TEXTES, QUELLES UNITÉS?</b> .....	103
<b>Autour du mot : formes graphiques, lemmes, catégories morphosyntaxiques</b> .....	104
La forme graphique .....	104
Le lemme .....	106
Les parties du discours .....	107
Aller en deçà du mot? .....	109
<i>Les N-grammes</i> .....	109
<i>Les racines</i> .....	109
<b>Au-delà des mots : des unités séquentielles</b> .....	109
Les segments répétés (SR) .....	110
Les associations de mots ou cooccurrences .....	113
<i>Principe</i> .....	113
<i>La cooccurrence contextuelle spécifique</i> .....	114
<i>La cooccurrence généralisée</i> .....	117
De nouvelles unités séquentielles à définir : les motifs et les routines .....	118
<b>Des unités « paradigmatiques » : les groupes de formes</b> .....	119

## CHAPITRE V

<b>QUELS OUTILS LOGICIELS ET POUR QUOI FAIRE?</b> .....	123
<b>Repères historiques et épistémologiques</b> .....	124
<b>Typologie logicielle</b> .....	125
Les outils essentiellement contrastifs et longitudinaux .....	125
Les outils proposant une approche structurante .....	125
Les outils reposant sur des ontologies ou des dictionnaires .....	131
Synthèse .....	133
<b>Principales méthodes et fonctionnalités (outils contrastifs et longitudinaux)</b> .....	135
Index hiérarchique et tableau lexical .....	135
Distributions (ou ventilations) de fréquences .....	139
Calcul des spécificités .....	141
Du nécessaire retour au texte : concordances et cartes topographiques .....	142
<i>Concordances</i> .....	142
<i>Topographie/Topologie textuelle et carte des sections</i> .....	146
<b>L'analyse factorielle des correspondances (principes généraux)</b> .....	152
<b>Synthèse : exploration méthodique d'un corpus</b> .....	155
Étape 1 – Conditions de recueil et de constitution du corpus .....	155
Étape 2 – Caractéristiques quantitatives .....	156
Étape 3 – Examen de l'index hiérarchique .....	157
Étape 4 – Affinités et contradictions lexicales : analyse factorielle des correspondances .....	157
Étape 5 – Spécificités .....	158
Autres explorations .....	159

CHAPITRE VI

<b>PROBLÉMATIQUES D'ANALYSE DU DISCOURS ET MÉTHODES</b> ....	163
Thèmes, événements, faits sociaux : de quoi parle le corpus et comment? ...	164
Approches déductives et inductives des thèmes .....	165
Approche inductive par la classification descendante .....	166
Approche inductive par la cooccurrence généralisée .....	168
<b>Le mot comme témoin. Les analyses du discours à entrée lexicale.</b> .....	172
Le mot comme témoin d'évolutions socio-historiques .....	172
Le mot comme témoin de prises de position .....	174
Le cotexte et la construction du sens .....	181
Les formes grammaticales et la typographie .....	186
<b>Quand le discours se fige : phraséologie, routines discursives, sloganisation</b> .....	188
Les segments répétés .....	189
Quelques pistes pour le repérage automatique de « formules » .....	190
Les segments semi-figés .....	191
<b>Comparer et caractériser des discours : la catégorie du genre</b> .....	193
<b>Discours médiatiques et saillance événementielle</b> .....	195
La densité fréquentielle .....	197
La diffusion d'un signifiant .....	198
La circulation d'une forme .....	199
<b>FICHES PRATIQUES ET APPROFONDISSEMENTS</b> .....	203
Fiche pratique 1 – Le langage des expressions régulières .....	203
Fiche pratique 2 – Commandes Unix pour manipuler des données .....	206
Fiche pratique 3 – Concaténation automatique de fichiers .....	213
Fiche pratique 4 – Construire un corpus avec Gromoteur .....	217
Fiche Approfondissement – En savoir plus sur l'analyse factorielle des correspondances .....	218
<b>BIBLIOGRAPHIE</b> .....	229
<b>INDEX</b> .....	239
<b>LES AUTEURS</b> .....	243