

TDs SLMD2 (Informatique)

Enseignant : Maria ZIMINA

CORDIAL : correcteur global et analyseur de la langue française

L'objectif de cette séance est d'utiliser la version "Universités" du logiciel CORDIAL 5 sur deux plans différents : la correction d'une part et l'analyse de données textuelles d'autre part.

Les principes de fonctionnement

La correction se déroule, sur trois niveaux : typographique, orthographique, grammatical. Chaque niveau est paramétrable. Un *premier passage* corrige les erreurs *typographiques* (doubles espaces, oubli de parenthèses, guillemets, ponctuation, etc.). Après avoir ainsi délimité et parfois restitué les unités syntaxiques que forment les phrases, un *second passage* s'attache au traitement des erreurs lexicales (fautes d'orthographe).

Lors d'une vérification orthographique, un logiciel compare la forme des mots du texte à la forme des mots de son *dictionnaire*, le moindre écart de forme produisant un message d'alerte. De la finesse et de la conception des algorithmes de recherche va dépendre le nombre de bonnes propositions.¹

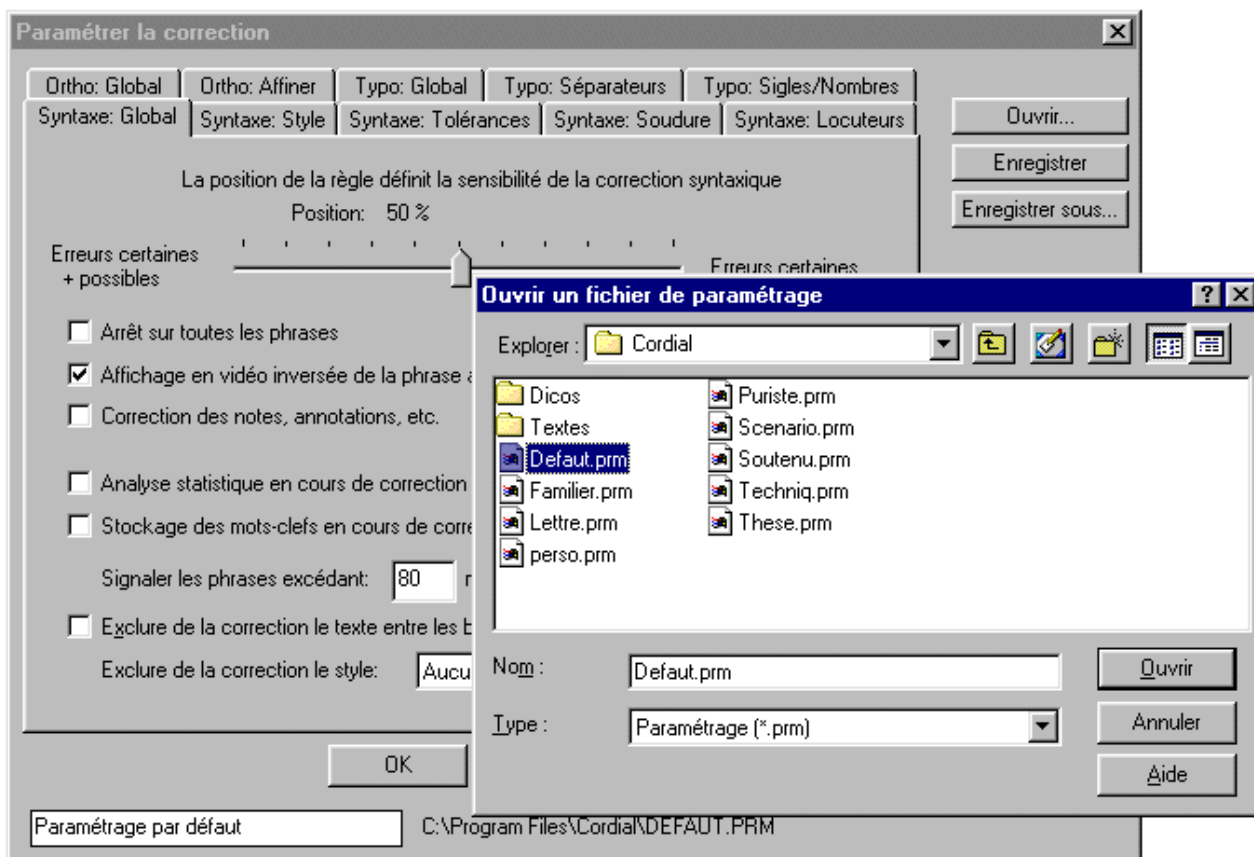
Le texte ainsi "nettoyé" peut alors être traité. Des opérations multiples (reconnaissance des formes ambiguës, découpe en propositions, délimitation des groupes, etc.) permettent de définir les fonctions et d'appliquer les règles.

Paramétrage de la correction

L'option "*Paramétrer la correction*" du menu "*Configuration*" permet d'atteindre **tous** les paramètres de correction sur lesquels vous pouvez agir. C'est une des plus importantes fonctionnalités de CORDIAL. Il existe trois grands domaines pouvant être paramétrés : la syntaxe, l'orthographe et la typographie.

Pour ouvrir une feuille correspondante, il suffit de cliquer sur l'un des onglets de la partie supérieure. Vous pouvez également effectuer le chargement du fichier de paramétrage de la correction avec lequel vous voulez travailler. Si vous n'avez pas encore créé de fichiers de paramétrage personnels, la commande "*ouvrir*" n'affiche que les huit fichiers livrés avec le logiciel : *défaut.prm*, *familier.prm*, *lettre.prm*, *puriste.prm*, etc. :

¹ On peut considérer que cette technique est maintenant au point et les meilleurs vérificateurs orthographiques se situent dans une fourchette de 80 à 90 % de propositions correctes sur les formes inconnues (lettre en trop, interversion, substitution, soudure, erreur d'accent, etc.) Par exemple, pour la forme fautive "souce" les formes proches vont apparaître ainsi "source, souche, soucie, sauce, suce, sou ce, soude". Le but consiste à faire monter une forme en première proposition et à l'isoler par une probabilité suffisamment élevée en proposition unique. Ce n'est pas toujours possible, ce ne l'est pas dans notre exemple.



Installation des dictionnaires

Configuration / Installer les dictionnaires ... : cette option permet à la fois d'installer ou de désinstaller des dictionnaires et d'obtenir des informations sur leurs caractéristiques.

En cliquant sur l'un des dictionnaires résidents, la zone du "*dictionnaire sélectionné*" fournit des informations sur celui-ci : le nom, le contenu, l'emplacement sur le disque, le nombre de mots etc. Vous pouvez **désactiver** un dictionnaire. Si, par mégarde, vous avez désactivé le dictionnaire des "*Noms communs*", les corrections orthographiques et syntaxiques deviennent irréalisables.

Vérification du texte : orthographe

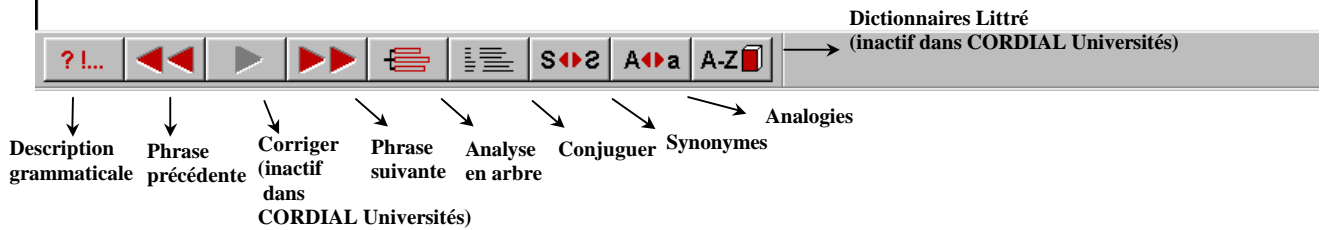
Après la saisie ou l'ouverture d'un texte, vous procédez à sa vérification orthographique par l'appel de l'item "*vérifier le texte*" du menu "*Orthographe*". Le logiciel compare les mots du texte **aux mots présents dans ses divers dictionnaires**.

Vérification du texte : syntaxe

Le menu "Syntaxe" commande les opérations de correction grammaticale. A l'appel de l'item "*corriger le texte*" s'ouvre la boîte spécifique à la correction. La partie inférieure de l'écran est occupée par la **zone des corrections**. Cette partie de correction est elle-même divisée en deux. Dans la première, s'affiche le corrigé de la phrase fautive. Les mots jugés fautifs sont colorés. Un "clic" avec le **bouton gauche** sur un mot non colorisé fait apparaître une bulle avec diverses informations sur le mot. Un **clic droit** développe les items d'aide à l'écriture de CORDIAL.

Chacun apprécie qu'on lui prête main forte car, s'il n'est pas à **cour** d'imagination, il lui arrive d'ignorer tel ou telle règle.

· cour : Dans ce contexte ce mot pourrait être confondu avec son homophone : court.



Occurrences des mots

La fonction "Occurrences orthographiques des mots" du menu "Orthographe" ou "Syntaxe" permet d'obtenir une liste de mots employés dans le texte et le nombre de leurs apparitions. Le menu "Syntaxe" fournit également les occurrences syntaxiques des mots d'un texte.

Etiquetage de textes

La fonction "Etiquetage de texte" du menu "Syntaxe" permet d'effectuer un étiquetage automatique. Etiquetez un texte, c'est associer à des segments de texte, le plus souvent les "mots", une ou plusieurs étiquette(s), le plus souvent leur catégorie grammaticale voire leur lemme. L'étiquetage se révèle souvent utile dans l'analyse de données textuelles. Par exemple, si les formes verbales sont rassemblées derrière un même lemme, il est plus aisé de les rechercher dans le texte, et de trier des conclusions de leurs fréquences cumulées.

Il existe plusieurs stratégies d'étiquetage. Cette diversité tient à l'utilisation envisagée du corpus mais aussi à son mode d'étiquetage (manuel ou automatique) ainsi que l'absence de consensus sur certaines catégories ou sur leur extension).

Exemple de manipulations à faire sous CORDIAL

1. Copier le fichier *corrigez.txt* (C:\DeugTAL\CORDIAL) dans votre répertoire de travail (C:\DeugTAL\Nom) puis effectuer la correction de ce fichier.

ATTENTION ! Pour enregistrer vos corrections dans CORDIAL Universités, créer un rapport de correction (menu Syntaxe) puis corriger le texte en fonction des erreurs signalées (figure 1) **OU** enregistrer les corrections en sortant du fichier à l'aide de la boîte de dialogue suivante (figure 2).

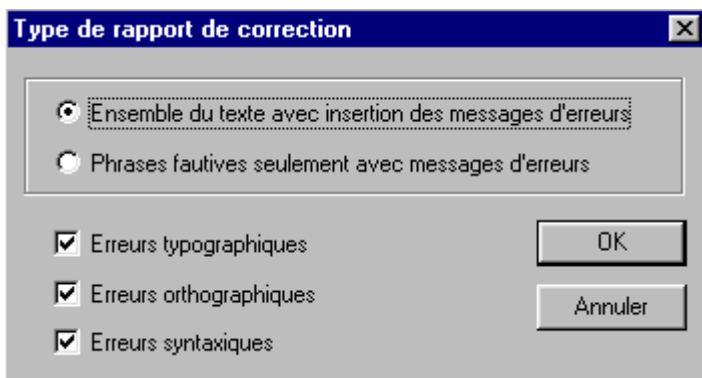


Figure 1



Figure 2

2. Accéder au fichier **dh89cor.txt** (C:\DeugTAL\CORDIAL). Copier ce fichier dans votre répertoire de travail (C:\DeugTAL\Nom).

3. En utilisant la fonction "*Occurrences des mots*", répondre aux questions suivantes :

- ?? Quel est le mot le plus fréquent de ce texte ?
- ?? Quelles sont les fréquences des mots "droit(s)", "citoyen", "homme" ?
- ?? Quels sont les contextes d'apparition du mot "droit(s)" ? Sauvegardez les contextes triés par ordre lexicographique des formes qui suivent le mot recherché.
- ?? Enregistré la liste de tous les noms communs triés par ordre alphabétique.
- ?? Quels sont les verbes le plus fréquemment utilisés dans ce texte ?
- ?? Sauvegarder la liste de tous les adverbes triés par ordre de fréquences décroissant.

4. Etiqueter le texte du fichier **dh89cor.txt** (C:\DeugTAL\CORDIAL). Enregistrer les résultats avec les options suivantes : **affichage du début des phrases, types grammaticaux et statistiques sur les catégories grammaticales.**

- ?? Quelle est la catégorie grammaticale la plus répandue dans le texte ?
- ?? Est-ce qu'il y a des erreurs dans l'attribution des types grammaticaux ? Consulter l'aide logicielle en ligne (**Aide/Index/Étiquetage de texte**) puis l'**Annexe** (pp. 5-6) pour obtenir des explications sur les codes utilisés dans CORDIAL.

5. En utilisant l'option "*Extraction de phrases du corpus...*", extraire toutes les phrases comportant les adjectifs du mot "droit(s)".

6. Si vous avez le temps, faites une analyse sémantique et stylistique du texte de la "Déclaration des droits de l'homme" (menu **Sémantique** / onglets "Sémantique" et "Stylistique"). Êtes-vous d'accord avec les résultats de l'analyse automatique ? Sauvegarder les résultats avec vos commentaires.

Rendre les résultats de cette séance de travail sur une disquette
--

ANNEXE. Type grammatical, codé de 0 à 201 ou en codage symbolique (facultatif), selon le tableau ci-dessous :

- | | |
|---|--|
| <p>0. AMS Adjectif Masculin Singulier</p> <p>1. AMP Adjectif Masculin Pluriel</p> <p>2. AFS Adjectif Féminin Singulier</p> <p>3. AFP Adjectif Féminin Pluriel</p> <p>4. AMIN Adjectif Masculin Invariant en Nombre</p> <p>5. AFIN Adjectif Féminin Invariant en Nombre</p> <p>6. ASIG Adjectif Singulier Invariant en Genre</p> <p>7. APIG Adjectif Pluriel Invariant en Genre</p> <p>8. AI Adjectif Invariant en Nombre et en Genre</p> <p>9. AD Adjectif Démonstratif</p> <p>10. AP Adjectif Possessif</p> <p>11. ANC Adjectif Numérique Cardinal</p> <p>12. ANO Adjectif Numérique Ordinal</p> <p>13. ADV Adverbe</p> <p>14. ADMS Article Défini Masculin Singulier</p> <p>15. ADFS Article Défini Féminin Singulier</p> <p>16. ADPIG Article Défini Pluriel Invariant en Genre</p> <p>17. AIMS Article Indéfini Masculin Singulier</p> <p>18. AIFS Article Indéfini Féminin Singulier</p> <p>19.</p> <p>20. COO Conjonction de Coordination</p> <p>21. SUB Conjonction de Subordination</p> <p>22. INT Interjection</p> <p>23. PREP Préposition</p> <p>24. NMS Nom Masculin Singulier</p> <p>25. NMP Nom Masculin Pluriel</p> <p>26. NFS Nom Féminin Singulier</p> <p>27. NFP Nom Féminin Pluriel</p> <p>28. NHMS Nom Masculin Singulier débutant par un "h" aspiré</p> <p>29. NHFS Nom Féminin Singulier débutant par un "h" aspiré</p> <p>30. NHSIG Nom Singulier Invariant en Genre débutant par un "h" aspiré</p> <p>31. NMIN Nom Masculin Invariant en Nombre</p> <p>32. NFIN Nom Féminin Invariant en Nombre</p> <p>33. NSIG Nom Singulier Invariant en Genre</p> <p>34. NPIG Nom Pluriel Invariant en Genre</p> | <p>35. NINom Invariant en Nombre et en Genre</p> <p>36. PPER1 S Pronom Personnel 1 e Personne du Singulier</p> <p>37. PPER2 S Pronom Personnel 2 e Personne du Singulier</p> <p>38. PPER3 S Pronom Personnel 3 e Personne du Singulier</p> <p>39. PPER1 P Pronom Personnel 1 e Personne du Pluriel</p> <p>40. PPER2 P Pronom Personnel 2 e Personne du Pluriel</p> <p>41. PPER3 P Pronom Personnel 3 e Personne du Pluriel</p> <p>42. PDS Pronom Démonstratif Singulier</p> <p>43. PDP Pronom Démonstratif Pluriel</p> <p>44. AINT Adjectif Interrogatif</p> <p>45. AIND Adjectif Indéfini</p> <p>46. NHMP Nom Masculin Pluriel débutant par un "h" aspiré</p> <p>47. PRMS Pronom Relatif Masculin Singulier</p> <p>48. PRFS Pronom Relatif Féminin Singulier</p> <p>49. AHFS Adjectif Féminin Singulier débutant par un "h" aspiré</p> <p>50. AHSIG Adjectif Invariant en Genre débutant par un "h" aspiré</p> <p>51. PRI Pronom Relatif Invariant en Genre et en Nombre</p> <p>52. PIMS Pronom Indéfini Masculin Singulier</p> <p>53. PIFS Pronom Indéfini Féminin Singulier</p> <p>54. PIMP Pronom Indéfini Masculin Pluriel</p> <p>55. PIFP Pronom Indéfini Féminin Pluriel</p> <p>56. PISIG Pronom Indéfini Singulier Invariant en Genre</p> <p>57. PIPIG Pronom Indéfini Pluriel Invariant en Genre</p> <p>58. PII Pronom Indéfini Invariant en Genre et en Nombre</p> <p>59. PPPronom Possessif</p> <p>60. NHMIN Nom Invariant en Nombre débutant par un "h" aspiré.</p> <p>100. Verbe à l'infinitif</p> <p>101. Verbe indicatif présent 1ère personne du singulier</p> <p>102. Verbe indicatif présent 2e personne du</p> |
|---|--|

singulier

- 103. Verbe indicatif présent 3e personne du singulier
- 104. Verbe indicatif présent 1ère personne du pluriel
- 105. Verbe indicatif présent 2e personne du pluriel
- 106. Verbe indicatif présent 3e personne du pluriel
- 107. Verbe indicatif imparfait 1ère personne du singulier
- 108. Verbe indicatif imparfait 2e personne du singulier
- 109. Verbe indicatif imparfait 3e personne du singulier
- 110. Verbe indicatif imparfait 1ère personne du pluriel
- 111. Verbe indicatif imparfait 2e personne du pluriel
- 112. Verbe indicatif imparfait 3e personne du pluriel
- 113. Verbe indicatif passé simple 1ère personne du singulier
- 114. Verbe indicatif passé simple 2e personne du singulier
- 115. Verbe indicatif passé simple 3e personne du singulier
- 116. Verbe indicatif passé simple 1ère personne du pluriel
- 117. Verbe indicatif passé simple 2e personne du pluriel
- 118. Verbe indicatif passé simple 3e personne du pluriel
- 119. Verbe indicatif futur simple 1ère personne du singulier
- 120. Verbe indicatif futur simple 2e personne du singulier
- 121. Verbe indicatif futur simple 3e personne du singulier
- 122. Verbe indicatif futur simple 1ère personne du pluriel
- 123. Verbe indicatif futur simple 2e personne du pluriel
- 124. Verbe indicatif futur simple 3e personne du pluriel
- 125. Verbe subjonctif présent 1ère personne du singulier
- 126. Verbe subjonctif présent 2e personne du singulier

- 127. Verbe subjonctif présent 3e personne du singulier
- 128. Verbe subjonctif présent 1ère personne du pluriel
- 129. Verbe subjonctif présent 2e personne du pluriel
- 130. Verbe subjonctif présent 3e personne du pluriel
- 131. Verbe subjonctif imparfait 1ère personne du singulier
- 132. Verbe subjonctif imparfait 2e personne du singulier
- 133. Verbe subjonctif imparfait 3e personne du singulier
- 134. Verbe subjonctif imparfait 1ère personne du pluriel
- 135. Verbe subjonctif imparfait 2e personne du pluriel
- 136. Verbe subjonctif imparfait 3e personne du pluriel
- 137. Verbe indicatif conditionnel 1ère personne du singulier
- 138. Verbe indicatif conditionnel 2e personne du singulier
- 139. Verbe indicatif conditionnel 3e personne du singulier
- 140. Verbe indicatif conditionnel 1ère personne du pluriel
- 141. Verbe indicatif conditionnel 2e personne du pluriel
- 142. Verbe indicatif conditionnel 3e personne du pluriel
- 143.
- 144. Verbe impératif présent 2e personne du singulier
- 145.
- 146. Verbe impératif présent 1ère personne du pluriel
- 147. Verbe impératif présent 2e personne du pluriel
- 148.
- 149. Verbe participe présent
- 150. Verbe participe passé masculin singulier
- 151. Verbe participe passé féminin singulier
- 152. Verbe participe passé masculin pluriel
- 153. Verbe participe passé féminin pluriel
- 201. Ponctuation faible
- 209. Ponctuation forte (fin de phrase)

