

## Cordial : Utilisation du logiciel dans le cadre d'analyses linguistico-communicationnelles.

Nous vous conseillons la lecture du chapitre dans ouvrage Panckhurst (1999), « Analyse linguistique assistée par ordinateur du courriel », in Anis J. (dir.), *Internet, communication et langue française*, Paris : Hermès, p. 55-70, dans lequel nous exposons une utilisation du logiciel Cordial. Le but de cette section consiste à montrer comment on peut procéder à des analyses morpho-syntaxiques, à l'aide d'outils comme Cordial ou Nomino, dans des domaines divers. Dans l'article cité, il s'agissait d'effectuer des recherches sur l'utilisation du courrier électronique par des étudiants et des enseignants.

Cette section sera divisée en deux parties : la correction grammaticale ; l'étiquetage morpho-syntaxique.

Pour lancer le logiciel, cliquer sur l'alias « Cordial Analyseur », qui se trouve sur le bureau.

Avant de lancer la correction grammaticale dans Cordial, nous vous conseillons de vérifier les paramètres dans le menu « Configuration ».

Plusieurs intercalaires permettent de régler la syntaxe (signalement de phrases incorrectes, phrases sans verbes, réglage de la sensibilité de la correction, etc.), l'orthographe et la typographie. Voyez l'intercalaire « Syntaxe : style » au sein du menu « Configuration/Paramétrer la correction ».

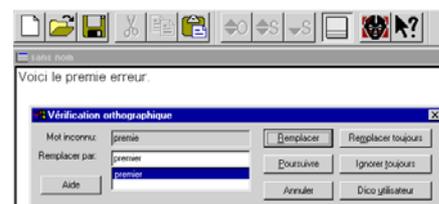
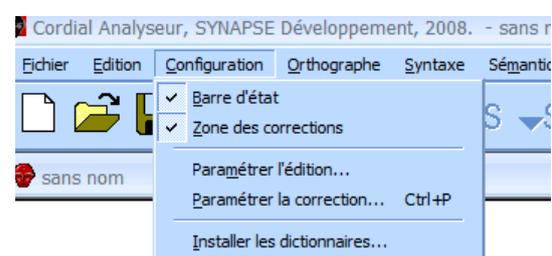
Plusieurs vérifications sont possibles pour l'orthographe :

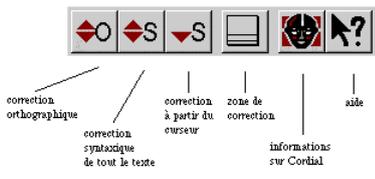
Le dernier élément du menu « Orthographe » permet de retenir tous les mots du texte et le nombre d'occurrences. Ce travail peut être utile dans le cas d'un repérage de concordances.

Dans l'exemple ci-contre, la phrase contient une erreur lexicale « premie » ; on lance la vérification orthographique en cliquant sur :

La fenêtre proposant une correction de dictionnaire s'affiche ensuite.

Une fois que la correction est lancée, l'utilisateur peut intervenir de plusieurs façons, en cliquant sur les icônes qui se trouvent en bas de l'écran, et ce, à partir des explications qui apparaissent dans la zone de correction.

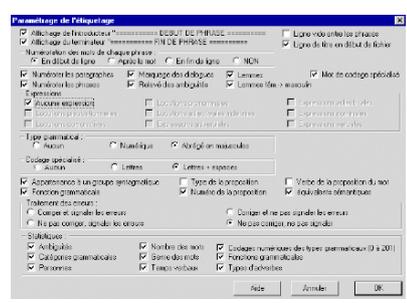




Après avoir réglé les paramètres pour que l'analyse soit la plus fine possible, procédez à une vérification de phrases. Remarque : dans cette version du logiciel, qui est plus basée sur l'analyse morpho-syntaxique que sur la correction, la possibilité d'effectuer la correction n'est pas toujours active.



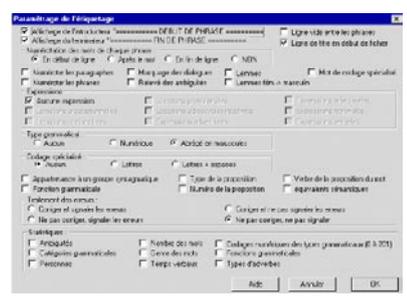
description grammaticale du mot sous le curseur / accepte la correction en cours et passe à la phrase suivante / refuse la correction de la phrase en cours et passe à la suivante / fenêtre d'analyse en arborescence / conjuguaisons / synonymes / analogies / recherche dans le Littré



Cordial propose une analyse morpho-syntaxique de la phrase. La version « Universités » de Cordial (qui est la version que nous utilisons) permet également plusieurs options supplémentaires : notamment l'étiquetage de texte et l'extraction de phrases sur occurrences et co-occurrences. Ces options se trouvent en bas du menu « Syntaxe » (« Étiquetage de texte », « Étiqueter un ensemble de fichiers », etc.) :



Pour l'étiquetage, le paramétrage par défaut est indiqué à la fenêtre ci-contre. Le choix de l'abrégié en majuscules pour le type grammatical (plutôt qu'un affichage numérique) permet de visualiser rapidement la catégorie syntaxique attribuée. On peut, bien entendu, maintenir ou désactiver les différentes options proposées en fonction de ce que l'on recherche.



À partir de la phrase « Voici un exemple d'étiquetage d'une phrase », nous avons fait un essai d'étiquetage. Après avoir introduit les réglages pour les paramètres, une fenêtre de dialogue classique, vous demandant le lieu de sauvegarde et le nom de votre fichier, s'affiche. Nous l'avons intitulé « exemple » (et son extension est .cnr ; le fichier est lisible dans Cordial, en demandant son ouverture depuis le menu « Fichier »). Les résultats apparaissent dans le fichier « exemple.cnr ».

Certains codages peuvent paraître flous dans un premier temps. Si on souhaite avoir uniquement accès aux simples catégories grammaticales, il suffit de décocher les différents options proposées dans la fenêtre de paramétrage. Dans cet exemple, nous n'avons retenu que la numérotation des mots de chaque phrase et le type grammatical en abrégé.

L'étiquetage proposé se lit de la manière suivante : voici (préposition), un (déterminant indéfini masculin singulier), exemple (nom commun masculin singulier), d' (préposition), étiquetage (nom commun masculin singulier), d' (préposition), une (déterminant indéfini féminin singulier), phrase (nom commun féminin singulier), « . » (ponctuation forte).

Pour des travaux en linguistique-informatique, notamment lorsqu'on souhaite travailler sur les résultats de fichiers étiquetés, il est souvent important d'avoir accès aux mots lemmatisés. Dans ce cas, à partir de la phrase « Voici un exemple d'étiquetage d'une phrase » contenant un codage lemmatisé, nous obtiendrons ce qui apparaît ci-contre.

Les occurrences des mots (ainsi que leurs catégories grammaticales) peuvent être consultées dans la fenêtre qui apparaît après avoir choisi « Syntaxe », « Occurrences des mots... »

La lemmatisation consiste à rapporter une unité à une forme canonique, ce qui aura pour effet de faciliter le travail ultérieur de la machine. Par exemple, un verbe fléchi sera généralement lemmatisé en verbe infinitif, un adjectif en masculin singulier, un nom en singulier, etc.

Dans l'exemple ci-contre, nous avons ajouté une phrase. Que remarquez-vous à propos de la lemmatisation ?

### Travail à faire pendant la séance :

Procédez à l'étiquetage d'un texte. Vérifiez les catégories grammaticales proposées par Cordial pour votre texte.

Quelles différences voyez-vous entre « occurrences orthographiques » et « occurrences syntaxiques » ?

L'étiquetage est-il satisfaisant ? Quels sont les types de problèmes qui se posent ?

Piste : pensez aux différents problèmes qui se posent en traitement automatique des langues, notamment, l'ambiguïté.

Faites des essais au sein des configurations (choisir, au sein du menu « Configuration », l'élément « Paramétrer la correction », puis l'onglet « Typo. Séparateurs ») : quelles différences remarquez-vous au niveau des résultats des occurrences lorsque vous choisissez soit « Retour-chariot est une fin de phrase », « Toujours », « Jamais », ou « Devant majuscule » ?

| N° | mot                       | Typegram |
|----|---------------------------|----------|
|    | ==== DEBUT DE PHRASE ==== |          |
| 1  | Voici                     | PREP     |
| 2  | un                        | DETIMS   |
| 3  | exemple                   | NCMS     |
| 4  | d'                        | PREP     |
| 5  | étiquetage                | NCMS     |
| 6  | d'                        | PREP     |
| 7  | une                       | DETIFS   |
| 8  | phrase                    | NCFS     |
| 9  | .                         | PCTFORTE |
|    | ==== FIN DE PHRASE ====   |          |

| N° | mot                       | lemme      | Typegram |
|----|---------------------------|------------|----------|
|    | ==== DEBUT DE PHRASE ==== |            |          |
| 1  | Voici                     | voici      | PREP     |
| 2  | un                        | un         | DETIMS   |
| 3  | exemple                   | exemple    | NCMS     |
| 4  | d'                        | de         | PREP     |
| 5  | étiquetage                | étiquetage | NCMS     |
| 6  | d'                        | de         | PREP     |
| 7  | une                       | un         | DETIFS   |
| 8  | phrase                    | phrase     | NCFS     |
| 9  | contenant                 | contenir   | VPARPRES |
| 10 | un                        | un         | DETIMS   |
| 11 | codage                    | codage     | NCMS     |
| 12 | lemmatisé                 | lemmatiser | VPARPMS  |
| 13 | .                         |            | PCTFORTE |
|    | ==== FIN DE PHRASE ====   |            |          |

