

TDs SLMD2 (Informatique)

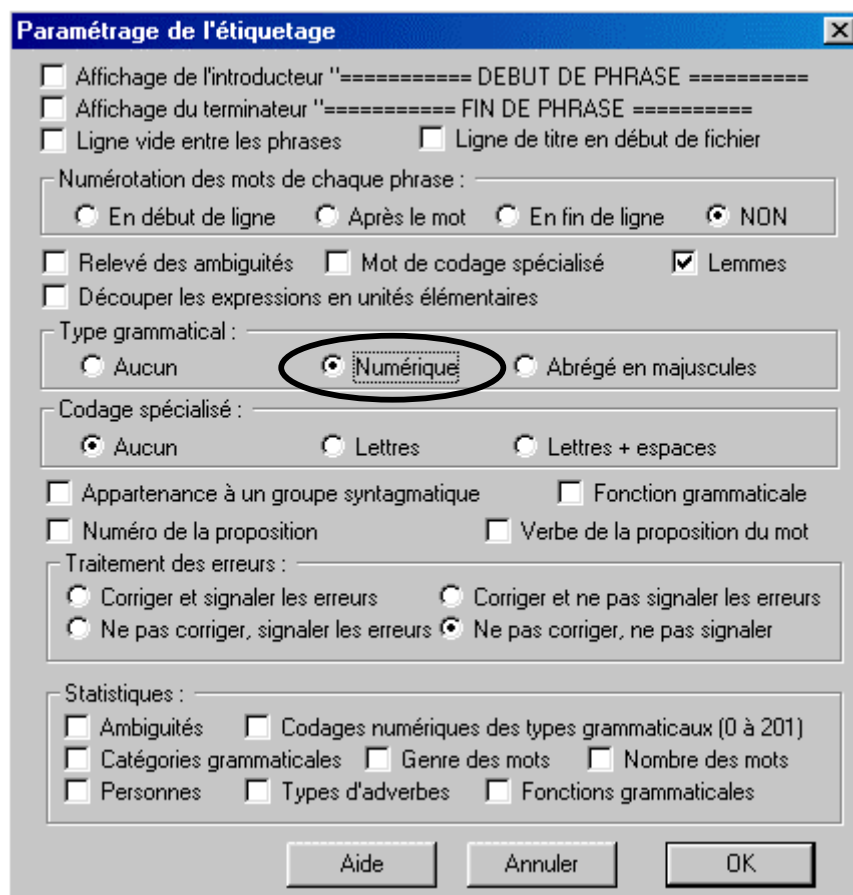
L'objectif de ces séances est d'utiliser *Microsoft Excel* pour l'analyse des résultats d'étiquetage morphosyntaxique de textes obtenues avec *CORDIAL* (cf. TD précédent).

Microsoft Excel

Visualiser le fichier **Excel_TDs.ppt** (C:\DeugTAL). Vous y trouverez des explications sur l'utilisation de Microsoft Excel.

Travail à faire sous Excel : Partie I

1. Importer **sous Excel** les résultats d'étiquetage morphosyntaxique du fichier **dh89cor.txt** enregistré avec les options suivantes (sous CORDIAL) :



2. Enregistrer le classeur Excel obtenu sous **C:\DeugTAL\VotreNom\Tableau1.xls**

3. Mettre en forme les données du classeur en rajoutant le titre "Analyse textuelle de la **Déclaration des droits de l'homme et du citoyen**", puis en rajoutant les intitulés des colonnes comme indiqué ci-dessous :

forme	lemme	code
Déclaration des droits de l'homme	déclaration des droits de l'homme	26
et	et	20
du	du	14
citoyen	citoyen	24
26	26	31
août	août	31
1789	1789	31

4. Trier le tableau obtenu par ordre alphabétique des formes, puis éliminer toutes les lignes correspondant aux signes de ponctuation. Enregistrer le tableau.

5. Créer une copie de la feuille de calcul obtenue dans le même classeur. Attribuer le nom "Analyse" à cette nouvelle feuille.

6. Dans la feuille "Analyse" créer un *filtre automatique* qui permet d'afficher tous les verbes utilisés dans la **Déclaration des droits de l'homme et du citoyen**.

Note : Utiliser les codes des catégories grammaticales du CORDIAL (cf. l'annexe du TD "prise en main du CORDIAL").

7. Parmi les verbes, lesquels sont employer à l'infinitif dans le texte ? Compléter le filtrage précédent avec un deuxième critère de sélection.

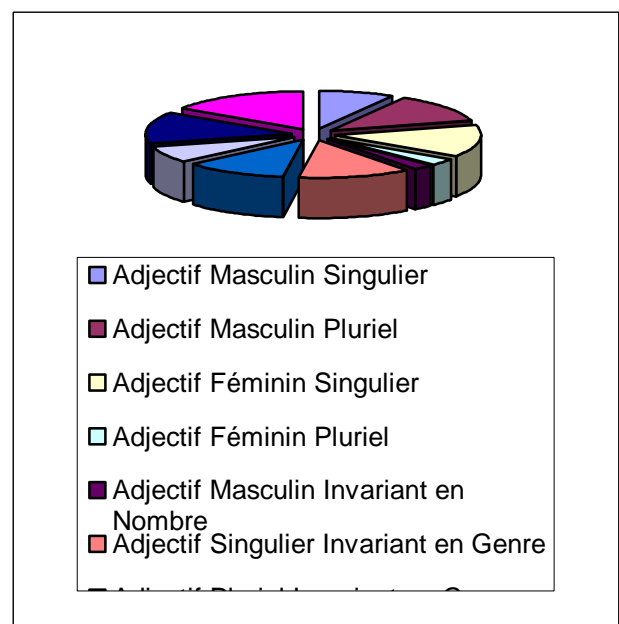
8. Dans la feuille de calcul "Analyse", créer un *filtre élaboré* qui permet d'afficher tous les verbes employés au pluriel. Copier les résultats de la sélection vers la feuille de calcul "Résultats1". Trier la liste des verbes par code.

9. Dans la feuille de calcul "Analyse", créer un *filtre élaboré* qui permet d'afficher tous les adjectifs utilisés dans le texte de la **Déclaration**. Copier les résultats de la sélection vers la feuille de calcul "Résultats2". Trier la liste des adjectifs par code.

10. En utilisant la fonctionnalité *Donnée/Sous-totaux*, calculer le nombre d'adjectifs dans chaque sous-catégorie grammaticale attribuée par **CORDIAL** (Adjectif Masculin Singulier, Adjectif Masculin Pluriel, Adjectif Féminin Singulier, Adjectif Féminin Pluriel etc.).

11. Créer un graphique type "Secteurs" pour visualiser la part de chaque sous-catégorie d'adjectifs au sein de cette catégorie grammaticale.

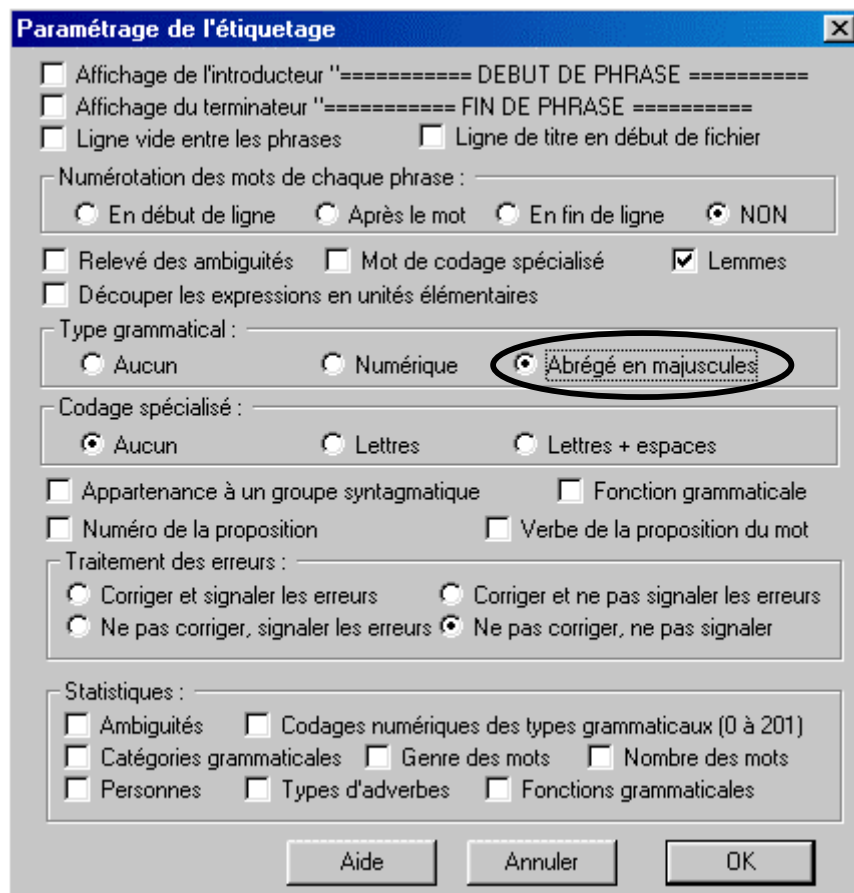
Note : Pour une meilleure lisibilité du graphique, pensez à écrire les libellés des codes en entier.



12. A partir du tableau dans la feuille "Analyse" Créer un *tableau croisé dynamique* qui permet d'afficher pour chaque *lemme* le nombre de *formes* du texte.

Partie II

13. Importer **sous Excel** les résultats d'étiquetage morphosyntaxique du fichier **dh89cor.txt** enregistré avec les options suivantes (sous CORDIAL) :



14. Enregistrer le classeur Excel obtenu sous **C:\DeugTAL\VotreNom\Tableau2.xls**

15. Mettre en forme les données du classeur.

16. Trier le tableau obtenu par ordre alphabétique des formes, puis éliminer toutes les lignes correspondant aux signes de ponctuation. Enregistrer le tableau.

17. Croiser les données des deux tableaux Excel : insérer la colonne avec les abrégés des codes grammaticaux en majuscules (classeur **Tableau2.xls**) dans la feuille de calcul du classeur **Tableau1.xls** Voici le tableau que vous allez avoir :

forme	lemme	code	code Lettres
a	avoir	103	VINDP3S
a	avoir	103	VINDP3S
a	avoir	103	VINDP3S
a	avoir	103	VINDP3S
a	avoir	103	VINDP3S
à	à	23	PREP

18. Créer un *filtre automatique* pour afficher tous les noms du texte employés au singulier.
19. Procéder à l'extraction automatique de tous les pronoms. Quelle est la proportion de pronoms par rapport au nombre total d'occurrences des mots du texte ? Créer un graphique pour l'illustrer.
20. Créer un graphique qui affiche la contribution de chaque catégorie grammaticale (noms, adjectifs, verbes, adverbes, déterminants, conjonctions, interjections, pronoms, prépositions) au nombre total d'occurrences des mots du texte. Quelle catégorie est-elle la plus répandue dans le texte ?

Note : Utilisez les filtres pour regrouper ensembles les sous-catégories au sein de chaque catégorie grammaticale. Vous pouvez comparer vos statistiques avec celles du CODIAL mais n'oubliez pas que nous avons exclu volontairement les signes de ponctuations de nos décomptes.

Partie III

21. Trouver un texte en français dont le nombre d'occurrences des mots est comparable à la **Déclaration des droits de l'homme et du citoyen**. Procéder à l'analyse morphosyntaxique de ce texte avec CORDIAL. Emporter les résultats d'étiquetage sous Excel, puis calculer pour chaque catégorie grammaticale (noms, adjectifs, verbes, adverbes, déterminants, conjonctions, interjections, pronoms, prépositions) sa contribution au total d'occurrences des mots (%).
22. Créer un graphique type "Radar" qui permet de comparer les données de l'analyse morphosyntaxique des deux textes : le votre et celui de la **Déclaration des droits de l'homme et du citoyen**.