## **UNIVERSITE DE LA SORBONNE NOUVELLE - PARIS 3**

#### Année universitaire 2002-2003

### 2ème SESSION

# **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) :

Paramétrage de l'étiquetage				
Affichage de l'introducteur ''====== DEBUT DE PHRASE ========     Affichage du terminateur ''======= FIN DE PHRASE =========     Ligne vide entre les phrases     Ligne de titre en début de fichier				
Numérotation des mots de chaque phrase : C En début de ligne C Après le mot C En fin de ligne © NON				
☐ Relevé des ambiguités ☐ Mot de codage spécialisé				
C Aucun				
Codage spécialisé : • Aucun         C Lettres         C Lettres         + espaces				
Appartenance à un groupe syntagmatique     Fonction grammaticale     Numéro de la proposition     Traitement des erreurs :				
C Corriger et signaler les erreurs C Ne pas corriger, signaler les erreurs Ne pas corriger, signaler les erreurs Ne pas corriger, ne pas signaler				
Statistiques : Ambiguités Codages numériques des types grammaticaux (0 à 201) Catégories grammaticales Genre des mots Nombre des mots Personnes Types d'adverbes Fonctions grammaticales				
Aide Annuler OK				

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/Soustotaux*, 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) :

<ul> <li>Affichage de l'introducteur " DEBUT DE PHRASE</li></ul>	Paramétrage de l'étiquetage 🔀				
Numérotation des mots de chaque phrase :         C En début de ligne       Après le mot       En fin de ligne       NON         Relevé des ambiguités       Mot de codage spécialisé       Lemmes         Découper les expressions en unités élémentaires         Type grammatical :       Aucun       Numérique       Abrégé en majuscules         Codage spécialisé :       Aucun       Lettres       Lettres + espaces         Appartenance à un groupe syntagmatique       Fonction grammaticale         Numéro de la proposition       Verbe de la proposition du mot         Traitement des erreurs :       Corriger et signaler les erreurs       Corriger et ne pas signaler les erreurs         Ne pas corriger, signaler les erreurs © Ne pas corriger, ne pas signaler       Statistiques :       Ambiguités         Ambiguités       Codages numériques des types grammaticaux (0 à 201)       Catégories grammaticales         Personnes       Types d'adverbes       Fonctions grammaticales	Affichage de l'introducteur "====== DEBUT DE PHRASE ========     Affichage du terminateur "======== FIN DE PHRASE =========     Ligne vide entre les phrases    Ligne de titre en début de fichier				
<ul> <li>Relevé des ambiguités Mot de codage spécialisé Lemmes</li> <li>Découper les expressions en unités élémentaires</li> <li>Type grammatical :</li> <li>Aucun</li> <li>Numérique</li> <li>Abrégé en majuscules</li> <li>Codage spécialisé :</li> <li>Aucun</li> <li>Lettres</li> <li>Lettres + espaces</li> <li>Appartenance à un groupe syntagmatique</li> <li>Fonction grammaticale</li> <li>Numéro de la proposition</li> <li>Verbe de la proposition du mot</li> <li>Traitement des erreurs Corriger et ne pas signaler les erreurs</li> <li>Corriger et signaler les erreurs</li> <li>Ne pas corriger, signaler les erreurs</li> <li>Ne pas corriger, signaler les erreurs</li> <li>Codages numériques des types grammaticaux (0 à 201)</li> <li>Catégories grammaticales</li> <li>Genre des mots</li> <li>Nombre des mots</li> </ul>	Numérotation des mots de chaque phrase : O En début de ligne O Après le mot O En fin de ligne O NON				
Aucun Numérique Abrégé en majuscules Codage spécialisé : Aucun Lettres Lettres + espaces Appartenance à un groupe syntagmatique Fonction grammaticale Numéro de la proposition Verbe de la proposition du mot Traitement des erreurs : Corriger et signaler les erreurs Corriger et ne pas signaler les erreurs Ne pas corriger, signaler les erreurs Ne pas corriger, ne pas signaler Statistiques : Ambiguités Codages numériques des types grammaticaux (0 à 201) Catégories grammaticales Genre des mots Nombre des mots Personnes Types d'adverbes Fonctions grammaticales	Relevé des ambiguités     Mot de codage spécialisé     Lemmes     Découper les expressions en unités élémentaires     Type grammatical :				
Codage specialise : <ul> <li>Aucun</li> <li>Lettres</li> <li>Lettres + espaces</li> </ul> <li>Appartenance à un groupe syntagmatique  Fonction grammaticale         <ul> <li>Numéro de la proposition</li> <li>Verbe de la proposition du mot</li> <li>Traitement des erreurs :             <ul> <li>Corriger et signaler les erreurs</li> <li>Corriger et signaler les erreurs</li> <li>Ne pas corriger, signaler les erreurs</li></ul></li></ul></li>	C Aucun C Numérique Abrégé en majuscules				
<ul> <li>Appartenance à un groupe syntagmatique</li> <li>Fonction grammaticale</li> <li>Numéro de la proposition</li> <li>Verbe de la proposition du mot</li> <li>Traitement des erreurs :</li> <li>Corriger et signaler les erreurs</li> <li>Corriger et signaler les erreurs</li> <li>Corriger et ne pas signaler les erreurs</li> <li>Ne pas corriger, signaler les erreurs</li> <li>Statistiques :</li> <li>Ambiguités</li> <li>Codages numériques des types grammaticaux (0 à 201)</li> <li>Catégories grammaticales</li> <li>Genre des mots</li> <li>Nombre des mots</li> <li>Personnes</li> <li>Types d'adverbes</li> </ul>	Aucun     C Lettres     C Lettres + espaces				
Statistiques : Ambiguités Codages numériques des types grammaticaux (0 à 201) Catégories grammaticales Genre des mots Nombre des mots Personnes Types d'adverbes Fonctions grammaticales	Appartenance à un groupe syntagmatique     Fonction grammaticale     Numéro de la proposition     Traitement des erreurs:     Corriger et signaler les erreurs     Corriger et ne pas signaler les erreurs				
Aide Annuler OK					

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
а	avoir	103	VINDP3S
а	avoir	103	VINDP3S
а	avoir	103	VINDP3S
а	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.**