

**M2SOL034 Corpus, ressources et
linguistique outillée**
TD 4 : Unitex

Ljudmila PETKOVIĆ

Sorbonne Université

Master « Langue et Informatique » (M1 ScLan)

UFR Sociologie et Informatique pour les Sciences Humaines

Semestre 2, 2024-2025, le 21 février 2025

Table des matières

1 Exercices	1
2 Solutions	2

1 Exercices

1. Manipulation de graphe

- compléter le graphe `je,tu,il...` pour qu'il reconnaisse un pronom personnel suivi d'un verbe (n'importe lequel, sous n'importe quelle forme)
- l'appliquer sur le texte

Les exercices suivants sont à faire sur le corpus *Tour du monde en 80 jours* de Jules Verne.

2. Recherches de motifs complexes

- toutes les occurrences des pronoms personnels (*je, tu, il...*)
- toutes les occurrences des pronoms personnels qui sont suivis par un verbe
- toutes les suites d'au moins 3 adjectifs (**A**) : qu'observez-vous de surprenant ?

- toutes les suites de noms. Pourquoi le motif `<N>*` produit-il une erreur ? Que faire pour l'éviter ?
- 3. **Recherches utilisant les informations grammaticales|flexionnelles|sémantiques**
 - tous les adjectifs au féminin pluriel
 - tous les mots possédant le trait sémantique « humain collectif »
 - tous les verbes à l'imparfait du langage courant
- 4. **Recherches complexes utilisant la concaténation et l'union**
 - tous les verbes, soit à l'imparfait, soit au présent ou à l'imparfait du subjonctif
- 5. **Recherches utilisant les négations**
 - tous les mots qui ne sont pas dans le dictionnaire
 - tous les mots qui ne sont pas écrits tout en minuscules
 - tous les noms non humains
- 6. **Recherches à l'aide de méta motifs**
 - tous les mots commençant par une majuscule
 - tous les mots qui possèdent le trait sémantique « concret »
- 7. **Recherches utilisant les filtres**
 - tous les mots qui commencent par *anti* ou *pro*, suivis par un tiret facultatif
 - tous les mots composés contenant un tiret
 - tous les mots qui ne sont pas dans le dictionnaire et qui se terminent par *es*
- 8. **Transformer le texte**

Créer un graphe qui réalise un début d'anonymisation du *Tour du monde en 80 jours*

 - remplacer toutes les occurrences de *Phileas Fogg* par **Pers1**
 - remplacer toutes les occurrences de *Passepartout* par **Pers2**
 - l'appliquer sur le texte et vérifier les résultats obtenus
- 9. **Annoter le texte**

Créer un graphe qui réalise un début d'annotation en entités nommées du *Tour de monde en 80 jours*

 - annoter toutes les occurrences de *Phileas Fogg* et de *Passepartout* par des balises `<pers></pers>`
 - appliquer le graphe sur le texte et vérifier les résultats obtenus

2 Solutions

Le contenu de cette présentation est sous licence **CC-BY-NC-SA 4.0**
 Utilisation non commerciale – Partage dans les mêmes conditions.

