

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 25 février 2025

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



Table des matières

1 Exercices	1
2 Solutions	3

1 Exercices

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

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

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 de 3 adjectifs (A) : qu'observez-vous de surprenant ?

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 métamotifs**

- 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. **Regex avancées**

- tous les verbes un peu ou très spécialisés, soit au participe passé, soit à l'infinitif
- toutes les séquences
 - commençant par le verbe *avoir*, et
 - se terminant par un verbe au participe passé
- tous les verbes au subjonctif passé ou présent, contenant *uiss*

9. **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

10. **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

1. Repérage des pronoms personnels suivis d'un verbe (Fig. 1).

je+tu+il+elle+on+nous+vous+ils+elles <V>

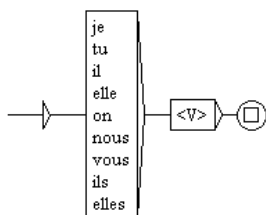


FIGURE 1 – Repérage des pronoms personnels suivis d'un verbe.

2. Repérage des pronoms personnels avec une seule balise, suivis d'un verbe (Fig. 3).

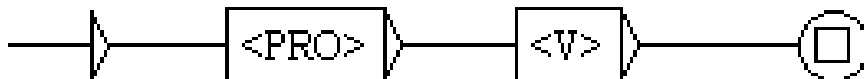


FIGURE 2 – Repérage des pronoms avec une seule balise, suivis d'un verbe.

Repérage des suites d'au moins 3 adjectifs (A).

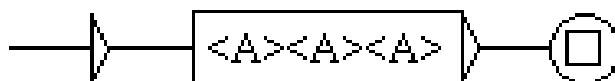


FIGURE 3 – Repérage des suites de trois adjectifs.

Nous remarquons également le repérage de certains faux positifs dans le résultat du concordancier (adverbes, locutions. ...).

3. Repérage de tous les adjectifs au féminin pluriel (Fig. 4).



FIGURE 4 – Repérage des adjectifs au féminin pluriel.

Repérage de tous les noms possédant le trait sémantique « humain collectif » (Fig. 5).

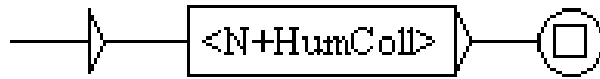


FIGURE 5 – Repérage de tous les noms possédant le trait sémantique « humain collectif ».

Repérage des verbes à l'imparfait du langage courant (Fig. 6).

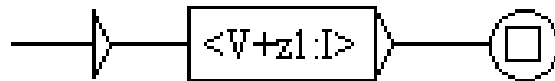


FIGURE 6 – Repérage des verbes à l'imparfait du langage courant.

4. Repérage de tous les verbes, soit à l'imparfait, soit au présent ou à l'imparfait du subjonctif (Fig. 7).

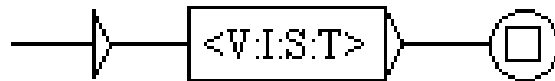


FIGURE 7 – Repérage de tous les verbes, soit à l'imparfait, soit au présent ou à l'imparfait du subjonctif.

5. Repérage de tous les mots qui ne sont pas dans le dictionnaire (Fig. 8).



FIGURE 8 – Repérage des mots hors dictionnaire.

Repérage de tous les mots qui ne sont pas écrits tout en minuscules (Fig. 9).

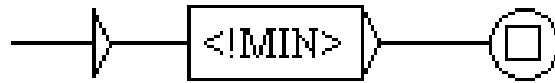


FIGURE 9 – Repérage de tous les mots qui ne sont pas écrits tout en minuscules.

Repérage de tous les noms non humains (Fig. 10).



FIGURE 10 – Repérage de tous les mots non humains.

6. Repérage de tous les mots commençant par une majuscule à l'aide des métamotifs (Fig. 11).



FIGURE 11 – Repérage de tous les mots commençant par une majuscule à l'aide des métamotifs.

Repérage de tous les mots qui possèdent le trait sémantique « concret » (Fig. 12).



FIGURE 12 – Repérage de tous les mots qui possèdent le trait sémantique « concret ».

7. Repérage de tous les mots qui commencent par *anti* ou *pro*, suivis par un tiret facultatif (Fig. 13).



FIGURE 13 – Repérage de tous les mots qui commencent par *anti* ou *pro*, suivis par un tiret facultatif.

Repérage de tous les mots composés contenant un tiret (Fig. 14).



FIGURE 14 – Repérage de tous les mots composés contenant un tiret.

Repérage de tous les mots qui ne sont pas dans le dictionnaire et qui se terminent par *es* (Fig. 15).



FIGURE 15 – Repérage de tous les mots qui ne sont pas dans le dictionnaire et qui se terminent par *es*.

8. Repérage de tous les verbes un peu ou très spécialisés, soit au participe passé, soit à l'infinitif (Fig. 16).

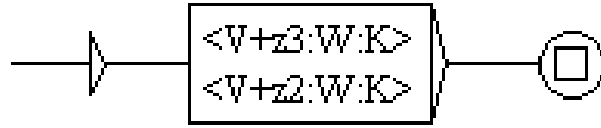


FIGURE 16 – Repérage de tous les verbes un peu ou très spécialisés, soit au participe passé, soit à l’infinitif.

Repérage de toutes les séquences commençant par le verbe *avoir* et se terminant par un verbe au participe passé (Fig. 17).

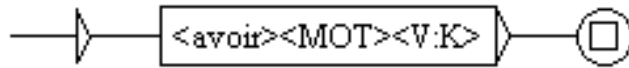


FIGURE 17 – Repérage de toutes les séquences commençant par le verbe *avoir* et se terminant par un verbe au participe passé.

Repérage de tous les verbes au subjonctif passé ou présent, contenant *uiss* (Fig. 18).

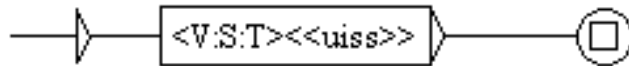


FIGURE 18 – Repérage de tous les verbes au subjonctif passé ou présent, contenant *uiss*.

9. Remplacer toutes les occurrences de *Phileas Fogg* par **Pers1** et de *Passepartout* par **Pers2** (Fig. 19).

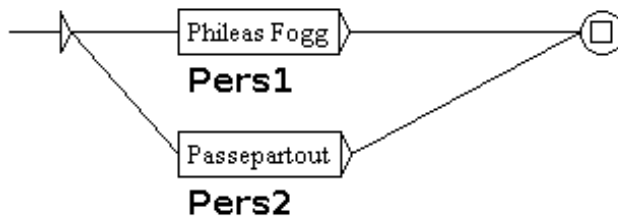


FIGURE 19 – Remplacer toutes les occurrences de *Phileas Fogg* par **Pers1** et de *Passepartout* par **Pers2**.

10. Annoter toutes les occurrences de *Phileas Fogg* et de *Passepartout* par des balises `<pers></pers>` (Fig. 20).

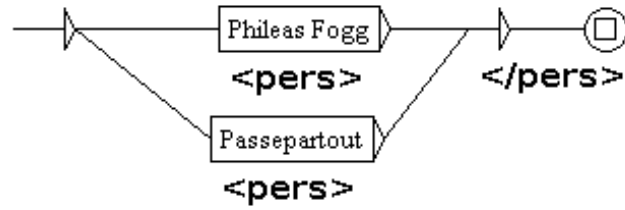


FIGURE 20 – Annoter toutes les occurrences de *Phileas Fogg* et de *Passepartout* par des balises `<pers></pers>`.