

TD – Langue et Informatique - Révisions générales

P. Mercuriali – L2 – Centre Tesnière – CRIT – UMLP

Année 2025-2026

Objectifs pédagogiques

- Réviser les notions clés du traitement automatique des langues (TAL) (langue et informatique)
- Manipuler à nouveau des outils concrets pour des tâches de
 - segmentation,
 - analyse,
 - traduction assistée par ordinateur (TAO)
- Comprendre les limites des outils modernes : LLM, chatbots
- Développer une réflexion critique sur les logiciels d'analyse automatique

1 Introduction

Quelques concepts clés...

1. Qu'est-ce que le TAL ?
2. Quelle est la différence entre traduction automatique et TAO ?
3. Qu'est-ce qu'un token ?
4. Quelle est la différence entre morphologie et syntaxe ?
- 5.

2 TAO et segmentation

Considérez la phrase suivante.

Time flies like an arrow, fruit flies like a banana.

2.1 Segmentation

1. Nous considérons que les tokens sont les mots. Donnez des critères pour déterminer un token dans la phrase ci-dessus.
2. Segmentez la phrase à la main et faites la liste de tous les tokens différents par ordre alphabétique.
3. Y a-t-il des tokens qui sont répétés dans la phrase ?

4. Y a-t-il des mots qui sont utilisés dans deux contextes différents ?
5. Quel(s) problème(s) cela peut-il poser dans le cas d'une traduction automatique ?

2.2 Traduction assistée par ordinateur (TAO)

1. Traduisez cette phrase « à la main », sans utiliser d'outils.
2. Quel(s) problème(s) peut rencontrer un outil de traduction automatique ?
3. Traduisez cette phrase à l'aide d'un outil : DeepL, Google Translate, etc.
4. Comparez votre traduction humaine et celle de l'outil. L'outil a-t-il rencontré le(s) problème(s) déterminé(s) ci-dessus ?

Bonus Demandez à un traducteur automatique de traduire la phrase suivante :

*Buffalo buffalo Buffalo buffalo buffalo buffalo Buffalo buffalo.*¹

qui est une phrase grammaticalement correcte en anglais !

3 Analyse morphologique

Le mot *incroyable* peut être décomposé de la manière suivante :

- *in-* (préfixe)
 - *croy* (racine)
 - *-able* (suffixe).
1. Donnez le rôle (= le sens, le contexte d'utilisation) des affixes dans le cas d'*incroyable*.
 2. Décomposez et analysez les rôles des affixes dans les mots suivants :
 - *internationalisation*
 - *incompréhension*
 - *défait*
 - *délavé*
 3. Que remarquez-vous à propos du suffixe *dé-* dans ces deux derniers cas ?

4 Arbres syntaxiques

Une phrase en français Considérez la phrase suivante.

Le mur murant Paris rend Paris murmurant.

1. Représentez la phrase sous forme d'un arbre syntaxique.
2. Donnez les règles grammaticales qui permettent de construire cette phrase.

Une phrase en anglais Considérez la phrase anglaise suivante.

Buffalo buffalo buffalo Buffalo buffalo.

1. Sachant qu'un *buffalo* est un type de bison d'Amérique du Nord, que le verbe *buffalo* est un synonyme d'*intimider*, et que *Buffalo* fait référence à une ville (Buffalo, dans l'État de New York), donnez des étiquettes grammaticales à chacun des mots de cette phrase.
2. Proposez un arbre syntaxique pour cette phrase.
3. Comment comprenez-vous cette phrase ?

1. Voir : Dmitri Borgmann, *Beyond Language : Adventures in Word and Thought*, 1967.

5 LLM et segmentation avancée

Considérez le texte suivant.

*A panda walks into a café. He orders a sandwich, eats it, then draws a gun and fires two shots in the air. "Why?" asks the confused waiter, as the panda makes towards the exit. The panda produces a badly punctuated wildlife manual and tosses it over his shoulder. "I'm a panda," he says at the door. "Look it up." The waiter turns to the relevant entry in the manual and, sure enough, finds an explanation. "Panda. Large black-and-white bear-like mammal, native to China. Eats, shoots and leaves."*²

1. Est-il toujours pertinent de segmenter par mots séparés par des espaces ici ? Donnez un exemple de segment (token) plus petit qu'un "mot", et un exemple de segment plus grand qu'un "mot".
2. Segmentez le texte thématiquement.
3. Demandez à un LLM de votre choix de segmenter le texte thématiquement.
4. Le LLM a-t-il repéré la subtilité ?
5. Comparez votre segmentation et celle du LLM.

2. Ursula Le Guin, *Steering the Craft*, 1998.