

TD – Langue et Informatique - Segmentation et tokenization par LLM

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

Année 2025-2026

Objectifs pédagogiques

- Comprendre comment les IA/LLM (type ChatGPT) découpent un texte
- Réfléchir aux usages en traduction, en résumé du texte, en analyse du discours
- Développer un regard critique sur les résultats produits

1 Rappels théoriques

(Rappels après les vacances!)

Questions/discussion

- Qu'est-ce que la segmentation ?
- Pourquoi ne pas découper en phrases ?
- Dans quels métiers la segmentation est-elle utile ?

Important la segmentation n'est pas une découpe mécanique.

Quelques exemples

- Traduction : découpe en unités de sens (qu'est-ce que le sens ?)
- Résumé : identification des idées d'un texte (qu'est-ce que les idées ?)
- Analyse du discours : structure argumentative
- SEO (Search Engine Optimization), rédaction web : amélioration de la lisibilité
- TAL/IA : prétraitement des données textuelles

2 Activité manuelle, segmentation humaine

Exemple de texte :

Mais, vous savez, moi je ne crois pas qu'il y ait de bonne ou de mauvaise situation. Moi, si je devais résumer ma vie aujourd'hui avec vous, je dirais que c'est d'abord des rencontres, des gens qui m'ont tendu la main, peut-être à un moment où je ne pouvais pas, où j'étais seul chez moi. Et c'est assez curieux de se dire que les hasards, les rencontres forment une destinée... Parce que quand on a le goût de la chose, quand on a le goût de la chose bien faite, le beau geste, parfois on ne trouve pas l'interlocuteur en face, je dirais, le miroir qui vous aide à avancer. Alors ce n'est pas mon cas, comme je le disais là, puisque moi au contraire, j'ai pu ; et je dis merci à la vie, je lui dis merci, je chante la vie, je danse la vie... Je ne suis qu'amour ! Et finalement, quand beaucoup de gens aujourd'hui me disent : « Mais comment fais-tu pour avoir cette humanité ? » Eh bien je leur réponds très simplement, je leur dis que c'est ce goût de l'amour, ce goût donc qui m'a poussé aujourd'hui à entreprendre une construction mécanique, mais demain, qui sait, peut-être simplement à me mettre au service de la communauté, à faire le don, le don de soi...

Consigne

1. Segmenter le texte en unités thématiques
2. Segmenter le texte en unités argumentatives
3. Donnez un titre pour chaque segment

Mise en commun

- Quelles sont les divergences ?
- Quels critères avez-vous utilisés ?
 - connecteurs ?
 - thèmes ?
 - ton ?
 - progression logique/argumentative ?

Conclusion La segmentation est subjective et dépend de l'interprétation !

3 Activité : segmentation par LLM

But : comparer les résultats de différents prompts sur un même LLM.

Exemple de LLM : ChatGPT, Claude, Mistral, Rufus (?), Gemini, etc.

Consigne Utilisez le prompt suivant :

Segmente ce texte en unités de sens cohérentes.

Donne un titre à chaque segment.

Justifie brièvement le découpage.

Choisissez un prompt chacun parmi les 4 prompts suivants :

- Segmentation thématique :

Segmente ce texte selon les thèmes abordés.

Donne un titre à chaque segment.

Justifie brièvement le découpage.

- Segmentation argumentative :

Segmente ce texte selon la structure argumentative.

Donne un titre à chaque segment.

Justifie brièvement le découpage.

- Segmentation pour la traduction :

Segmente ce texte en unités optimales pour la traduction.

Donne un titre à chaque segment.

Justifie brièvement le découpage.

(Bonus : comparer avec et sans spécifier la langue !)

- Segmentation rhétorique :

Segmente ce texte selon sa progression rhétorique.

Donne un titre à chaque segment.

Justifie brièvement le découpage.

Analyse

1. Comparez les segments produits, les titres, et les logiques de découpages.
2. L'IA segmente-t-elle comme les humains ?
3. Où se trompe-t-elle ? Où est-elle meilleure ?
4. La logique de découpage correspond-elle à la logique que l'IA explique ? (i.e. peut-on faire confiance à ses explications ?)

4 Activité : effets du prompt

Consigne

1. Testez le prompt suivant :

Segmente strictement ce texte en 5 parties.
Donne un titre à chaque segment.
Justifie brièvement le découpage.

2. Testez le prompt suivant :

Segmente librement ce texte selon sa logique interne.
Donne un titre à chaque segment.
Justifie brièvement le découpage.

3. Comparez.

5 Discussion

Questions

- L'IA segmente-t-elle "objectivement" ?
- Peut-elle imposer une lecture ?
- Quels sont les risques en traduction ?
- Quels sont les risques en analyse du discours ?