



# **Recueil et structuration de corpus – TD 2**

Achille Falaise – Alexandre Roulois

# Plan du TD

- Correction TD 1
  - Format
  - URLs
  - Variables
- Couverture d'un corpus
- XML
  - Théorie : boîtes et balises/attributs
  - Exemple : *Corpus 14* (correspondances de [Poilus](#))
  - Pratique : formatage d'un document

# Format

**Importer**

Jeu de caractères :

Langue :

À partir de la ligne :  - +

**Options de séparateur**

Largeur fixe  Séparé par

Tabulation  Virgule  Point-virgule  Espace  Autre

Fusionner les séparateurs  Espaces superflus Séparateur de chaîne de caractères :

**Autres options**

Formater les champs entre guillemets comme texte  Détecter les nombres spéciaux

**Champs**

Type de colonne :

|   | Standard   |
|---|--|
| 1 | titre URL genre textuel codage format licence taille nombre de 'vaccination  |
| 2 | Covid-19 : « pas de raison » de reconfiner, efficacité du vaccin, immunité collect   |
| 3 | Covid chez l'enfant : un danger plus grand que prévu ? Derniers chiffres et derni  |
| 4 | Coronavirus : quelles sont les séquelles du CoVid-19 ? <a href="https://www.doctissimo.fr/">https://www.doctissimo.fr/</a>                             |
| 5 | Covid-19 : le risque de la fatigue démocratique <a href="https://www.lemonde.fr/idees/artic">https://www.lemonde.fr/idees/artic</a>                    |
| 6 | Covid-19 : les difficultés pour accélérer le rythme de la campagne de vaccination  |
| 7 | Maladie à coronavirus 2019 <a href="https://fr.wikipedia.org/wiki/Maladie_%C3%A0_coronavir">https://fr.wikipedia.org/wiki/Maladie_%C3%A0_coronavir</a> |
| 8 | Pandémie de Covid-19 en France <a href="https://fr.wikipedia.org/wiki/Pand%C3%A9mie_de_Cov">https://fr.wikipedia.org/wiki/Pand%C3%A9mie_de_Cov</a>     |

# Éditeur hexadécimal

Valeur hexadécimale  
(= base 16)

Caractère ASCII

File Edit View Windows Help

|          |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |     |         |   |                     |             |            |  |
|----------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|-----|---------|---|---------------------|-------------|------------|--|
| 00000000 | EF | BB | BF | 74 | 69 | 74 | 12 | 65 | 20 | 20 | 20 | 55 | 52 | 4C | 20 | 67 | 65 | 6E | 72 | 65 | 20 | 74 | 65 | 78 | 74 | 75 | 65 | 6C | 20 | ... | titre   | URL   | genre               | textuel     |            |  |
| 0000001D | 20 | 20 | 63 | 6F | 64 | 61 | 67 | 65 | 20 | 20 | 66 | 6F | 72 | 6D | 61 | 74 | 20 | 20 | 6C | 69 | 63 | 65 | 6E | 63 | 65 | 20 | 74 | 61 | 69 | ... | codage  | format  | licence             | taille      |            |  |
| 0000003A | 6C | 6C | 65 | 20 | 20 | 6E | 6F | 6D | 62 | 72 | 65 | 20 | 64 | 65 | 20 | E2 | 80 | 98 | 76 | 61 | 63 | 63 | 69 | 6E | 61 | 74 | 69 | 6F | 6E | ... | nombre  | de  | ...                 | vaccination |            |  |
| 00000057 | E2 | 80 | 99 | 20 | 6E | 6F | 6D | 62 | 72 | 65 | 20 | 64 | 65 | 20 | 27 | 76 | 61 | 63 | 63 | 69 | 6E | 27 | 0D | 0A | 43 | 6F | 76 | 69 | 64 | ... | nombre  | de  | 'vaccin'            | ..Covid     |            |  |
| 00000074 | 2D | 31 | 39 | E2 | 80 | 89 | 3A | 20 | C2 | AB | E2 | 80 | 89 | 70 | 61 | 73 | 20 | 64 | 65 | 20 | 72 | 61 | 69 | 73 | 6F | 6E | E2 | 80 | 89 | ... | pas     | de  | raison...           |             |            |  |
| 00000091 | C2 | BB | 20 | 64 | 65 | 20 | 72 | 65 | 63 | 6F | 6E | 66 | 69 | 6E | 65 | 72 | 2C | 20 | 65 | 66 | 66 | 69 | 63 | 61 | 63 | 69 | 74 | C3 | A9 | ... | de      | reconfiner  | , efficacit..       |             |            |  |
| 000000AE | 20 | 64 | 75 | 20 | 76 | 61 | 63 | 63 | 69 | 6E | 2C | 20 | 69 | 6D | 6D | 75 | 6E | 69 | 74 | C3 | A9 | 20 | 63 | 6F | 6C | 6C | 65 | 63 | 74 | ... | du      | vaccin  | , immunit.. collect |             |            |  |
| 000000CB | 69 | 76 | 65 | E2 | 80 | A6 | 20 | 4C | 65 | 20 | 72 | C3 | A9 | 63 | 61 | 70 | E2 | 80 | 99 | 20 | 64 | 75 | 20 | 32 | 36 | 20 | 6A | 61 | 6E | ... | ive...  | Le r..cap...  | du                  | 26 jan      |            |  |
| 000000E8 | 76 | 69 | 65 | 72 | 20 | 20 | 68 | 74 | 74 | 70 | 73 | 3A | 2F | 2F | 77 | 77 | 77 | 2E | 6C | 61 | 2D | 63 | 72 | 6F | 69 | 78 | 2E | 63 | 6F | ... | vier    | <a href="https://www.la-croix.com/Sciences-et-ethique/Covid-19-pas-raison-reconfiner-efficacite-vaccin-immunite-collective-Le-recap-26-janvier-2021-01-26-1201137155">https://www.la-croix.com/Sciences-et-ethique/Covid-19-pas-raison-reconfiner-efficacite-vaccin-immunite-collective-Le-recap-26-janvier-2021-01-26-1201137155</a> | article             | de          | p          |  |
| 00000105 | 6D | 2F | 53 | 63 | 69 | 65 | 6E | 63 | 65 | 73 | 2D | 65 | 74 | 2D | 65 | 74 | 68 | 69 | 71 | 75 | 65 | 2F | 43 | 6F | 76 | 69 | 64 | 2D | 31 | ... | resse   | utf-8   | HTML                | .. Bay      |            |  |
| 00000122 | 39 | 2D | 70 | 61 | 73 | 2D | 72 | 61 | 69 | 73 | 6F | 6E | 2D | 72 | 65 | 63 | 6F | 6E | 66 | 69 | 6E | 65 | 72 | 2D | 65 | 66 | 66 | 69 | 63 | ... | ard     | Presse  | 2003                | 5           | 17.        |  |
| 0000013F | 61 | 63 | 69 | 74 | 65 | 2D | 76 | 61 | 63 | 63 | 69 | 6E | 2D | 69 | 6D | 6D | 75 | 6E | 69 | 74 | 65 | 2D | 63 | 6F | 6C | 6C | 65 | 63 | 74 | ... | .Covid  | chez  | l'enfant            | : un dan    |            |  |
| 0000015C | 69 | 76 | 65 | 2D | 4C | 65 | 2D | 72 | 65 | 63 | 61 | 70 | 2D | 32 | 36 | 2D | 6A | 61 | 6E | 76 | 69 | 65 | 72 | 2D | 32 | 30 | 32 | 31 | 2D | ... | ger     | plus  | grand               | que         | pr..vu ? D |  |
| 00000179 | 30 | 31 | 2D | 32 | 36 | 2D | 31 | 32 | 30 | 31 | 31 | 33 | 37 | 31 | 35 | 35 | 20 | 61 | 72 | 74 | 69 | 63 | 6C | 65 | 20 | 64 | 65 | 20 | 70 | ... | erniers | chiffres  | et                  | derni       | ro         |  |
| 00000196 | 72 | 65 | 73 | 73 | 65 | 20 | 20 | 75 | 74 | 66 | 2D | 38 | 20 | 20 | 20 | 48 | 54 | 4D | 4C | 20 | 20 | 20 | 20 | C2 | A9 | 20 | 42 | 61 | 79 | ... |         |   |                     |             |            |  |
| 000001B3 | 61 | 72 | 64 | 20 | 50 | 72 | 65 | 73 | 73 | 65 | 20 | 20 | 20 | 20 | 32 | 30 | 30 | 33 | 20 | 20 | 20 | 20 | 35 | 20 | 20 | 20 | 31 | 37 | 0D | ... |         |   |                     |             |            |  |
| 000001D0 | 0A | 43 | 6F | 76 | 69 | 64 | 20 | 63 | 68 | 65 | 7A | 20 | 6C | 27 | 65 | 6E | 66 | 61 | 6E | 74 | 20 | 3A | 20 | 75 | 6E | 20 | 64 | 61 | 6E | ... |         |   |                     |             |            |  |
| 000001ED | 67 | 65 | 72 | 20 | 70 | 6C | 75 | 73 | 20 | 67 | 72 | 61 | 6E | 64 | 20 | 71 | 75 | 65 | 20 | 70 | 72 | C3 | A9 | 76 | 75 | 20 | 3F | 20 | 44 | ... |         |   |                     |             |            |  |
| 0000020A | 65 | 72 | 6E | 60 | 65 | 72 | 73 | 20 | 63 | 68 | 60 | 66 | 66 | 72 | 65 | 73 | 20 | 65 | 74 | 20 | 64 | 65 | 72 | 6E | 60 | C3 | A8 | 72 | 65 | ... |         |   |                     |             |            |  |

Valeur décimale  
(= base 10)

|                  |              |                  |               |                |          |
|------------------|--------------|------------------|---------------|----------------|----------|
| Signed 8 bit:    | 32           | Signed 32 bit:   | 1428168736    | Hexadecimal:   | 20       |
| Unsigned 8 bit:  | 32           | Unsigned 32 bit: | 1428168736    | Octal:         | 040      |
| Signed 16 bit:   | 8224         | Signed 64 bit:   | 1428168736    | Binary:        | 00100000 |
| Unsigned 16 bit: | 8224         | Unsigned 64 bit: | 1428168736    | Stream Length: | 8        |
| Float 32 bit:    | 1,100374e+13 | Float 64 bit:    | 5,673161e+188 |                |          |

Show little endian decoding  Show unsigned and float as hexadecimal

Offset: 0x8

# URLs

Généralement, on peut enlever tout ce qui se trouve après le caractère « ? ».

- **<https://pdf.sciencedirectassets.com/272642/1-s2.0-S0222961720X00110/1-s2.0-S0222961720302038/main.pdf?X-Amz-Security-Token=IQoJb3JpZ2luX2VjEEoaCXVzLWVhc3QtMSJGMEQCIDQmE%2BWzTvNOBZfBkh9uq5CNYIVSWzE0mx3ORUdLmLH5AiAUANbFOT%2BZG9AOi%2FCNSDJO6HUfC536gJqHHNwaovMg3yq0AwgzEAMaDDA1OTAwMzU0Njg2NSIM2oRLxQcg%2Bn7zEYx8KpEDAj1C7VI9ma0JdGgJGk8CfFQ9Tm0J7y14nyNpigf3x7r7TCwLLcwrwPJhSDKIOI65w2BRqNnY6FGYd>**

# Licences



Neuropsychiatrie de l'Enfance et de l'Adolescence

Volume 69, Issue 1, January 2021, Pages 1-3



Compte-rendu  
Actes de journées

## COVID 19 : Comment la pédopsychiatrie s'adapte en ces temps d'adversité ?

## COVID 19: How does child psychiatry adapt in these times of adversity? ☆

N. Catheline <sup>a</sup>, E. Berthaut <sup>b</sup>, M. Bon-Saint-Côme <sup>c</sup>, E. Carretier <sup>d</sup>, J.M. Coq <sup>e</sup>, V. Delvenne <sup>f</sup>, E. Deroin-Thévenin <sup>g</sup>, L. Gicquel <sup>h</sup>, F. Labarraque <sup>i</sup>, C. Potel <sup>j</sup>, J.P. Raynaud <sup>k</sup>

Show more ▾

+ Add to Mendeley Share Cite

<https://doi.org/10.1016/j.neurenf.2020.10.005>

Get rights and content

Previous article

Next article

\* Compte rendu de la Journée SFPEADA du 18 septembre.

View full text

## Maladie à coronavirus 2019

Cet article concerne cette [maladie virale](#) chez l'humain. Pour la [pandémie](#) correspondante, voir [Pandémie de Covid-19](#). Pour l'agent pathogène, voir [SARS-CoV-2](#). Pour l'effet de cet agent chez les autres animaux, voir [SARS-CoV-2 chez les animaux non humains](#).

**Wikipédia ne donne pas de conseils médicaux ou sanitaires.** Cet article est susceptible de contenir des informations obsolètes ou inexactes. Seul un professionnel de santé est apte à vous fournir un avis **⚠** médical, et seules les autorités sanitaires de votre pays sont compétentes pour donner des consignes de santé publique relatives à la pandémie de Covid-19 : [France \[archive\]](#), [Canada \[archive\]](#), [Belgique \[archive\]](#), [Maroc \[archive\]](#), [Algérie \[archive\]](#), [Suisse \[archive\]](#), [Tunisie \[archive\]](#), [Rép. dém. du Congo \[archive\]](#).



### Le sujet de cet article est lié à un événement récent ou en cours.

Les informations peuvent manquer de recul, ne pas prendre en compte des développements récents ou changer à mesure que l'événement progresse. N'hésitez pas à l'améliorer en veillant à **citer vos sources**.

La dernière modification de cette page a été faite le 28 janvier 2021 à 14:37.

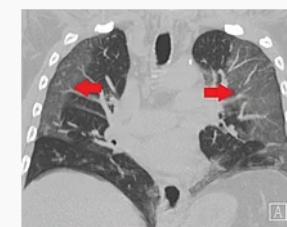
La **maladie à coronavirus 2019** <sup>🔊</sup> [Écouter](#), ou la ou le **Covid-19**

<sup>🔊</sup> [Écouter](#) (acronyme anglais de *coronavirus disease 2019*), est une maladie infectieuse émergente de type zoonose virale causée par la souche de coronavirus SARS-CoV-2. Les symptômes les plus fréquents sont la fièvre, la toux, la fatigue et la gêne respiratoire. Dans les formes les plus graves, l'apparition d'un syndrome de détresse respiratoire aiguë peut entraîner la mort, notamment chez les personnes plus fragiles du fait de leur âge ou en cas de comorbidités. Une autre complication mortelle est une réponse exacerbée du système immunitaire inné (choc cytokinique).

Une perte brutale de l'odorat (anosmie), associée ou non à une perte du goût (agueusie), est une manifestation relativement fréquente et parfois révélatrice de l'infection par le SARS-CoV-2.

Le taux de formes asymptomatiques est estimé à près de 20 % des personnes infectées. La transmission interhumaine se fait surtout par gouttelettes respiratoires et aérosolisation<sup>2,3</sup>, lorsqu'une personne respire dans un même lieu clos ou à proximité immédiate d'autres personnes. La transmission augmente en intérieur mal ventilé et lorsque la personne infectée tousse, éternue, parle ou chante. La transmission indirecte par surface contaminée, bien que possible, n'a

### Maladie à coronavirus 2019 (Covid-19)



Pneumonie aiguë bilatérale, les flèches indiquent un infiltrat en verre dépoli dans les deux poumons.

**Spécialité** Infectiologie, pneumologie, virologie, épidémiologie et médecine d'urgence

**CIM-10** U07.1 et U07.2

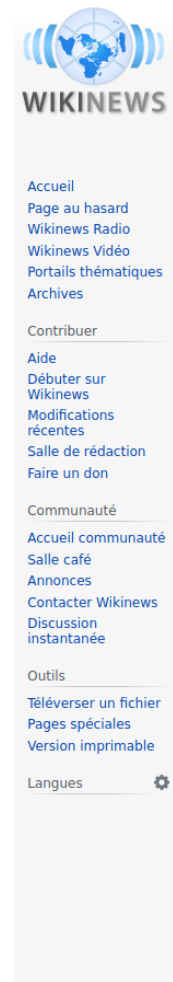
**DiseasesDB** 60833

**eMedicine** 2500114

# Licences



The screenshot shows the Le Monde website interface. At the top, there's a navigation bar with categories like ACTUALITÉS, ÉCONOMIE, VIDÉOS, OPINIONS, CULTURE, and M LE M. The main content area is titled "PLANÈTE" and features a large article header "CORONAVIRUS ET PANDÉMIE DE COVID-19". Below the header, there's a sub-header "Tous nos articles sur l'épidémie due au coronavirus SARS-CoV-2." and three buttons: "L'évolution de l'épidémie en France et dans le monde", "Pandémie", and "Journal de crise des blouses blanches". There are two image thumbnails: one showing people in a laboratory setting and another showing a person in a hospital bed. A sidebar on the right contains a "Les pl" section with a large number "1" and the text "Oliv la si mes « T'e".



The screenshot shows the Wikinews sidebar. It features the Wikinews logo at the top, followed by a list of navigation links: Accueil, Page au hasard, Wikinews Radio, Wikinews Vidéo, Portails thématiques, Archives, Contribuer, Aide, Débuter sur Wikinews, Modifications récentes, Salle de rédaction, Faire un don, Communauté, Accueil communauté, Salle café, Annonces, Contacter Wikinews, Discussion instantanée, Outils, Téléverser un fichier, Pages spéciales, Version imprimable, and Langues.



The screenshot shows the Wikinews search results page. At the top, there's a search bar with the query "Coronavirus : le nombre de cas par pays" and a "Rechercher" button. Below the search bar, there's a "Recherche avancée" section with options for "Trier par pertinence" and "Rechercher dans". The main content area displays the search results for the query "Coronavirus : le nombre de cas par pays". It includes a summary of the article, the number of words (4 Kio), and the date of the last update (15 septembre 2020 à 18:12). There are also links to related articles, such as "Syrie : la pandémie de Covid-19 est « hors de contrôle », selon une ONG" and "Coronavirus : plus d'un million de cas recensés dans le monde".

# Variables

- Thème Covid-19
  - ~~Nombre d'occurrences de xxx~~
  - Date (mois)
  - Situation nationale / internationale
    - ou pays, continent, etc.
  - Épidémiologie (texte sur le nb de cas, les conséquences sociales, etc.) / médecine (effets de la maladie)
  - Etc.



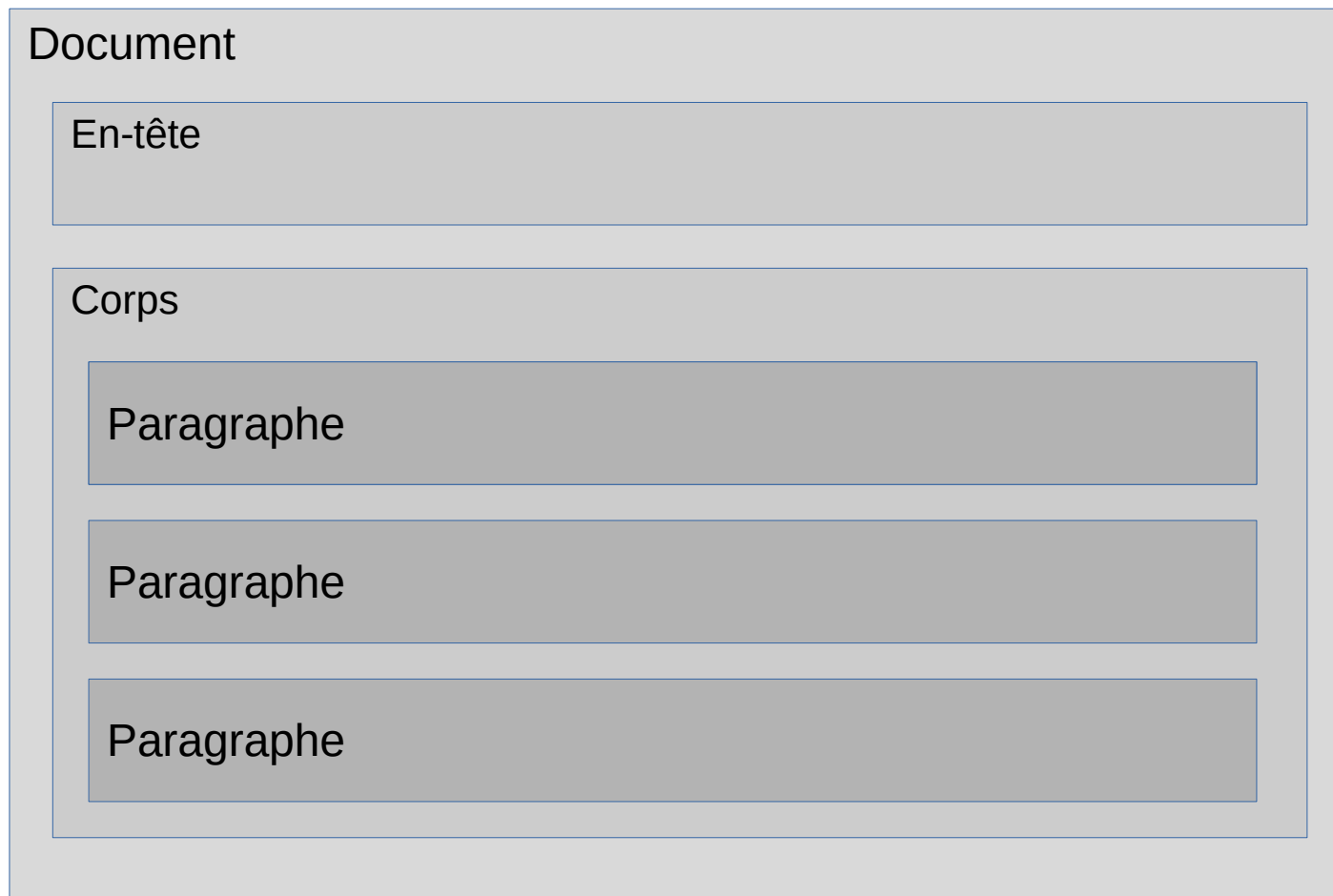
# Couverture d'un corpus

- Couverture de chaque intersection variable 1 / variable 2
  - Pas d'espace vide !
- Équilibrage du nombre de mots pour chaque intersection
- Ex. pour des variables *Date (mois) / Genre*

|   | A       | B       | C       | D       | E     | F   | G    | H       | I    | J         | K       | L        | M        |
|---|---------|---------|---------|---------|-------|-----|------|---------|------|-----------|---------|----------|----------|
| 1 |         | Janvier | Février | Mars    | Avril | Mai | Juin | Juillet | Août | Septembre | Octobre | Novembre | Décembre |
| 2 | Presse  |         | 5000    |         |       |     |      | 2000    |      |           |         |          |          |
| 3 | Encyclo |         |         |         | 6000  |     |      |         |      |           |         |          |          |
| 4 | Science |         | 3000    |         |       |     | 1000 |         |      |           |         |          |          |
| 5 | Blog    |         |         | 1000000 |       |     |      |         |      | 10000     |         |          |          |

# Principe de XML

- Encapsulation



# Principe de XML

- Encapsulation

```
<TEI>
```

```
<teiHeader>  
...  
</teiHeader>
```

```
<body>
```

```
<p> ... </p>
```

```
<p> ... </p>
```

```
<p> ... </p>
```

```
</body>
```

```
</TEI>
```

# Principe de XML

- Encapsulation (exemple de XML-TEI)

```
<TEI>
  <teiHeader>
    ...
  </teiHeader>
  <text>
    <p> ... </p>
    <p> ... </p>
    <p> ... </p>
  </text>
</TEI>
```

*balise ouvrante*

*balise fermante*

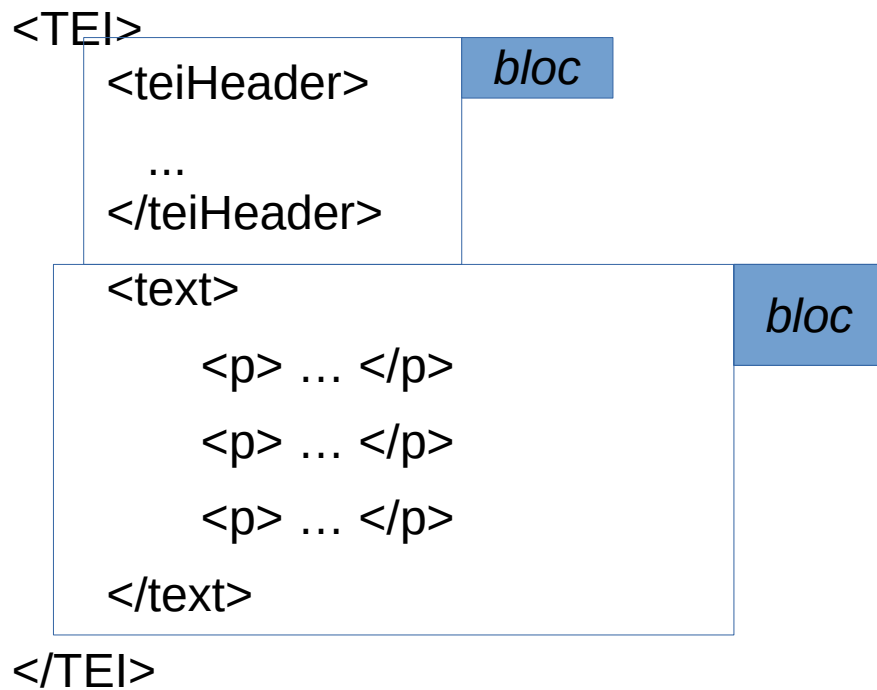
# Principe de XML

- Encapsulation (exemple de XML-TEI)

```
<TEI>  
  <teiHeader> bloc  
    ...  
  </teiHeader>  
  <text>  
    <p> ... </p>  
    <p> ... </p>  
    <p> ... </p>  
  </text>  
</TEI>
```

# Principe de XML

- Encapsulation (exemple de XML-TEI)



# Principe de XML

- Encapsulation (exemple de XML-TEI)

```
<TEI>
  <teiHeader>
    ...
  </teiHeader>
  <text>
    <p> ... </p>
    <p> ... </p>
    <p> ... </p>
  </text>
</TEI>
```

The diagram shows a hierarchical structure of XML elements. The root element is <TEI>. Inside <TEI>, there is a <teiHeader> element followed by an ellipsis and then </teiHeader>. Below this is a <text> element. The <text> element is enclosed in a large blue box labeled 'bloc'. Inside this <text> element, there are three <p> elements, each followed by an ellipsis and </p>. Each <p> element is enclosed in a smaller blue box labeled 'bloc'. The <text> element is closed with </text>. Finally, the <TEI> element is closed with </TEI>.

# Principe de XML

- Encapsulation (exemple de XML-TEI)

```
<TEI>
  <teiHeader>
    ...
  </teiHeader>
  <text xml:lang="fr"> attribut = valeur
    <p> ... </p>
    <p> ... </p>
    <p> ... </p>
  </text>
</TEI>
```



# Principe de XML

- Encapsulation (exemple de XML-TEI)

```
<TEI>
  <teiHeader>
    ...
  </teiHeader>
  <text xml:lang="fr">
    <p> ... <lb/> ... </p>
    <p> ... <lb/> ... </p>
    <p> ... </p>
  </text>
</TEI>
```

*balise auto-fermante*

# Principe de XML

- Encapsulation (exemple de XML-TEI)

```
<TEI>
  <teiHeader>
    ...
  </teiHeader>
  <text xml:lang="fr">
    <p> ... <lb/> ... </p>
    <p> ... <lb/> ... </p>
    <!-- Un commentaire !!! -->
    <p> ... </p>
  </text>
</TEI>
```

*Comme son contenu l'indique...*

# Exemple concret

- Encapsulation (exemple de XML-TEI)

```
<?xml version="1.0" encoding="UTF-8"?>
```

*Métadonnées*

```
<TEI>  
  <teiHeader>  
  
  ...  
  </teiHeader>  
  <text xml:lang="fr">  
    <p> ... <lb/> ... </p>  
    <p> ... <lb/> ... </p>  
    <p> ... </p>  
  </text>  
</TEI>
```

# Exemple concret

<https://pro.aiakide.net/cours/Corpus2021a/exempleCorpus14.xml>

```
exempleCorpus14.xml x
1 <?xml version="1.0" encoding="UTF-8"?><?xml-stylesheet type="text/xsl" href="custom.xsl"?><TEI xmlns="http://www.tei-c.org/ns/1.0">
2 <!-- Projet Corpus 14, Praxiling - représentation d'une correspondance de poilu suivant le modèle:
3 <teiHeader> ... <sourceDoc> ... <body> <opener> <p> ... <closer> </body>
4 -->
5 <teiHeader xml:lang="fr">
6 <fileDesc>
7 <titleStmt>
8 <title>Claudia Tarest 31-12-1915</title>
9 <author>Claudia Tarest</author>
10 <editor>Giancarlo Luxardo</editor>
11 </titleStmt>
12 <editionStmt>
13 <edition>Version 2.0<date>février 2015</date>
14 </edition>
15 <respStmt>
16 <resp>Transcrit par </resp>
17 <name>Manon PENGAM</name>
18 </respStmt>
19 </editionStmt>
20 <publicationStmt>
21 <publisher>
22 <name>Praxiling, UMR 5267</name>
23 </publisher>
24 <date when="2018-07-27">27-07-2018</date>
25 <pubPlace>
26 <address>
27 <addrLine>Université Paul-Valéry Montpellier</addrLine>
28 <addrLine>34199 Montpellier Cedex 5</addrLine>
29 </address>
30 </pubPlace>
31 <idno type="tei">http://textometrie.univ-montp3.fr/txm/xml/claudia-1-151231.xml</idno>
32 <availability status="restricted">
33 <p>Version 2.0</p>
34 <licence target="http://creativecommons.org/licenses/by-nc-sa/3.0/fr">
35 <p>
36 <hi>Conditions d'utilisation de ce fichier</hi> : Ce fichier numérique est mis à disposition selon
37 les termes de la licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Partage à
38 l'Identique 3.0 France </p>
39 </licence>
40 </availability>
41 </publicationStmt>
```

# Exemple concret

<https://pro.aiakide.net/cours/Corpus2021a/exempleCorpus14.xml>

```
<text xml:lang="fr" facs="#">
  <!-- L'élément <opener> (optionnel) regroupe une ligne de date (et lieu) <dateline> et une formule de politesse <salute> -->
  <body>
    <opener>Condrieu Le 31 decembre 1915<lb/> bien cher Oncle<lb/>
    </opener>
    <p>Bien éloigner de vous avec ma carte je viens<lb/> auprès de vous pour vous souhaité une bonheur
    heureuse<lb/> une parfaite santé et que cette nouvelle année<lb/> de 1916 qui vous rapproche de nous
    bientôt<lb/> qu'elle soit pour vous plus favorable que l'année 1915<lb/>
    </p>
    <p>cher Oncle<lb/>
    </p>
    <p>Nous somme tous en bonne santé nous esperons<lb/> que vous en éte de mêmes<lb/>
    </p>
    <closer>Votre Niece bien affectueuse<lb/> Claudia <unclear>Tarest ?</unclear>
    <lb/>
    </closer>
  </body>
```

# Pratique

- Prenez 2 documents différents de votre corpus (pas issus du même site !)
- Formatez-les en XML avec votre éditeur de texte préféré (*pas* Word ou Writer – dans le doute utilisez *SublimeText*), en suivant le schéma page suivante

# Pratique

Le cas échéant, utilisez aussi les blocs suivants :

- `<title> ... </title>` (*titres*)
- `<q> ... </q>` (*citations*)

Voir aussi : [https://tei-c.org/release/doc/tei-p5-exemplars/html/tei\\_lite.doc.html](https://tei-c.org/release/doc/tei-p5-exemplars/html/tei_lite.doc.html)

```
<TEI>
  <teiHeader>
    <titleStmt>
      <title> Titre du document </title>
      <author> Auteur du document </author>
      <editor> Éditeur du document (c'est vous !) </editor>
    </titleStmt>
    <publicationStmt>
      <availability>
        <licence> Licence du document </licence>
      </availability>
    </publicationStmt>
  </teiHeader>
  <text>
    <body>
      <p>...</p>
      <p>...</p>
      <p>...</p>
    </body>
  </text>
</TEI>
```

Validez votre document XML sur :

- <https://www.xmlvalidation.com>



# Prochaines séances

- Mise en forme de votre corpus au format XML-TEI
- Chemins XPath
- Récupération automatique de corpus depuis Internet en utilisant XML