

Université de Neuchâtel
Master en littérature

Édition critique et génétique numérique

Élodie Paupe, chargée d'enseignement
elodie.paupe@unine.ch

Editions critiques et génétiques numérique

Non sécurisé — vangoghletters.org

001 (001, 1): To Theo van Gogh. The Hague, Sunday, 29 September 1872. - Vincent van Gogh Letters

Van Gogh Museum huysens ing

Vincent van Gogh
The Letters

by period
by correspondent
by place
with sketches

Search
keyword or number(s) >>
Advanced search
Search results

Van Gogh as a letter-writer
Correspondents
Biographical & historical context
Publication history

About this edition
Chronology
Concordance, lists, bibliography
Book edition

001 001a >>

To Theo van Gogh. The Hague, Sunday, 29 September 1872.

SEARCH THIS LETTER
PRINT

original text + line endings facsimile translation notes artworks

001
Br. 1990: 001 | CL: 1
From: Vincent van Gogh
To: Theo van Gogh
Date: The Hague, Sunday, 29 September 1872
more...

original text + line endings facsimile translation notes artworks

1' [Den Haag, 29 september 1872].

[Waarde] Theo,
Dank voor je brief, het deed mij genoegen dat je weer goed
aangekomen zijt. Ik heb je de eerste dagen gemist & het was
mij vreemd je niet te vinden als ik s'middags t' huis kwam.
Wij hebben prettige dagen samen gehad, en tusschen de
droppeltjes door ² toch nog al eens gewandeld & het een
en ander gezien.
Wat vreesselijk weer, je zult het wel *benauwd* hebben op je
wandelingen naar Ois[ter]wijk. ¹ Gisteren is het
harddraverij geweest ter gelegenheid van de
tentoonstelling, ² maar de illuminatie & het vuurwerk zijn
uit gesteld, om het slechte weer. ³ het is dus maar goed dat
je niet gebleven zijt om die te zien. Groeten van de familie
Haanebeek ⁴ & Roos. ⁵ Steeds

je liefh.
Vincent

top

original text + line endings facsimile translation notes artworks

1' [Den Haag, 29 september 1872].

2' [Waarde] Theo,
Dank voor je brief,
het deed mij genoegen dat je
weer goed aangekomen zijt.
Ik heb je de eerste dagen ge-
mist & het was mij vreemd
je niet te vinden als ik s' mid-
dags t' huis kwam.
Wij hebben prettige dagen sa-
men gehad, en tusschen de
droppeltjes door ² toch nog al
eens gewandeld & het een en
ander gezien.
Wat vreesselijk weer, je zult
het wel *benauwd* hebben
op je wandelingen naar
Ois[ter]wijk. ¹ Gisteren is het hard-
draverij geweest ter gelegenheid van
de tentoonstelling, ² maar de illumi-
natie & het vuurwerk zijn uit
gesteld, om het slechte weer, ³ het
is dus maar goed dat je niet
gebleven zijt om die te zien. Groeten
van de familie Haanebeek ⁴ & Roos. ⁵
Steeds

je liefh.
Vincent

1-2 Text lost due to damaged edge of the paper. Part of the top of the letter has been torn off.
1 [Den Haag, 29 september 1872] < Only the final digit, the 2, is legible, because the top of the sheet has been torn off. We have added the place and date on the basis of Van Gogh's usual way of heading his letters in this period.

<http://vangoghletters.org/vg/letters/let001/letter.html>

\wedge *extensible*

M *markup*

L *Language*

Un (méta)langage de balisage:

- Pas un langage de programmation: un document XML n'exécute aucune fonction, il n'y a pas d'algorithme.
- Un langage qui sert à ajouter des métadonnées sur un contenu qui n'est pas mis en forme = un langage informatique de représentation des données.
- Un langage ouvert (= extensible) qui peut s'adapter aux besoins de l'utilisateur.
- 1996: premiers travaux
- 1998: XML 1.0
- Objectif: permettre l'interopérabilité

Concept de base

Transformer un texte brut (= une chaîne de caractères) en base de données à l'aide d'un langage informatique (= XML) qui isole des portions de texte (= sous-chaînes de caractères) en fonction des informations que l'éditeur souhaite mettre en évidence.

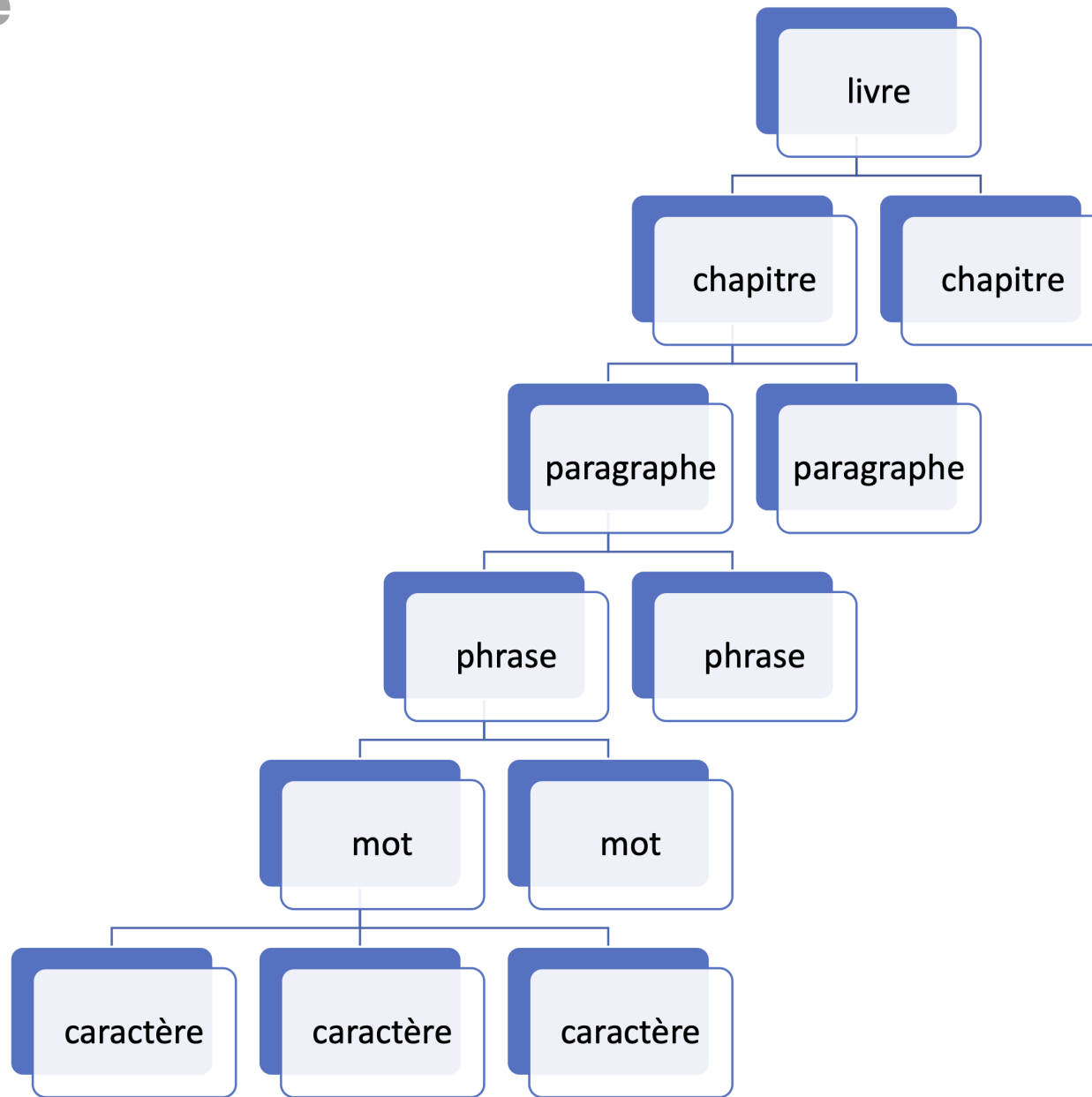
Structuration d'un texte brut

- Autrement dit, le XML va apporter de la structure à un texte brut qui est considéré comme des données non mises en forme.
- Il faut penser le texte comme une architecture:
 - livre
 - chapitres
 - paragraphes
 - phrases
 - mots
 - caractères



Le document XML comme des poupées russes

```
<livre>
  <chapitre>
    <paragraphe>
      <phrase>
        <mots>
          <caractère>
            A
          </caractère>
          <caractère>
            V
          </caractère>
          <caractère>
            E
          </caractère>
        </mot>
        <mot>popule</mot>
      </phrase>
      <phrase> </phrase>
    </paragraphe>
    <paragraphe> </paragraphe>
  </chapitre>
  <chapitre> </chapitre>
</livre>
```

Fonctionnement de base

Ce langage de balisage fonctionne de manière simple

```
<élément attribut="valeur">données</élément>
```

1. Les données sont la chaîne de caractères (= le texte brut).
2. On appelle "balise" les `<élément>` qui sont toujours écrits entre chevrons.
3. Une `<balise>` s'ouvre et doit impérativement être fermée `</balise>`
4. Des balises ne peuvent pas se chevaucher, c'est le principe des poupées russes:
 - Correct: `<baliseA>` `<baliseB>` données `</baliseB>` `</baliseA>`
 - Fautif: `<baliseA>` `<baliseB>` données `</baliseA>` `</baliseB>`
5. Une `<balise/>` peut être auto-fermante: elle ne contient donc pas de données.
6. Un `<élément>` peut porter un `@attribut` (noté avec un `@`)
7. L' `@attribut` a une "valeur" (entre guillemets)
 - `<élément @="valeur de l'attribut">`

Des balises pour donner du sens

En résumé:

De facto, l'édition du *Lai de l'Ombre* réalisée par Joseph Bédier en 1913 est considérée comme le point de départ de la *méthode* bédierisme.

Les italiques employés ci-dessus sont-ils identiques?

```
<document>
  <phrase>
    En résumé:
    <locutionétrangère>de facto</locutionétrangère>,
    l'édition du <titre>Lai de l'Ombre</titre> réalisée par Joseph Bédier en 1913
    est considérée comme le point de départ de la <emphase>méthode</emphase> bédierisme.
  </phrase>
</document>
```

Les retours à la ligne ne sont pas signifiant:

```
<document>
  <phrase>
    En résumé:
    <locutionétrangère>de facto</locutionétrangère>,
    l'édition du <titre>Lai de l'Ombre</titre> réalisée par Joseph Bédier en 1913
    est considérée comme le point de départ de la <emphase>méthode</emphase> bédierisme.
  </phrase>
</document>
```

```
<document>
  <phrase>
    En résumé:
    <locutionétrangère>de facto</locutionétrangère>, l'édition du <titre>Lai de l'Ombre</titre> réalisée par Joseph Bédier en 1913 est considérée comme le point de départ de la <emphase>méthode</emphase> bédierisme.
  </phrase>
</document>
```

Exercice(s) 1(-2) [ici](#)

XML et normalisation

```
<document>
  <phrase>
    <locutionétrangère>De facto</locutionétrangère>,
    l'édition du <titre>Lai de l'Ombre</titre> réalisée par Joseph Bédier en 1913
    est considérée comme le point de départ de la <emphase>méthode</emphase> bédieriste.
  </phrase>
</document>
```

- Si la personne qui encode l'extrait est germanophone? anglophone? italophone?...

Le langage XML ne propose pas un jeu prédéfini et fermé de balises, mais une syntaxe et des règles qui permettent de "bien former" un document pour qu'il soit compréhensible par un ordinateur.

Un document XML [...] est dit bien formé s'il respecte la syntaxe [XML], avec balises ouvrantes et fermantes présentes et correctement imbriquées.

– Burnard, Lou. "La TEI et le XML" in: *Qu'est-ce que la Text Encoding Initiative?* [en ligne]. Marseille : OpenEdition Press, 2015, §6, Disponible sur Internet : <http://books.openedition.org/oep/1298>.

- Comment garantir l'interopérabilité et le partage des données?
- ... vers une normalisation.

Exercice 3: [ici](#)

XML-TEI

E = *Encoding*

I = *Initiative*

- Version standardisée du XML pour représenter un texte sous une forme numérique:

The Text Encoding Initiative (TEI) is a consortium which collectively develops and maintains a standard for the representation of texts in digital form. Its chief deliverable is a set of Guidelines which specify encoding methods for machine-readable texts, chiefly in the humanities, social sciences and linguistics.

- Recommandations émises par un consortium:

The TEI Consortium is a nonprofit membership organization composed of academic institutions, research projects, and individual scholars from around the world.

Source: <https://tei-c.org>

XML et TEI

- La TEI est donc une normalisation du langage XML utilisée par une communauté.
- Cette normalisation implique d'utiliser les balises définies par les guidelines établies et continuellement mises à jours.
- La TEI est en anglais: un *titre* sera mis en évidence par la balise `title` (`<title>le titre</title>`)
- Comme on l'a déjà vu, la TEI encode le sens d'une chaîne de caractère et pas de sa mise en forme: on parle d'**encode sémantique**.
- La TEI n'est pas le seul vocabulaire tiré du XML, il en existe d'autres, notamment
 - l'EAD (*Encoded Archival Description*) pour les archivistes
 - le DC (*Dublin Core*) pour les bibliothécaires

Encodage sémantique

De facto, l'édition du *Lai de l'Ombre* réalisée par Joseph Bédier en 1913 est considérée comme le point de départ de la "méthode" bédieriste.

Encodage sémantique XML:

```
<locutionétrangère>De facto</locutionétrangère>,  
l'édition du <titre>Lai de l'Ombre</titre> réalisée par Joseph Bédier en 1913  
est considérée comme le point de départ de la méthode bédieriste.
```

Autre proposition d'encodage sémantique XML:

```
<latin>De facto</latin>,  
l'édition du <titre>Lai de l'Ombre</titre> réalisée par Joseph Bédier en 1913  
est considérée comme le point de départ de la méthode bédieriste.
```

Encodage sémantique TEI:

```
<foreign xml:lang="la">De facto</foreign>,  
l'édition du <titre>Lai de l'Ombre</titre> réalisée par Joseph Bédier en 1913  
est considérée comme le point de départ de la méthode bédieriste.
```

Quelques concepts méthodologiques

Modélisation

Modélisation, subst. fém.: Opération par laquelle on établit le modèle d'un système complexe, afin d'étudier plus commodément et de mesurer les effets sur ce système des variations de tel ou tel de ses éléments composants.

– TLFi, <https://www.cnrtl.fr/definition/modélisation>

Il s'agit de définir un modèle adapté :

1. aux documents que l'on édite
2. à nos questions de recherche
3. aux moyens (techniques, financiers...) dont on dispose

Modélisation et démarche philologique

- Quel type d'édition je souhaite réaliser?
 - une édition diplomatique nécessitera que la matérialité du document et sa structure soit prise en compte (retour à la ligne, page, notes marginales, etc.)
- Qu'est-ce qui m'intéresse dans le texte et dont je dois rendre le sémantisme?
 - une édition critique mettra en évidence les variantes, les corrections, etc.
 - la constitution d'index sera facilitée si les données associées ont été encodées. Dans ce cas, il sera utile d'identifier les personnes, les lieux, etc.
 - ai-je besoin d'encoder des phrases? des vers? des mots? des figures de style? etc.
- Quel type d'intertextualité est-ce que je souhaite dans mon édition électronique?

n résumé:

De *facto*, l'édition du *Lai de l'Ombre* réalisée par Joseph Bédier en 1913 est considérée comme le point de départ de la *méthode* bédierisme.

```

<document>
  <phrase>
    En résumé:
    <locutionétrangère>de facto</locutionétrangère>,
    l'édition du <titre>Lai de l'Ombre</titre> réalisée par Joseph Bédier en 1913
    est considérée comme le point de départ de la <emphase>méthode</emphase> bédieriste.
  </phrase>
</document>

```

OU

```

<document>
  <pb n="1"/>
    En résumé:
  <pb n="2">
    <locutionétrangère>de facto</locutionétrangère>,
    l'édition du <titre>Lai de l'Ombre</titre> réalisée par Joseph Bédier en 1913
    est considérée comme le point de départ de la <emphase>méthode</emphase> bédieriste.
  </phrase>
</document>

```

Granularité

Lors de la modélisation, on définit la **granularité** de son encodage, autrement dit la plus petite unité qui va être encodée.

- Un granularité faible (= de gros grains, donc de gros morceaux de document) implique un balisage moins lourd.
- Une granularité élevée (= de petits grains, donc de petits morceaux de documents) implique une multiplication de balise.

Plus la granularité augmente, plus il devient difficile de lire le document.

```
<document>
  <paragraphe>
    En résumé:
    de facto,
    l'édition du Lai de l'Ombre réalisée par Joseph Bédier en 1913
    est considérée comme le point de départ de la méthode bédieriste.
  </phrase>
</paragraphe>
</document>
```

```
<document>
  <paragraphe>
    <phrase>
      En résumé:
      <locutionétrangère>de facto</locutionétrangère>,
      l'édition du <titre>Lai de l'Ombre</titre> réalisée par Joseph Bédier en 1913
      est considérée comme le point de départ de la <emphase>méthode</emphase> bédieriste.
    </phrase>
  </paragraphe>
</document>
```

```
<document>
  <paragraphe>
    <phrase>
      <w lemma="en" pos="PRP">En</w>
      <w lemma="résumé" pos="NOM">résumé</w>
      <pc unit=">:</pc>
      ...
```

Exercices

Exercice 4: [ici](#)

Sources

Burnard, Lou, "La TEI et le XML" in: *Qu'est-ce que la Text Encoding Initiative?* [en ligne], Marseille: OpenEdition Press, 2015. Disponible sur Internet : <http://books.openedition.org/oep/1298>. DOI : <https://doi.org/10.4000/books.oep.1237>.

Gabay, Simon (éd.), *Encoder, Analyser - Introduction à la philologie numérique*, Genève: Université de Genève, 2020, https://github.com/gabays/Cours_Edition_Geneve.

Verlaine, Paul, *Poèmes saturniens*, Paris: Alph. Lemerre, 1867. Disponible sur Internet: <https://gallica.bnf.fr/ark:/12148/bpt6k71276f>