

Approches Sociales et Historiques des Littératures des Suds

Séance du 05 février 2021

**Faire émerger le réseau littéraire
francophone haïtien à l'ère du numérique**

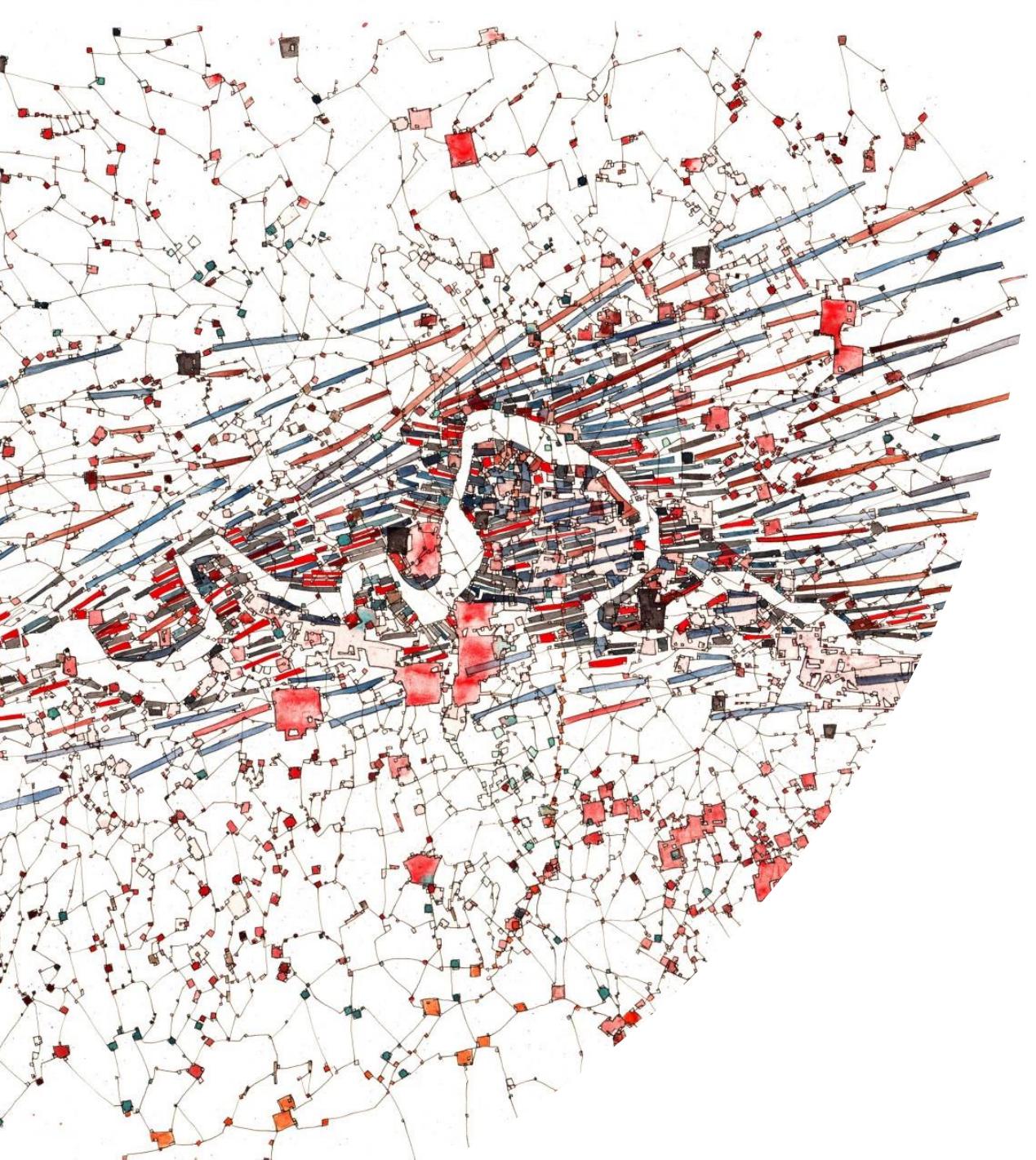


Anaïs CHAMBAT - ENS de Lyon
Célian RINGWALD - Université Lumière Lyon 2

Approches Sociales et Historiques des Littératures des Suds

Faire émerger le réseau littéraire francophone haïtien à l'ère du numérique

- I. Introduction
- II. Croiser Web sémantique et contenus participatifs
- III. Normalisation du corpus et structuration en XML-TEI
- IV. Exploitation des données
- V. Conclusions



I. Introduction



Des profils complémentaires fédérés autour d'un double cursus en Humanités Numériques

- **Inès Burri** : Infographiste 3D de formation, elle est titulaire d'une licence de lettres et d'un master en histoire ancienne effectués au sein de l'Université Jean Moulin Lyon 3. Elle travaille actuellement au sein du GIRPAM (Groupement International de Recherche sur la Poésie Antique et Médiévale).
- **Anaïs Chambat** : Diplômée d'un master en histoire des mondes germaniques, elle a étudié à l'ENS de Lyon les applications du numérique et des méthodes quantitatives à l'histoire. Elle est actuellement gestionnaire de données de recherche au sein du service d'histoire de la médecine de la Bibliothèque InterUniversitaire de Santé.
- **Célian Ringwald** : *Datascientist* et spécialiste en traitement automatique du langage. Il est aujourd'hui éditeur bénévole pour la revue *Programming Historian* portant sur les humanités computationnelles et la méthodologie de l'histoire numérique.

Qu'est-ce que le Web littéraire ?

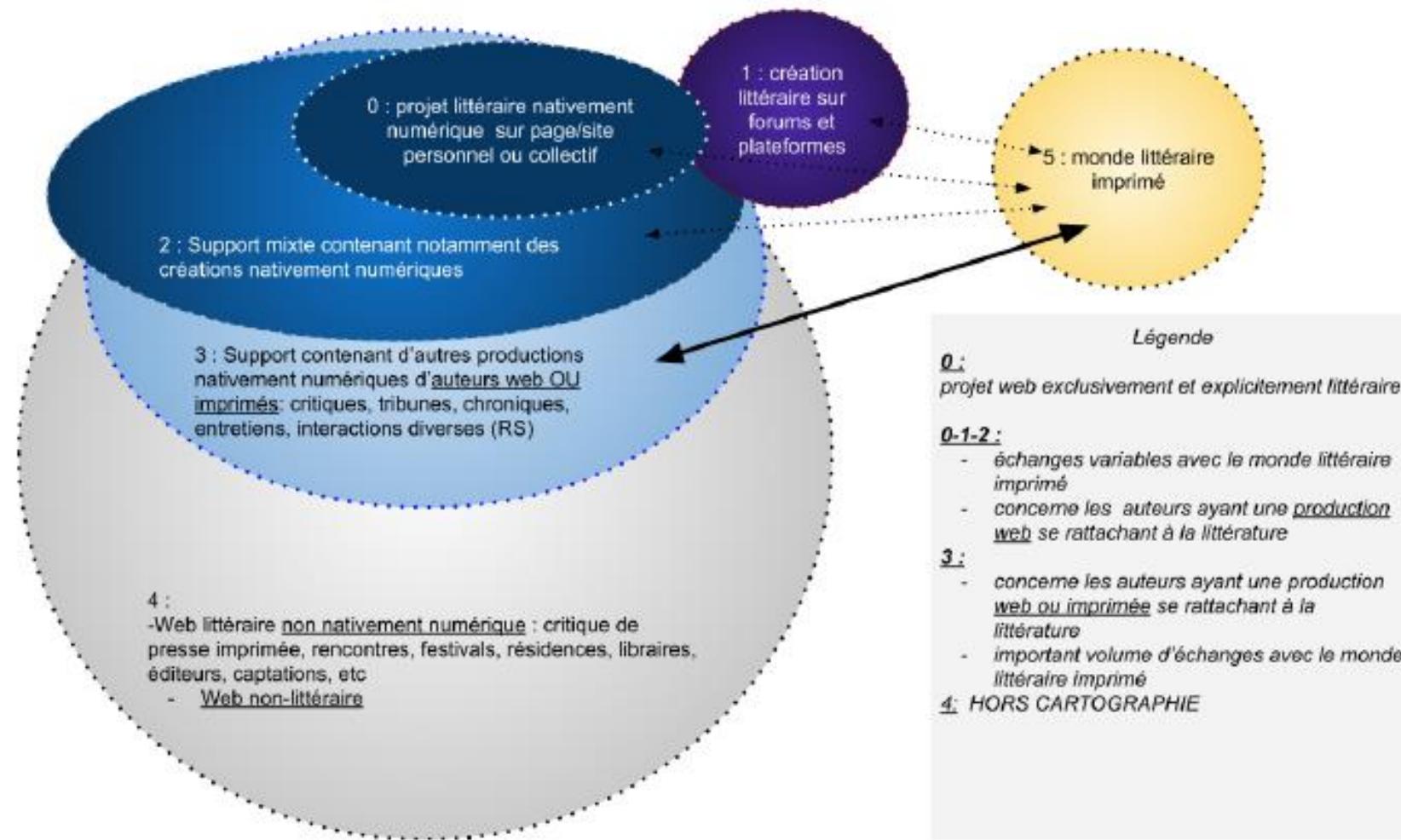
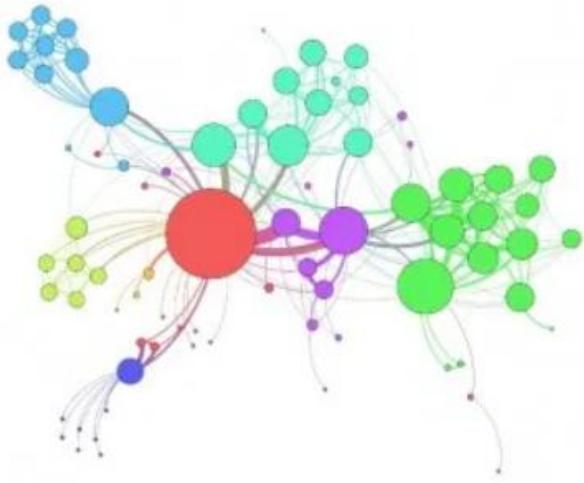


Schéma élaboré par l'équipe MARGE (Université Jean Moulin, Lyon 3) dans le cadre de l'IDEX CartoWeb.

Cartographier le Web et le visualiser...

Réseaux d'acteurs ?



Des cartes ...

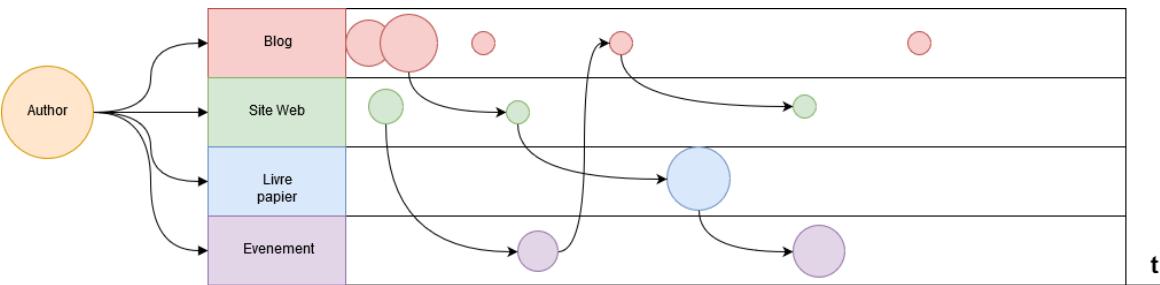


Contenus, producteurs et autres en milieu Internet

Lecture distante de contenus ?

plume toutes femmes cœur soleil
lieu de corps rebonds sur le plancher
instruments de lune ciel à pollution
doux retard frissons
cloches d'une cathédrale manière cadence de queue
rêve heures de falaise sofa
oubli couleurs poème belle aventure
châtiments le ramassage ciel grand harmonie silences déverrouillés
vie femmes vêtues roses absence sens
nuit cauchemar audie à la requête paraphrase inédite
femmes fièvre à refaire pouls
cauchemar fenêtres de prose
présence
partage en pâturage faux azur bleu caricature d'une voix
oui-dire d'une femme terre les femmes coeurs forme indéterminée
secret regard sans retour soulèvement de cheveu

Parcours de construction littéraire ?





Communication du 22 janvier 2020
« Ébauche de visualisation des productions littéraires francophones d'Haïti et retour d'expérience »

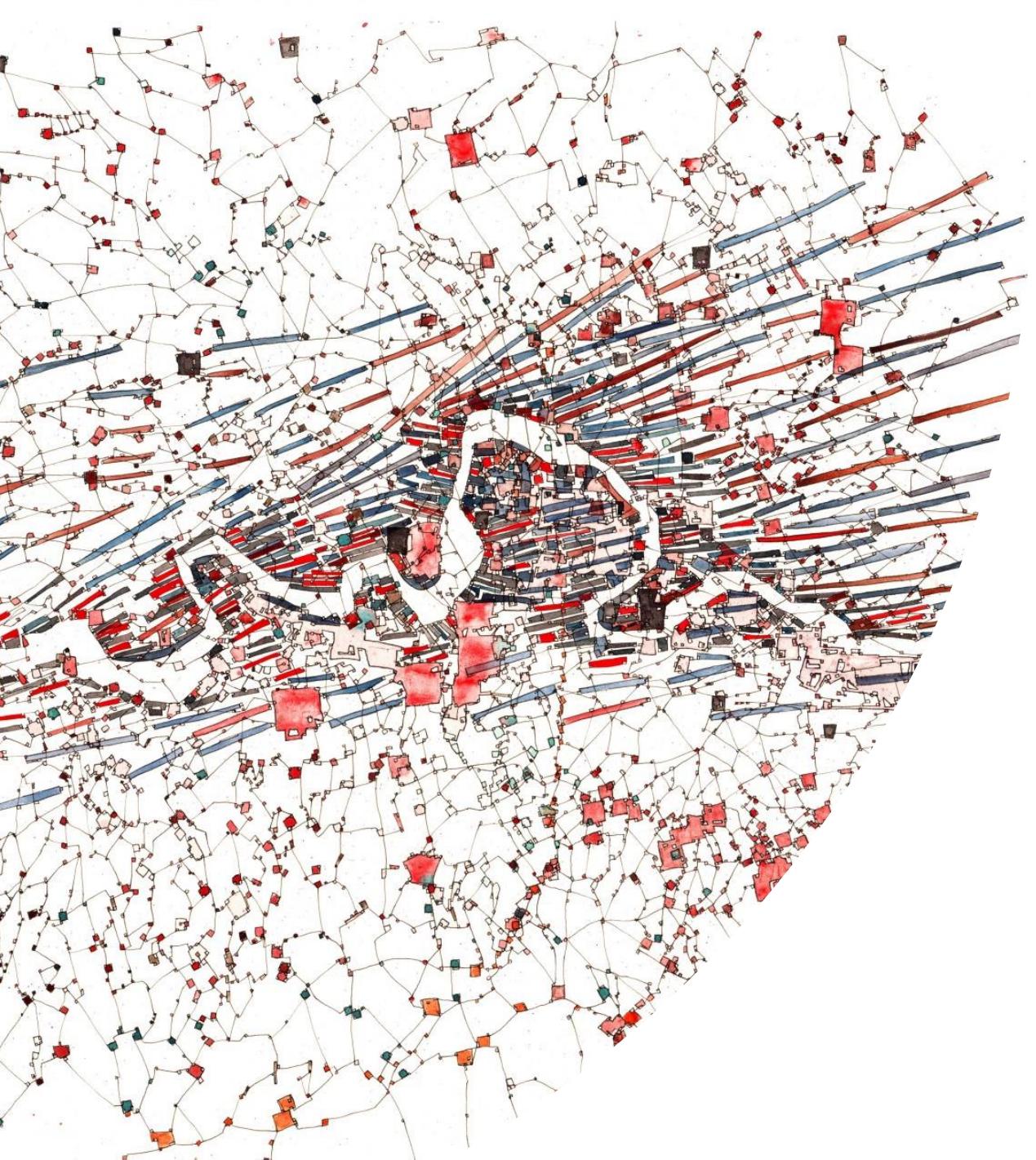


marge



<https://webtv.univ-lyon3.fr/videos/?video=MEDIA200205101828908>





II. Croiser Web sémantique et contenus participatifs

Sur la base d'une bibliothèque de liens Zotéro



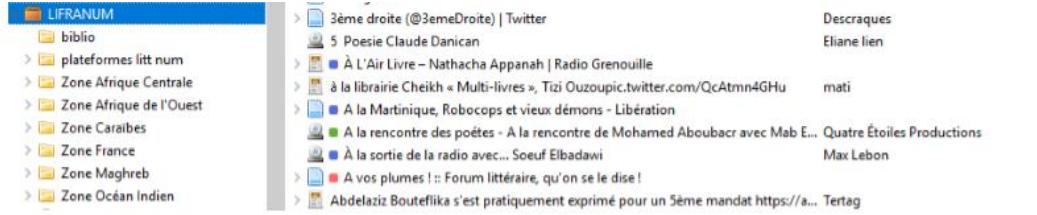
Bibliothèque Zotéro Organisée géographiquement



Extrait RDF des ressources :
Titre site, type, date, auteur,
droits, id, publications,
contributeur, couverture



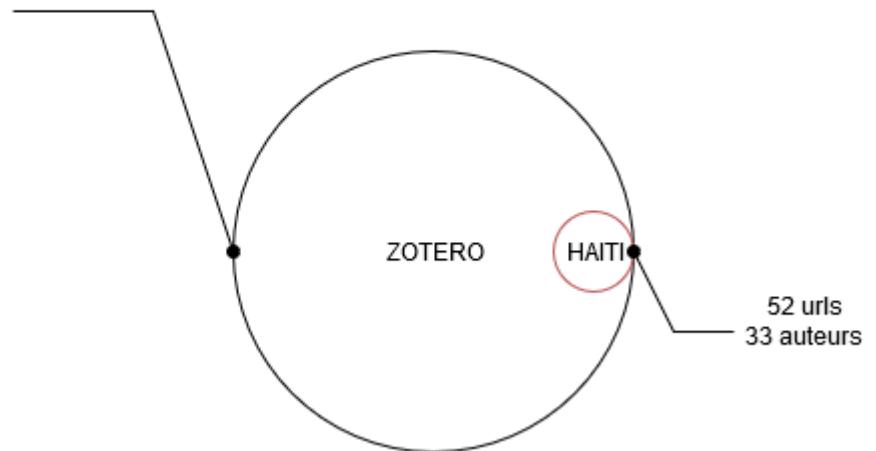
Corpus limité à Haïti



```
krdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:dc="http://purl.org/dc/elements/1.1/">
<rdf:Description rdf:about="https://masiwakomor.com/fr-km/
culture/2019-01-17-la-redemption-des-iguanes-est-une-veritable-foire-un-souk-ou-chacun-peut-trouver">
    <dc:title>La Rédemption des Iguanes est une véritable foire. Un souk où chacun peut trouver sa sauce » | Masiwa komori</dc:title>
    <dc:type>webpage</dc:type>
</rdf:Description>
<rdf:Description rdf:about="https://www.facebook.com/
DrMohamedAHMEDBACAR?ref=profile&_tn_=K2Cdk2CP-R&amp;eid=ARBxm1RoMMY23WaBgrOHpTL-3SgVm9U_n2nYTRILBwBFFF7C6BMrw5fRR4_v6NZ4SwkbjatT-w94mMT">
    <dc:title>Mohamed Ahmed Bacar</dc:title>
    <dc:type>webpage</dc:type>
</rdf:Description>
<rdf:Description rdf:about="https://www.youtube.com/channel/UC603XzimhMR9buKJdhQzABQ">
    <dc:title>Chaine Youtube de Mohamed Ahmed Bacar Rezida</dc:title>
    <dc:type>webpage</dc:type>
    <dc:creator>Mohamed Ahmed Bacar Rezida</dc:creator>
</rdf:Description>
<rdf:Description rdf:about="http://patpantin.over-blog.com/">
    <dc:title>Le blog de Monique MERABET</dc:title>
    <dc:type>webpage</dc:type>
    <dc:creator>MERABET, Monique</dc:creator>
</rdf:Description>
```

Premières observations

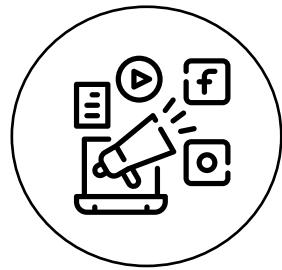
1658 urls
592 ayant un titre
586 typés
240 auteurs identifiés
240 datés



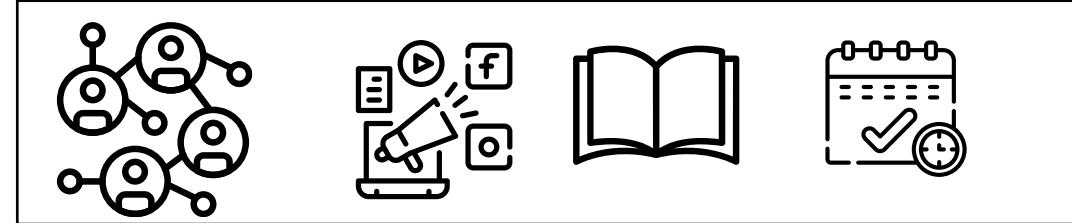
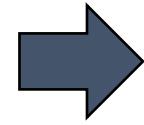
- Visualiser des contenus ?
 - Sources hétérogènes : poèmes, articles de blog ou à caractère journalistique, etc.
- Risque d'un corpus « brouhaha »



Changement de paradigme

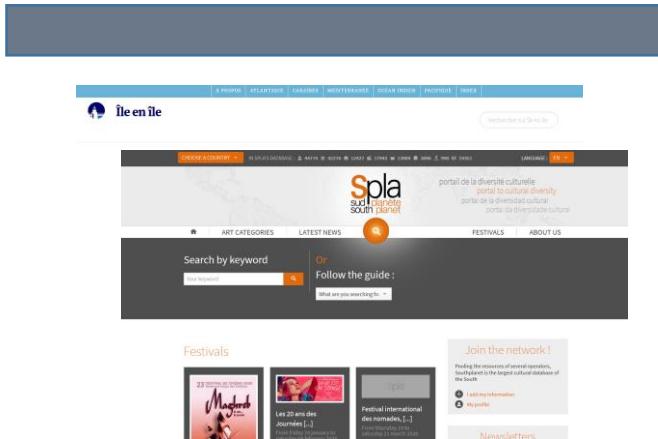


Contenu Web

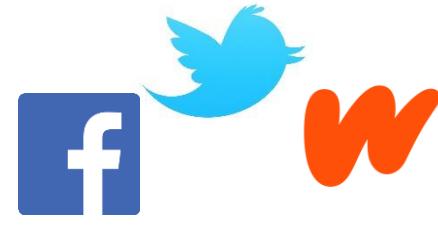


Réseaux d'*auteurs* → production Web + imprimés + évènements

Scraping de sites de « référence »:
Ile-en-île.org, Spla.pro



Données du Web social



...

Données du Web sémantique

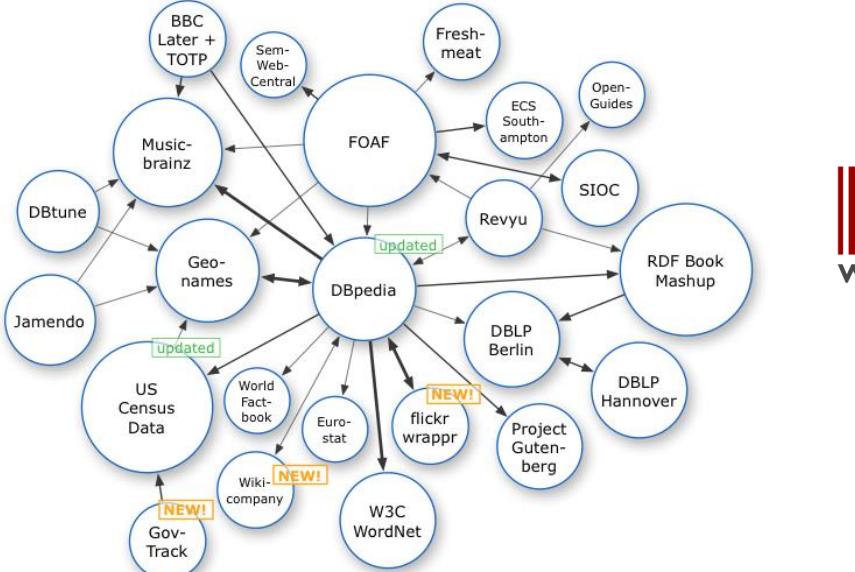
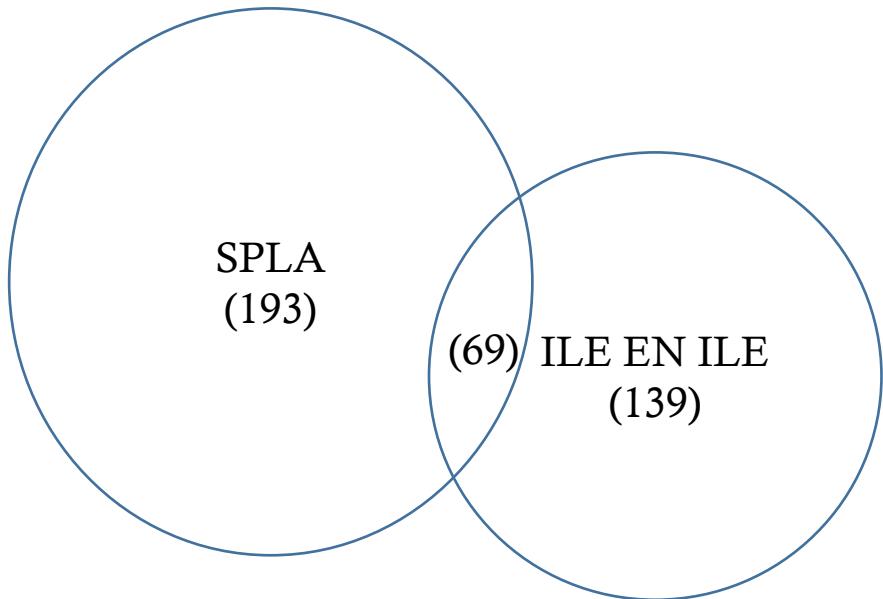




Les sources potentielles et leurs caractéristiques

| | <i>Scraping de sites de « référence »</i> | Données du Web social | Données du Web sémantique |
|--------------------------|---|--|--|
| Qualité de formalisation | - Très peu formatées - Souvent textuelles | - Formatées | - Formatées + vérifiées |
| Informations disponibles | - Profils auteurs - Bibliographies - Références Web | - Profils auteurs - Productions - Évènements + liens | - Profils auteurs - Bibliographies - Liens sémantiques |
| Accès à l'information | - Moteur de recherche | - API + M.R. | - API + M.R. |
| Garant éditorial | - Garant « amateur » | - L'auteur et son réseau | - Garant « expert » |

Informations obtenues via les sites de « référence »



Meta-auteurs Excel : 31 auteurs

- 17 VIAF
- 14 BNF

Meta-final : 274 auteurs

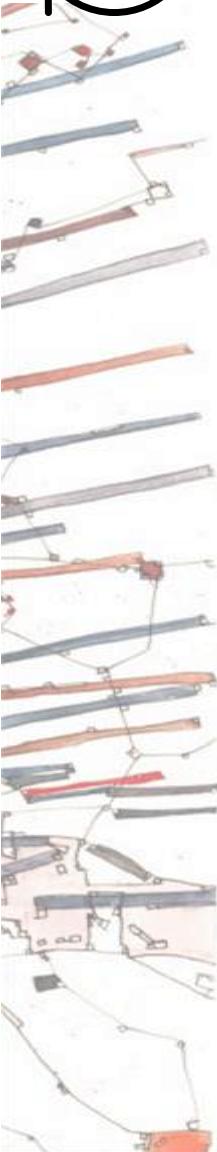
- 74 VIAF
- 86 BNF

MAIS...

- Présence de 44 doublons
 - Et parfois peu d'informations
- nécessité d'avoir une description minimale et normalisée, indépendante de la langue

Notre proposition : interroger le Wikidata via « OpenRefine » et d'une requête SPARQL

[Introduction aux données liées avec OpenRefine](#)



Explicitation de la requête Wikidata - SPARQL



Wikidata Query Service Exemples Aide Davantage d'outils français

```
1 #Auteurs, écrivains et poètes... né après 1920 à Haïti
2 SELECT distinct ?s
3 WHERE
4 {
5   {?s wdt:P106 wd:Q36180 .} UNION {?s wdt:P106 wd:Q482980 .} UNION {?s wdt:P106 wd:Q49757 .} UNION {?s wdt:P106 wd:Q49757 .}
6   {?s wdt:P19 wd:Q790 .} UNION {?s wdt:P19 wd:Q34261 .} UNION {?s wdt:P19 wd:Q923362 .} UNION {?s wdt:P19 wd:Q986301 .} UNION {?s wdt:P19 wd:Q921827 .}
7   ?s wdt:P570 ?time0 .
8   FILTER(?time0 >= "1920-01-01T00:00:00Z"^^xsd:dateTime)
9
10 }
11 }
```

Occupation : écrivain, poète, essayiste...

Né à Haïti

Né après 1920

<https://query.wikidata.org/>

Critères retenus :

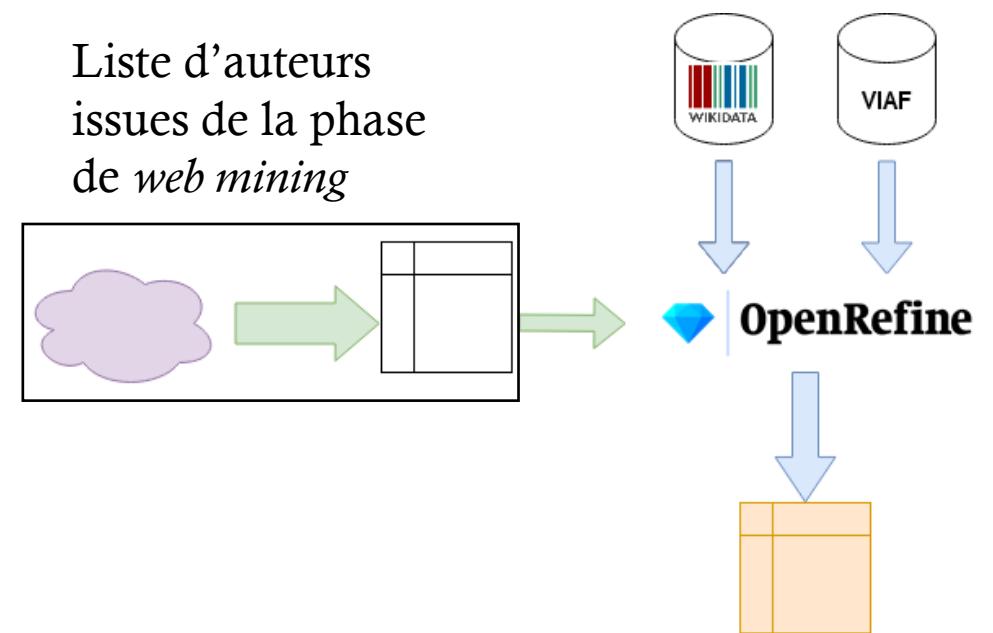
- être né après 1920,
- être originaire d'Haïti,
- occuper une fonction littéraire.

→ 146 résultats dont 25 nouveaux auteurs pertinents

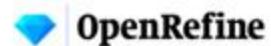


Consolidation des données

OpenRefine permet de réaliser des recherches assistées dans des bases de données Ouvertes et Liées mais aussi à partir de répertoires d'autorités



Insertion de la liste d'écrivains haïtiens



A power tool for working with messy data.

Create Project

Open Project

Import Project

Language Settings

Create a project by importing data. What kinds of data files can I import?

TSV, CSV, *SV, Excel (.xls and .xlsx), JSON, XML, RDF as XML, and Google Data documents are all supported. Support for other formats can be added with OpenRefine extensions.

Get data from

This Computer

Web Addresses (URLs)

Clipboard

Database

Google Data

Paste data from clipboard here:

Stanley Péan
Ignace Nau
Jean Métellus
Lenelle Moïse
Edmond Laforest
Raymond Chassagne
Gary Klang
Frédéric Marcelin
Antoine Innocent
Yanick Jean
Dantès Bellegarde
Georges Anglade
Jean-Max Calvin
Hervé Fanini-Lemoine
Émile Ollivier
Luckner Desvignes
Frédéric Marcelin
Jean Durosier Desrivières
Yanick Lahens
Roland Morisseau
Gesner Jean Gilles
Wébert Charles
Marc Léo Laroche
Rodolphe Matburin
Dany Laferrière
Gérard Chanet

Next »



Regroupements d'informations via Wikidata



OpenRefine clipboard Permalink

Facet / Filter Undo / Redo 0 / 0

Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started?
[Watch these screencasts](#)

243 rows Show as: **rows** records

Reconcile column "Column 1"

Pick a Service or Extension on Left

Services

- Wikidata (old) (en)
- Getty Vocabularies Reconciliation Service
- Wikidata (en)

Facet

Text filter

Edit cell

Edit column

Transpose

Sort...

View

Reconcile

Start reconciling...

Facets

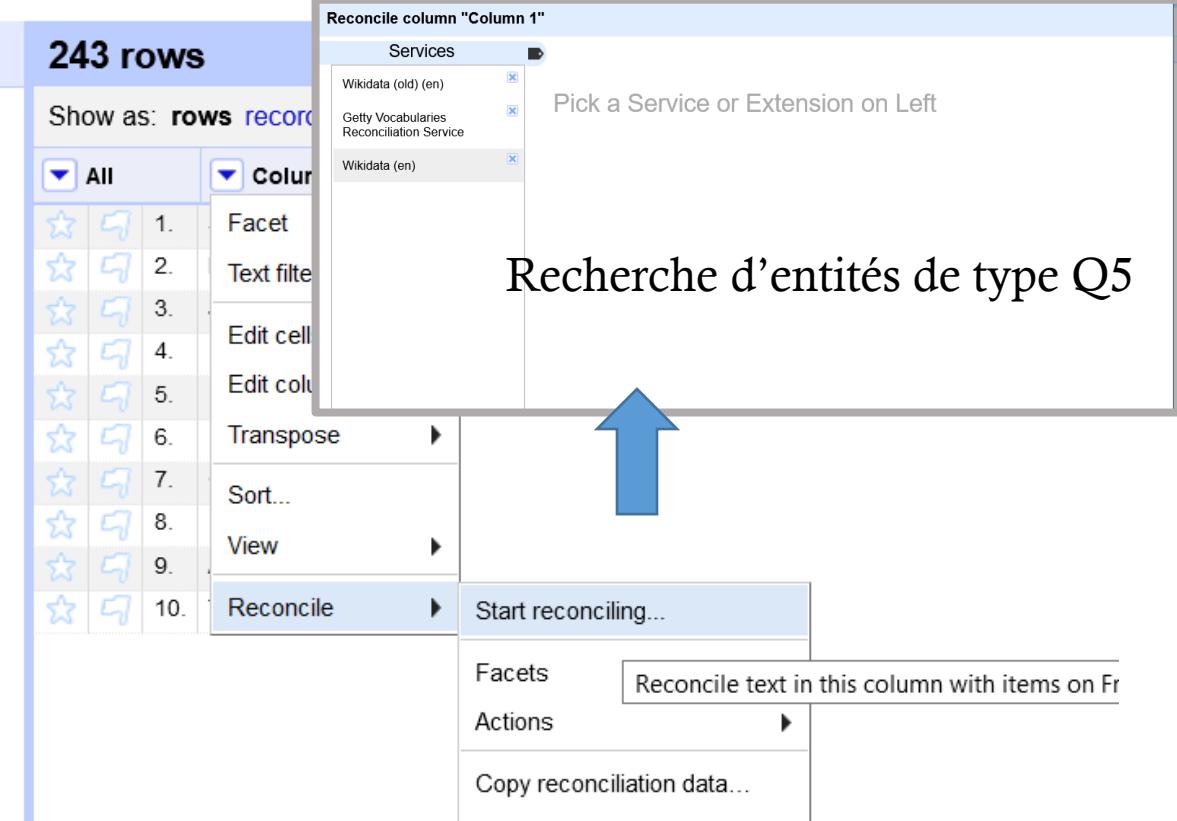
Reconcile text in this column with items on Fr

Actions

Copy reconciliation data...

Use values as identifiers

Recherche d'entités de type Q5



Regroupements d'informations via le VIAF



D'autres services sont disponibles sur <http://refine.codefork.com/> tels que :

- ORCID
- Open Library

Reconcile each cell to an entity of one of these types:

Person
[/people/person](#)

Work
[/book/book](#)

Expression
[/book/book edition](#)

Also use relevant details from other columns:

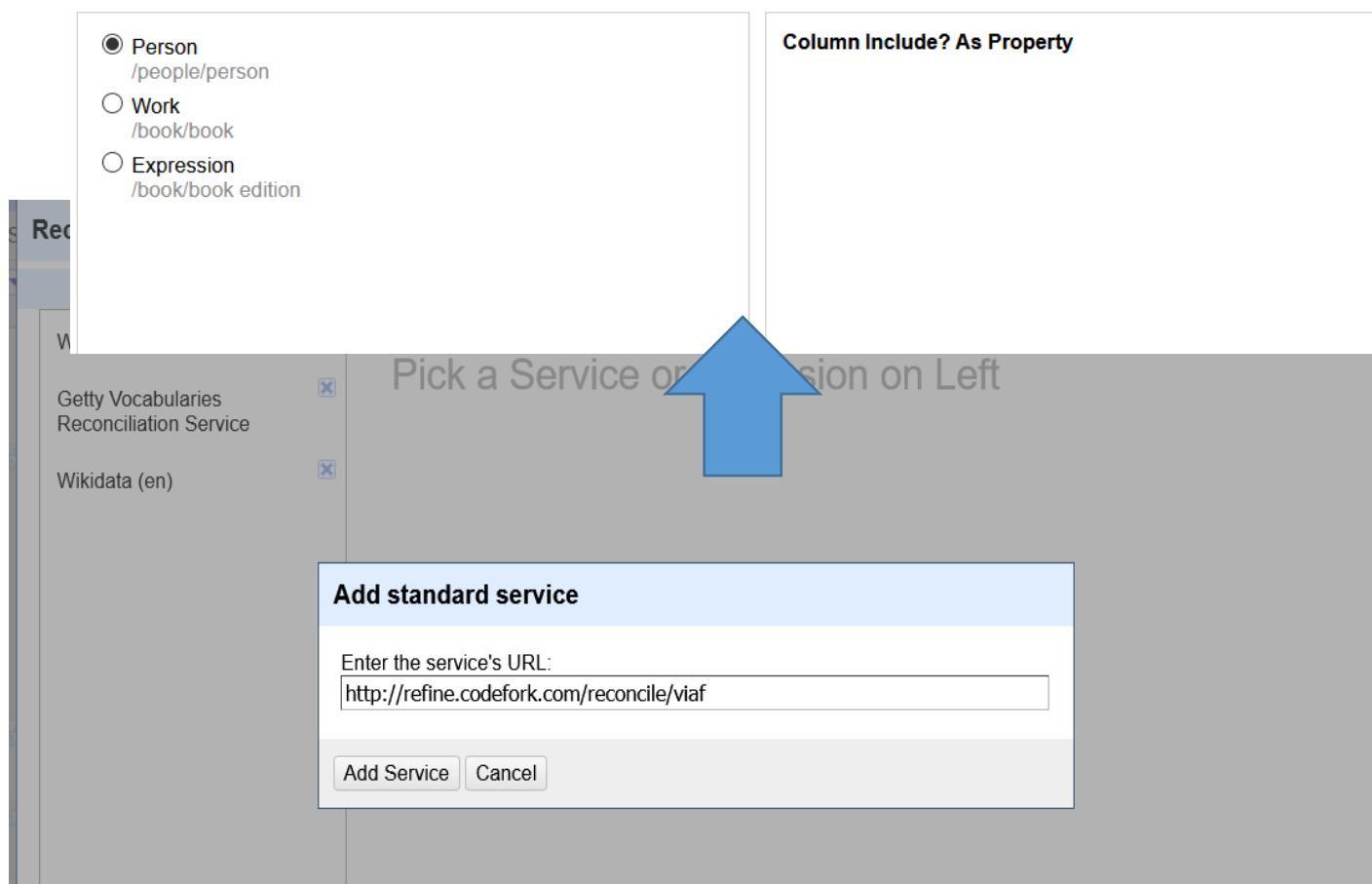
Column Include? As Property

Pick a Service or Reconciliation on Left

Add standard service

Enter the service's URL:

Add Service Cancel



Récupération d'informations clefs

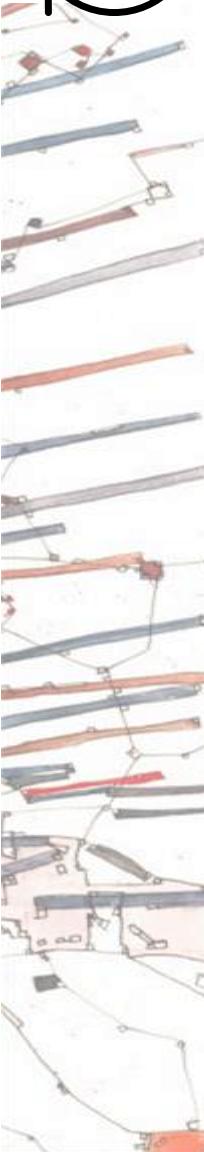


| 207 matching rows (262 total) | | | | | | | |
|---|--|--|--|---|--|--|-------------------------------|
| Show as: rows records | | Show: 5 10 25 50 rows Sort ▾ | | | | | |
| <input type="checkbox"/> All | | <input type="checkbox"/> Column 1 | <input type="checkbox"/> place of birth | <input type="checkbox"/> country of citizenship | <input type="checkbox"/> date of birth | <input type="checkbox"/> date of death | <input type="checkbox"/> VIAF |
| 190. | | Facet ► | Mayotte Choose new match | France Choose new match | 1965-01-06T00:00:00Z | | |
| 182. | | Text filter | | Haiti Choose new match | 1967-01-01T00:00:00Z | | |
| 150. | | Edit cells ► | | | 1945-08-30T00:00:00Z | | |
| 175. | | Edit column ► | Split into several columns... | | | | |
| 186. | | Transpose ► | Join columns... | | | | |
| 98. | | Sort ► | Add column based on this column... | | 1963-01-01T00:00:00Z | | |
| 205. | | View ► | Add column by fetching URLs... | | 1963-01-01T00:00:00Z | | |
| Anthony Phelps | | Reconcile ► | Add columns from reconciled values... | | 1850-10-18T00:00:00Z | 1911-09-19T00:00:00Z | |
| | | | Rename this column | | 1928-08-25T00:00:00Z | | |
| | | | Remove this column | | 1874-01-01T00:00:00Z | 1960-01-01T00:00:00Z | |
| | | | Move column to beginning | | | | |
| | | | Move column to end | | | | |
| | | | Move column left | | | | |
| | | | Move column right | | | | |
| 107. | | | Port-au-Prince Choose new match | Haiti Choose new match | 1776-01-01T00:00:00Z | 1806-01-01T00:00:00Z | |
| | | | | | 1973-03-06T00:00:00Z | | |

Ainsi nous récupérons via Wikidata :

Date de naissance
Date de décès
Lieu de naissance
Lieu de décès
Nationalité
Professions

....



Filtrage du corpus



1) Suppression des doublons

2) Normalisation du corpus

→ 232 individus uniques

3) Application de filtres plus restrictifs:

- Débuts d'Internet (à destination du public et des universitaires) : années 1990-2000
- Littéraires ayant pu avoir recours à Internet pour diffuser leurs travaux et faire leur promotion.

Choix corpus : littéraires nés avant 1920, décédés après 2000

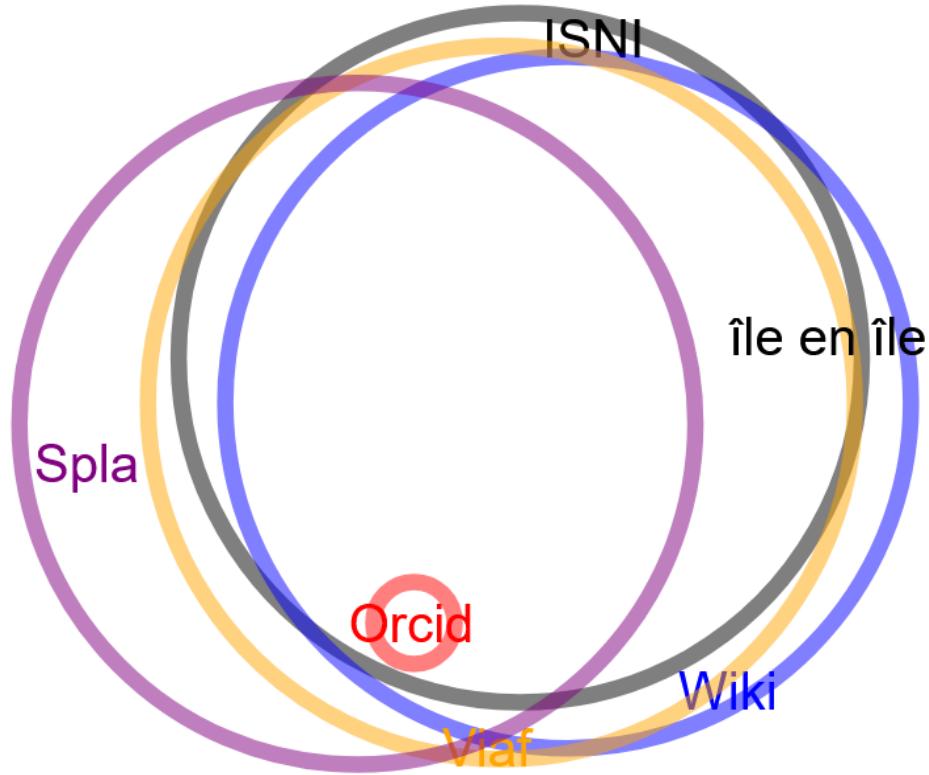
Si la consolidation des données via Wikidata s'est avérée efficace, il demeure que :

- 22 dates de naissances pu être récupérées manuellement dans les notices biographiques

- Et 8 autres grâce au récolement du VIAF

→ Il convient donc de mobiliser diverses sources d'enrichissement.

Le corpus final en quelques chiffres clés



190 individus dont :

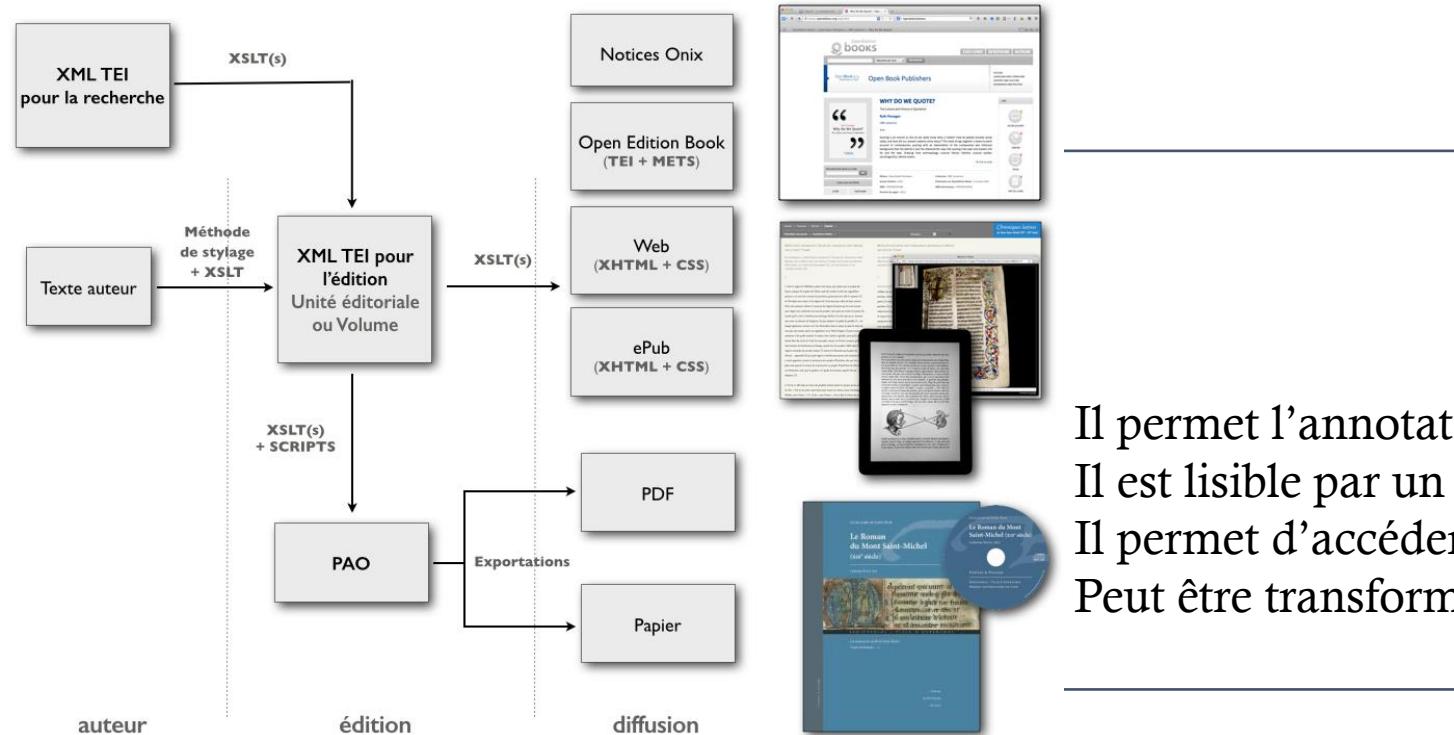
- 151 sont renseignés par une notice VIAF
- 142 Wikipédia
- 141 ISNI
- 138 Spla
- 108 Ile-en-île
- 2 ORCID

| Informations | Nbre de données obtenues |
|----------------------------|--------------------------|
| Sexe/genre | 190 |
| Nationalité | 190 |
| Langue | 190 |
| Activités professionnelles | 190 |
| Date de naissance | 163 |
| Lieu de naissance | 129 |
| Date de décès | 35 |
| Lieu de décès | 26 |



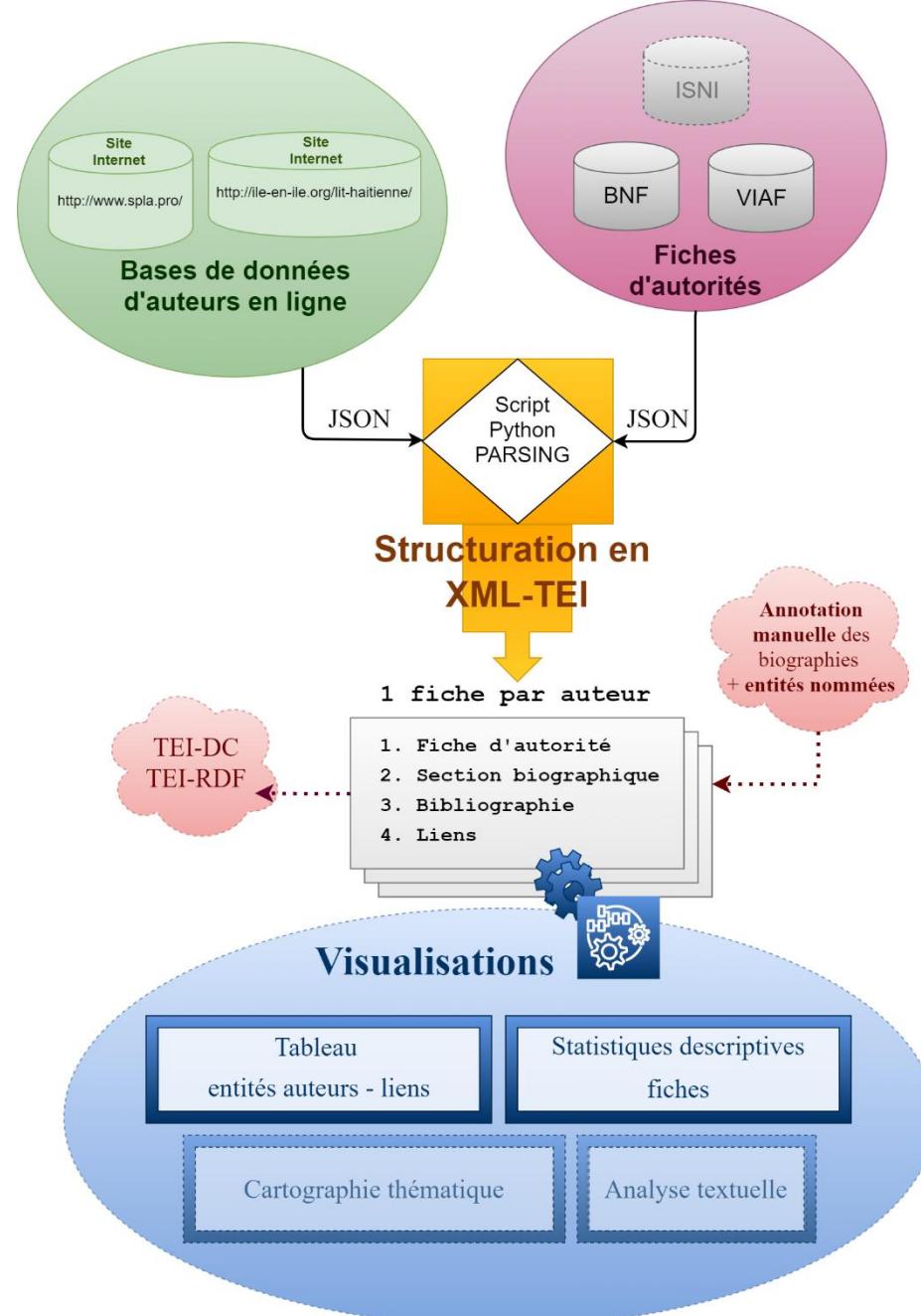


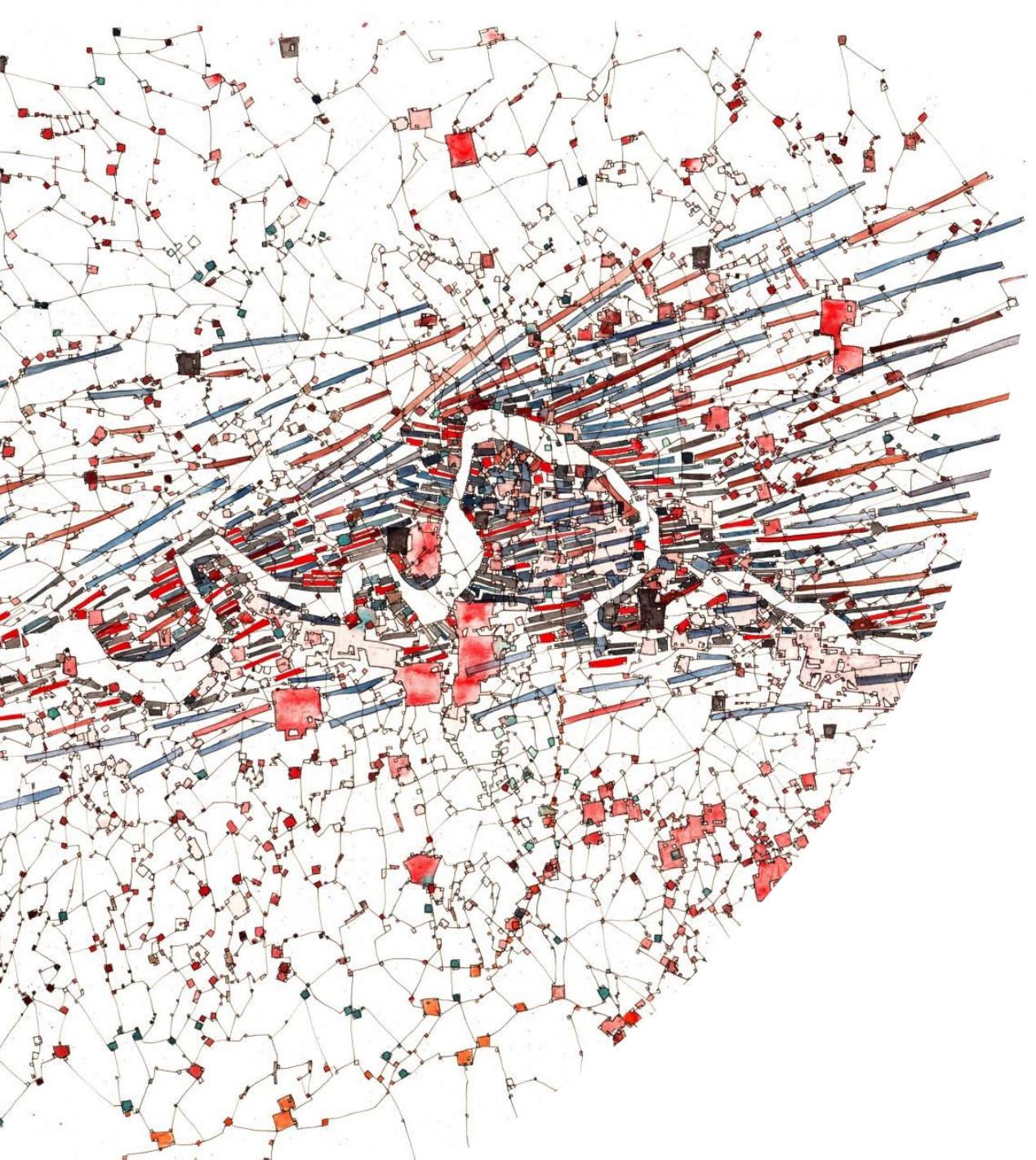
Agréger et normaliser le tout ?



Il permet l'annotation de document à même le contenu
Il est lisible par un humain
Il permet d'accéder à un écosystème orienté documents
Peut être transformé dans de nombreuses « formes »

Schéma de la chaîne éditoriale (modèle du Single Source Publishing)





III. Normalisation du corpus et structuration en XML-TEI

- Avant toute chose ...

Un contenu textuel

```
<persName>Bureau Bruno</persName>
```

Une balise : ouvrante `<balise>` et fermante `</balise>`

```
<persName>
  <surname>Bureau</surname>
  <forename>Bruno</forename>
</persName>
```

Des balises dans des balises

```
<persName ana="bureau_bruno">Il Maestro</persName>
```

Un attribut typé avec un texte

Un texte qui contient des annotations numériques en vue de l'enrichir et sans changer son contenu.

Structure du fichier XML-TEI

TEI

```

1  <?xml version="1.0" encoding="UTF-8"?> <!-- déclaration du fichier TEI et de son encodage -->
2  <TEI xmlns:xi="http://www.w3.org/2001/XInclude" xml:lang="fr" xml:id="Marie-Celie_Agnant">
3
4      <!-- 1 fichier TEI par entité auteur -->
5      <!-- xml:id= correspondant à celui du fichier master.-->
6      <!-- Normalisation de l'id: prenom_nom de entité auteur -->
7
8  <teiHeader> [21 lines] → 1. Un teiHeader
30
31  <div xml:id="Marie-Celie_Agnant_fiche" n="1"> [65 lines]
97
98  <div xml:id="Marie-Celie_Agnant_bio" n="2"> [73 lines]
172
173  <div xml:id="Marie-Celie_Agnant_biblio" n="3"> [81 lines]
255
256  <div xml:id="Marie-Celie_Agnant_web" n="4"> [11 lines]
268
269  </TEI>

```

→ Un identifiant *xml:id* unique et normalisé

1. Un teiHeader
2. Le corps



Le <teiHeader>

```
<teiHeader>
  <!-- déclaration des métadonnées -->
  <fileDesc>
    <titleStmt>
      <title>Notice de Marie-Célie Agnant - corpus Haïti</title>
      <!-- prénom, nom de l'auteur -->
      <author>Inès Burri</author>
      <author>Anaïs Chambat</author>
      <author>Célian Ringwald</author>
    </titleStmt>
    <publicationStmt>
      <orgName>Université Jean Moulin Lyon 3</orgName>
      <orgName>Université Lumière Lyon 2</orgName>
      <orgName>ENS de Lyon</orgName>
      <pubPlace>Lyon</pubPlace>
      <date>17.01.2020</date>
      <!-- date de l'extraction -->
    </publicationStmt>
    <sourceDesc>Notice et contenus associés à l'entité auteur dont les données sont issues
      de sites de références et de répertoires d'autorités</sourceDesc>
  </fileDesc>
</teiHeader>
```

Description du fichier, ...

... des données éditoriales

... et des sources

Des <div>

TEI

```

1  <?xml version="1.0" encoding="UTF-8"?> <!-- déclaration du fichier TEI et de son encodage -->
2 ▷ <TEI xmlns:xi="http://www.w3.org/2001/XInclude" xml:lang="fr" xml:id="Marie-Celie_Agnant">
3
4      <!-- 1 fichier TEI par entité auteur -->
5      <!-- xml:id= correspondant à celui du fichier master -->
6      <!-- Normalisation de l'id: prenom_nom de entité auteur -->
7
8 ▷ <teiHeader> [21 lines]
30
31 ▷ <div xml:id="Marie-Celie_Agnant_fiche" n="1"> [65 lines] → Un numéro
97
98 ▷ <div xml:id="Marie-Celie_Agnant_bio" n="2"> [73 lines]
172
173 ▷ <div xml:id="Marie-Celie_Agnant_biblio" n="3"> [81 lines]
255
256 ▷ <div xml:id="Marie-Celie_Agnant_web" n="4"> [11 lines]
268
269 </TEI>

```

Un numéro

Un identifiant unique

<div> fiche d'autorité (1)



```
<div xml:id="Marie-Celie_Agnant_fiche" n="1">
    <!-- xml:id= prenom_nom de entité auteur_type de fichier (ici fiche)-->
    <!-- n=1 pour fiche -->

    <person ana="Marie-Celie_Agnant" sex="f">
        <!-- ana= prenom_nom de l'auteur -->
        <!-- sex= f pour féminin; m pour masculin -->
        <persName>Agnant Marie-Célie</persName> → Nom et prénom

        <!-- Identifiants liés aux répertoires d'autorités disponibles -->
        <!-- Il est possible d'indiquer pour chaque balise, la provenance de l'élément grâce à
            <idno type="BNF"><g ref="https://catalogue.bnf.fr/ark:/12148/cb145884081"
                />FRBNF14588408</idno>
            <idno type="VIAF"><g ref="http://viaf.org/viaf/17470529"/>17470529</idno>
            <idno type="DBPEDIA"><g ref="https://fr.wikipedia.org/wiki/Marie-C%C3%A9lie_Agnant"
                />Marie-Célie_Agnant</idno>
            <!-- DB de Wikipédia -->

            <birth source="BNF">
                <date>1953</date>
                <placeName>Haïti</placeName>
            </birth>
            <death>
                <date/>
                <placeName/>
            </death>

            <langKnowledge source="BNF">
                <!-- pour le code de la langue utilisé en XML-TEI voir : http://www.iana.org/assign
                <langKnown tag="en">anglais</langKnown>
                <langKnown tag="fr">français</langKnown>
                <langKnown tag="ht">haitien-creole</langKnown>
            </langKnowledge>
        </!-- Identifiants des répertoires autorité -->
        <!-- Naissance et décès -->
        <!-- Langues parlées -->
    </person>
</div>
```

Nom et prénom

Identifiants des répertoires autorité

Naissance et décès

Langues parlées

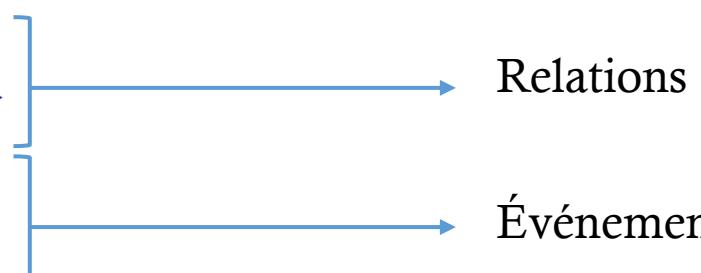
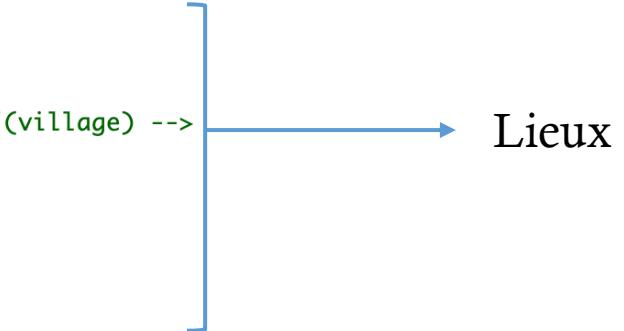
<div> fiche d'autorité (2)



```
<residence>
  <placeName>
    <settlement/>
    
    <country source="BNF">Haïti</country>
    
    <country source="BNF">Canada</country>
  </placeName>
</residence>

<occupation source="île-en-île">romancière</occupation>
<occupation source="île-en-île">poète</occupation>
<occupation source="île-en-île">nouvelliste</occupation>
<occupation source="île-en-île">conteuse</occupation>
<occupation source="île-en-île">traductrice</occupation>
<occupation source="île-en-île">interprète</occupation>
<occupation source="île-en-île">assistante de recherche</occupation>
<!-- Le contenu textuel des balises "residence" et "occupation" ne pourra être rempli que par une personne physique ou morale. Les deux balises sont donc vides dans cet exemple. --&gt;

&lt;listRelation&gt;
  <!-- Liste des noms associés à l'auteur --&gt;
  &lt;relation ana="" source="BNF"&gt;François Gardant&lt;/relation&gt;
&lt;/listRelation&gt;
&lt;listEvent&gt;
  <!-- Liste des événements associés à l'auteur --&gt;
  &lt;event type="prix_littéraire" when="01/01/2020"/&gt;
&lt;/listEvent&gt;
&lt;/person&gt;
<!-- Outre la BNF et le VIAF, l'ISNI peut s'avérer complémentaire en cas de déficit d'information --&gt;</pre>
```



Partie 1

```

<div xml:id="Marie-Celie_Agnant_fiche" n="1">
    <!-- xml:id= prenom_nom de entité auteur_type de fichier (ici fiche)-->
    <!-- n=1 pour fiche -->

    <person ana="Marie-Celie_Agnant" sex="f">
        <!-- ana= prenom_nom de l'auteur -->
        <!-- sex= f pour féminin; m pour masculin -->
        <persName>Agnant Marie-Célie</persName>

        <!-- Identifiants liés aux répertoires d'autorités disponibles -->
        <!-- Il est possible d'indiquer pour chaque balise, la provenance de
            la donnée -->
        <idno type="BNF"><g ref="https://catalogue.bnf.fr/ark:/12148/cb14588408"/>FRBNF14588408</idno>
        <idno type="VIAF"><a ref="http://viaf.org/viaf/17470529">17470529</a></idno>
        <idno type="DBPEDIA"><g ref="https://fr.wikipedia.org/wiki/Marie-C%C3%A9lie_Agnant">Marie-Célie_Agnant</idno>
        <!-- DB de Wikipédia -->

        <birth source="BNF">
            <date>1953</date>
            <placeName>Haïti</placeName>
        </birth>
        <death>
            <date/>
            <placeName/>
        </death>

        <langKnowledge source="BNF">
            <!-- pour le code de la langue utilisé en XML-TEI voir : http://
            <langKnown tag="en">anglais</langKnown>
            <langKnown tag="fr">français</langKnown>
            <langKnown tag="ht">haitien-creole</langKnown>
        </langKnowledge>

```

Partie 2

```

<residence>
    <placeName>
        <settlement/>
        <!-- balise qui admet un type="city"(ville) et "town"(village) -->
        <region source="BNF">Québec</region>
        <!-- région -->
        <country source="BNF">Haïti</country>
        <!-- pays -->
        <country source="BNF">Canada</country>
    </placeName>
</residence>

<occupation source="ile-en-île">romancière</occupation>
<occupation source="ile-en-île">poète</occupation>
<occupation source="ile-en-île">nouvelliste</occupation>
<occupation source="ile-en-île">conteuse</occupation>
<occupation source="ile-en-île">traductrice</occupation>
<occupation source="ile-en-île">interprète</occupation>
<occupation source="ile-en-île">assistante de recherche</occupation>
<!-- Le contenu textuel des balises "residence" et "occupation" ne pouvait pas être affiché dans la partie 1 -->

<listRelation>
    <!-- Liste des noms associés à l'auteur -->
    <relation ana="" source="BNF">François Gardant</relation>
</listRelation>
<listEvent>
    <!-- Liste des événements associés à l'auteur -->
    <event type="prix_litteraire" when="01/01/2020"/>
</listEvent>
</person>
<!-- Outre la BNF et le VIAF, l'ISNI peut s'avérer complémentaire en cas de
    recherche -->

```

<div> bio

```
<div xml:id="Marie-Celie_Agnant_bio" n="2">
    <!-- xml:id= prenom_nom de entité auteur_type de fichier (ici bio)-->
    <!-- n=2 pour bio -->

    <text ana="Marie-Celie_Agnant" source="ile-en-ile" ref="http://ile-en-ile.org/agnant/" hand="Lucie Lequin">
        <!-- ana= prenom_nom de l'auteur -->
        <!-- ref= adresse de la bio -->
        <!-- hand = dernier paragraphe de la bio -->

        <p>Poète, nouvelliste, romancière, Marie-Célie Agnant quitte Haïti pour le Canada en 1970. Traductrice et interprète, elle détient un diplôme en enseignement du français et a enseigné plusieurs années tout en travaillant comme assistante de recherche. C'est ainsi que son premier roman, La dot de Sara, publié en 1995, sera « construit à partir de récits de vie de grands-mères d'origine haïtienne, transmis en créole dans le cadre d'une recherche en sociologie ». </p>
        <p>Écrivaine présente et attentive au monde qui l'entoure, son écriture porte à la fois le sceau de la poésie et de la violence issue des sociétés postcoloniales qui naviguent entre misère criante et opulence indécente. Ses textes, dont certains ont été traduits en plusieurs langues, trouvent leur ancrage dans la réalité sociale contemporaine ; elle aborde les thèmes tels l'exclusion, la solitude, le racisme, l'exil. La condition des femmes, le rapport au passé et à la mémoire font aussi partie de son champ d'exploration. </p>
    </text>
</div>
```

La source

L'adresse Web

Le nom de l'auteur de la biographie

Contenu de la biographie dans des balises <p> (paragraphe)



<div> biblio



```
<div xml:id="Marie-Celie_Agnant_biblio" n="3">
  <!-- xml:id= prenom_nom de l'auteur_type de fichier (ici biblio)-->
  <!-- n=3 pour biblio -->

  <div ana="Marie-Celie_Agnant" source="ile-en-ile" ref="http://ile-en-ile.org/agnant/" n="3.1">
    <!-- ana= prenom_nom de l'auteur -->
    <!-- source= nom du site à partir duquel les données ont été extraites -->
    <!-- ref= adresse de la biblio -->
    <!-- n= numéro itératif qui permet de hiérarchiser suivant le nombre de sites cons-->

    <!-- Chaque item = une référence unique (les romans ont été développés en exemple)
    <!-- POUR LES LIST : subtype= titre de la section principale -->
    <!-- POUR LES ITEM : type= titre de la section principale ; subtype= titres des soi
    <!-- Les références changent d'une entité auteur à l'autre: l'information extraite

    <list subtype="romans"> [10 lines]
      <list subtype="nouvelles"> [2 lines]
      <list subtype="poésie"> [2 lines]
      <list subtype="romans pour la jeunesse"> [2 lines]
      <list subtype="contes"> [2 lines]
      <list subtype="textes publiés dans des ouvrages collectifs"> [2 lines]
      <list subtype="distinctions littéraires"> [2 lines]

      <!-- œuvres traduites -->
      <!-- pour le code de la langue utilisé en XML-TEI voir : http://www.iana.org/assign
      <list subtype="catalan" tag="ca"> [2 lines]
      <list subtype="anglais" tag="en"> [2 lines]
      <list subtype="espagnol" tag="es"> [2 lines]
      <list subtype="italien" tag="it"> [2 lines]
      <list subtype="néerlandais" tag="nl"> [2 lines]

      <list subtype="ouvrage collectif"> [2 lines]
      <list subtype="articles sélectionnés"> [2 lines]
      <list subtype="entretiens"> [2 lines]
    </list>
  </div>

  <div ana="Marie-Celie_Agnant" source="VIAF" n="3.2"> [4 lines]
</div>
```

Un numéro

L'adresse Web

La source

Des balises <list> avec un attribut *subtype*
Dans celles-ci des <item>
pour chaque entrée bibliographique

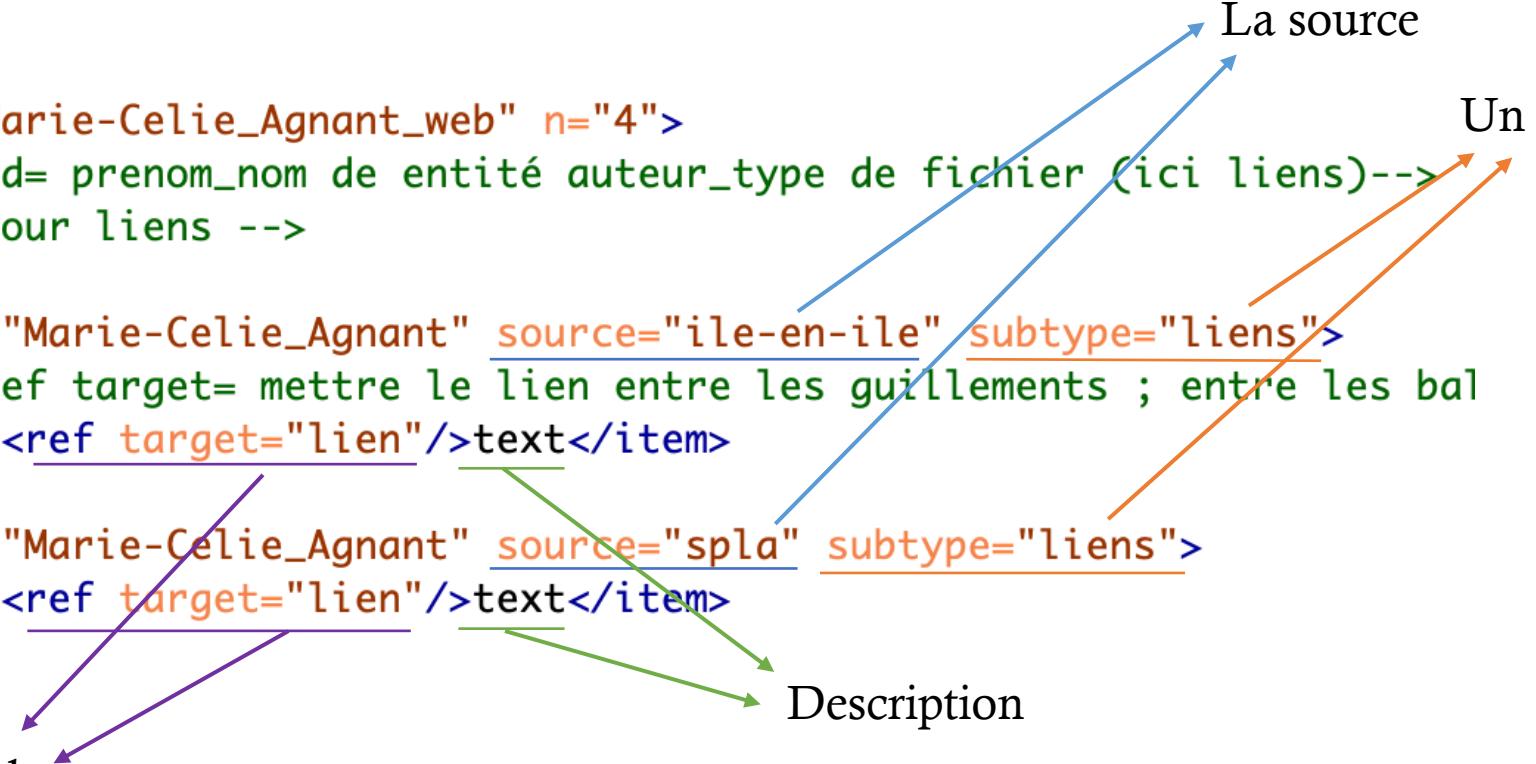
<div> adresses Web

```

<div xml:id="Marie-Celie_Agnant_web" n="4">
    <!-- xml:id= prenom_nom de entité auteur_type de fichier (ici liens)-->
    <!-- n=4 pour liens -->

    → <list ana="Marie-Celie_Agnant" source="ile-en-ile" subtype="liens">
        <!-- ref target= mettre le lien entre les guillements ; entre les bal
            ↗ <item><ref target="lien"/>text</item>
        </list>
    → <list ana="Marie-Celie_Agnant" source="spla" subtype="liens">
        ↗ <item><ref target="lien"/>text</item>
    </list>
</div>

```



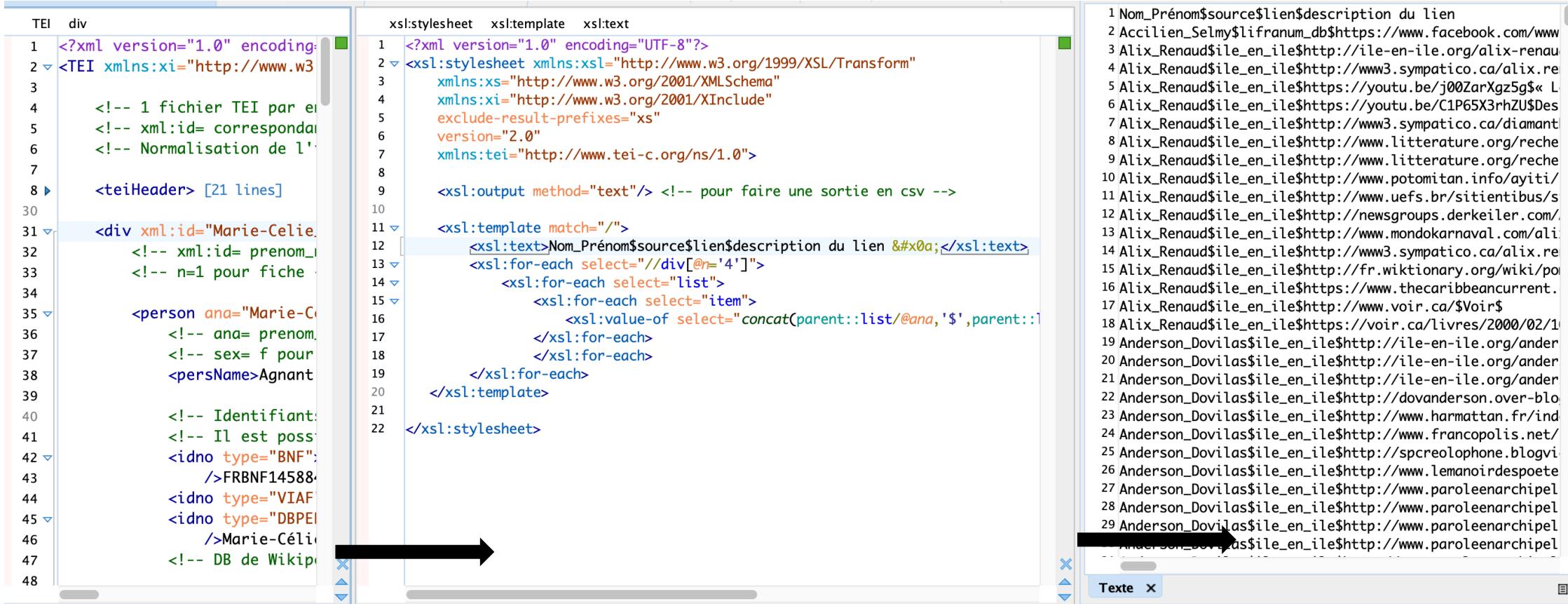
La source

Un sous-type

Description

L'adresse Web

Les feuilles de transformations XSL



The diagram illustrates the XSLT transformation process:

1. Corpus XML (represented by the leftmost window showing TEI XML code)
- +
2. Le fichier XSL (represented by the middle window showing XSLT code)
- =
3. Un résultat (.xml, .html, .txt, etc) (represented by the rightmost window showing a list of transformed results)

Corpus XML (Left Window):

```
TEI div
1 <?xml version="1.0" encoding="UTF-8"?>
2 <TEI xmlns:xi="http://www.w3.org/2001/XInclude"
3
4     <!-- 1 fichier TEI par exemple
5     <!-- xml:id= correspondant
6     <!-- Normalisation de l'ensemble
7
8 <teiHeader> [21 lines]
30
31 <div xml:id="Marie-Celie...
32     <!-- xml:id= prenom_id
33     <!-- n=1 pour fiche
34
35     <person ana="Marie-Celie...
36         <!-- ana= prenom_id
37         <!-- sex= f pour femme
38         <persName>Agnant...
39
40         <!-- Identifiant
41         <!-- Il est possible
42         <idno type="BNF">
43             />FRBNF14588...
44         <idno type="VIAF">
45         <idno type="DBPEI">
46             />Marie-Célie...
47         <!-- DB de Wikipedia
```

XSLT_stylesheet (Middle Window):

```
1 <xsl:stylesheet xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
2     xmlns:xs="http://www.w3.org/2001/XMLSchema"
3     xmlns:xi="http://www.w3.org/2001/XInclude"
4     exclude-result-prefixes="xs"
5     version="2.0"
6     xmlns:tei="http://www.tei-c.org/ns/1.0">
7
8     <xsl:output method="text"/> <!-- pour faire une sortie en csv -->
9
10    <xsl:template match="/">
11        <xsl:text>Nom_Prenom$source$lien$description du lien &#xA;</xsl:text>
12        <xsl:for-each select="//div[@n='4']">
13            <xsl:for-each select="list">
14                <xsl:for-each select="item">
15                    <xsl:value-of select="concat(parent::list/@ana, '$', parent::...
16                </xsl:for-each>
17            </xsl:for-each>
18        </xsl:for-each>
19    </xsl:template>
20
21 </xsl:stylesheet>
```

Résultat (Right Window):

```
1 Nom_Prenom$source$lien$description du lien
2 Accilien_Selmy$lifranum_db$https://www.facebook.com/www...
3 Alix_Renaud$ile_en_ile$http://ile-en-ile.org/alix-re...
4 Alix_Renaud$ile_en_ile$http://www3.sympatico.ca/alix.re...
5 Alix_Renaud$ile_en_ile$https://youtu.be/j00ZarXgz5g$< L...
6 Alix_Renaud$ile_en_ile$http://youtu.be/C1P65X3rhZU$Des...
7 Alix_Renaud$ile_en_ile$http://www3.sympatico.ca/diamant...
8 Alix_Renaud$ile_en_ile$http://www.litterature.org/reche...
9 Alix_Renaud$ile_en_ile$http://www.litterature.org/reche...
10 Alix_Renaud$ile_en_ile$http://www.potomitan.info/ayiti/...
11 Alix_Renaud$ile_en_ile$http://www.uefs.br/sitientibus/s...
12 Alix_Renaud$ile_en_ile$http://newsgroups.derkeiler.com/...
13 Alix_Renaud$ile_en_ile$http://www.mondokarnaval.com/ali...
14 Alix_Renaud$ile_en_ile$http://www3.sympatico.ca/alix.re...
15 Alix_Renaud$ile_en_ile$http://fr.wiktionary.org/wiki/po...
16 Alix_Renaud$ile_en_ile$https://www.thecaribbeancurrent....
17 Alix_Renaud$ile_en_ile$http://www.voir.ca/$Voir$...
18 Alix_Renaud$ile_en_ile$https://voir.ca/livres/2000/02/1...
19 Anderson_Dovilas$ile_en_ile$http://ile-en-ile.org/ander...
20 Anderson_Dovilas$ile_en_ile$http://ile-en-ile.org/ander...
21 Anderson_Dovilas$ile_en_ile$http://ile-en-ile.org/ander...
22 Anderson_Dovilas$ile_en_ile$http://dovanderson.over-blog...
23 Anderson_Dovilas$ile_en_ile$http://www.harmattan.fr/ind...
24 Anderson_Dovilas$ile_en_ile$http://www.francopolis.net/...
25 Anderson_Dovilas$ile_en_ile$http://spcreolophone.blogvi...
26 Anderson_Dovilas$ile_en_ile$http://www.lemanoirdespoete...
27 Anderson_Dovilas$ile_en_ile$http://www.paroleenarchipel...
28 Anderson_Dovilas$ile_en_ile$http://www.paroleenarchipel...
29 Anderson_Dovilas$ile_en_ile$http://www.paroleenarchipel...
Anderson_Dovilas$ile_en_ile$http://www.paroleenarchipel...
```

1. Corpus XML



2. Le fichier XSL



3. Un résultat (.xml, .html, .txt, etc)

Les feuilles de transformations XSL

Exemple : création d'un tableau avec *entité auteur/ lien/ source du lien/ description du lien*

```
<?xml version="1.0" encoding="UTF-8"?>
<xsl:stylesheet xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
  xmlns:xs="http://www.w3.org/2001/XMLSchema"
  xmlns:xi="http://www.w3.org/2001/XInclude"
  exclude-result-prefixes="xs"
  version="2.0"
  xmlns:tei="http://www.tei-c.org/ns/1.0">

  <xsl:output method="text"/> <!-- pour faire une sortie en csv --&gt;

  &lt;xsl:template match="/"&gt;
    &lt;xsl:text&gt;Nom_Préénom$source$lien$description du lien &amp;#xA;&lt;/xsl:text&gt;
    &lt;xsl:for-each select="//div[@n='4']"&gt;
      &lt;xsl:for-each select="list"&gt;
        &lt;xsl:for-each select="item"&gt;
          &lt;xsl:value-of select="concat(parent::list/@ana, '$', parent::list/@source, '$', ref/@target, '$', ., '$', '&amp;#xA;')"/&gt;
        &lt;/xsl:for-each&gt;
        &lt;/xsl:for-each&gt;
      &lt;/xsl:for-each&gt;
    &lt;/xsl:template&gt;
&lt;/xsl:stylesheet&gt;</pre>

Une sortie .txt



Création des noms de colonnes



Remplissage des lignes du tableau

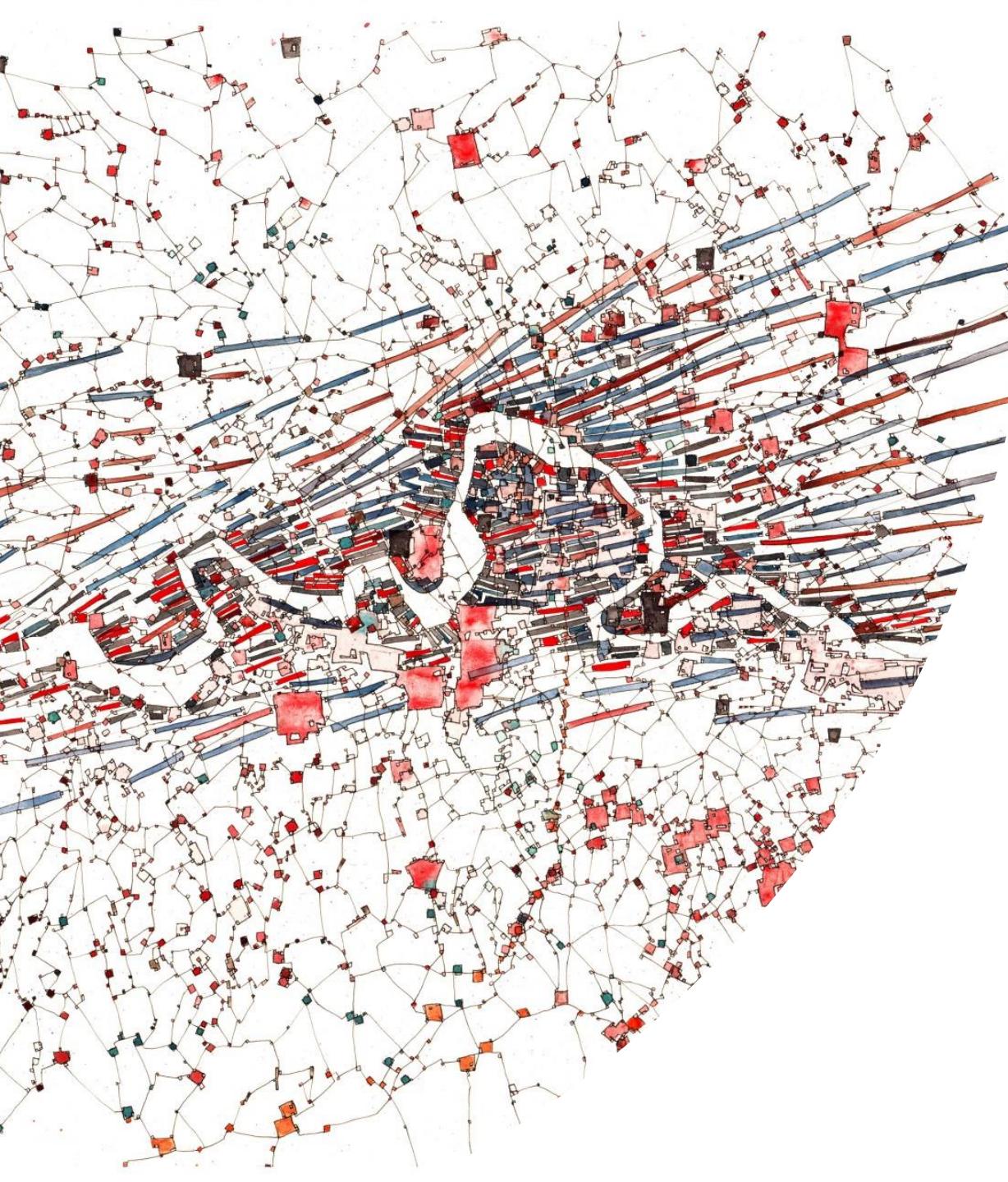

```



Pour obtenir :

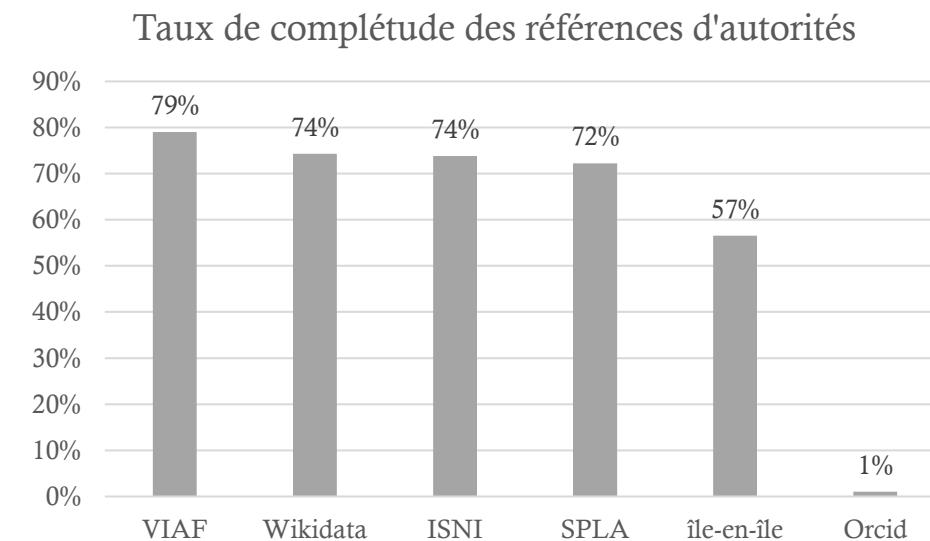
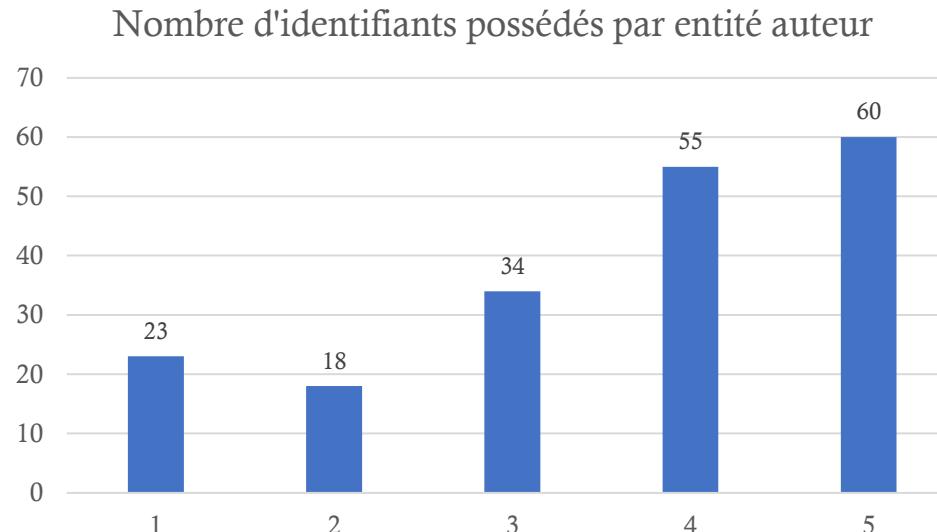
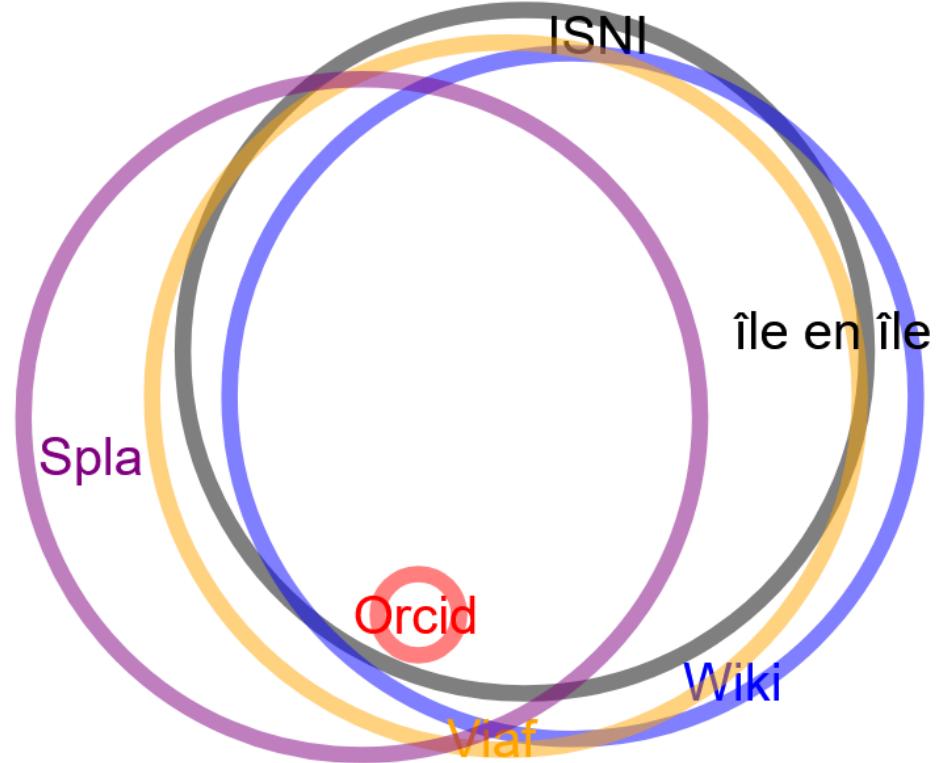


| A | B | C | | D |
|----|------------------|-------------|---|--|
| 1 | Nom_Prenom | source | lien | description du lien |
| 2 | Accilien_Selmy | lifranum_db | https://www.facebook.com/www.revedemuse.livre/ | Selmy Accilien poète |
| 3 | Alix_Renaud | ile_en_ile | http://ile-en-ile.org/alix-renaud-poemes/ | Ménage à trois |
| 4 | Alix_Renaud | ile_en_ile | http://www3.sympatico.ca/alix.renaud/ | Alix Renaud, ou le plaisir du texte |
| 5 | Alix_Renaud | ile_en_ile | https://youtu.be/j00ZarXgz5g | « Le Mot et la chose », poème de Lataignant |
| 6 | Alix_Renaud | ile_en_ile | https://youtu.be/C1P65X3rhZU | Des genoux au vertige |
| 7 | Alix_Renaud | ile_en_ile | http://www3.sympatico.ca/diamantbleu/accueil.html | À la recherche du Petit Prince |
| 8 | Alix_Renaud | ile_en_ile | http://www.litterature.org/recherche/ecrivains/renaud-alix-389/ | Présentation sur L'île |
| 9 | Alix_Renaud | ile_en_ile | http://www.litterature.org/recherche/ecrivains/renaud-alix-389/ | Présentation sur L'île |
| 10 | Alix_Renaud | ile_en_ile | http://www.potomitan.info/ayiti/renaud.php | Présentation sur Potomitan |
| 11 | Alix_Renaud | ile_en_ile | http://www.uefs.br/sitientibus/sitientibus_9/entrevista.pdf | Entrevista |
| 12 | Alix_Renaud | ile_en_ile | http://newsgroups.derkeiler.com/Archive/Soc/soc.culture.haiti/2005-11/msg00159.html | Le don de jouer avec les mots |
| 13 | Alix_Renaud | ile_en_ile | http://www.mondokarnaval.com/alix_chronique_2016.html | Les Chroniques d'Alix Renaud |
| 14 | Alix_Renaud | ile_en_ile | http://www3.sympatico.ca/alix.renaud/grac_esp.html | Ofrenda de Gracias |
| 15 | Alix_Renaud | ile_en_ile | http://fr.wiktionary.org/wiki/pompion | Pompion |
| 16 | Alix_Renaud | ile_en_ile | https://www.thecaribbeancurrent.com/quebec-awards-alix-renaud-for-excellence-in-arts-and-cult | Quebec Awards Alix Renaud for Excellence in Arts and Culture |
| 17 | Alix_Renaud | ile_en_ile | http://www.voir.ca/ | Voir |
| 18 | Alix_Renaud | ile_en_ile | https://voir.ca/livres/2000/02/10/alix-renaud-ovation/ | Ovation |
| 19 | Anderson_Dovilas | ile_en_ile | http://ile-en-ile.org/anderson-dovilas-pwezi-pefomans/ | Anderson Dovilas, Pwezi pèfòmans |
| 20 | Anderson_Dovilas | ile_en_ile | http://ile-en-ile.org/anderson-dovilas-la-nuit-nest-pas-le-premier-secret-des-baisers-bleus/ | La nuit n'est pas le premier secret des baisers bleus |
| 21 | Anderson_Dovilas | ile_en_ile | http://ile-en-ile.org/anderson-dovilas-trois-poemes/ | Trois poèmes |
| 22 | Anderson_Dovilas | ile_en_ile | http://dovanderson.over-blog.com/ | Le blog de Dovilas Anderson |
| 23 | Anderson_Dovilas | ile_en_ile | http://www.harmattan.fr/index.asp?Navig=catalogue&Obj=Livre&No=43605 | Mémoire d'Outre-Monde |
| 24 | Anderson_Dovilas | ile_en_ile | http://www.francopolis.net/revues/Dovilas-mai2012.html | Notes sur « Vingt poèmes pour traverser la nuit » |
| 25 | Anderson_Dovilas | ile_en_ile | http://spcreolophone.blogvie.com/ | Sosyete Powèt Kreyolofòn |
| 26 | Anderson_Dovilas | ile_en_ile | http://www.lemanoirdespoetes.fr/poemes-anderson-dovilas.php | « Je vole », « Je voyage » et « Point » |
| 27 | Anderson_Dovilas | ile_en_ile | http://www.paroleenarchipel.org/ | Parole en Archipel |
| 28 | Anderson_Dovilas | ile_en_ile | http://www.paroleenarchipel.com/article-vingt-poemes-pour-traverser-la-nuit-104458858.html | Vingt poèmes pour traverser la nuit |
| 29 | Anderson_Dovilas | ile_en_ile | http://www.paroleenarchipel.com/article-en-hollande-la-france-devient-franoise-104871758.htm | En Hollande, la France devient Françoise |
| 30 | Anderson_Dovilas | ile_en_ile | http://www.paroleenarchipel.com/article-les-iles-en-accent-aigu-105328874.html | Les îles en accent aigu |
| 31 | Anderson_Dovilas | ile_en_ile | http://www.paroleenarchipel.com/article-le-sang-de-l-oubli-le-sang-de-l-avenir-105976958.html | Le sang de l'oubli, le sang de l'avenir |
| 32 | Anderson_Dovilas | ile_en_ile | http://www.paroleenarchipel.com/article-liminasyon-parcours-d-un-voyageur-106118692.html | Liminasyon: Parcours d'un voyageur |
| 33 | Anderson_Dovilas | ile_en_ile | http://issuu.com/pallaikaroly/docs/vents_aliz_s_-_komansman?mode=window&backgroundColor=000000 | Trois poèmes |
| 34 | Anderson_Dovilas | ile_en_ile | http://www.potomitan.info/ayiti/dovilas.php | Pour en finir avec dérogation littéraire en Haïti |



IV. Exploitation des données

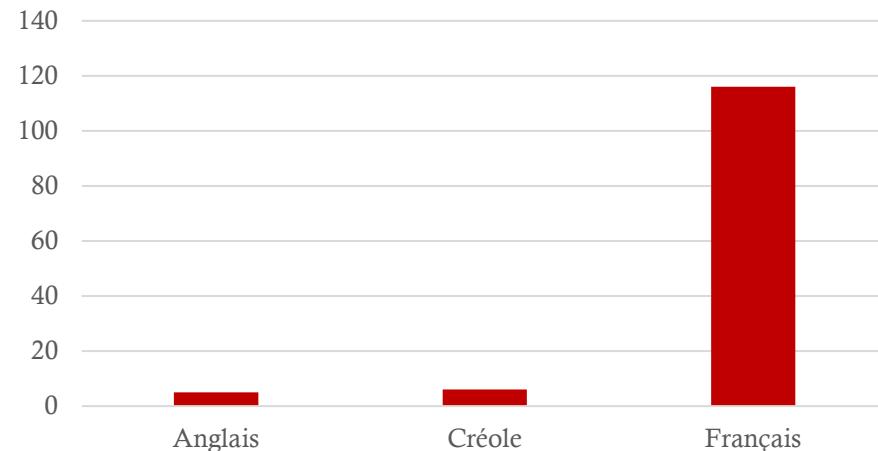
Le spectre couvert par les répertoires d'autorités



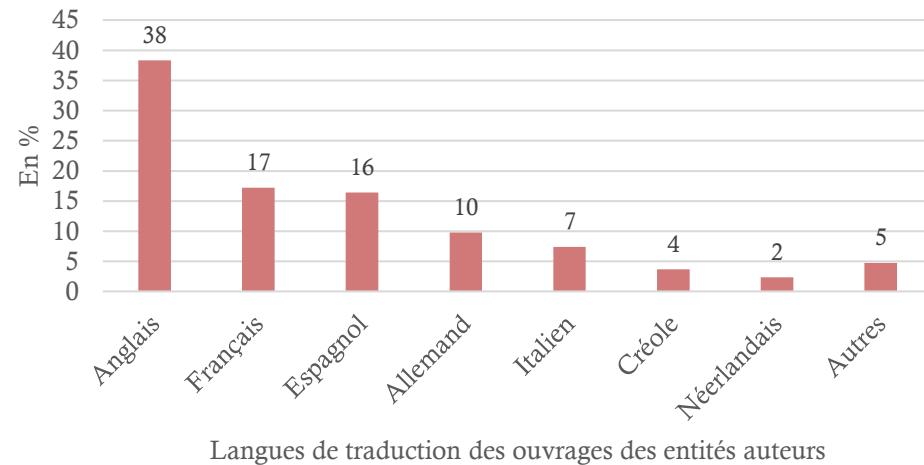


Haïti : une littérature francophone ?

Mentions de langues écrites, parlées ou signées indiquées dans les notices

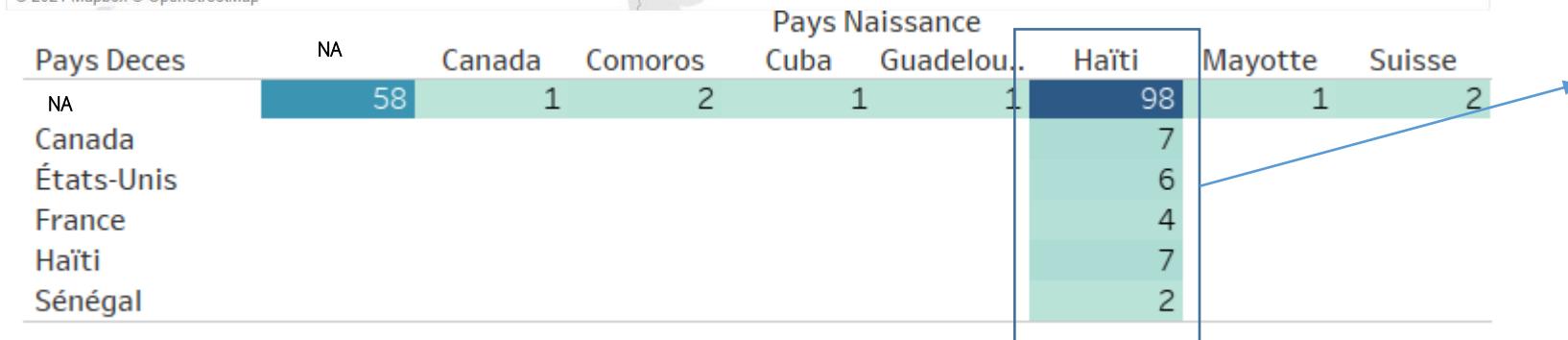


Langue de traduction des ouvrages des entités auteurs présents sur *île-en-île*

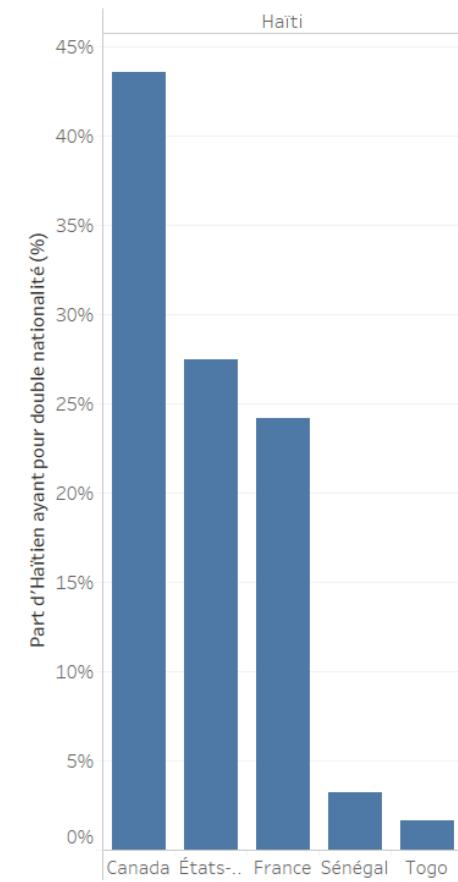


- Biais des répertoires d'autorités et/ou des articles Wikipédia qui ne renseignent que la langue d'écriture de la notice
- Synthèse bibliographique : omission d'ouvrages.

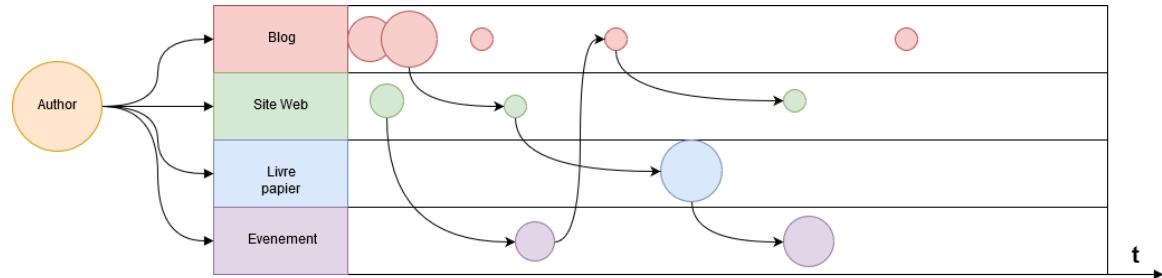
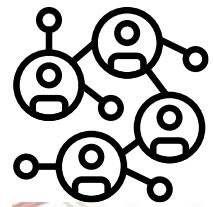
Haïti : un réseau international



Nationalités



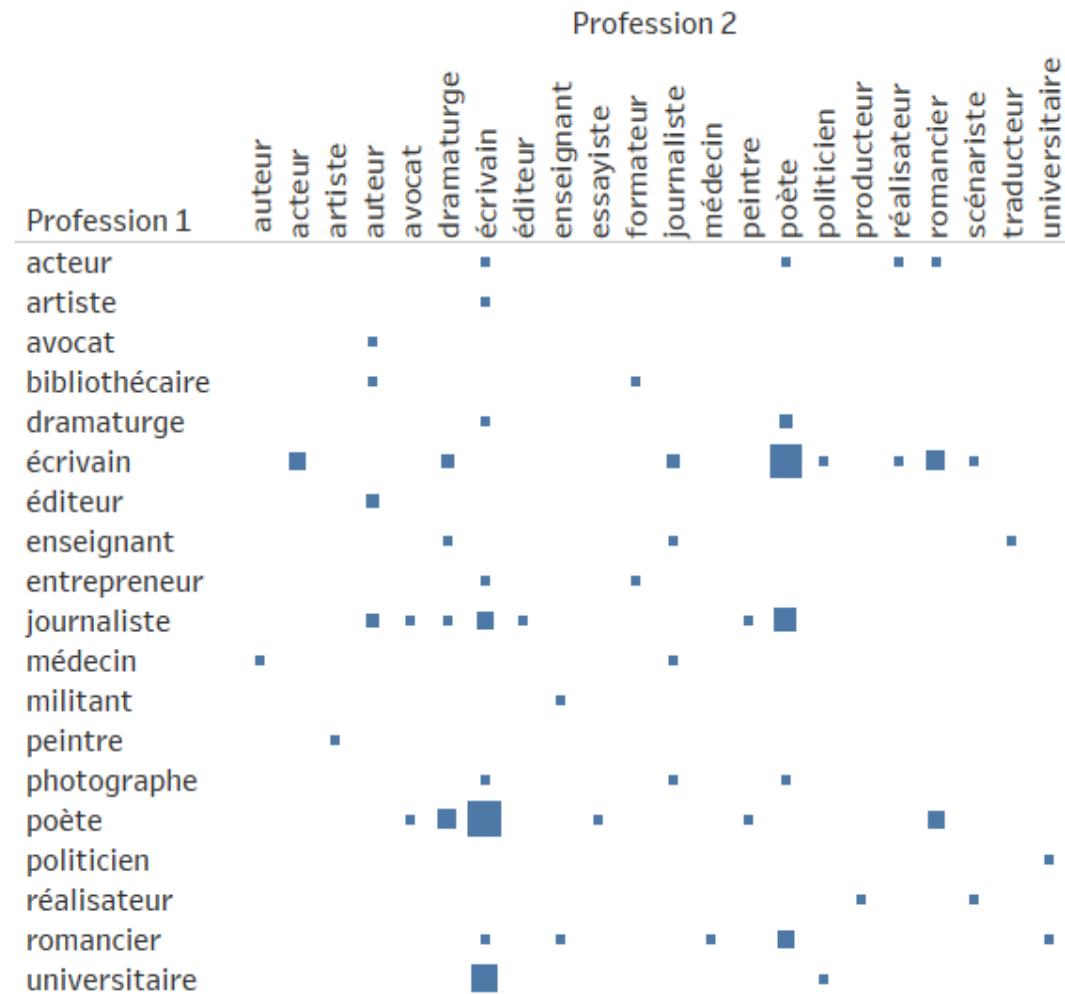
Des professions aux « parcours de vie »



- Un individu peut avoir exercé plusieurs professions au cours de sa vie.
- L'extraction ne nous permet pas d'adopter une approche diachronique des entités auteurs et ne peut qu'en rendre compte quantitativement.

Quelques faits intéressants toutefois :

- « auteur » n'apparaît qu'en seconde profession.
- « écrivain » et « poète » sont deux notions très liées.
- Les métiers de la presse, de l'art et de l'enseignement sont des professions pour lesquels l'activité littéraire est secondaire.

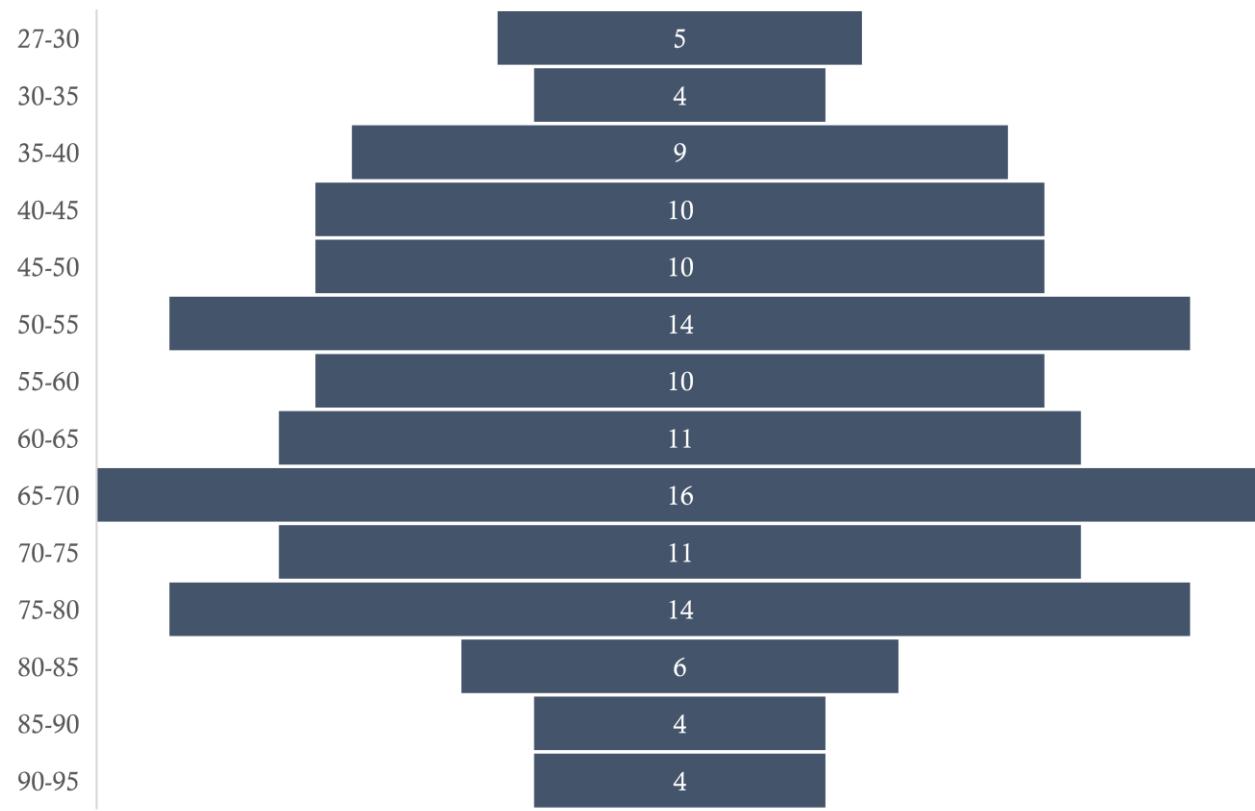


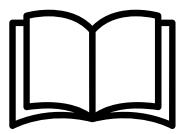
Un corpus d'acteurs « institutionnalisés »



128 dates de naissance
156 auteurs non décédés

Pyramide des âges des acteurs littéraires



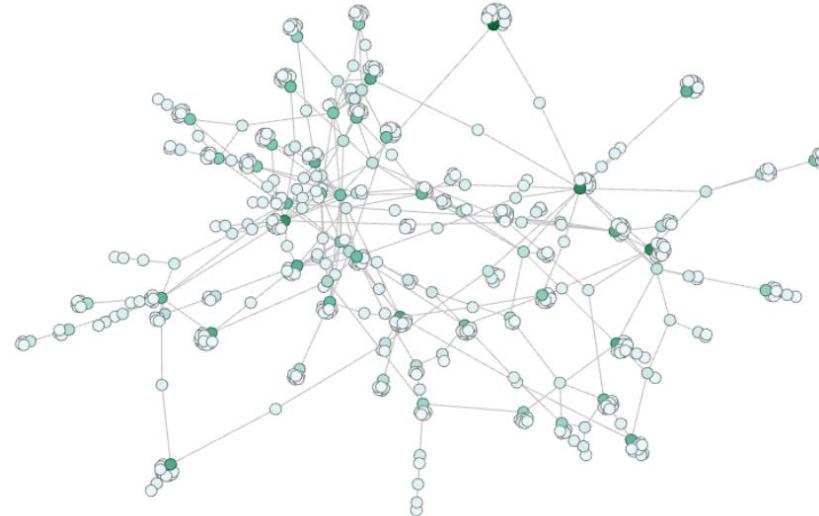


Vers l'identification de nouveaux auteurs : le co-autorat

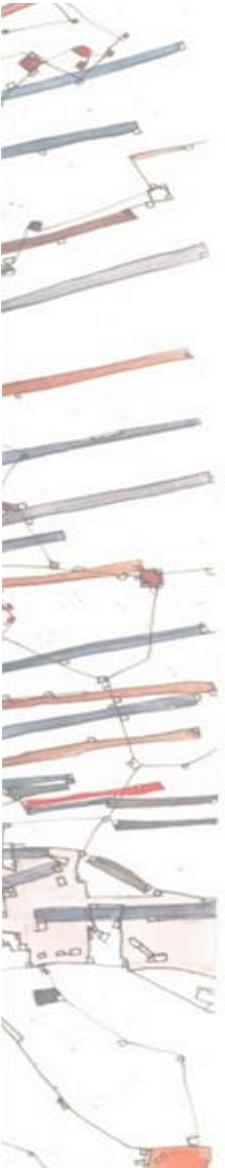
- Analyse textuelle des notices biographiques
- Analyses en réseaux des liens de collaborations identifiés au sein des notices biographiques et des répertoires d'autorités

Réseau des liens de collaborations obtenu sur Gephi

```
<listRelation>
  <relation ana="Tapia_Veronica" source="VIAF">Tapia Veronica</relation>
  <relation ana="Saint_Eloi_Rodney" source="VIAF">Saint-Éloi Rodney</relation>
  <relation ana="Zekina_Daniela" source="VIAF">Zekina Daniela</relation>
  <relation ana="Tiga" source="VIAF">Tiga</relation>
  <relation ana="Pean_Stanley" source="VIAF">Péan Stanley</relation>
  <relation ana="Cousineau_Normand" source="VIAF">Cousineau Normand</relation>
  <relation ana="Mata_Barreiro_Carmen" source="VIAF">Mata Barreiro Carmen</relation>
  <relation ana="Barroux" source="VIAF">Barroux</relation>
  <relation ana="Wiener_Jacqueline" source="VIAF">Wiener Jacqueline</relation>
  <relation ana="Ellis_Zilpha" source="VIAF">Ellis Zilpha</relation>
  <relation ana="Celie_Agnant" source="ile_en_ile_auto">Célie Agnant</relation>
  <relation ana="Alexis_D_Haiti" source="ile_en_ile_auto">Alexis d'Haïti</relation>
</listRelation>
```



- Relations de co-autorat et de liens de collaborations.
→ Près de 326 entités nommées ont été extraites automatiquement depuis île-en-île, 506 du VIAF.
- Sur ces 832 entités, on note 583 uniques soit 86% du corpus.
- 65 apparaissent à deux reprises (10%) et 18 à trois reprises (3%).



Les littéraires haïtiens et le Web

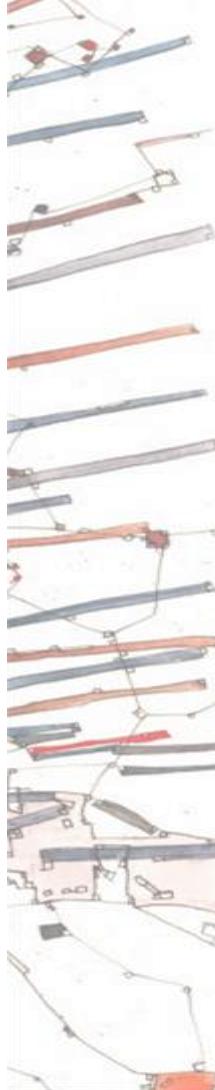


| Nom de domaine | Effectifs |
|-------------------|-----------|
| ile-en-ile.org | 360 |
| lenouvelliste.com | 154 |
| web.archive.org | 51 |
| youtube.com | 35 |
| potomitan.info | 30 |
| tanbou.com | 18 |
| csmagazine.org | 16 |
| lematinhaiti.com | 15 |
| hpnhaiti.com | 13 |
| alterpresse.org | 12 |

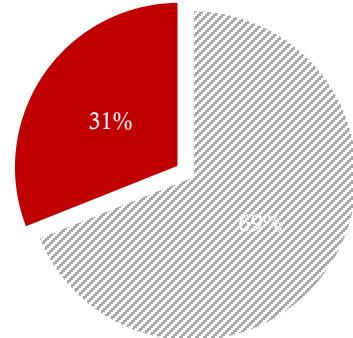
- La majorité des entités auteurs ne renvoient qu'à une à quatre adresses URLs (40 %).
 - 5 liens issus de la base de données RDF d'origine ;
 - 26 liens issus du Spla et 1258 d'*île-en-île*.
 - Les noms de domaines uniques et sites personnels représentent près de 73 % des liens.
- Noms de domaines récurrents
 - Les renvois vers *île-en-île* sont naturellement à écarter (360).
 - *Lenouvelliste* (154) est quant à lui le seul quotidien publié à Port-au-Prince.
 - Site d'archives en ligne (51) et la plateforme Youtube (35).
 - *potomitan* (30) et *tanbou* (18), des sites de promotion des cultures et langues créoles.



Pages personnelles issues de Spla - Un voyage dans le temps

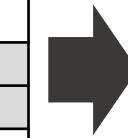


Sites référencés



Sur 9 références, 6 peuvent être observés :

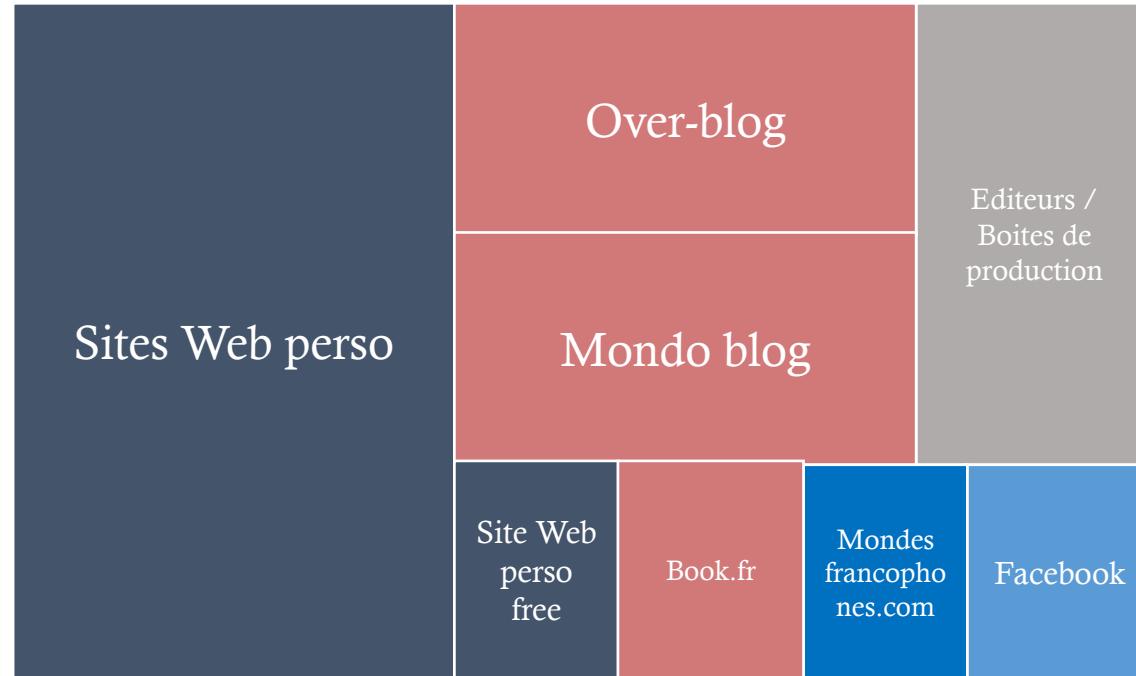
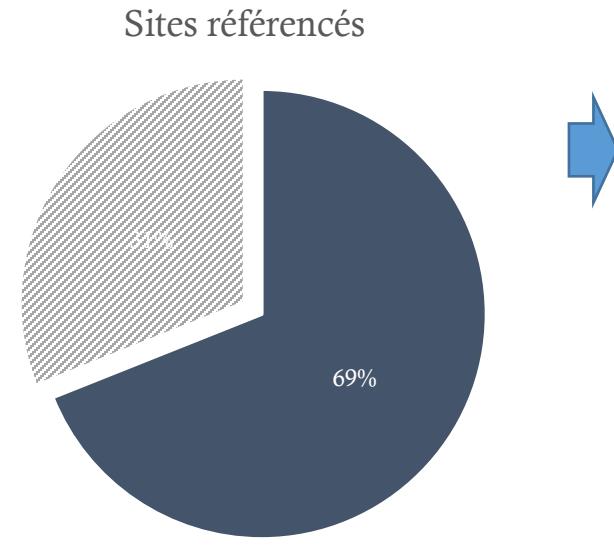
| Site Web | Période d'activité |
|---|--------------------|
| http://josaphat-robert-large | [2004;2014] |
| http://www.franketienne.com | [2005;2008] |
| http://www.elifleurant.com | [2013;2017] |
| http://www.margaretmitchellarmand.com/ | [2013;2018] |
| http://www.elsieaugustave.com/ | [2013;2017] |
| http://jmlominy.com/ | [2013;2017] |



INTERNET ARCHIVE
WayBack Machine

- Clôture des pages liée à l'âge ?
 - décès de certains *auteurs*
 - d'autres actuellement très âgés (plus de 70 ans)
 - mais aussi des jeunes qui ont cessé d'alimenter leur sites afin de se consacrer à d'autres activités.
- Pas seulement : les pages peuvent également disparaître pour des raisons techniques (obsolescences des plugins utilisés non-conformes avec une nouvelle mise à jour).
- Sites très descriptifs.
- « billets d'humeurs », mais ne diffusant de contenu littéraire inédit.

18 pages personnelles toujours en activité

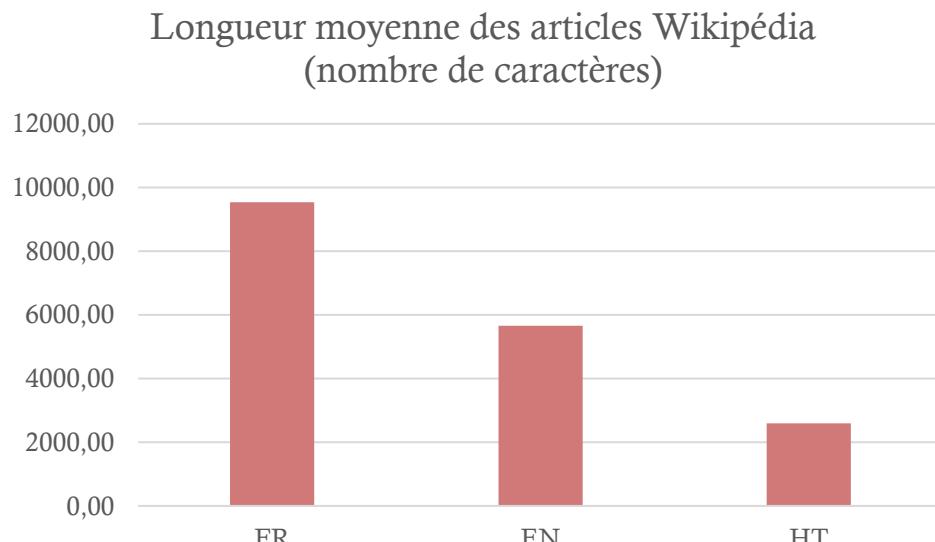
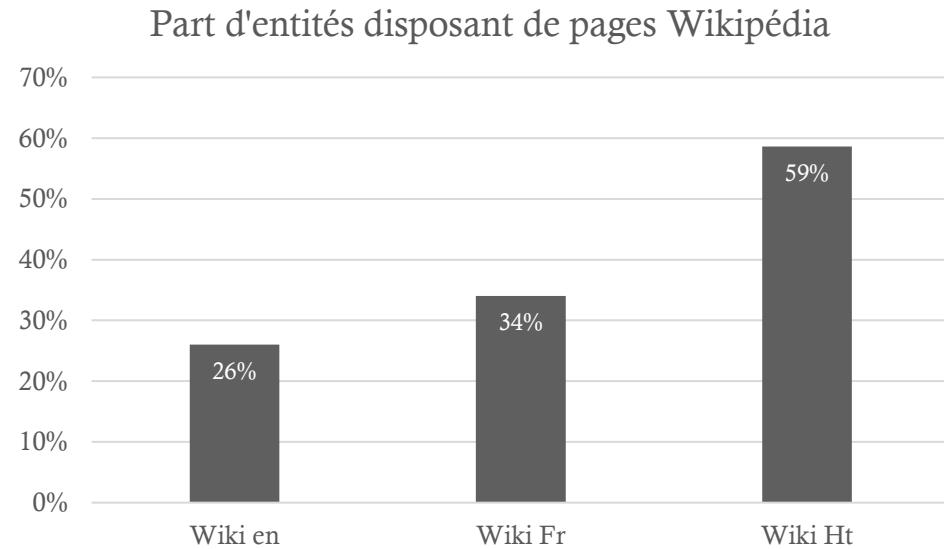


- Les *auteurs* ayant des sites renseignés ont entre 35 et 70 ans (soit un âge moyen de 54 ans)
- 3 sites été intégralement rédigés en anglais, 5 en français, 8 sont plurilingues (haïtien + au moins anglais et français)
- 8 ont un format blog, où il est possible « d'échanger » → peu d'échanges à travers ce biais
- Nombre égal de pages promotionnelles et de blogs littéraires
- La moitié sont des sites n'ayant pas été alimentés depuis au moins un an (2013-2107)

Des entités Wikidata aux pages Wikipédia



- 142 entités possèdent un identifiant Wikidata
- 124 au moins une page Wikipédia. Si ce n'est pas le cas, l'entité Wikidata a été identifiée dans une base de données LOD (*Linked Open Data*).



Wikipédia, un espace à géométrie variable

Des contributeurs différents par langue :

88% des articles haïtiens ont été créés par le même contributeur

→ l'administrateur de Wikipédia Haïti

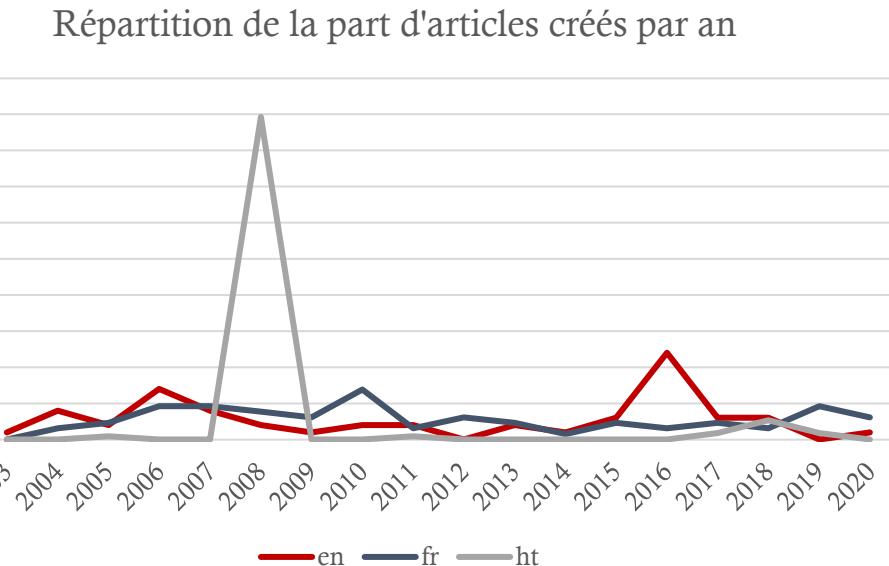
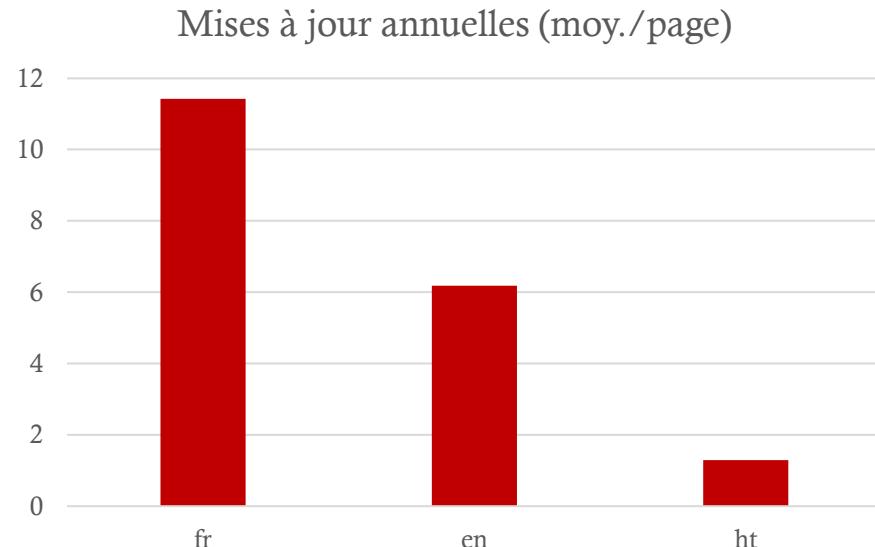
4 contributeurs ont écrit 30% des pages en français

→ visibilité de la francophonie : éditeur belges et québécois

→ souvent des amateurs d'art, de littérature et des universitaires

2 contributeurs sont à l'origine de 30% des pages en anglais :

→ contributeurs canadiens et étudiants haïtiens.

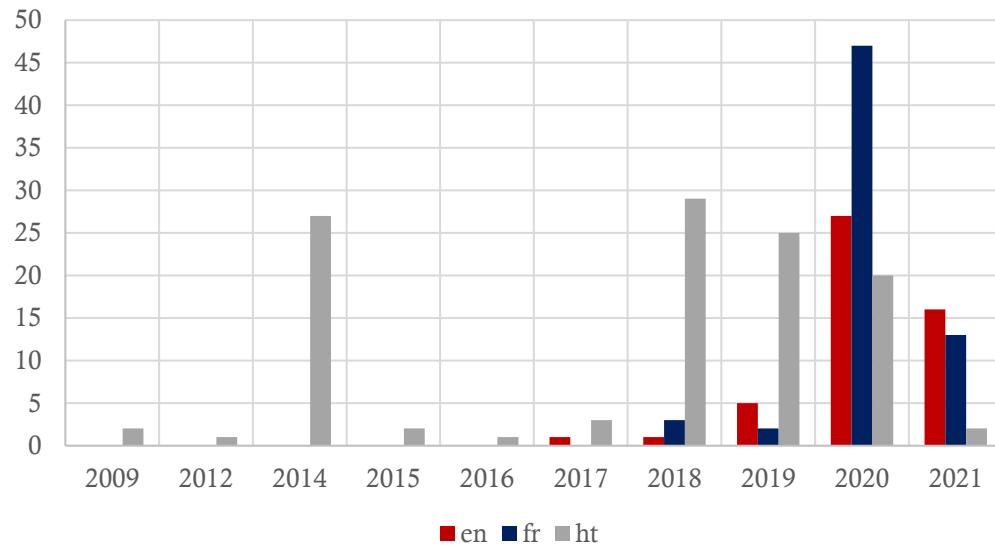




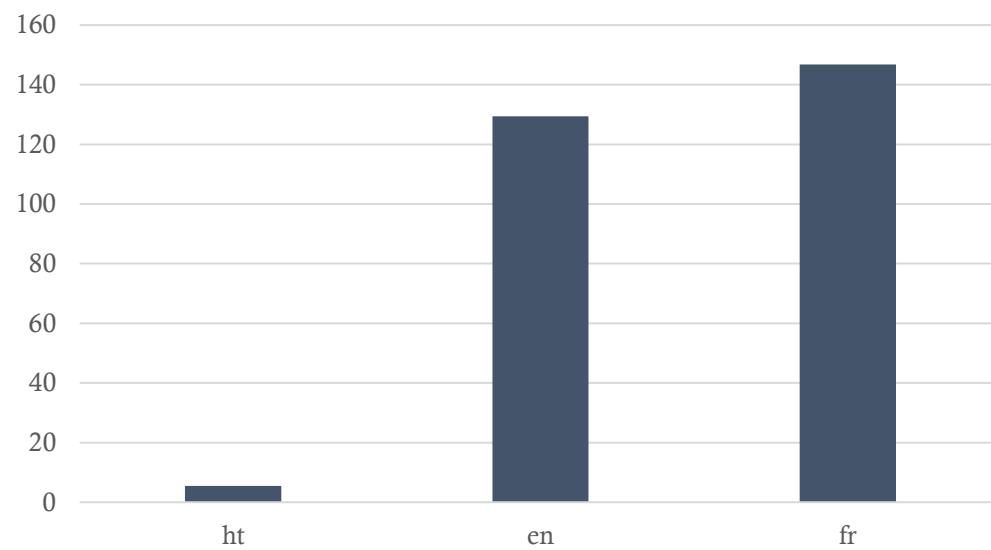
WIKIPEDIA

- Concernant le référencement dans Wikipédia :
 - Aucun article haïtien n'est pointé par d'autres articles
 - Le Wikipédia français réfère plus amplement les articles liés aux entités de notre corpus
- Vis-à-vis du nombre de vues et du nombre de mises à jour, les pages en français sont les plus dynamiques, suivies des pages anglaises, enfin haïtiennes.
- Les articles datent en moyenne d'il y a 9 à 10 ans, seulement 5% n'ont pas été modifiés depuis.

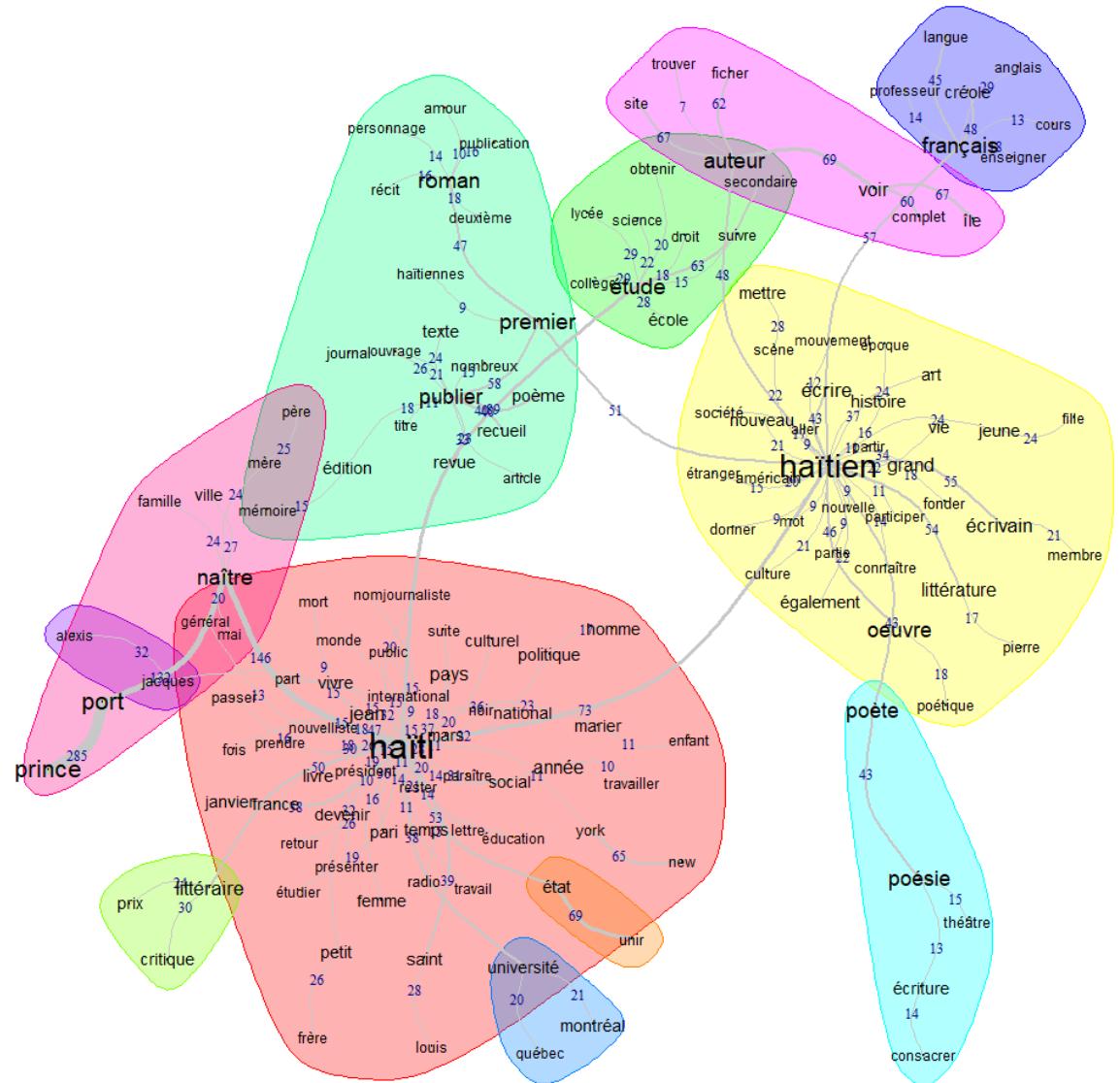
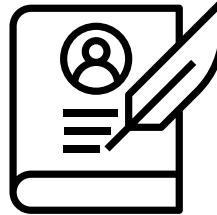
Dernières dates de mise à jours des articles par années et langues



Nombre de vues moyen lors du dernier mois des pages par langues



Les espaces biographiques du Spla et d'*île-en-île*



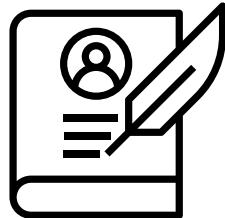
Analyse de similitude du corpus entier
réalisée sous IRaMuTeQ, fréquence ≥ 50
soit 150 formes actives, indice de
cooccurrence, halos communautaires

Une description des « parcours de vie » :

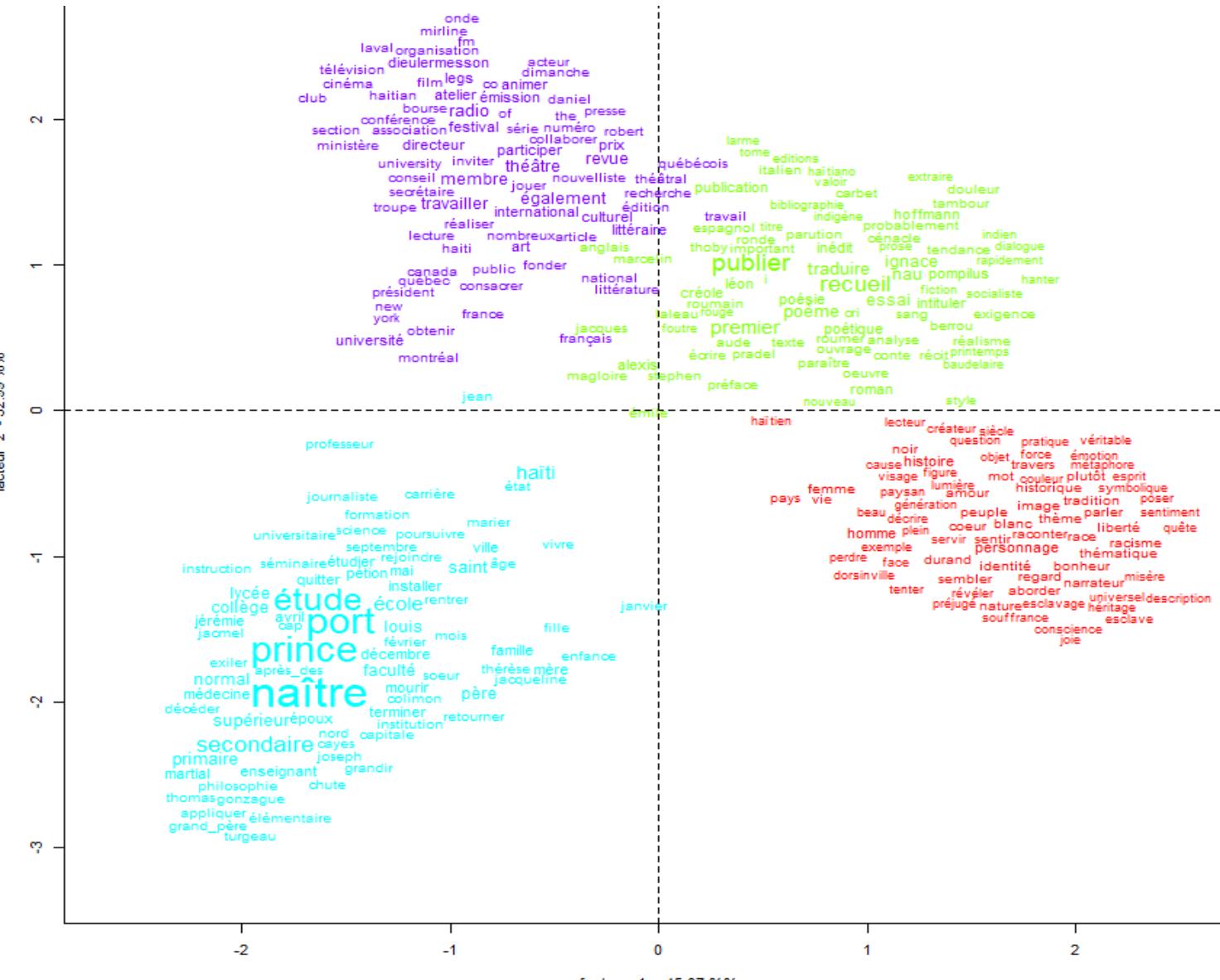
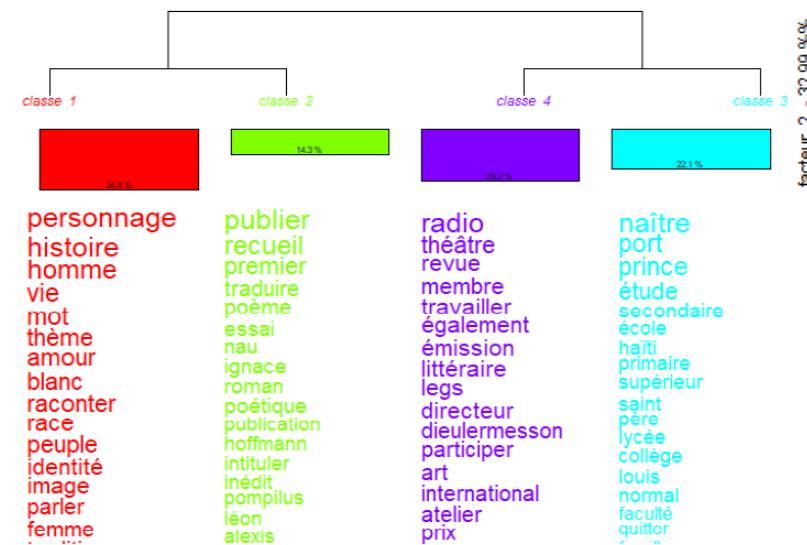
- Le contexte de naissance
 - Le parcours académique
 - Les conditions d'accès à la vie littéraire
 - Spécialisations



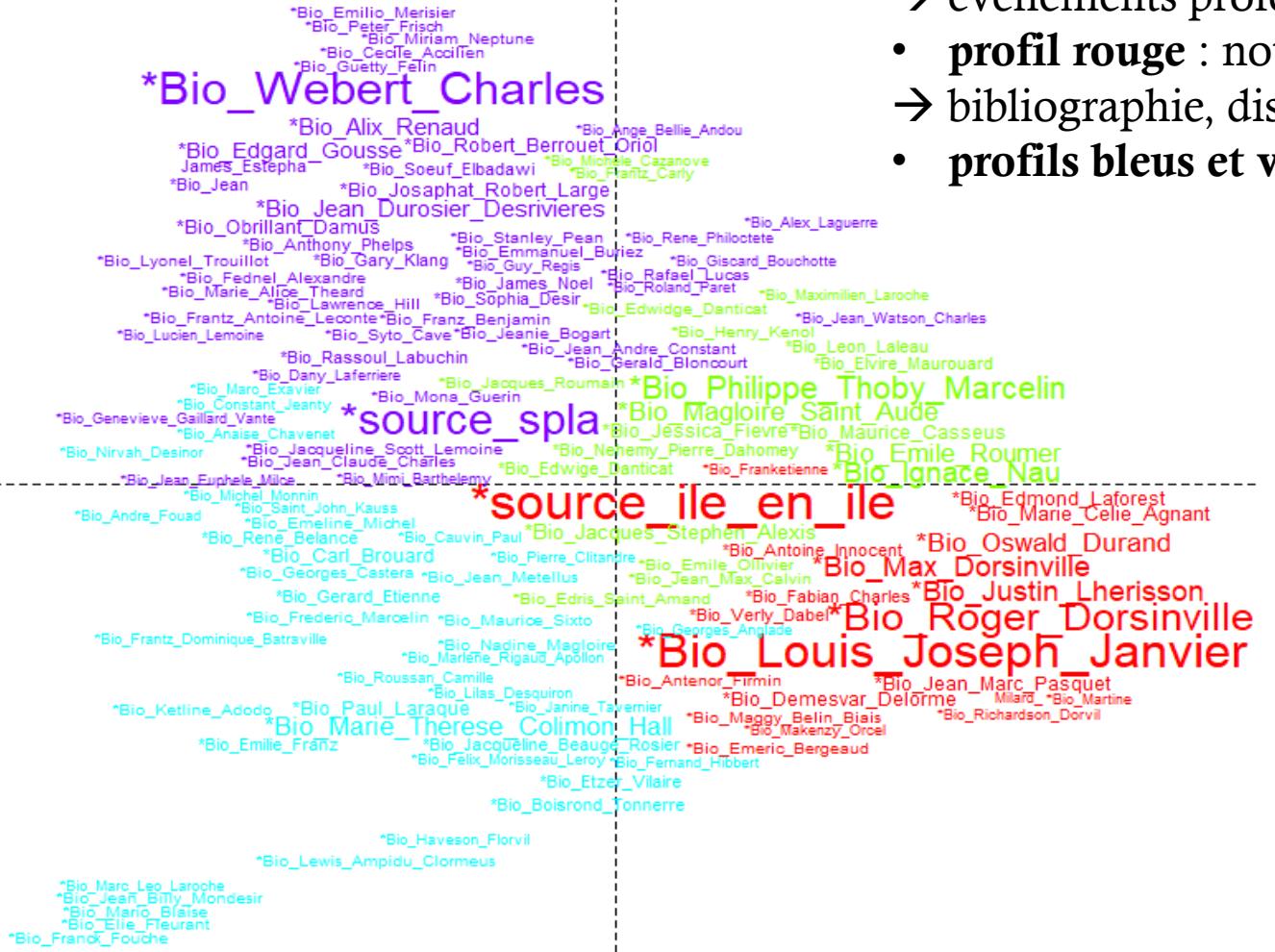
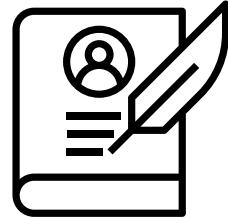
Analyse factorielle des notices



Classification de Max Reinert et AFC sur variables actives



Des profils variés



- **profil violet** : notices Spla, courtes.
→ événements professionnels mis en exergues.
- **profil rouge** : notices île-en-île, moyennes et longues.
→ bibliographie, distinctions, parcours académique, etc.
- **profils bleus et verts** se mélangent au sein des notices.



Mieux décrire les auteurs émergeants

- Une littérature en train de se faire, peu normalisée...
- Un « crawl » est un flot de données qui risque sans indexation de ne pas permettre d'analyses
- Une multitude de canaux, une multitude de discours...
- Une notice biographique traduit un intérêt par les pairs. Elle contient par ailleurs de nombreuses informations textuelles telles que des dates de naissance et de décès.
- Extraire des routines discursives ?

Machiner une variétés de textes... pour induire des informations

Quels peuvent être les apports du TAL?

- Repérage d'entités nommées : lieux, noms propres
- Exploration contextuelle : à partir de la recherche des dates d'état civil au sein des notices biographiques



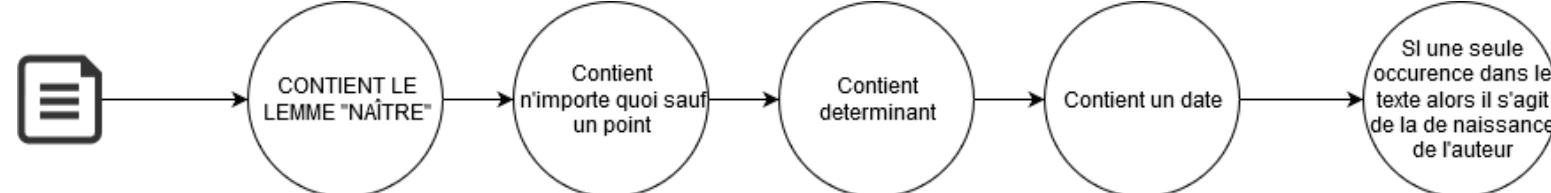
Obtenir des données biographiques...

Utiliser les structures récurrentes induites par le Web pour en extraire des informations :

Cléante Desgraves naît le 13 janvier 1891 à Port-au-Prince (Haïti) d'une mère américaine, née Alice Cunningham, et d'un père haïtien, Hector Desgraves, qui était pharmacien et également pianiste.

Bonel Auguste, né à Port-au-Prince, Haïti, le 6 mars 1973, est poète, bibliothécaire et animateur culturel. Il a étudié la linguistique à l'Université d'Etat d'Haïti et a reçu une formation en histoire de l'art à l'Institut français d'Haïti. Fondateur de l'atelier de poésie « Dimanche en poésie » à la bibliothèque Etoile filante, et animateur des « Vendredis littéraires » de l'université Caraïbe, il écrit en français et en créole.

...





Lier les biographies sur des entités plus vastes...

Utiliser des systèmes « entraînés » pour en détecter d'autres:

Bonel Auguste, né à Port-au-Prince, Haïti, le 6 mars 1973, est poète, bibliothécaire et animateur culturel. Il a étudié la linguistique à l'Université d'État d'Haïti et a reçu une formation en histoire de l'art à l'Institut français d'Haïti. Fondateur de l'atelier de poésie « Dimanche en poésie » à la bibliothèque Étoile filante, et animateur des « Vendredis littéraires » de l'université Caraïbe, il écrit en français et en créole.

Mais....

Né en Haïti en 1994, Jean D'Amérique est écrivain, auteur de Petite fleur du ghetto, recueil qui lui vaut une mention spéciale du Prix René Philoctète 2015 et une sélection au Prix Révélation de Poésie 2016 de la Société des Gens de Lettres.

Beaucoup d'ambiguités,
Demande d'être validé et
mis en contexte....

Un espace encore à explorer

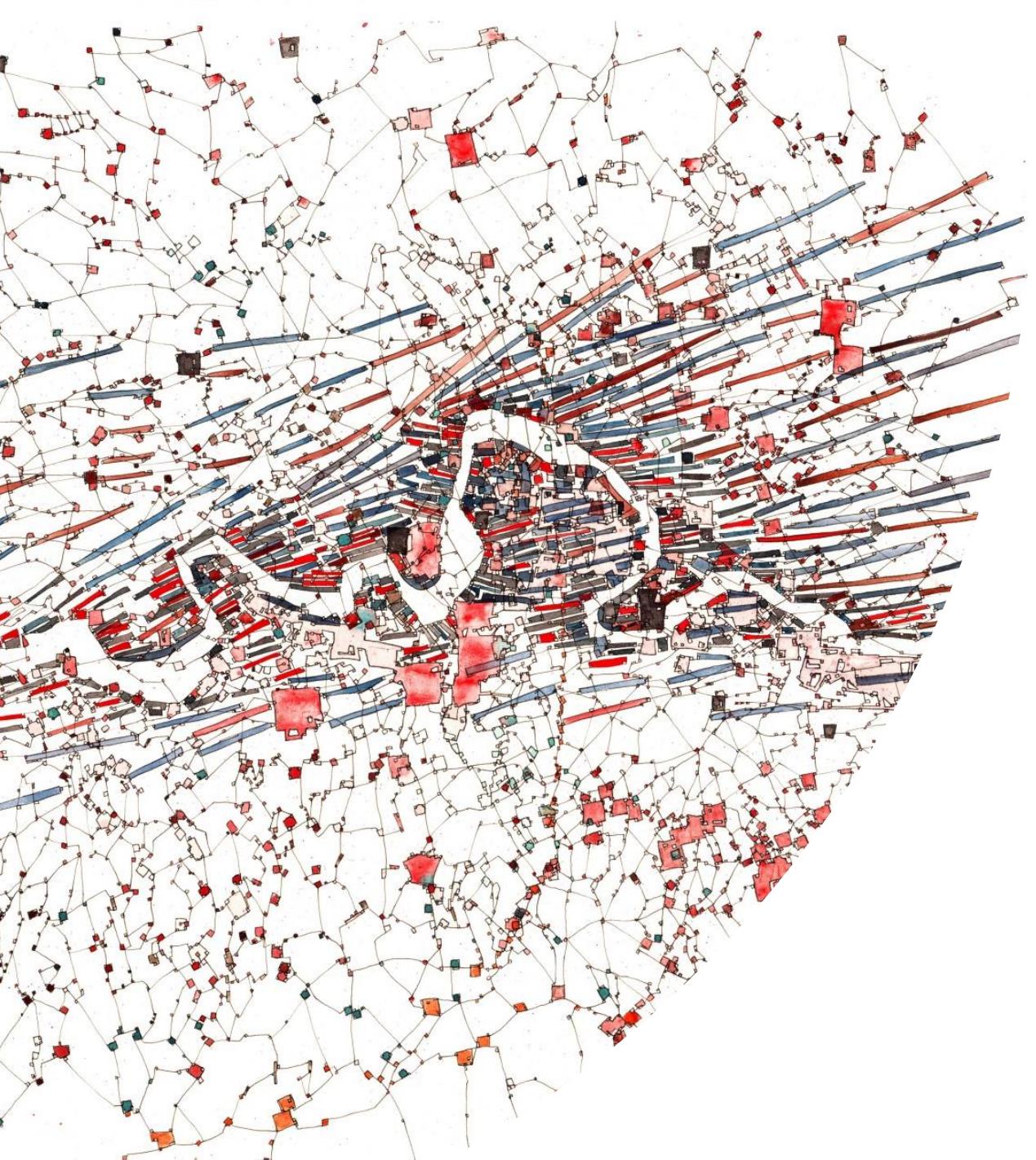


- La porte fermée de Facebook
- Twitter, un espace littéraire à 280 caractère ? [#covidpoème](#)
- Inkitt / Wattpad des plateformes littéraires sociales

C'est ici que notre travail s'arrête et que commence le projet LIFRANUM...



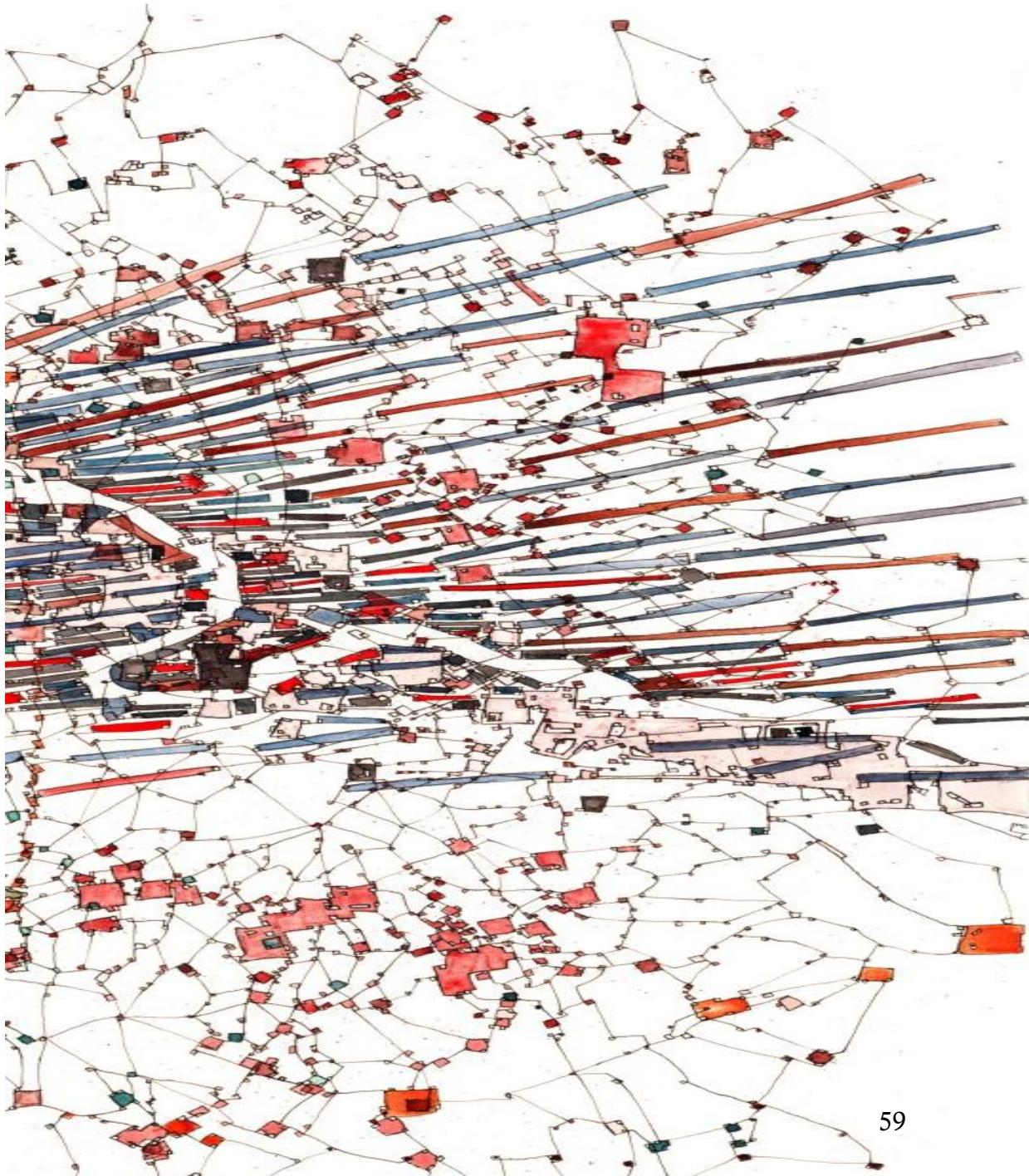
<https://projet-lifranum.univ-lyon3.fr/projet>

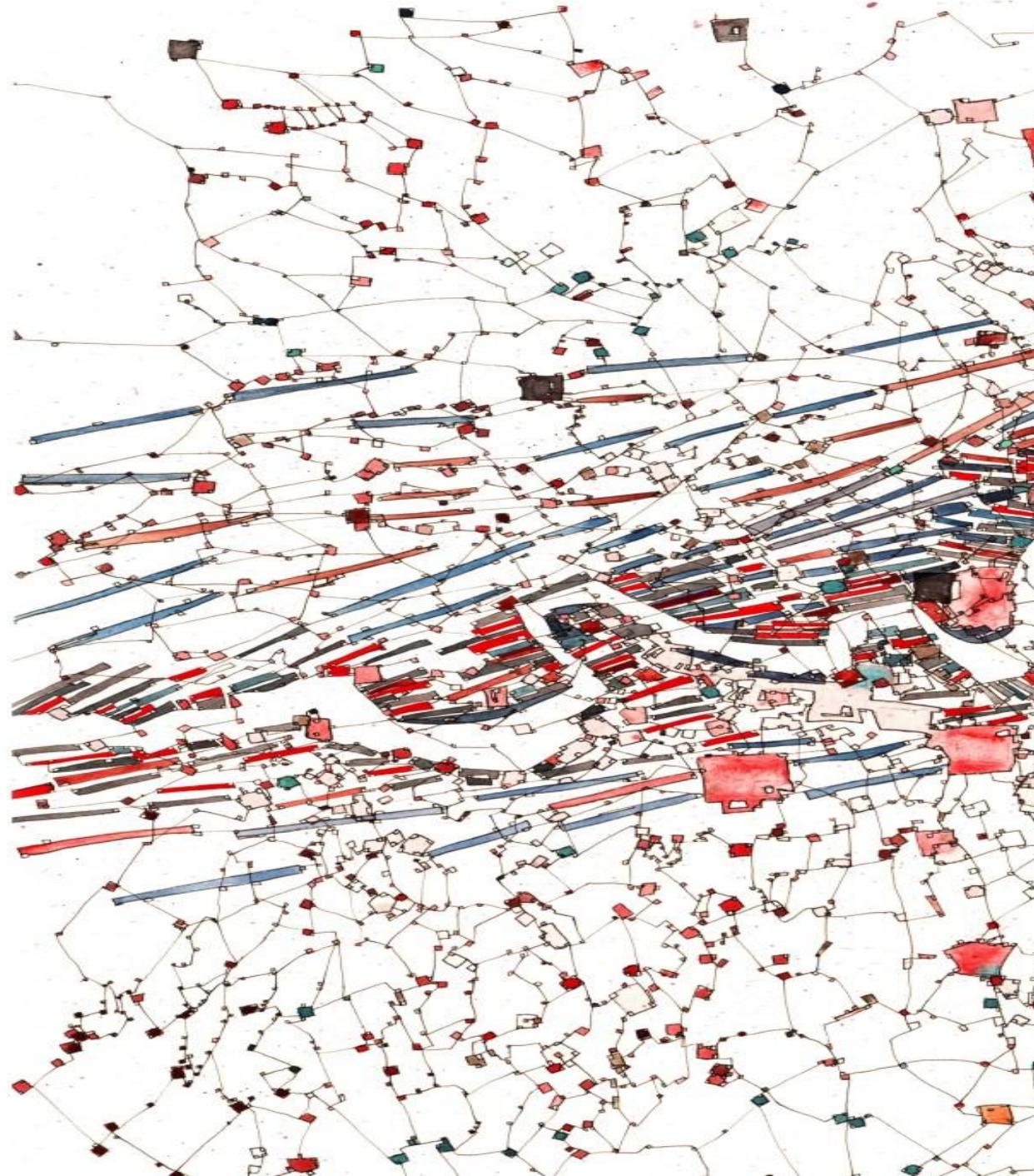


V. Conclusions

Le cas d'Haïti

- Repérage de certaines particularités du réseau : variétés linguistiques, problématiques d'émigration
- Les répertoires d'autorités permettent d'atteindre des individus reconnus par leurs pairs, qui ont souvent déjà publié un premier ouvrage par les canaux officiels.
- Les sites de référence regroupent quant à eux des individus aux profils variés mettant la littérature et les arts au service de la culture haïtienne.
- Les différents degrés de normalisation de ces deux types de sources s'est avérée complémentaire, et préférable, notamment avec un corpus plus restreint.
- Le processus de « Wikipédiaisation » de ces acteurs est a géométrie variable.



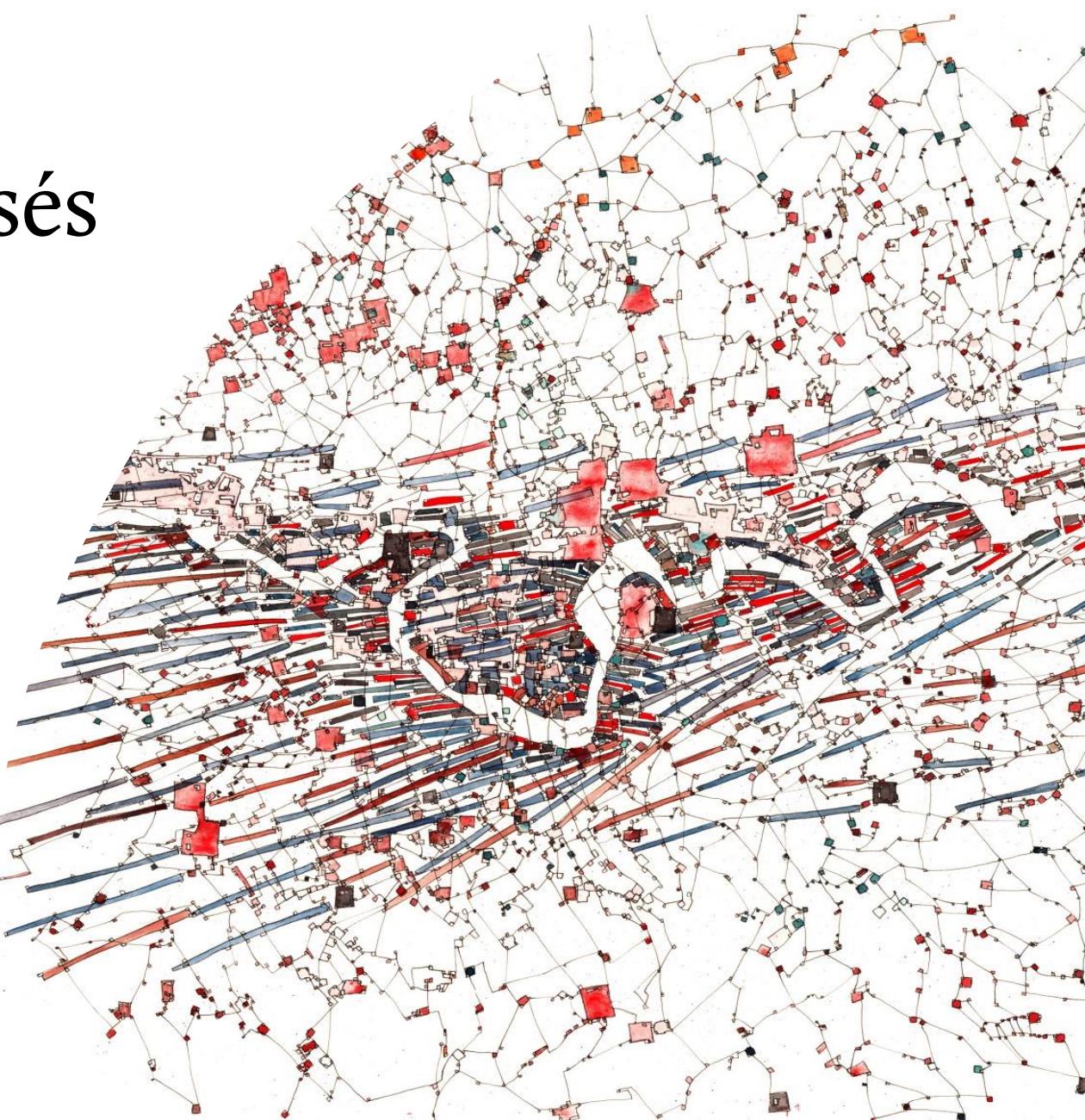


Plus largement

- Une méthodologie numérique adaptable, à une autre échelle, du local au global.
- Modèle flexible.
- Possibilité d'ajouter de nouvelles sources d'informations, notamment issues du Web social.
- Identifications de nouveaux auteurs grâce au processus d'enrichissement.
- Mais aussi, la nécessité de mobiliser un large spectre de sources hétérogènes afin de capter le plus d'informations possible.
- Repérage de certains biais induits par ces sources

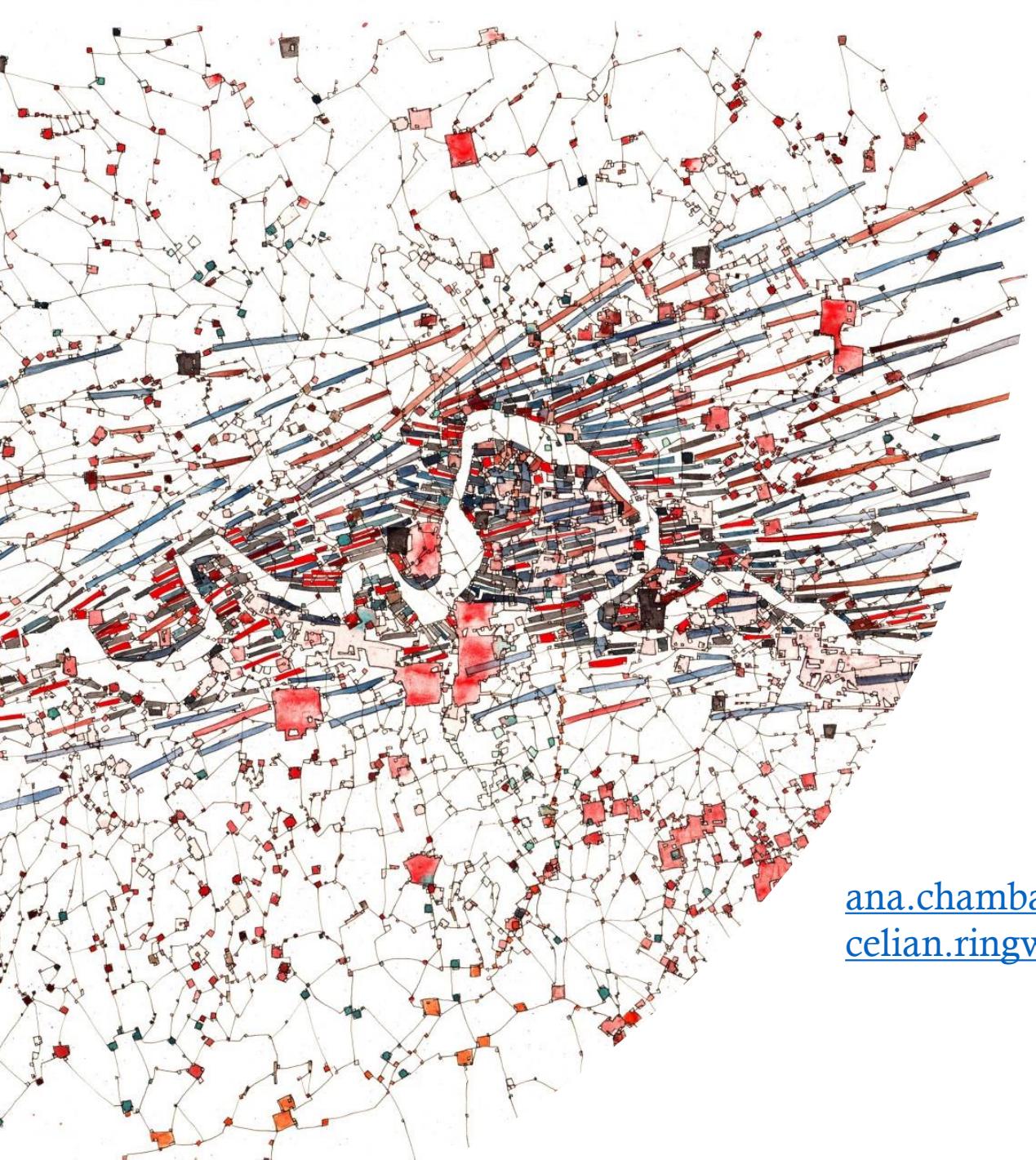
Outils numériques mobilisés

- OpenRefine : <https://openrefine.org/>
- TXM : <http://textometrie.ens-lyon.fr/>
- IRaMuTeQ : <http://www.iramuteq.org/>
- Gephi : <https://gephi.org/>
- Oxygen XML Editor :
<https://www.oxygenxml.com>
- Python : <https://www.python.org/>
 - RDFlib,
 - LXML,
 - BeautifulSoup,
 - Scikit-learn,
 - Spacy, etc.
- Venn JS : benfred.github.io/venn.js



Bibliographie

- BARTHES Roland, « The death and return of the author », Edinburgh : Edinburgh university press , 1998
- BEN HAZZ Slim, DECLES Jean Pierre, MINEL Jean-Luc, « Modèle d'exploration contextuelle pour l'analyse sémantique des textes », *HAL*, 2006
- BÉLISLE Claire, « Du papier à l'écran : lire se transforme », *Lire dans un monde numérique*, Villeurbanne, 2011, pp. 112-162.
- BONNET Gilles, « L'hypertexte. Poétique de la relecture dans l'œuvre numérique de François Bon », in *Poétique*, n° 175, Paris, 2014, pp. 21-34.
- -, « L'autoblographie. Écritures numériques de soi », in *Poétique*, n° 177, Paris, 2015, pp. 131-143.
- BUREAU Bruno, « Méthodologies de construction de corpus », Université Lyon 3, Cours Humanité Numériques, 2019
- COMPAGNON Antoine, Qu'est-ce qu'un auteur ?, Cours Université de Paris IV-Sorbonne.
- CASATI Roberto, « Contre le colonialisme numérique, manifeste pour continuer à lire », Albin Michel 2013
- LEJEUNE Philippe, *Signes de vie. Le pacte autobiographique*, 2, Paris, 2005.
- MERCKLÉ Pierre, *Sociologie des réseaux sociaux*, Paris : La Découverte, 2004, 2011, 2016.
- MONAT Jean-Baptiste, *A la recherche du Web littéraire*, mémoire de master 2 Humanités Numériques soutenu à l'Université Lumière Lyon 2 sous la direction de Jérôme DARMONT et de Gilles BONNET, 2019, 93 p.
- GEFEN Alexandre, « On ne sait jamais rien du sort d'un livre », Théorie et pratique des best-sellers, Ghent University & ENS, 2017.
- THOMAS Sue *et al.*, « Transliteracy : Crossing Divides », *First Monday*, vol. 12, n° 12, décembre 2007,



Merci de votre
attention, place
à l'échange !

ana.chambat@gmail.com
celian.ringwald1@univ-lyon2.fr