

Une infrastructure numérique pour étudier les anciens manuscrits scientifiques : un cas d'utilisation appliqué aux manuscrits de Saussure



UNIVERSITÉ DE GENÈVE
FACULTÉ DES LETTRES

Luka Nerima, Gilles Falquet, Claire Forel, Giuseppe Cosenza, Sahar Aljalbout, Massimo Brero, Jérémie Morel, Maria Andueza Rodriguez
Département de linguistique - Unité d'informatique & CUI, Université de Genève
{luka.nerima, gilles.falquet, claire.forel, giuseppe.cosenza, sahar.aljalbout}@unige.ch



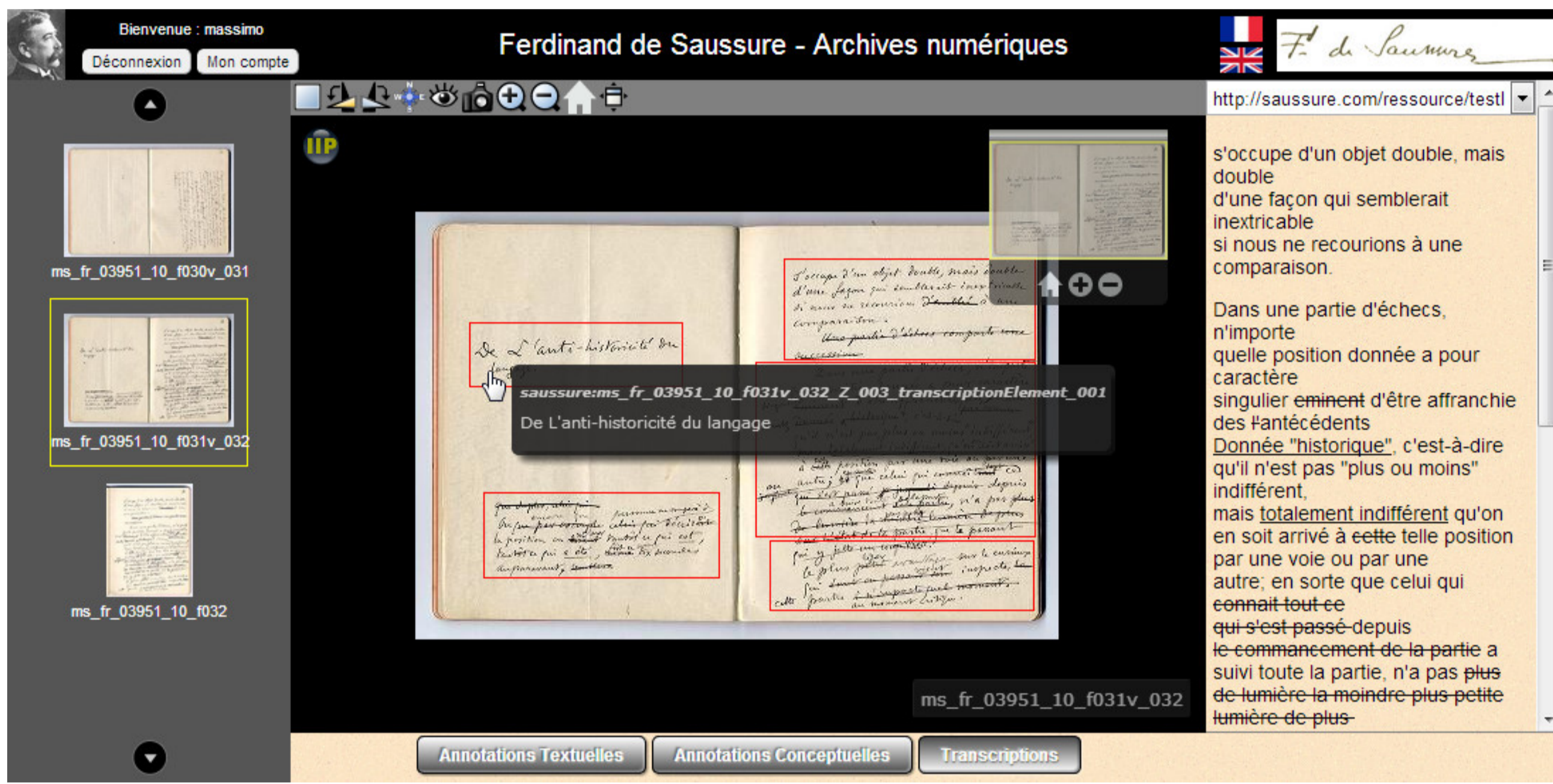
Contexte

- Ferdinand de Saussure, scientifique genevois (1857 - 1913)
- Considéré comme le fondateur de la linguistique moderne
 - N'a jamais publié la théorie développée dans son *Cours de linguistique générale* (CLG)
 - CLG édité en 1916 par Charles Bally et Albert Sechehaye sur la base des notes de cours d'étudiants
 - Fonds de Saussure: ~50'000 feuillets manuscrits déposés dans les Bibliothèques de Genève (BGE), Paris et Harvard
 - Images du Fonds de Saussure de la BGE numérisées depuis janvier 2017

Enjeux

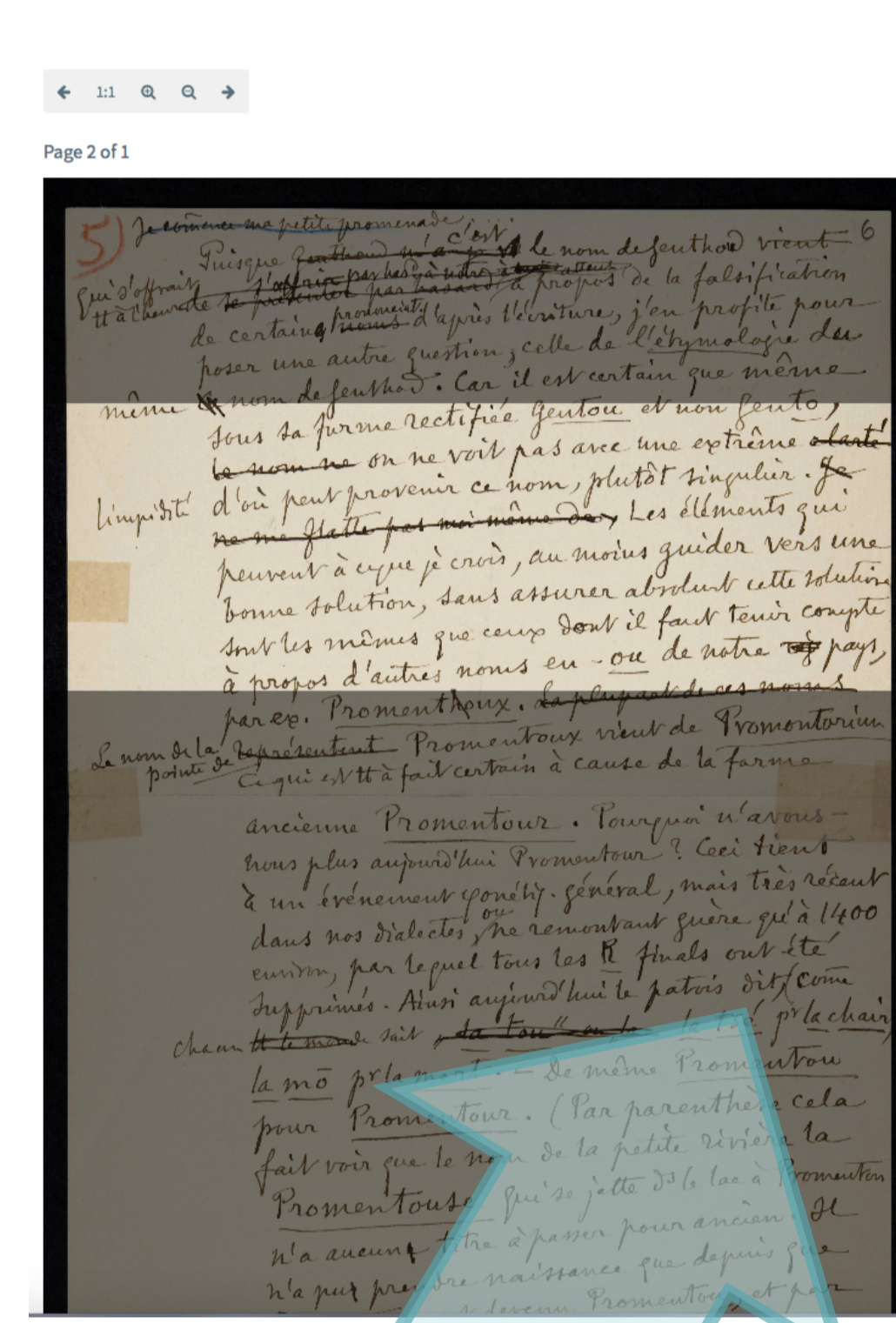
Le travail des scientifiques sur les manuscrits c'est :

- **Retrouver et accéder** - visualiser, classer par thématiques
- **Comprendre** - composer avec des terminologies instables, évolutives
- **Dater** - attribuer une date, retrouver l'ordre chronologique des manuscrits
- **Publier** - révéler, diffuser le travail de l'auteur, sous différentes présentations



pybossa Community Projects Create About Sign In

Transcription manuscrits Ferdinand de Saussure: Contribute



Crowdsourcing

- 1^{ère} campagne sur les noms de lieu de la région genevoise et sur les patois romands
- ~600 manuscrits - 3'000 tâches

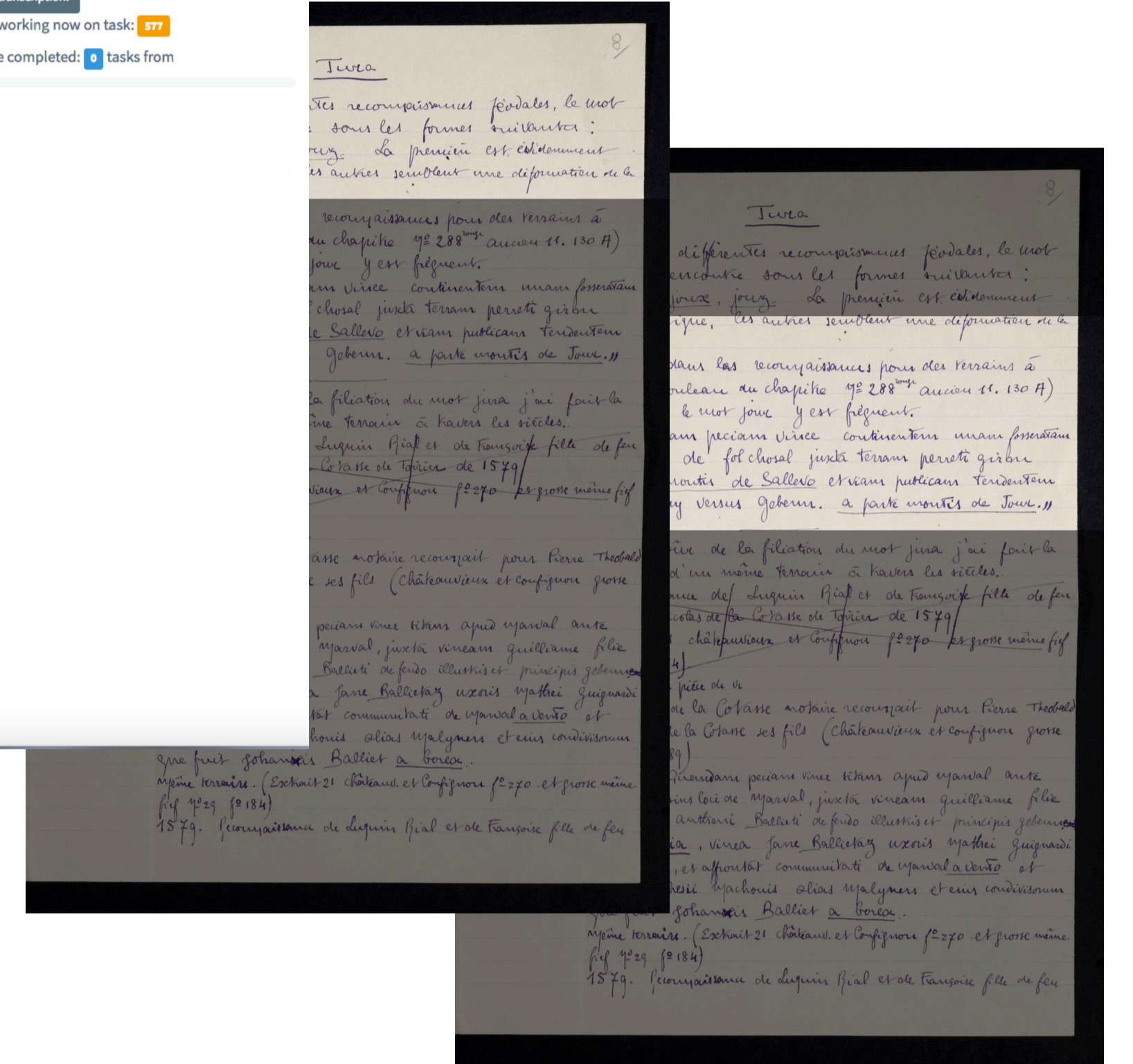
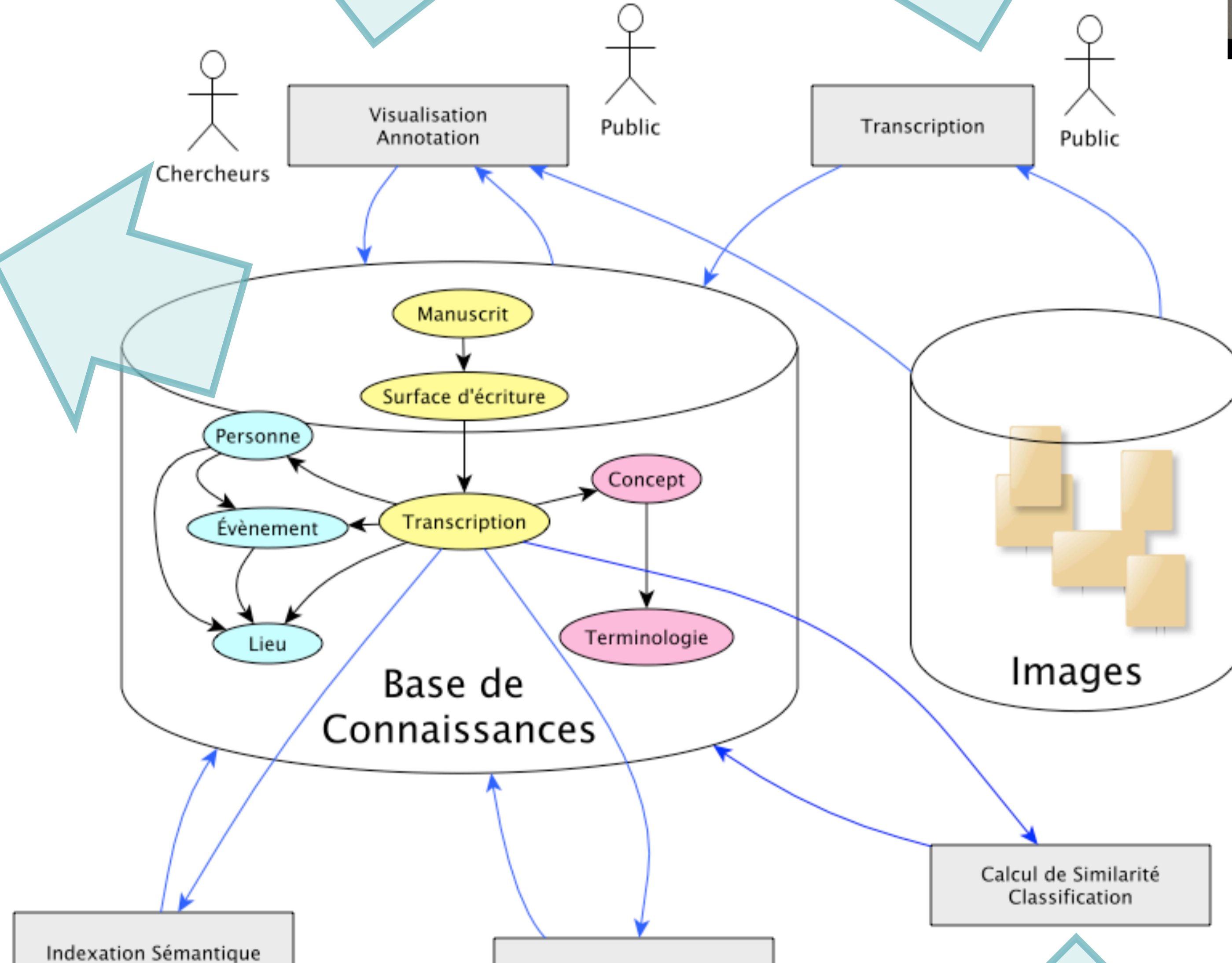
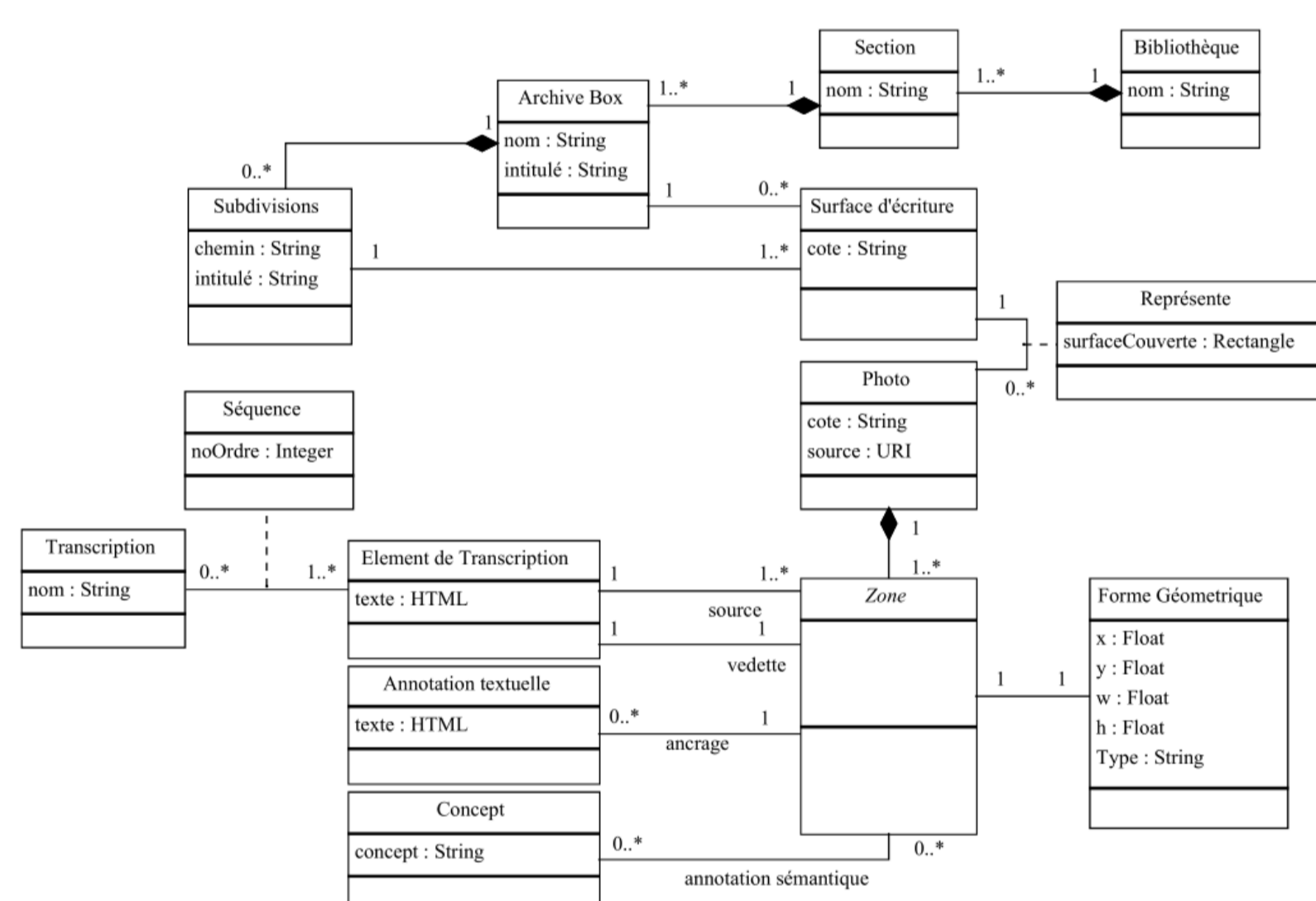


Schéma conceptuel des manuscrits



Stockage RDF - notation RDF Turtle

```
s:cours_univ_phasArchiveBox s:cours_univ_432 ; s:cours_univ_434 ; s:cours_univ_433 ;
s:cours_univ_435 ; s:cours_univ_761 ; s:cours_univ_846 ; s:cours_univ_824 ;
s:cours_univ_762 ;
a s:Section .

s:cours_univ_761 a s:ArchiveBox ;
p:hasSubdivisions s:cours_univ_761_04 ; s:cours_univ_761_03 ;
s:cours_univ_761_02 ; s:cours_univ_761_01 .

s:cours_univ_761_02 a s:Subdivisions ;
p:hasSurfaceEcriture s:cours_univ_761_02_p086 ;
s:cours_univ_761_02_relo06 ; s:cours_univ_761_02_relo03 ;
s:cours_univ_761_02_relo05 ; s:cours_univ_761_02_p097 ; s:cours_univ_761_02_relo01 ;
s:cours_univ_761_02_p098 ; s:cours_univ_761_02_p094 .

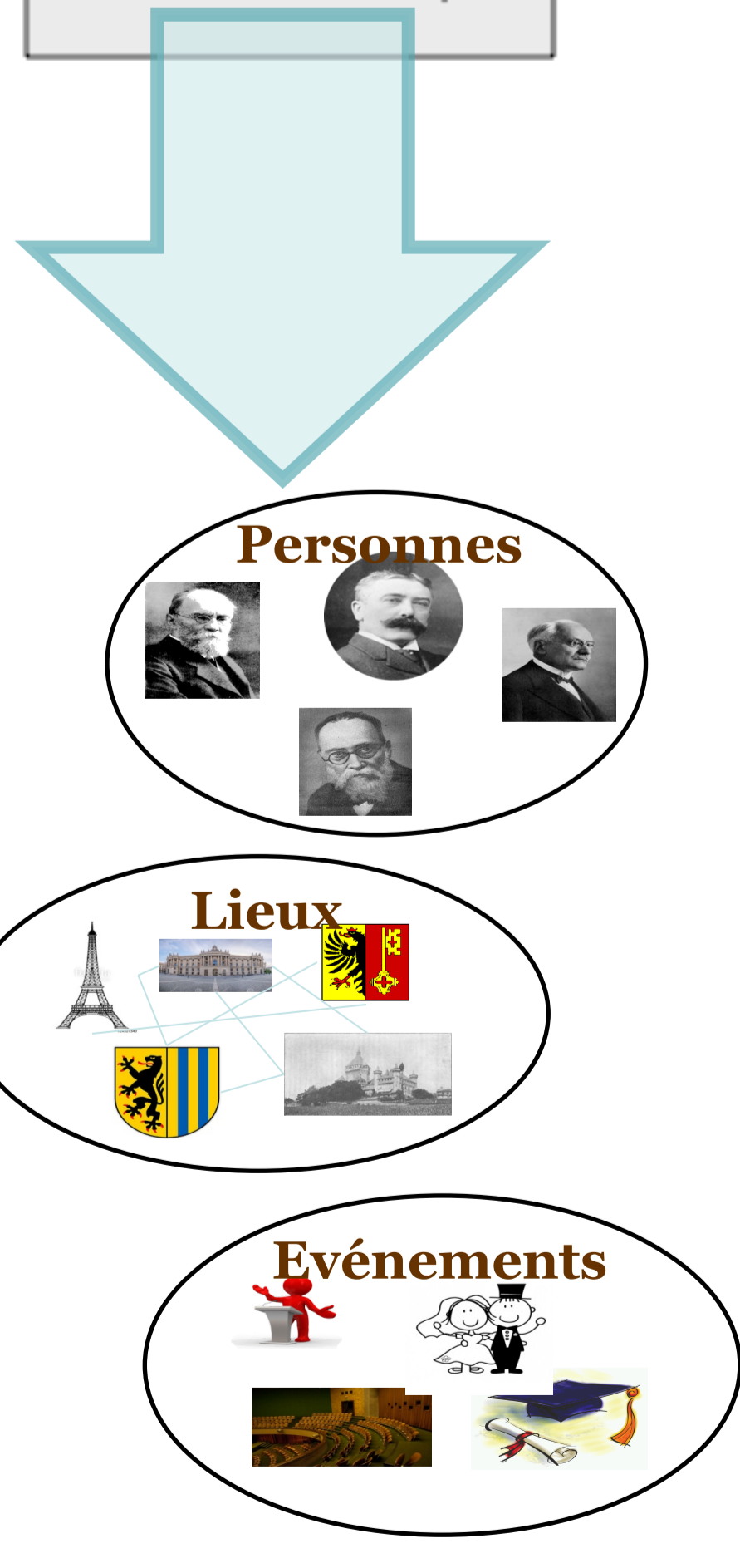
s:cours_univ_761_02_p096 a s:Surface_d_écriture ;
p:hasCote "cours_univ_761_02_p096" ;
p:hasPhoto s:cours_univ_761_02_p096-DOT-jp2 ;
p:in stestGroup .

s:cours_univ_761_02_p096-DOT-jp2 a s:Photo ;
p:hasCote "cours_univ_761_02_p096" ;
p:content _node1bm0v7ahix2376 ; _node1bm0v7ahix5106 ;
p:hasFilename "cours_univ_761_02_p096.jp2" ;
p:hasSource "http://saussure.unige.ch/bge/manuscript/
cours_univ_761_02_p096.jp2" .

_node1bm0v7ahix1011 a s:Zone ;
p:hasForme _node1bm0v7ahix1012 ;
p:hasPermission _node1bm0v7ahix1013 ;
p:hasTranscriptionElement _node1bm0v7ahix1014 .

_node1bm0v7ahix1014 a s:TranscriptionElement ;
p:hasPermission _node1bm0v7ahix1015 ;
p:hasCreationDate "2017-08-11 11:53:45" ;
p:hasCreatorName "gfalquet" ;
p:hasEditionDate "2017-08-11 11:53:45" ;
p:hasEditorName "gfalquet" ;
p:rawText "L'analogie a pour premier effet: de sauver les éléments de formes en
les reprenant toujours pour les transformations nouvelles." ;
p:texte "L'analogie a pour premier effet: de sauver les éléments de formes en
les reprenant toujours pour les transformations nouvelles." .

Notation RDF Turtle
```



Collocation candidate	Type colloc	#occ	score
sujet parlant	Nom-Adj	28	254.3
étymologie populaire	Nom-Adj	20	178.6
phénomène phonétique	Nom-Adj	28	110.2
point de vue	Nom-prep-Nom	16	97.5
moindre effort	Adj-Nom	8	91.4
création analogique	Nom-Adj	11	74.8
cavité buccal	Nom-Adj	7	67.8
formation analogique	Nom-Adj	11	63.7
vieux allemand	Adj-Nom	6	55.5
ère chrétien	Nom-Adj	4	55.5
physiologie phonologique	Nom-Adj	6	54
degré de fermeture	Nom-prep-Nom	6	44.7
syllabe final	Nom-Adj	6	38.8
allemand moderne	Nom-Adj	4	37.8
lien grammatical	Nom-Adj	5	35.6
unité inféneur	Nom-Adj	5	32.5
hypothèse de continuité	Nom-prep-Nom	3	31.8
conscience de sujet	Nom-prep-Nom	6	30.6
avoir importance	Verbe-obj	3	29.7

Matrice de similarité des manuscrits (similarité si score > 0.033)

