

Présents

Anaïs Chambat
Jorge Fins
Benjamin Galliot
Vincent Paillusson
Axel Renon
Christophe Rey
Nathalie Rousseau

Ordre du jour

- rappel des partenariats mis en place (réseau METALEX et CLUNL)
- présentation des enjeux de l'atelier : proposer un recensement des projets d'édition numérique de documents lexicographiques (achevés et en cours) sous la forme d'une interface Scenari. Une mutualisation avec l'atelier « textes antiques » est envisagée.

Discussion autour des premiers critères proposés

1. Acquisition

Numérisation (image interrogeable ou non)
Édition papier pré-existante
Corpus oraux
Images

2. Production

Indexation (lien texte/image, audio)
Plein-texte (relu ou non)
Texte enrichi (balisage)
Édition imprimée
Ressource numérique native

3. Opérations réalisées et outils numériques associés

Transcription (ex. eScriptorium, Transkribus, ABBYY FineReader)
Conversion (ex. Odette, TEIGarage)
Composition (ex. LaTeX)
Balisage, annotation (ex. GROBID-Dictionaries, expressions régulières, modèles de langue)
Vérification et correction (ex. plateforme TACT, Pyrrha)
Métadonnées (ex. Mynkl)
Affichage (ex. TXM, TEIPublisher)
Impression (ex. InDesign)

4. Modalités de consultation du corpus et d'accès aux données produites

Accès ouvert à la consultation d'une interface
Accès payant à la consultation d'une interface
Données consultables, non téléchargeables
Données téléchargeables

Statut du projet - À insérer dans la notice :

En développement

Plus mis à jour, mais toujours accessible

Archivé

5. Fonctionnalités proposées

Recherche par lemmes

Recherche avancée

Comparaison d'éditions d'un même ouvrage

Mise en relation de plusieurs ouvrages

Export disponible

6. Langues

Allemand (deu)

Anglais (eng)

Espagnol (esp)

Français (fra)

Grec ancien (grc)

Italien (ita)

Latin (lat)

Portugais (prt)

Autre (norme ISO-639-3)

7. Volumétrie

→ Pas toujours pertinent, écarté pour le moment

Volume unique

Deux à trois

Quatre à dix

Dix à vingt

Plus de trente

Proposition d'un usage de la section « parcourir la taxonomie » pour la classification disciplinaire (sujet et support du corpus étudié dans le cadre du projet)

Éditions numériques**Textes lexicographiques****Lexicographie générale***Lexicographie monolingue*

Dictionnaire

Dictionnaire encyclopédique, encyclopédie

Glossaire, lexique, index

Autre

Lexicographie bilingue

Dictionnaire

Dictionnaire encyclopédique, encyclopédie

Glossaire, lexique, index

Autre

Lexicographie multilingue

Dictionnaire
Dictionnaire encyclopédique, encyclopédie
Glossaire, lexique, index
Autre

Lexicographie de spécialité

Lexicographie monolingue

Dictionnaire
Dictionnaire encyclopédique, encyclopédie
Glossaire, lexique, index
Autre

Lexicographie bilingue

Dictionnaire
Dictionnaire encyclopédique, encyclopédie
Glossaire, lexique, index
Autre

Lexicographie multilingue

Dictionnaire
Dictionnaire encyclopédique, encyclopédie
Glossaire, lexique, index
Autre

Quels liens avec le questionnaire élaboré pour les textes antiques ?

– Lien vers le questionnaire (à ne pas diffuser, en cours de construction) :

<https://framaforms.org/questionnaire-ariane-gt3-outils-et-pratiques-editoriales-editions-numeriques-de-corpus-antiques-0>

– Axel va faire une proposition d’alignement questionnaire-notice pour la prochaine fois.

2024/10/18 - Réunion de l'atelier LEX

Présents

Anaïs Chambat
Benjamin Galliot
Christophe Rey
Nathalie Rousseau
Ana Salgado

Ordre du jour

Cette réunion a été l'occasion de passer en revue les critères déjà établis et de revenir sur l'utilisation de l'onglet « parcourir la taxonomie » pour interroger les notices.

I. Reprise des critères déjà établis

1. Acquisition (modifié)

Numérisation (image interrogeable ou non)
Édition papier pré-existante
Corpus oraux
Images

Ce critère a été remplacé par :

Édition papier pré-existante

Oui
Non

Il s'agit de pouvoir établir la **distinction entre un projet de rétroconversion et la production d'une ressource numérique native**. Il faudrait réfléchir à un sous-titre pour expliciter davantage, tel que « support : édition papier pré-existante » ?

2. Objectifs et productions (modifié)

Indexation (lien texte/image, ~~audio~~)

Plein-texte (relu ou non)

Texte enrichi (balisage)

Édition ~~imprimée~~ imprimable

~~Ressource numérique native~~

→ Modalité retirée pour ne pas faire doublon avec le critère précédent.

3. Opérations réalisées et outils numériques associés (modifié, longuement discuté)

Transcription (ex. eScriptorium, Transkribus, ABBYY FineReader)

Conversion (ex. Odette, TEIGarage)

Composition (ex. LaTeX)

Balisage, annotation (ex. GROBID-Dictionaries, expressions régulières, modèles de langue)

Vérification et correction (ex. plateforme TACT, Pyrrha)

Métadonnées (ex. Mynkl)

Affichage (ex. TXM, TEIPublisher)

Impression (ex. InDesign)

Opérations réalisées et outils associés (nouvelle version)

Segmentation (ex. algorithmes de *layout analysis*, ROI)
Reconnaissance du texte (ex. eScriptorium, Rescribe, Transkribus, ABBYY FineReader)
Post-correction (ex. Pyrrha)
Balisage (ex. GROBID-Dictionaries, Oxygen)
Métadonnées (ex. Mynkl)
Conversion (ex. Odette, TEIGarage)
Composition (ex. LaTeX, InDesign)
Affichage (ex. TXM, TEIPublisher)
Vérification participative, mise en conformité et augmentations (ex. plateforme TACT)

Il reste à trancher sur l'intitulé des modalités et sur le choix des exemples :

- **« Affichage » ou « Publication » ?**

Le terme choisi doit refléter au mieux l'action (temporaire ou formelle).

- **« Vérification participative, mise en conformité et augmentations »** : formulation trop longue. Une version plus concise pourrait clarifier cette phase de fin/suivie de projet centrée sur les sciences participatives.

L'**ordonnement des modalités** a été revu pour suivre une **logique éditoriale**. Les exemples sont intégrés aux descriptions, rendant les critères plus lisibles.

- La consultation du livrable de l'atelier « chaînes éditoriales » pourrait être utile pour aider à finaliser ces choix. Il sera fourni d'ici la prochaine réunion.

4. Modalités de consultation du corpus et d'accès aux données produites (inchangé)

Accès ouvert à la consultation d'une interface
Accès payant à la consultation d'une interface
Données consultables, non téléchargeables
Données téléchargeables

5. Fonctionnalités proposées (inchangé, reste à discuter)

Recherche par lemmes
Recherche avancée
Comparaison d'éditions d'un même ouvrage
Mise en relation de plusieurs ouvrages
Export disponible

6. Langues (inchangé)

Allemand (deu)
Anglais (eng)
Espagnol (esp)
Français (fra)
Grec ancien (grc)
Italien (ita)
Latin (lat)
Portugais (prt)
Autre (norme ISO-639-3)

Ce critère a fait l'objet d'une discussion approfondie, notamment concernant le choix d'utiliser la norme **ISO-639-3** plutôt que **ISO-639-2**. La norme **ISO-639-3** a été retenue, car elle permet une meilleure représentativité des langues rares, répondant ainsi à l'objectif d'inclusivité dans la description et la gestion des langues moins courantes.

II. Utilisation de l'onglet « parcourir la taxonomie » dans l'interrogation des projets

Rappel de la proposition initiale mettant l'accent sur une classification disciplinaire

Niveau 0 : Éditions numériques

Niveau 1 : Textes antiques

Niveau 1 : Documents lexicographiques

Niveau 2 (- distinction par sujet)

Lexicographie générale

Niveau 3 (- distinction par nombre de langues)

Lexicographie monolingue

Niveau 4 (- distinction par support)

Dictionnaire

Dictionnaire encyclopédique, encyclopédie

Glossaire, lexique, index

Autre

Lexicographie bilingue

Dictionnaire

Dictionnaire encyclopédique, encyclopédie

Glossaire, lexique, index

Autre

Lexicographie multilingue

Dictionnaire

Dictionnaire encyclopédique, encyclopédie

Glossaire, lexique, index

Autre

Lexicographie de spécialité

Lexicographie monolingue

Dictionnaire

Dictionnaire encyclopédique, encyclopédie

Glossaire, lexique, index

Autre

Lexicographie bilingue

Dictionnaire

Dictionnaire encyclopédique, encyclopédie

Glossaire, lexique, index

Autre

Lexicographie multilingue

Dictionnaire

Dictionnaire encyclopédique, encyclopédie

Glossaire, lexique, index

Autre

Une simplification de la proposition a été discutée.

Niveau 0 : Projets numériques

Niveau 1 : Textes antiques

Niveau 1 : Ressources lexicographiques

Niveau 2 (- distinction par sujet)

Lexicographie de langue générale

Niveau 3 (- distinction par support)

Dictionnaire

Encyclopédie

Glossaire, lexique, index

Autre

Lexicographie de langue de spécialité

Dictionnaire

Encyclopédie

Glossaire, lexique, index

Autre

La distinction entre « monolingue », « bilingue » et « multilingue » n'a plus été jugée nécessaire, étant donné le maintien du critère de langue.

- Ana se propose d'y réfléchir pour la prochaine réunion.

Prochain ordre du jour (réunion mi/fin novembre)

- Recueillir les observations concernant le prototype
- Débuter la réflexion sur la structure des notices et la formulation d'un questionnaire permettant d'importer les réponses dans l'instance
 - **Quels liens avec le questionnaire élaboré pour les textes antiques ?**
Lien vers le questionnaire (à ne pas diffuser, en cours de construction) :
<https://framaforms.org/questionnaire-ariane-gt3-outils-et-pratiques-editoriales-editions-numeriques-de-corpus-antiques-0>

2024/11/21 - Réunion de l'atelier LEX

Présents

Anaïs Chambat
Benjamin Galliot
Ana Salgado

Ordre du jour

Cette réunion a été l'occasion de passer en revue les critères déjà établis et de revenir sur l'utilisation de l'onglet « parcourir la taxonomie » pour interroger les notices.

I. Discussion de la proposition d'Ana pour préciser l'utilisation de l'onglet « parcourir la taxonomie » dans l'interrogation des notices

Niveau 0 : Projets numériques

Regroupe tous les types de ressources numériques développées

Niveau 1 : Thématiques principales

- A. Textes antiques [Corpus ou éditions numériques de textes antiques]
- B. Ressources lexicographiques ou terminologiques [Projets contemporains liés à la description et à l'analyse lexicale ou terminologique]
- C. ... [Autres types de textes]

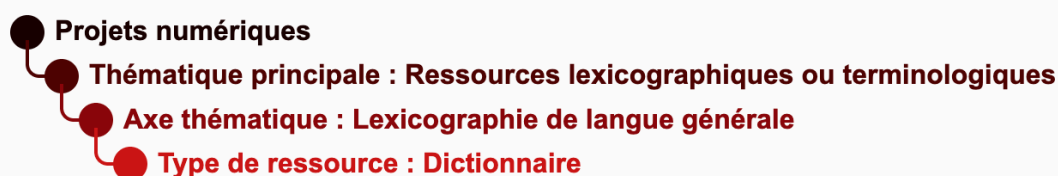
Niveau 2 : Axes thématiques

- A. Lexicographie de langue générale [Inclut les ressources traitant de la langue dans son ensemble, sans spécialisation disciplinaire]
- B. Lexicographie de langue de spécialité [Inclut les ressources orientées vers des disciplines spécifiques ou des domaines de spécialité]

Niveau 3 : Types de ressources

Dictionnaire
Encyclopédie
Glossaire/lexique/index
Autres ressources ou documentaires

Classification



Taxons

- I Nénufar
- I Trésor de la Langue Française informatisé

II. Reprise des critères déjà établis

1. Édition papier pré-existante

Titre modifié, remplacé par : « **Numérisation d'une ressource physique** »

2. Objectifs et productions

Texte modifié, remplacé par : « **Objectifs numériques** »

3. Opérations réalisées et outils associés

Deux modalités ont été modifiées

Affichage

→ **Disponibilisation** (ex. TXM, TEIPublisher)

Vérification participative, mise en conformité et augmentations

→ **Validation et améliorations collaboratives** (ex. plateforme TACT)

4. Fonctionnalités proposées (entièrement revu)

Recherche avancée (ex. présence de filtres et de tris : par lemmes, dates, mots-clés, etc.)

Mise en relation diachronique ou synchronique (ex. d'éditions d'un même ouvrage, de ressources de nature différente)

Multimédia (ex. audio, vidéo, image)

Exportation

5. Modalités de consultation et d'accès (entièrement revu)

Consultation seulement.

Accès fermé : Les données sont consultables via une interface soumise à des restrictions (ex. abonnement, enregistrement). Pas de téléchargement possible.

Consultation et téléchargement uniquement.

Accès semi-fermé : Les données sont accessibles pour consultation et téléchargement sous certaines conditions (ex. licence restrictive, abonnement, enregistrement).

Consultation, téléchargement et réutilisation.

Accès ouvert : Les données sont librement consultables, téléchargeables et réutilisables, conformément aux conditions d'une licence ouverte.

6. Langues

Autre (norme ISO-639-3)

Un commentaire a été ajouté : « Pour les langues rares, voir la norme ISO-15924 ou le catalogue *Glottolog* ».

III. Nouvelle discussion des premiers critères par les membres de l'atelier dédié aux éditions numériques de textes antiques (10/12/2024)

1. Numérisation d'une ressource physique (inchangé)

2. Objectifs numériques et scientifiques (modifié)

~~Indexation (lien texte/image)~~ Lien avec le fac-similé de la source

Plein texte (relu ou non)

Texte balisé

Texte annoté (insertion de commentaires de l'éditeur)

Texte critique (établi par l'éditeur à partir de différentes versions)

~~Édition imprimable~~

3. Opérations réalisées et outils associés (modifié)

Segmentation **du document** (ex. algorithmes de layout analysis, ROI)

Transcription et reconnaissance automatique du texte (ex. eScriptorium, Rescribe, Transkribus, ABBYY FineReader)

Post-correction (ex. Pyrrha)

Lemmatisation (ex. Pie-Extended, ctk)

Balisateur (ex. Oxygen **XML Editor**, **XMLMind XML Editor**, GROBID-Dictionaries)

Métadonnées (ex. Mynkl)

Conversion (ex. Odette, TEIGarage)

Composition (ex. LaTeX, InDesign, **MaX**, TEIPublisher).

Mise à disposition (ex. **MaX**, TXM, TEIPublisher)

Validation et améliorations collaboratives (ex. plateforme TACT, **PLUCO dans XXE**)

4. Fonctionnalités proposées (modifié)

Recherche avancée (ex. présence de filtres et de tris : par lemmes, dates, mots-clés, etc.)

Mise en relation diachronique ou synchronique (ex. d'éditions d'un même ouvrage, de ressources de nature différente)

Multimédia (ex. audio, vidéo, image)

Exportation **des données de l'édition**

Édition imprimable

5. Modalités de consultation et d'accès (non discuté)

6. Langues (non discuté)

Prochaines actions

- La consultation du livrable de l'atelier « chaînes éditoriales » pourrait être utile pour aider à finaliser les modalités du critère 3 (<https://hal.science/hal-04734959v1>).
- La prochaine réunion de l'atelier dédié aux éditions numériques de textes antiques aura lieu le mardi 14 janvier 2025 (16h15-18h). Elle sera notamment l'occasion de terminer la discussion des critères déjà établis par notre atelier. Le compte rendu détaillé de la réunion est proposé ici :
 - o <https://mypads2.framapad.org/mypads/?/mypads/group/ariane-3lbqo9v1/pad/view/2024-ant-cr-4gbso9zt>