

The CNRS logo is a white circle containing the lowercase letters 'cnrs' in a dark blue, sans-serif font. The background of the slide is a dark blue network of interconnected nodes and lines, with a bright yellow light source on the left side. There are also clusters of small purple plus signs in the top right and bottom right corners.

cnrs

Webinaire 3

Exploration et analyse avancée de corpus avec Lodex

Mathilde Huguin

`mathilde.huguin@inist.fr`

Ingénieure de recherche Istex

→ Institut de l'information scientifique et technique

Webinaire 1

Découverte et fonctions de base de LODEX

Lucile Bourguignon

- Chargement des données
- Paramétrage d'un modèle
- Traitements minimaux sur les données
- Configuration du site
- Export et suppression de la base de données



Webinaire 2

La fouille de textes avec LODEX

Valérie Bonvallot

- Introduction théorique au TDM
- Découverte d'Istex TDM
- Enrichissement d'un corpus HAL
- Représentation graphique des enrichissements
- Présentation du rôle de contributeur-annotateur



Sommaire

- 01 Introduction
- 02 Import d'un corpus multi-sources
- 03 Enrichissement des données avec Lodash
- 04 Création et gestion de sous-ressources
- 05 Enrichissement via des requêtes SPARQL
- 06 Visualisation avancée avec Vega-Lite



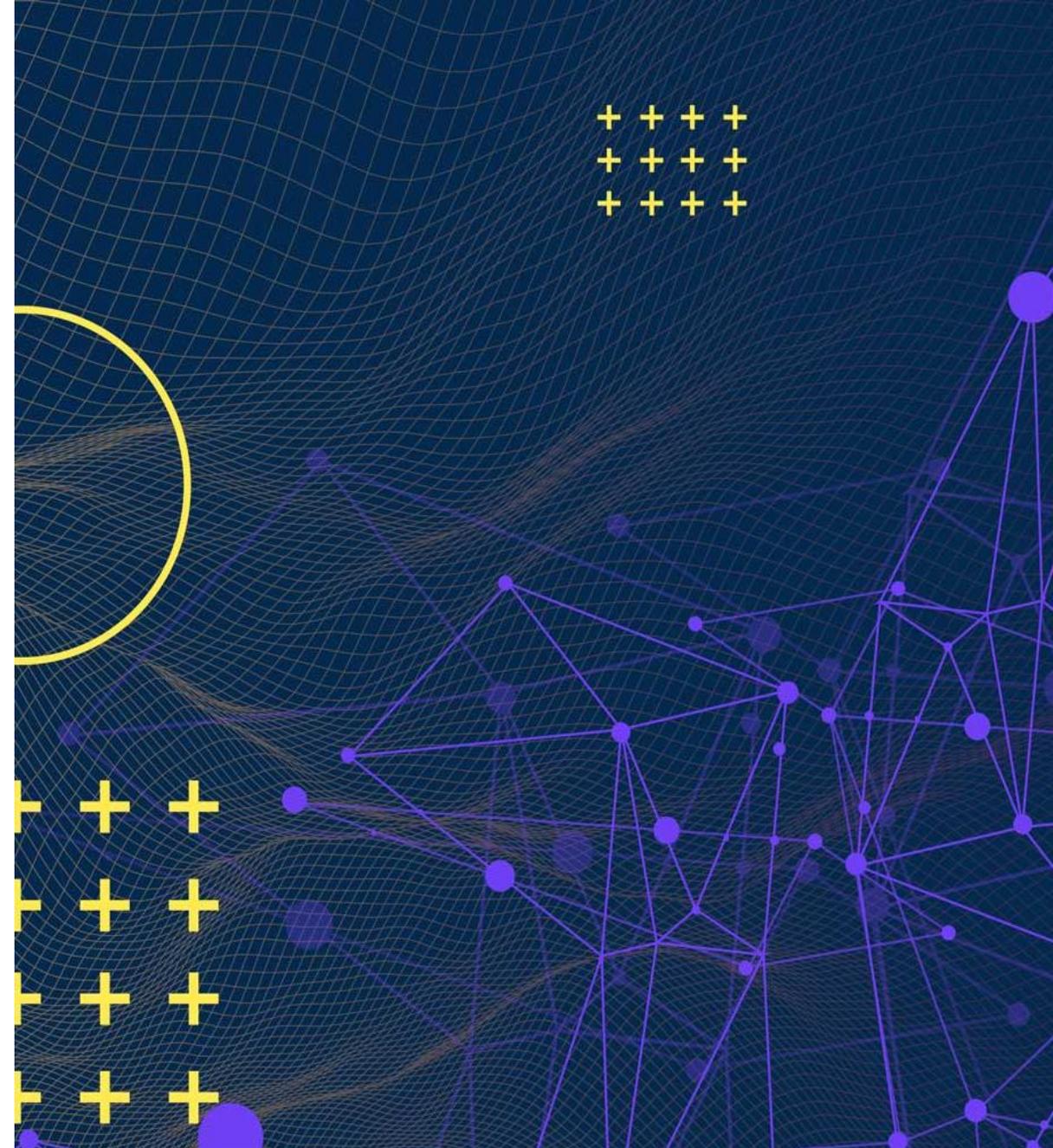
01 Introduction

Introduction

Objectif

Construire et explorer un corpus sur les différentes langues signées en SHS

Quelles sont les langues les plus étudiées ? Pourquoi ? Quelles sont les analyses effectuées (syntaxe, sémantique, sociolinguistique) ?



DÉTECTER LES NOMS DE LANGUES SIGNÉES

Comment enrichir des données avec Lodash

2

1

IMPORTER 2 JEUX DE DONNÉES DIFFÉRENTS

Comment modifier un loader

3

CRÉER DES PAGES POUR CHAQUE LANGUE DÉTECTÉE

Comment créer des sous-ressources



4

ENRICHIR LES PAGES CONCERNANT LES LANGUES SIGNÉES

Comment utiliser des requêtes SPARQL

5

CRÉER UN GRAPHIQUE SUR L'ÉVOLUTION DES LANGUES SIGNÉES (HISTOGRAMME EMPILÉ)

Comment créer d'autres graphiques avec Vega-Lite



02

Import d'un corpus multi-sources

Import d'un corpus multi-sources

Extraction des publications

- Requête [Istex Search](#) **2 498 documents**: (title:("sign language" "sign languages" "signed language" "signed languages" "visuo-gestural language" "visuo-gestural languages" "visual-gestural language" "visual-gestural languages" "iconic language" "iconic languages..."
- Requête [Conditor](#) **417 documents**: ((title.en:("sign language" "signed language" "visuo-gestural language" "visual-gestural language" "iconic language" "spatial language" "fingerspelling alphabet") abstract.en:("sign language" "sign languages" "signed language" "signed languages..."



Données Istex Search

Format JSON (Zip Istex Search)

```
{
  "corpusName": "elsevier",
  "keywords": {
    "teeft": [
    ]
  },
  "subject": [
  ],
  "language": [
    "eng"
  ],
  "originalGenre": [
    "Full-length article"
  ],
  "qualityIndicators": {
  },
  "title": "An empirical investigation of exper
  "pii": [
    "S0024-3841(12)00168-4"
  ],
  "refBibs": [
    {
      "host": {
        "author": [
          {
            "name": "A Aikhenvald"
          }
        ]
      },
      "publisher": "Oxford University Press",
    }
  ]
}
```

Données Conditor

Format JSONL

```
{
  "abstract": {
    "default": "Pathogenic variants in DDX3X are as
  },
  "abstract": {
    "default": "Dynamic semantics provides a mechan
  },
  "abstract": {
    "default": "AbstractIn recent years, research d
  },
  "authors": [
    {
      "forename": "Alexis",
      "isni": [],
      "researcherId": [
    ]
  ],
  "abstract": {
    "default": "Signing avatars are increasingly us
  },
  "abstract": {
    "default": "The topic of this paper is machine
  },
  "abstract": {
    "default": "This paper describes KAZOO, a web a
  },
  "abstract": {
    "default": "Existing work on the animation of s
  },
  "abstract": {
    "default": "Natural animation of Sign Language
  },
  "abstract": {
    "default": "AbstractProfessional Sign Language
  },
  "abstract": {
    "default": "The goal of this paper is to compar
  },
  "authors": [
    {
      "forename": "Laia",
      "isni": [],
      "researcherId": [
    ]
  ],
  "abstract": {
    "default": "ASL (American Sign Language) can ex
  },
  "abstract": {
    "default": "In several sign languages, plurals
  }
}
```

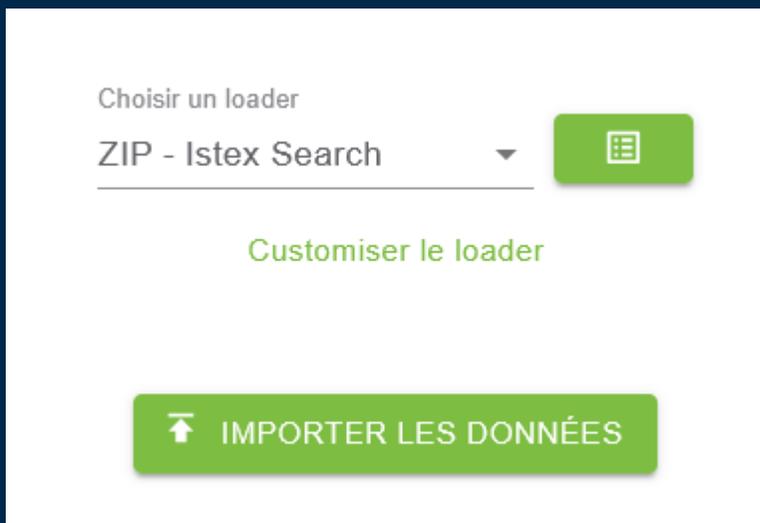
JSON (JavaScript Object Notation – Notation Objet issue de JavaScript) est un format léger d'échange de données.



Données Istex Search

Format JSON (Zip Istex Search)

↔ loader Lodex adapté aux données



Chosir un loader
ZIP - Istex Search

Customiser le loader

↑ IMPORTER LES DONNÉES

Données Conditor

Format JSONL

↔ loader Lodex adapté aux données



Chosir un loader
JSON - format text JSON Lines

Customiser le loader

↑ IMPORTER LES DONNÉES

Le **loader** reformate un jeu de données vers un format utilisable dans Lodex. Il fait appel à un fichier de configuration .ini définissant le traitement à appliquer à chaque champ du jeu de données.



Données Istex Search

Données Conditor

Les deux jeux de données peuvent être chargés. Toutefois, les noms des champs ne sont pas les mêmes (ex. *Auteur(s)* vs *authors*) et les valeurs ne sont pas du même type (ex. *tableau de chaîne de caractères* vs *tableau d'objets*).

Il faut adapter le loader pour copier la structure de l'un ou l'autre jeu de données.

COLONNES	FILTRES	DENSITÉ	AJOUTER			
uri	Titre	Auteur(s)	Affiliation(s)	Revue ou monogra...	Auteur(s) mono	
<input type="checkbox"/>	"ark:/67375/6H6-MXFI	"An empirical investig	["Inge Zwitserlood", "P	["Radboud University	"Lingua"	[]
<input type="checkbox"/>	"ark:/67375/6H6-ZNQI	"How grammar can co	["Carlo Geraci", "Marta	["Dipartimento di Psic	"Cognition"	[]
<input type="checkbox"/>	"ark:/67375/WNG-JK0	"From Gesture to Sign	["Chloë R. Marshall", "i	["Institute of Educatio	"Topics in Cognitive S	[]
<input type="checkbox"/>	"ark:/67375/6H6-8969	"Writing the smile: Lar	["Erika Hoffmann-Dillo	["Oberlin College, De	"Language and Comr	[]
<input type="checkbox"/>	"ark:/67375/6H6-7RZC	"Modality and meaninç	["Roland Pfau", "Marku	["University of Amster	"Lingua"	[]
<input type="checkbox"/>	"ark:/67375/VQC-8LW	"Language justice for :	["Sarah C. E. Batterbu	["Centre for Deaf Stud	"Language Policy"	[]
<input type="checkbox"/>	"ark:/67375/6H6-0G3E	"Right dislocated pron	["Felix Yim Binh Sze"]	["Centre for Sign Ling	"Journal of Pragmatic	[]



COLONNES	FILTRES	DENSITÉ	AJOUTER		
uri	abstract	authors	business	classifications	
<input type="checkbox"/>	"uid:/KFWxdLCnC"	{"default":"Pathogenic	{["forename":"Elana", "	{"duplicateGenre":"Art	{"hal":{"code":"sdv.ne
<input type="checkbox"/>	"uid:/n3mNrDCje"	{"default":"Dynamic se	{["forename":"Jeremy", "	{"duplicateGenre":"Ch	{"hal":{"code":"scco.lit
<input type="checkbox"/>	"uid:/8eyzQJjRm"	{"default":"AbstractIn r	{["forename":"Sylvie", "	{"duplicateGenre":"Ch	{"hal":{"code":"tef","dewey
<input type="checkbox"/>	"uid:/cib6Pymkj"	{"default":"This paper	{["forename":"Alexis", "	{"duplicateGenre":"Art	{"hal":{"code":"info","e
<input type="checkbox"/>	"uid:/jm4Tq4M9J"	{"default":"Signing ava	{["forename":"Sylvie", "	{"duplicateGenre":"Art	{"hal":{"code":"info","e
<input type="checkbox"/>	"uid:/g78qSddDt"	{"default":"The topic of	{["forename":"Michael", "	{"duplicateGenre":"Art	{"hal":{"code":"info","e
<input type="checkbox"/>	"uid:/Q69PpnjDV"	{"default":"This paper	{["forename":"Annelie", "	{"duplicateGenre":"Art	{"hal":{"code":"info","e
<input type="checkbox"/>	"uid:/EB8ekdHLx"	{"default":"Existing wo	{["forename":"Lucie", "i	{"duplicateGenre":"Art	{"hal":{"code":"scco.c

Import d'un corpus multi-sources

Adaptation du loader

Solution : customiser le loader lors de l'import du second jeu de données pour copier la structure (et les types) du premier.

```
CUSTOM - Loader customisé
1  append = pack
2  label = json-conditor
3
4  # load some plugins to activate some statements
5  [use]
6  plugin = basics
7
8  # Toggle ezs traces (see server stderr log)
9  [debug]
10 ezs = false
11
12 # Parse input file
13 [unpack]
14
15 [replace]
16
17 path = Titre
18 value = get("title.default")
19
20 path = fix('Auteur(s)')
21 value = get("authors").map(item => (item.fullname))
```

SUPPRIMER ANNULER SAUVEGARDER



Atelier Lodash



Documentation sur les loaders



Cours EZS



Documentation EZS

Import d'un corpus multi-sources

Adaptation du loader

Navigation: DONNÉES | AFFICHAGE | ANNOTATIONS

Options: COLONNES | FILTRES | DENSITÉ | **+ AJOUTER**

Table headers: | uri | Titre | Ajouter plus de données

Table content:

<input type="checkbox"/>	"ark:/67375/6H6-MXFI	"An empirical investig: ["Inge Zwitterlood",
<input type="checkbox"/>	"ark:/67375/6H6-ZNQL	"How grammar can co ["Carlo Geraci", "Ma

Étape 1 : importer le ZIP Istex Search en utilisant le loader du même nom puis cliquer sur *Ajouter* dans le jeu de données.

Glisser un fichier sur cette page ou cliquer ici

Conditor.jsonl

Choisir un loader
CUSTOM - Loader customisé

Customiser le loader

IMPORTER LES DONNÉES

Étape 2 : charger le fichier Conditor, sélectionner un loader puis cliquer sur *Customiser le loader*.

Import d'un corpus multi-sources

Adaptation du loader

CUSTOM - Loader customisé

```
1 append = pack
2 label = json-conditor
3
4 # load some plugins to activate some statements
5 [use]
6 plugin = basics
7
8 # Toggle ezs traces (see server stderr log)
9 [debug]
10 ezs = false
11
12 # Parse input file
13 [unpack]
14
15 [replace]
16
17 path = Titre
18 value = get("title.default")
19
20 path = fix('Auteur(s)')
21 value = get("authors").map(item => (item.fullname))
```

SUPPRIMER

ANNULER

SAUVEGARDER

Étape 3 : modifier le loader.

Spécifie que les documents transformés seront regroupés
append = pack

Nom de l'étiquette pour le loader (utile pour le suivi ou les logs)
label = json-conditor

Activation des plugins nécessaires
[use]

plugin = basics # Plugin de base pour les fonctions EZS (get, map, etc.)

Débogage EZS (logs visibles côté serveur)
[debug]

ezs = false # false = pas de trace dans les logs

Parsing des objets JSON d'entrée
[unpack]

...

Import d'un corpus multi-sources

Adaptation du loader

Exemple 1 : adapter le *titre* pour copier la structure du ZIP Istex Search importé dans Lodex.

Jsonl Conditor	Données Istex	Instruction du loader
<pre>"title":{"default":"A Dynamic Semantics for Multimodal Communication"},"en":"A Dynamic Semantics for Multimodal Communication"}</pre>	<p>Titre</p> <pre>"An empirical investig</pre>	<pre>[replace] path = Titre value = get("title.default")</pre>
↪ Objet (structure composée de paires clé / valeur)	↪ Chaîne de caractères	↪ L'instruction [replace] permet de remplir la colonne Titre avec la valeur de la clé default du champ title

Import d'un corpus multi-sources

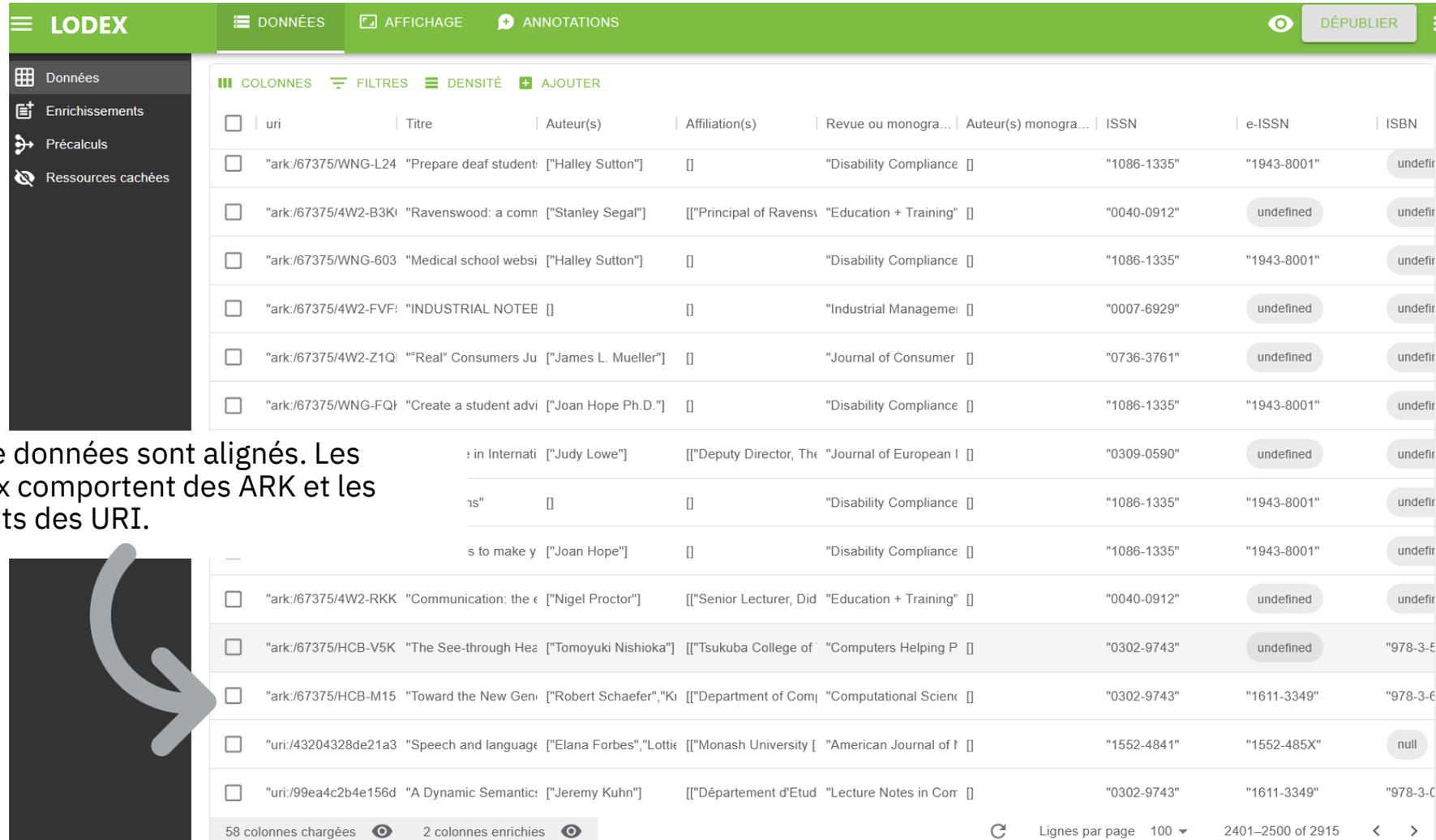
Adaptation du loader

Exemple 2 : adapter les *noms d'auteurs* pour copier la structure du zip Istex Search importé dans Lodex.

Jsonl Conditor	Données Istex	Instruction du loader
<pre>"authors": [{"forename": "Alexis", "isni": [], "researcherId": [], "surname": "Heloir", "affiliations": [], "halAuthorId": [], "orcid": [], "fullname": "Alexis Heloir", "idHal": [], "idRef": [], "viaf": [], "rnsr": ["201220427F"]}]}...</pre>	<pre> Auteur(s) ----- : ["Inge Zwitserlood", "P;</pre>	<pre>[replace] path = fix('Auteur(s)') value = get("authors").map(item => (item.fullname))</pre>
<p>↪ Tableau ([]) d'objets (structure composée de paires clé / valeur)</p>	<p>↪ Tableau de chaîne de caractères</p>	<p>↪ L'instruction [replace] permet de remplir la colonne Auteur(s) avec la valeur de la clé fullname du champ authors. map permet de parcourir chaque item du tableau.</p>

Import d'un corpus multi-sources

Résultat



LODEX

DONNÉES AFFICHAGE ANNOTATIONS DÉPUBLIER

COLONNES FILTRES DENSITÉ AJOUTER

<input type="checkbox"/>	uri	Titre	Auteur(s)	Affiliation(s)	Revue ou monogra...	Auteur(s) monogra...	ISSN	e-ISSN	ISBN
<input type="checkbox"/>	"ark:/67375/WNG-L24"	"Prepare deaf student"	["Halley Sutton"]	[]	"Disability Compliance"	[]	"1086-1335"	"1943-8001"	undefir
<input type="checkbox"/>	"ark:/67375/4W2-B3K"	"Ravenswood: a com"	["Stanley Segal"]	[["Principal of Ravens	"Education + Training"	[]	"0040-0912"	undefined	undefir
<input type="checkbox"/>	"ark:/67375/WNG-603"	"Medical school websi"	["Halley Sutton"]	[]	"Disability Compliance"	[]	"1086-1335"	"1943-8001"	undefir
<input type="checkbox"/>	"ark:/67375/4W2-FVF"	"INDUSTRIAL NOTEE"	[]	[]	"Industrial Managemen"	[]	"0007-6929"	undefined	undefir
<input type="checkbox"/>	"ark:/67375/4W2-Z1Q"	"Real Consumers Ju"	["James L. Mueller"]	[]	"Journal of Consumer"	[]	"0736-3761"	undefined	undefir
<input type="checkbox"/>	"ark:/67375/WNG-FQ"	"Create a student advi"	["Joan Hope Ph.D."]	[]	"Disability Compliance"	[]	"1086-1335"	"1943-8001"	undefir
<input type="checkbox"/>			in Internati	["Judy Lowe"]	[["Deputy Director, The	"Journal of European I"	"0309-0590"	undefined	undefir
<input type="checkbox"/>			rs"	[]	[]	"Disability Compliance"	"1086-1335"	"1943-8001"	undefir
<input type="checkbox"/>			s to make y	["Joan Hope"]	[]	"Disability Compliance"	"1086-1335"	"1943-8001"	undefir
<input type="checkbox"/>	"ark:/67375/4W2-RKK"	"Communication: the c"	["Nigel Proctor"]	[["Senior Lecturer, Did	"Education + Training"	[]	"0040-0912"	undefined	undefir
<input type="checkbox"/>	"ark:/67375/HCB-V5K"	"The See-through Hea"	["Tomoyuki Nishioka"]	[["Tsukuba College of	"Computers Helping P"	[]	"0302-9743"	undefined	"978-3-5
<input type="checkbox"/>	"ark:/67375/HCB-M15"	"Toward the New Gen"	["Robert Schaefer", "Ki	[["Department of Com]	"Computational Scienc"	[]	"0302-9743"	"1611-3349"	"978-3-6
<input type="checkbox"/>	"uri:/43204328de21a3"	"Speech and language"	["Elana Forbes", "Lottie	[["Monash University ["American Journal of I"	[]	"1552-4841"	"1552-485X"	null
<input type="checkbox"/>	"uri:/99ea4c2b4e156d"	"A Dynamic Semantic:"	["Jeremy Kuhn"]	[["Département d'Etud	"Lecture Notes in Corr"	[]	"0302-9743"	"1611-3349"	"978-3-0

58 colonnes chargées 2 colonnes enrichies Lignes par page 100 2401-2500 of 2915

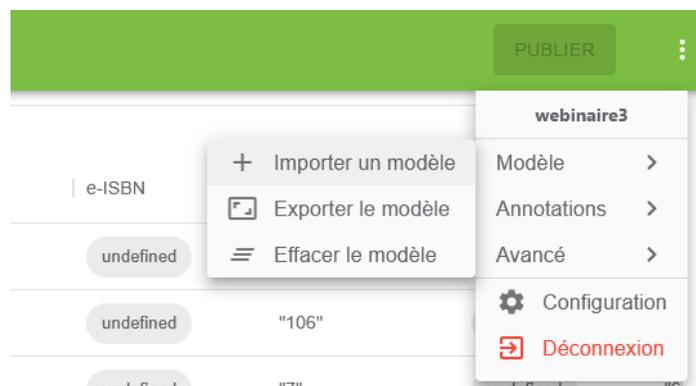
Les deux jeux de données sont alignés. Les documents Istex comportent des ARK et les autres documents des URI.

Import d'un corpus multi-sources

Import du modèle

Pour importer un modèle

Cliquer sur le menu en haut à droite > *Modèle* > *Importer un modèle* et sélectionner le fichier à importer.



Pour en savoir plus

→ [Le modèle / Import d'un modèle](#)

→ [Le modèle / Création d'un modèle](#)

Le **modèle** rassemble toutes les instructions décrivant la manière dont les données, préalablement importées dans l'instance sous forme tabulaire, vont être mises en forme à des fins de visualisation, de stockage, etc.

Import d'un corpus multi-sources

Résultat

LODEX DONNÉES AFFICHAGE ANNOTATIONS

PAGE

DATASET - Titre DATASET - Description

[ydSX] Titre

[fwzP] Image

Interface administrateur (l'onglet *Affichage* comporte désormais les champs à afficher).

ISTEX Webinaire Lodex
Exploration et analyse avancée de corpus

Langues signées

La diversité des langues signées dans la recherche scientifique



Comment naviguer dans le site ?

Retrouvez dans le menu en bas à gauche les différents éléments du corpus.

Accès aux représentations visuelles produites à partir des données.

Accès à chacune des publications composant le corpus.



Interface utilisateur (les champs sont affichés).

03

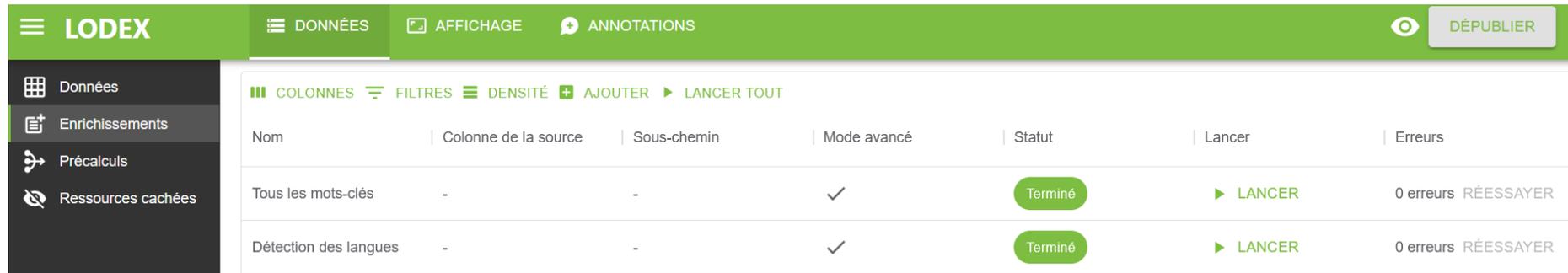
Enrichissement des données avec Lodash

Enrichissement des données avec Lodash

Qu'est-ce qu'un enrichissement ?

Les enrichissements permettent :

- D'utiliser **des web services de TDM** (cf. [Webinaire 2](#)) ;
- De créer **des traitements en mode avancé** permettant la curation, l'homogénéisation ou l'enrichissement des données



The screenshot shows the LODEX interface with a green header. The main menu on the left includes 'Données', 'Enrichissements', 'Précalculs', and 'Ressources cachées'. The 'Enrichissements' section is active, displaying a table of enrichment processes. The table has columns for 'Nom', 'Colonne de la source', 'Sous-chemin', 'Mode avancé', 'Statut', 'Lancer', and 'Erreurs'. Two processes are listed: 'Tous les mots-clés' and 'Détection des langues', both with a status of 'Terminé' and a 'LANCER' button.

Nom	Colonne de la source	Sous-chemin	Mode avancé	Statut	Lancer	Erreurs
Tous les mots-clés	-	-	✓	Terminé	▶ LANCER	0 erreurs RÉESSAYER
Détection des langues	-	-	✓	Terminé	▶ LANCER	0 erreurs RÉESSAYER

Interface d'accès aux différents enrichissements.

44

[Web services](#)

43

[Recettes d'enrichissements utiles en mode avancé](#)

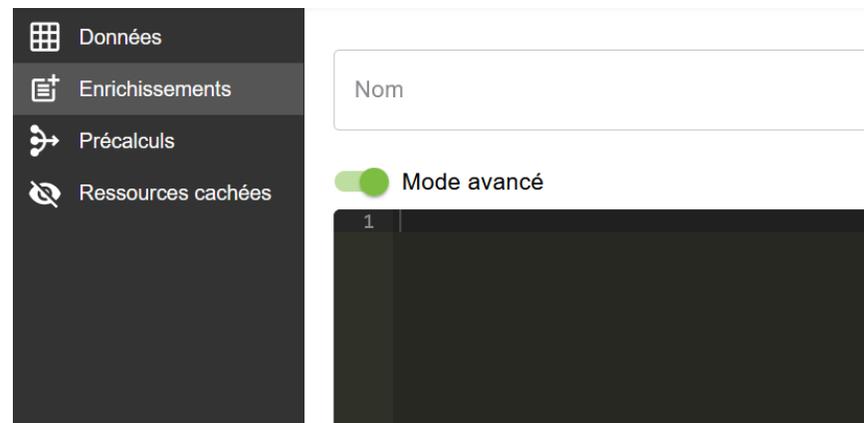
Enrichissement des données avec Lodash

Annoter les documents selon la langue citée

Solution : utiliser un enrichissement qui va appeler un TSV distant contenant une liste de langues signées et leur identifiant sur Wikidata, détecter ces noms de langues dans les mots-clés d'auteurs et les mots-clés [Teeft](#).

```
1 Langue Infos
2 Adamorobe Sign Language Q27709
3 Afghan Sign Language Q4689093
4 Albanian Sign Language Q4709168
5 Albarradas Sign Language Q110736970
6 Algerian Sign Language Q3135421
7 Alipur Sign Language Q4726836
8 Al-Sayyid Bedouin Sign Language Q2915457
9 Amami O Shima Sign Language
```

Fichier TSV contenant les noms de langues signées répertoriés dans [Glottolog](#).



Enrichissement en *Mode avancé*.

Les **documents Istex** sont enrichis par un programme (Teeft) qui détecte les termes les plus spécifiques dans le texte intégral.

Enrichissement des données avec Lodash

Annoter les documents selon la langue citée

Étape 1 : fusionner les mots-clés d'auteur et les mots-clés Teeft dans une seule colonne. Créer un nouvel enrichissement en cliquant sur *Ajouter*, puis cliquer sur *Mode avancé* et coller le [code suivant](#). Cliquer ensuite sur *Sauvegarder* et *Lancer*.

```
1 [assign]
2 path = value
3 value = fix(self.value["Mots-clés d'auteur"], self
  .value["Mots-clés (teeft)"].join(',').toLowerCase().replace
    (/ \(\.\.\)/g, "").split(',')
```

[assign] Méthode pour attribuer une valeur

path = Chemin vers le jeu de données

value = fix(... Colonnes à considérer dans le jeu de donnée

join(,) Transformation des 2 tableaux en des chaînes de caractères séparés par des virgules

toLowerCase() Transformation des mots-clés en minuscules

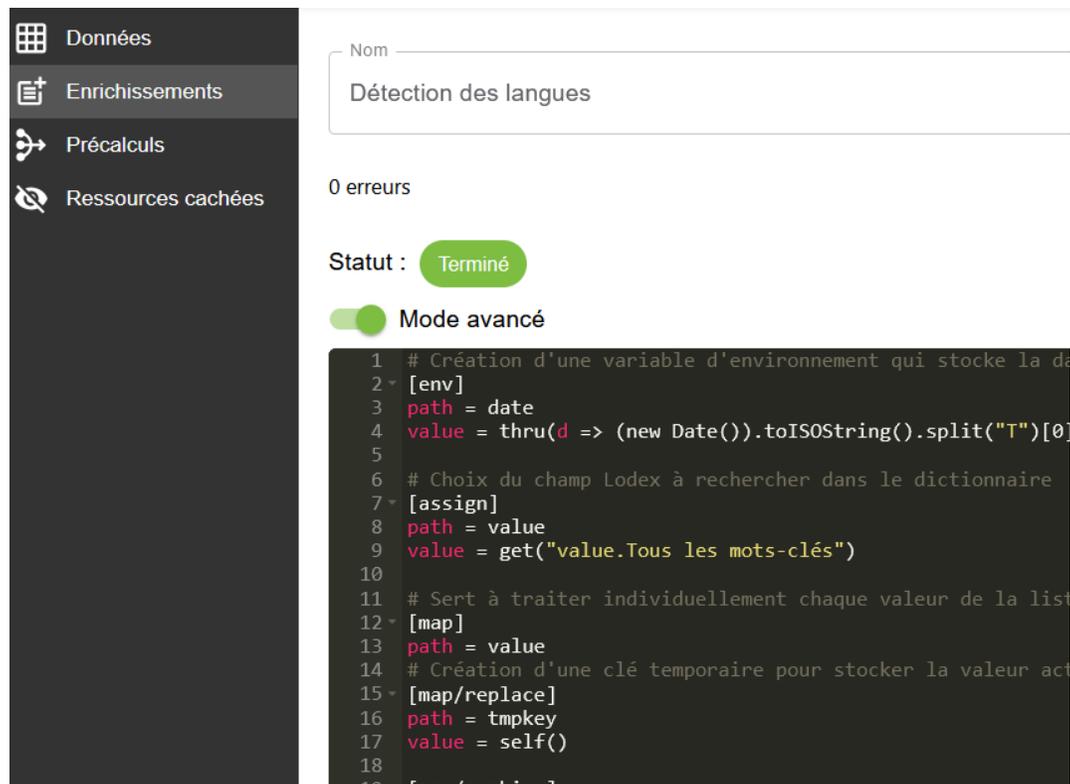
replace(/ \(\.\.\)/g, "") Expression régulière pour supprimer les sigles derrière les noms de langue

split(`,`) Création d'un tableau JSON

Enrichissement des données avec Lodash

Annoter les documents selon la langue citée

Étape 2 : détecter les noms de langues grâce au [code suivant](#) (enrichissement 31).



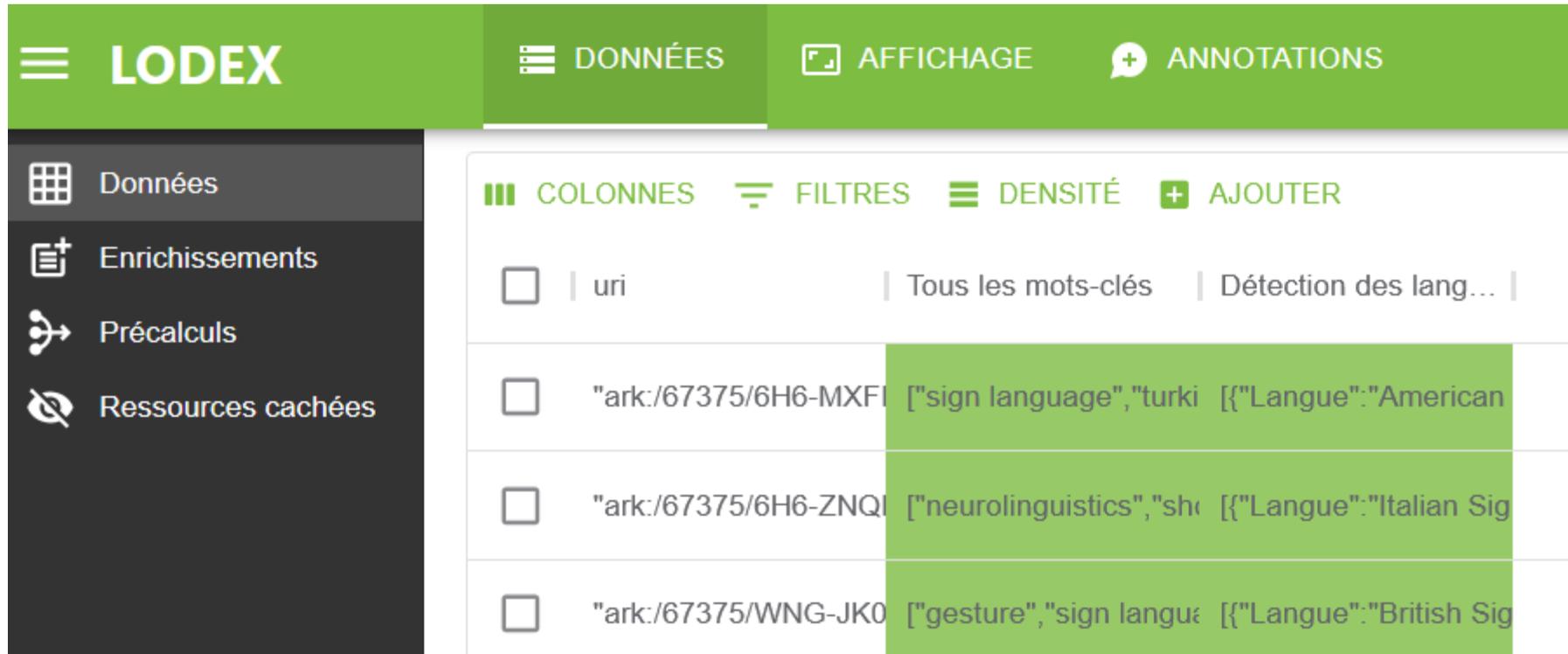
The screenshot displays the 'Enrichissements' (Enrichments) section of a software interface. On the left, a sidebar contains menu items: 'Données', 'Enrichissements' (selected), 'Précalculs', and 'Ressources cachées'. The main area shows a configuration for 'Détection des langues' (Language Detection). It includes a 'Nom' (Name) field with the value 'Détection des langues', a 'Statut' (Status) of 'Terminé' (Completed), and a 'Mode avancé' (Advanced mode) toggle switch that is turned on. Below this, a code editor shows the following JavaScript code:

```
1 # Création d'une variable d'environnement qui stocke la date
2 [env]
3 path = date
4 value = thru(d => (new Date()).toISOString().split("T")[0])
5
6 # Choix du champ Lodex à rechercher dans le dictionnaire
7 [assign]
8 path = value
9 value = get("value.Tous les mots-clés")
10
11 # Sert à traiter individuellement chaque valeur de la liste
12 [map]
13 path = value
14 # Création d'une clé temporaire pour stocker la valeur actuelle
15 [map/replace]
16 path = tmpkey
17 value = self()
18
19 [map/combine]
```

Enrichissement des données avec Lodash

Résultat

Les deux résultats des enrichissements **sont visibles dans les colonnes** du jeu de données.



The screenshot displays the LODEX application interface. The top navigation bar includes 'LODEX', 'DONNÉES', 'AFFICHAGE', and 'ANNOTATIONS'. A left sidebar contains 'Données', 'Enrichissements', 'Précalculs', and 'Ressources cachées'. The main area shows a table with columns 'COLONNES', 'FILTRES', 'DENSITÉ', and 'AJOUTER'. The table lists data rows with checkboxes and text, including URIs and enriched content like '["sign language", "turki [{"Langue": "American...']' and '["neurolinguistics", "sho [{"Langue": "Italian Sig...']'. The enriched content columns are highlighted in green.

<input type="checkbox"/>	COLONNES	FILTRES	DENSITÉ	AJOUTER
<input type="checkbox"/>	uri	Tous les mots-clés	Détection des lang...	
<input type="checkbox"/>	"ark:/67375/6H6-MXF			["sign language", "turki [{"Langue": "American
<input type="checkbox"/>	"ark:/67375/6H6-ZNQ			["neurolinguistics", "sho [{"Langue": "Italian Sig
<input type="checkbox"/>	"ark:/67375/WNG-JK0			["gesture", "sign languæ [{"Langue": "British Sig

04

Création et gestion de sous-ressources

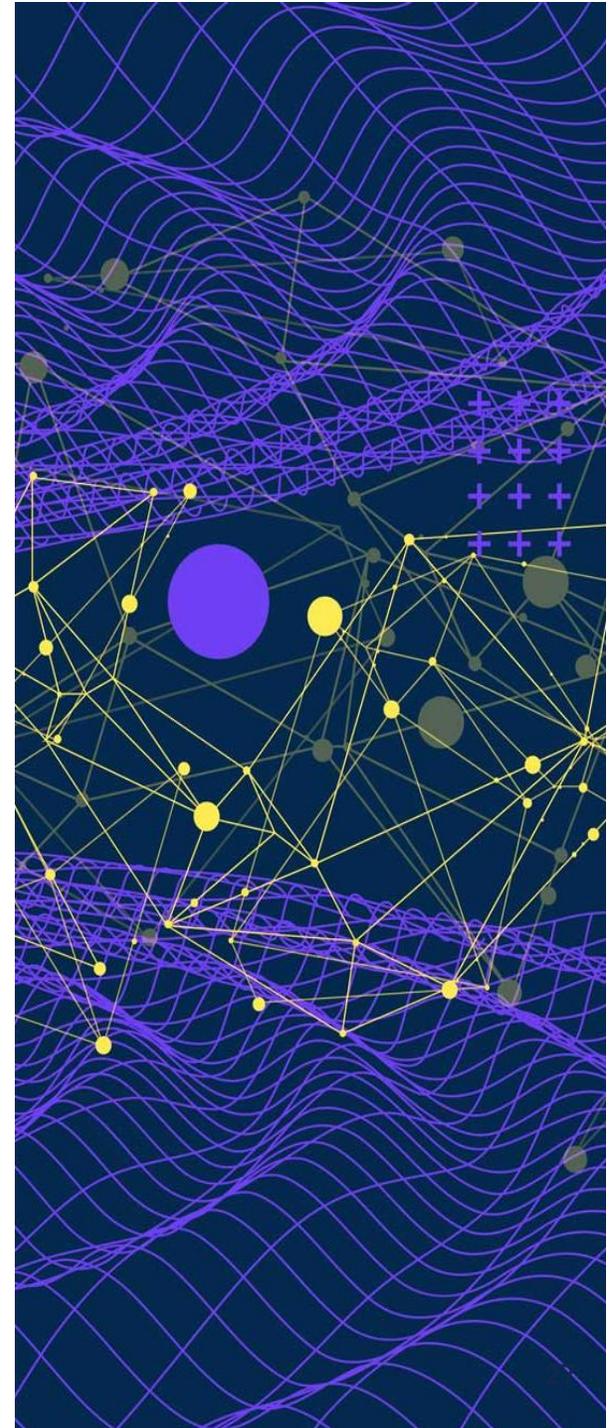
Création et gestion de sous-ressources

Qu'est-ce qu'une sous-ressource ?

Une **sous-ressource dans Lodex** est une donnée ajoutée pour enrichir une ressource principale. En d'autres termes, une sous-ressource sert à enrichir les données d'une ressource principale en ajoutant des couches d'information supplémentaires, utilisant un système de liaison et de navigation intégré dans Lodex.

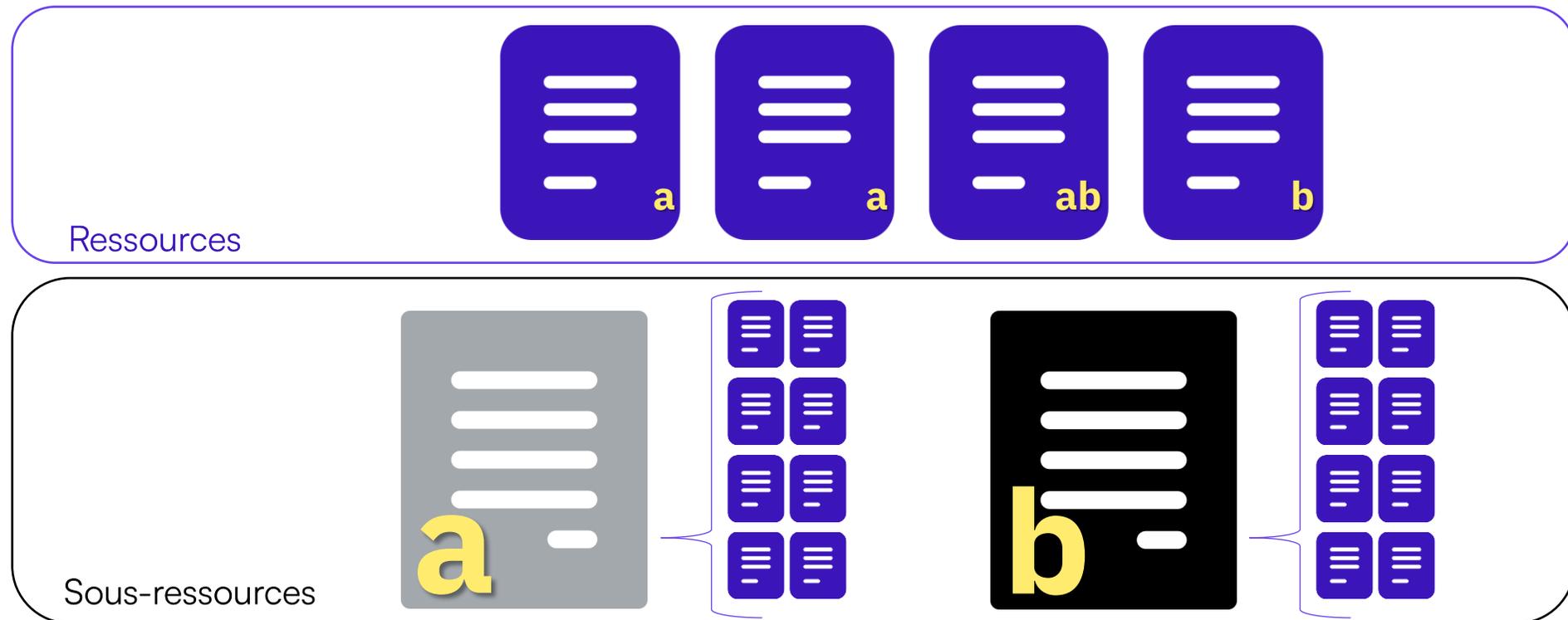
Elle permet de structurer et d'organiser des informations associées à une ressource, facilitant ainsi la navigation entre une ressource principale et ses détails ou éléments connexes.

Par exemple, dans un corpus scientifique, une ressource principale pourrait être un article, et une sous-ressource pourrait être la liste des espèces détectées dans cet article.



Création et gestion de sous-ressources

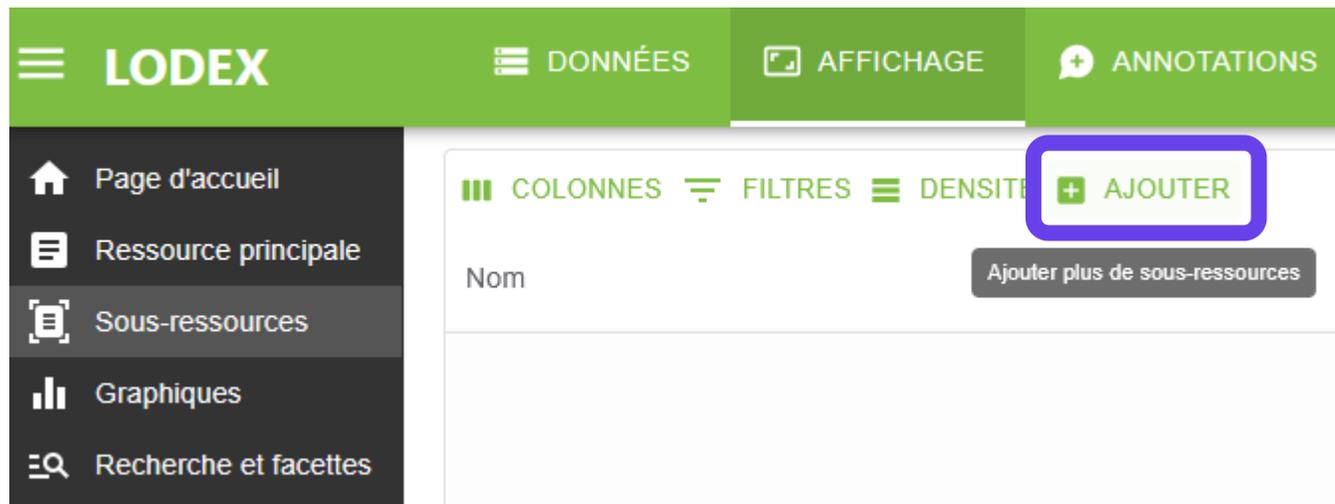
Qu'est-ce qu'une sous-ressource ?



Création et gestion de sous-ressources

Comment créer une sous-ressource ?

Étape 1 : dans l'interface de gestion de l'affichage, cliquer sur *Sous-ressources* puis cliquer sur *Ajouter*.



The screenshot displays the LODEX interface. At the top, there is a green navigation bar with the LODEX logo and three tabs: 'DONNÉES', 'AFFICHAGE', and 'ANNOTATIONS'. The 'AFFICHAGE' tab is currently selected. On the left side, a dark sidebar contains a menu with five items: 'Page d'accueil', 'Ressource principale', 'Sous-ressources', 'Graphiques', and 'Recherche et facettes'. The 'Sous-ressources' item is highlighted. In the main content area, there is a toolbar with three icons: 'COLONNES', 'FILTRES', and 'DENSITÉ', followed by a red button with a plus sign and the text 'AJOUTER'. Below this toolbar, a table is visible with a header row containing the text 'Nom'. A tooltip with the text 'Ajouter plus de sous-ressources' is positioned over the 'AJOUTER' button.

Création et gestion de sous-ressources

Comment créer une sous-ressource ?

Étape 2 : sélectionner les données qui permettront de construire la sous-ressource en remplissant le formulaire.



Nom

Chemin

Identifiant

- Langue
- Infos

Nom : nom de votre sous-ressource (arbitraire)

Chemin : colonne de votre jeu de données qui contient les données permettant de créer la sous ressource

Identifiant : nom du sous-champ sur lequel pointer dans les données en JSON

Détection des lang...

["Langue": "American Sign Language"]
["Langue": "Italian Sign Language"]

```
"root" : [ 1 item
  0 : { 2 items
    "Langue" : string "Turkish Sign Language"
    "Infos" : string "Q36885"
  }
]
```

Création et gestion de sous-ressources

Comment créer une sous-ressource ?

Étape 3 : paramétrer l'affichage de la sous-ressource (sensiblement identique au paramétrage de l'affichage d'une ressource),



Nom: Langues signées

Chemin: Détection des langues

Identifiant: Langue

SUPPRIMER

PAGE

DONNÉES PUBLIÉES

SOUS-RESSOURCE - T. ▾

+ NOUVEAU CHAMP

[Uv9E]

Titre

L'ajout de *Nouveaux champs* permet de paramétrer l'affichage des fiches sous-ressources.



Création et gestion de sous-ressources

Comment créer une sous-ressource ?

Étape 4 : lier la sous-ressource à la ressource principale. Pour créer ce lien, il faut créer deux champs dans la *Ressource principale* :

(1) un champ non visible contenant **l'étiquette utilisée** pour renvoyer vers la sous-ressource (identique à la configuration de la sous-ressource). Dans *Affichage* décocher le bouton *Visible*.

The screenshot displays the Inist interface for creating a sub-resource. The main panel is titled "GÉNÉRAL" and contains the following elements:

- Etiquette:** A text field containing "Langue signée".
- icône(s) du champ:** A set of icons for field configuration.
- Nom interne:** A text field containing "Etiquette pour lien vers sous ressource".
- Source de la valeur:** A selection menu with options: "VALEUR ARBITRAIRE", "CHOIX DE LA ROUTINE", "DONNÉE PRÉCALCULÉE", "COLONNE(S) EXISTANTE(S)" (selected), and "DEPUIS UNE SOUS-RESSOURCE".
- Colonne(s) existante(s):** A dropdown menu with "Détection des langues" and "Saisir colonne(s) existante(s)".
- Opérations de transformation:** A section with a "TOUT SUPPRIMER" button and a "GET Langue" button.
- Ajouter une opération:** A dashed green box at the bottom with the text "AJOUTER UNE OPÉRATION".

On the right side, a secondary panel titled "Aperçu de la valeur*" shows the "AFFICHAGE" tab. It displays a list of language labels, such as ["Italian Sign Language", "American Sign Language"], ["American Sign Language", "British Sign Language"], and ["American Sign Language", "Spanish Sign Language"]. A "Visible" toggle is shown as checked.

Création et gestion de sous-ressources

Comment créer une sous-ressource ?

Étape 4 : lier la sous-ressource à la ressource principale. Pour créer ce lien, il faut créer deux champs dans la *Ressource principale* :

- (1) un champ contenant le nom de la sous-ressource. Ce champ sera visible et comportera l'**hyperlien** vers la sous-ressource.
- (2) un champ contenant le lien vers la sous-ressource. Ce champ sera visible et comportera l'**hyperlien** vers la sous-ressource. Pour cela, sélectionner *Depuis une sous-ressource* dans *source de la valeur*. Dans l'affichage choisir *URL — Lien interne* et choisir comme étiquette le code alphanumérique du champ créé en (1).

GÉNÉRAL AFFICHAGE ANNOTATIONS SÉMANTIQUE

Étiquette
Langue détectée

icône(s) du champ

Nom interne
Lien vers sous ressource

Source de la valeur

VALEUR ARBITRAIRE CHOIX DE LA ROUTINE DONNÉE PRÉCALCULÉE COLONNE(S) EXISTANTE(S) DEPUIS UNE SOUS-RESSOURCE

Sous-ressource
Langues signées

Colonne(s) existante(s)

Aperçu de la valeur*

Langue détectée

["uid:/acc5598f0a38c073a08f31d53d9cdd5...]

["uid:/a55493c6db49b23e15e077c203d202...]

["uid:/a55493c6db49b23e15e077c203d202...]

["uid:/d0a0b5ab507199594f22ab2d8ef86f9...]

["uid:/a55493c6db49b23e15e077c203d202...]

["uid:/a55493c6db49b23e15e077c203d202...]

["uid:/a55493c6db49b23e15e077c203d202...]

GÉNÉRAL AFFICHAGE ANNOTATIONS SÉMANTIQUE

Paramètres du format

Appliquer un format
URL - Lien interne

Paramètres d'affichage

Comment l'étiquette est-elle définie ?
L'étiquette est le contenu d'une autre colonne

La colonne
tDvN

ANNULER CONFIRMER

Ce champ contient l'URL qui renvoie vers la sous-ressource

mais l'affichage choisi est le nom de la langue.

Création et gestion de sous-ressources

Bilan

The screenshot shows the LODEX interface for creating sub-resources. The main content area contains the following fields:

- [Jzko] Éditeur
- [koIO] Type de publication
- [Bj2d] Type de document
- [Fb8z] Mots-clés d'auteur(s)
- [YktD] Publications similaires (mots-clés)
- [qE59] Note
- [aRmn] Catégories Scopus
- [moX1] Catégories Scopus
- [Ls8V] Auteurs cités
- [qiNz] Langue détectée (Lien vers sous ressource)
- [tDvN] Langue signée (Etiquette pour lien vers sous ressource)

La ressource principale contient deux nouveaux champs. Le champ *Langue détectée* permet d'afficher l'hyperlien en front. Il contient le lien qui s'affichera sous forme verbalisée grâce à l'étiquette définie dans le champ *Langue signée*.

Création et gestion de sous-ressources

Bilan

Résultat depuis une page ressource

A Complete System for the Specification and the Generation of Sign Language Gestures



Auteur(s)

1. T. Lebourque
2. S. Gibet

Affiliation(s)

1. LIMSI-CNRS, Université Paris-Sud, BP 133, 91403, Orsay, France
2. LIMSI-CNRS, Université Paris-Sud, BP 133, 91403, Orsay, France

Résumé

This paper describes a system called GeSsyCa which is able to produce synthetic sign language gestures from a high level specification. This specification is made with a language based both on a discrete description of space, and on a movement decomposition inspired from sign language gestures. Communication gestures are represented through symbolic commands which can be described by qualitative data, and introduced in the spatio-temporal targets driving a generation system. Such an approach is possible for the class of gestures controlled through key-point information. The generation model used in our approach is composed of a set of sensori-motor servo-loops which solves in real time the inversion of the servo-loop, from the direct specification of location targets, while satisfying psychophysical constraints. The whole control system is applied to the synthesis of communication and sign language gestures, and a validation of the generated movements is presented.

Année de publication

1999

Revue de bibliographie

Gesture-based Communication in Human-Computer Interaction

Langue détectée

[French Sign Language](#)

Langue détectée

[French Sign Language](#)

Exemple d'une page sous-ressource

French Sign Language

Acronyme

Alias :

LSF

EN

Nombre de signeur-euses

Number of speakers :

283000

Document associés

Titre ↑

Id

[A Combined Semantic and Motion Capture Database for Real-Time Sign Language Synthesis](#)

ark:/67375/HCB-SS0K7R6C-L

[A Complete System for the Specification and the Generation of Sign Language Gestures](#)

ark:/67375/HCB-M0SF8KV6-8

05

Enrichissement via des requêtes SPARQL

Enrichissement via des requêtes SPARQL

Qu'est-ce qu'une requête SPARQL ?

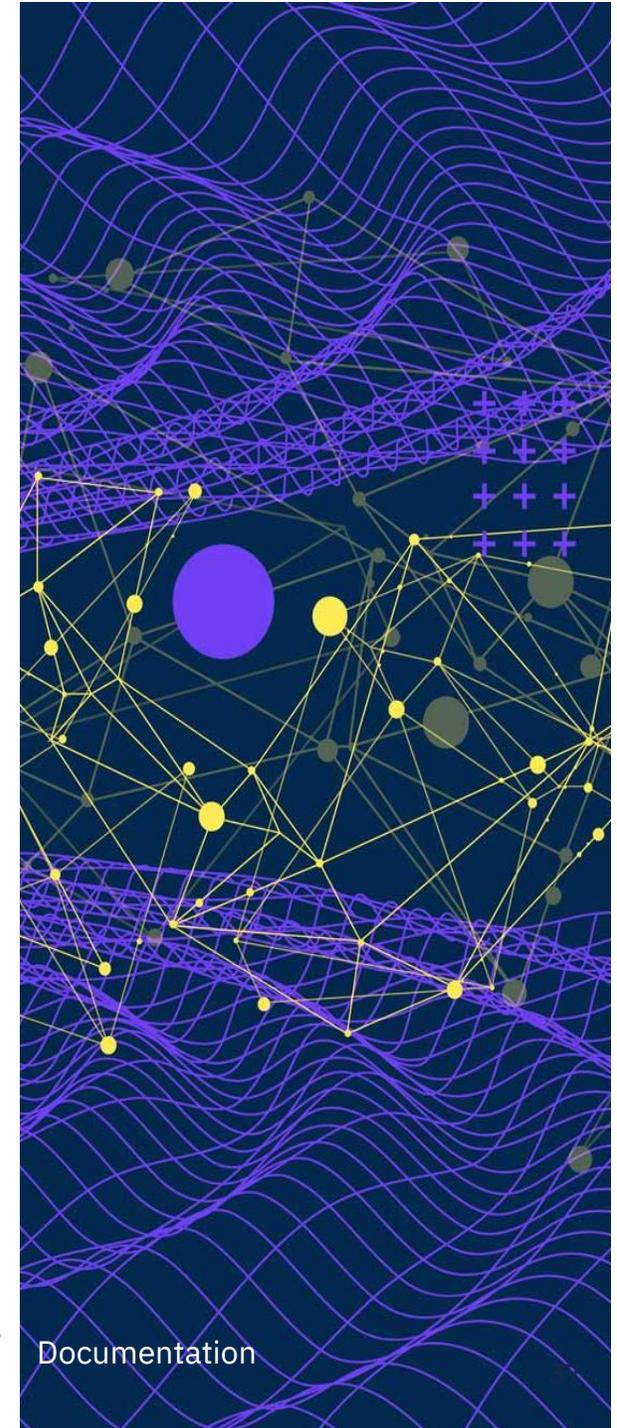
SPARQL est un langage de requête utilisé pour interroger des bases de données RDF (Resource Description Framework). RDF est un modèle de données utilisé dans le **Web sémantique** pour représenter des informations sous forme de triplets :

- Sujet (ce dont on parle)
- Prédicat (la propriété ou relation)
- Objet (la valeur ou la cible de la relation)

SPARQL permet de poser des questions sur des graphes RDF. Un **SPARQL endpoint** (ou **point d'accès SPARQL**) est une **interface web** qui permet aux utilisateur·ices et aux programmes d'interroger une base de données RDF via des requêtes **SPARQL**.

```
@prefix wd: <http://www.wikidata.org/entity/> .
@prefix wdt: <http://www.wikidata.org/prop/direct/> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .

wd:Q14759 a wd:Q34228 ;
  rdfs:label "American Sign Language"@en ;
  wdt:P31 wd:Q34228 ;          # instance of sign language (language)
  wdt:P424 "ase" ;           # code ISO 639-3 pour ASL
  wdt:P1098 "250000"^^<http://www.w3.org/2001/XMLSchema#integer> .
```



Enrichissement via des requêtes SPARQL

Ajouter le nombre de signeur·euses dans les sous-ressources

Étape 1 : dans la sous-ressource, créer un nouveau champ à partir d'une colonne existante et sélectionner la valeur de *Infos* (contenant l'identifiant Wikidata).

The screenshot shows the LODEX interface with a green header and a dark sidebar. The main content area is divided into four tabs: GÉNÉRAL, AFFICHAGE, ANNOTATIONS, and SÉMANTIQUE. The GÉNÉRAL tab is active, showing a form to create a new field. The field name is 'Nombre de signeur-euses'. Below the name, there are icons for field types: home, list, selected (highlighted), funnel, and bar chart. The 'Source de la valeur' section has four options: VALEUR ARBITRAIRE, CHOIX DE LA ROUTINE, DONNÉE PRÉCALCULÉE, and COLONNE(S) EXISTANTE(S) (highlighted in green). Below this, a dropdown menu shows 'Infos' selected. At the bottom, there is a button 'AJOUTER UNE OPÉRATION' and a 'TOUT SUPPRIMER' button. On the right, a preview panel shows the field name and several Wikidata identifiers in quotes.

LODEX DONNÉES AFFICHAGE ANNOTATIONS

GÉNÉRAL AFFICHAGE ANNOTATIONS SÉMANTIQUE

Étiquette
Nombre de signeur-euses

Icône(s) du champ
 Nom interne

Source de la valeur

VALEUR ARBITRAIRE CHOIX DE LA ROUTINE DONNÉE PRÉCALCULÉE COLONNE(S) EXISTANTE(S)

Colonne(s) existante(s)
Infos

Opérations de transformation TOUT SUPPRIMER

Aperçu de
Nombre de sig
["Q375619"]
[]
["Q36885"]
["Q375619"]
["Q375619"]
[]
["Q14759"]
[]
* aperçu de la valeur e
opérations de transfor
appliquées

AJOUTER UNE OPÉRATION

Enrichissement via des requêtes SPARQL

Ajouter le nombre de signeur·euses dans les sous-ressources

Étape 2 : dans *Affichage* sélectionner *SPARQL - Texte*, renseigner l'adresse du point d'accès Wikidata et écrire la requête.

Paramètres du format

Appliquer un format

SPARQL - Texte

Paramètres des Données

Endpoint SPARQL *

https://query.wikidata.org/sparql

Requête SPARQL

```
SELECT ?numberOfSpeakers WHERE {  
  wd:?? wdt:P1098 ?numberOfSpeakers.  
}
```

Valideur SPARQL

Valeur maximum à afficher

1

Cacher le champ de la ressource

Ajouter le séparateur pour transformer en liste

`SELECT ?numberOfSpeakers` : récupérer la valeur de la variable `?numberOfSpeakers`

`WHERE { ... }` : chercher les triplets qui correspondent à la forme donnée

`wd:??` : identifiant Wikidata d'une entité précise

`wdt:` : préfixe pour accéder à la **valeur directe** d'une propriété Wikidata

`P1098` : propriété "**nombre de locuteurs**" dans Wikidata (Number of speakers)

`?numberOfSpeakers` : variable qui contiendra la valeur liée à la propriété `P1098`

Enrichissement via des requêtes SPARQL

Ajouter le nombre de signeur·euses dans les sous-ressources

Étape 3 : paramétrer l’affichage du résultat de la requête.

Paramètres d’affichage ^

+ Ajouter un sous-format

Attribut ×
?numberOfSpeakers

Appliquer un format
Texte - Paragraphe ▾ 📄

Paramètres d’affichage ^

Pourcentage de la largeur
100 % ▾

Couleur
Chaîne de code couleur
#2b83ba
■

Enrichissement via des requêtes SPARQL

Bilan

French Sign Language

Acronyme

Alias :

LSF

EN

Nombre de signeur-euses

Number of speakers :

283000

Document associés

Titre ↑

Id

[A Combined Semantic and Motion Capture Database for Real-Time Sign Language Synthesis](#)

[ark:/67375/HCB-SS0K7R6C-L](https://nfdi.ub.uni-leipzig.de/ark:/67375/HCB-SS0K7R6C-L)

[A Complete System for the Specification and the Generation of Sign Language Gestures](#)

[ark:/67375/HCB-M0SF8KV6-8](https://nfdi.ub.uni-leipzig.de/ark:/67375/HCB-M0SF8KV6-8)

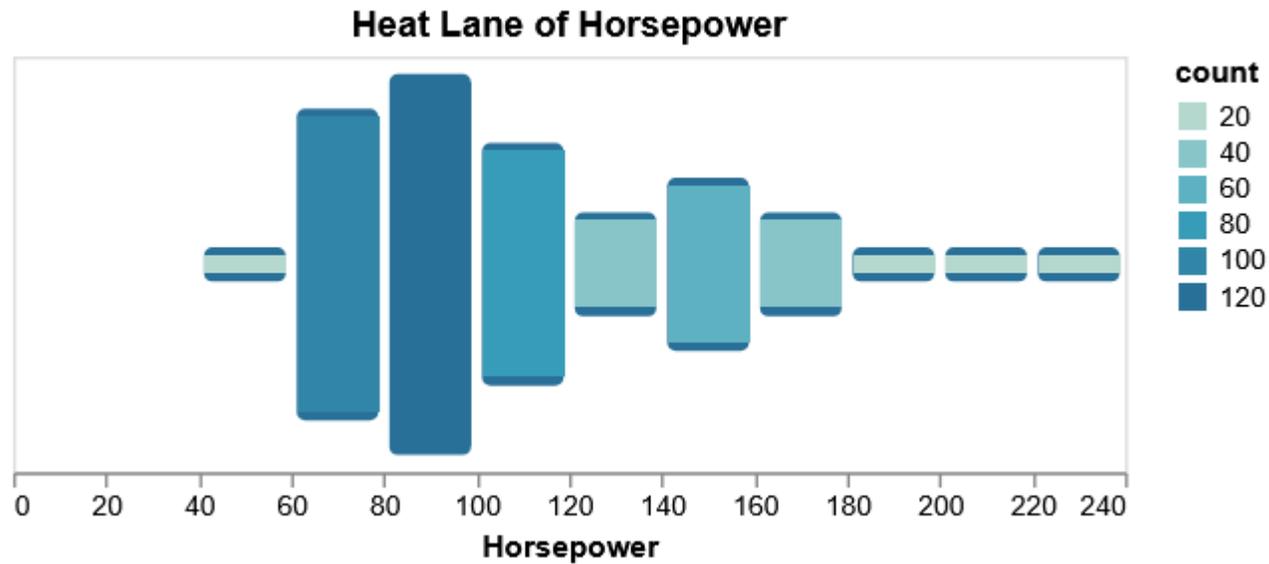
06

Visualisation avancée avec Vega-Lite

Visualisation avancée avec Vega-Lite

Qu'est-ce que Vega-Lite ?

Vega-Lite est un **langage déclaratif de visualisation de données**. Il permet de **décrire des graphiques (barres, lignes, cartes, etc.) en JSON**, de manière simple, lisible, et **facilement intégrable** dans des outils (Notebook, dashboards web).



Visualisation avancée avec Vega-Lite

Comment créer un graphique Vega-Lite ?

Étape 1 : créer un nouveau champ depuis l'onglet *Graphiques*. Donner un nom au nouveau graphique et sélectionner la routine *pairing-with* qui va croiser les données de deux champs (source, target, weight). Ici on souhaite obtenir le nombre de langues signées citées par années dans un histogramme empilé.

Étiquette

Évolution des études sur les langues signées

Icône(s) du champ

Nom interne

Source de la valeur

VALEUR ARBITRAIRE

CHOIX DE LA ROUTINE

DONNÉE PRÉCALCULÉE

COLONNE(S) EXISTANTE(S)

Choisir une routine

pairing-with

Champs de la routine

Champ N°1

Année de publication [zODA]

Champ N°2

Langue signée [tDvN]

Visualisation avancée avec Vega-Lite

Comment créer un graphique Vega-Lite ?

Étape 2 : paramétrer l’affichage du graphique. Dans l’onglet *Affichage* en sélectionnant le format *Graphique* — *Syntaxe Vega-Lite*, on peut accéder à une console capable d’interpréter la syntaxe Vega-Lite. Coller le [code souhaité](#). Il est possible d’utiliser le *Valideur Vega* en ligne pour améliorer le rendu du graphique.



Paramètres du format

Appliquer un format

Graphique - Syntaxe Vega-Lite

Paramètres des Données

Paramètres du Graphique

Valideur Vega

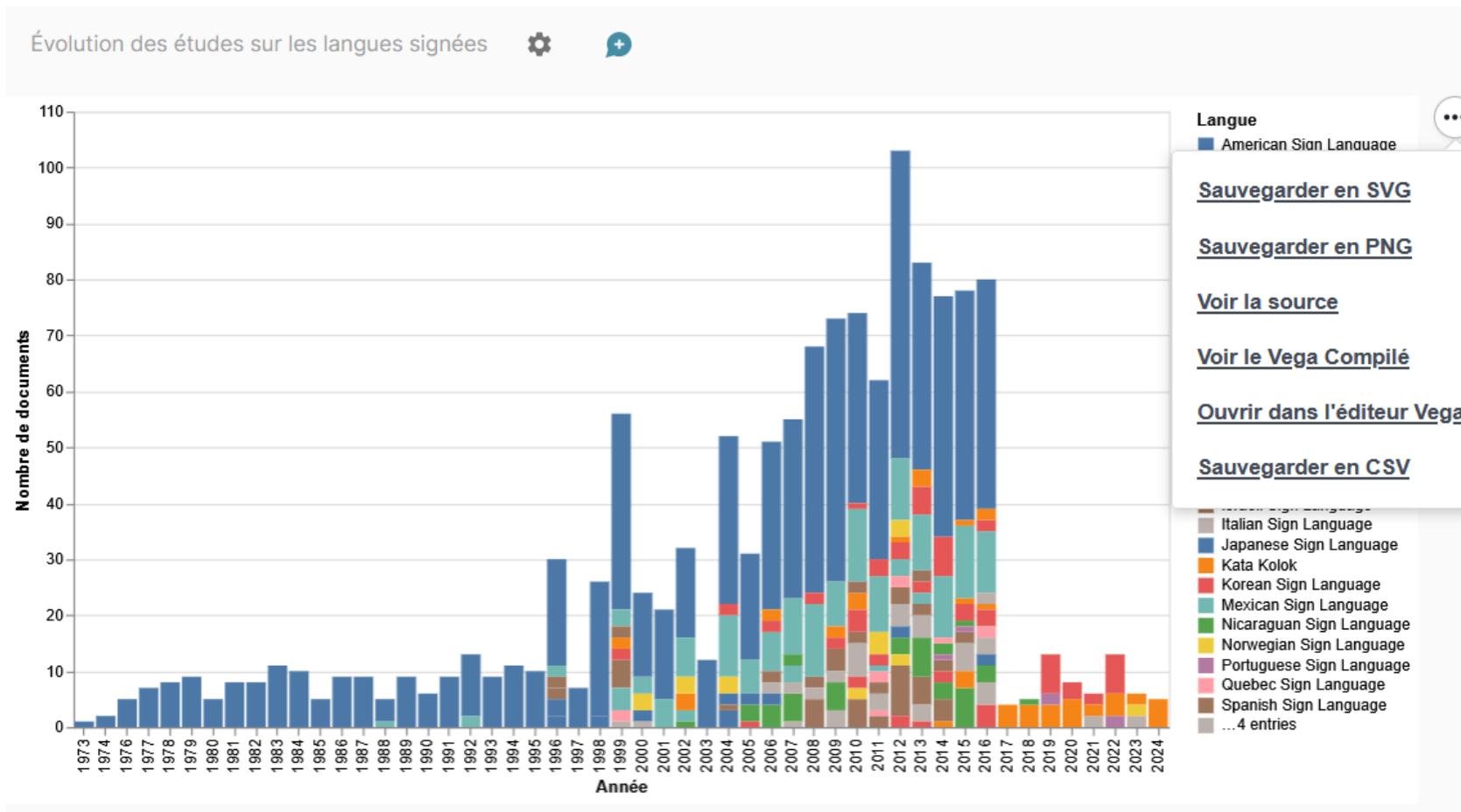
Vous pouvez utiliser ces variables pour définir la taille de manière dynamique :

- `{[_LODEX_WIDTH_]}` (Remplacée par la largeur du conteneur du graphique)
- `{[_LODEX_HEIGHT_]}` (Remplacée par la hauteur du conteneur du graphique)
- `container` (Variable Vega-Lite pour 'width' et 'height' permettent de fixer la taille du graphique à la même taille que le conteneur parent)

```
6-   "values": [  
7-     {"source": "2016", "target": "French Sign Language", "weight": 1},  
8-     {"source": "2006", "target": "French Sign Language", "weight": 1},  
9-     {"source": "2006", "target": "Israeli Sign Language", "weight": 1}  
10-  ],  
11- },  
12- },  
13- "mark": "bar",  
14- "encoding": {  
15-   "x": {  
16-     "field": "source",  
17-     "type": "ordinal",  
18-     "title": "Année",  
19-     "sort": "ascending"
```

Visualisation avancée avec Vega-Lite

Résultat



Pour conclure....

Exploration et analyse avancée de corpus avec Lodex



[Accéder à la documentation](#)



[Demander un hébergement](#)



[S'inscrire à la liste de discussion](#)



[Me contacter](#)



Merci pour votre attention !

Un rendez-vous
pour manipuler
lors d'un atelier

