



« Consignes d'écriture à suivre pour entrer une nouvelle entité dans la base de connaissance LaCAS data »

Auteurs : Elisabeth de Pablo, Christine Ho,
Raphaëlle Hervé.

Paris, Inalco, MAJ mai 2024

TABLE DES MATIERES

Introduction	3
1. Remarques préliminaires sur la création des entités.....	5
2. Consignes générales à respecter pour entrer et nommer une entité	8
2.1. Intitulés	8
2.2. Noms communs	8
2.3. Evènements historiques datés	9
2.4. Noms géographiques	12
2.5. Noms propres	14
2.6. Noms d'institution.....	15
Annexe	16

INTRODUCTION

Nous présentons ici quelques consignes d'écriture ainsi que des règles typographiques et orthotypographiques en usage à l'Imprimerie Nationale, dans le but de standardiser au mieux la **dénomination des entités**, exprimant un concept ou une notion, de **la base de connaissances du Domaine LaCAS**.

Le comité terminologique vérifie la forme des vedettes, contrôle l'orthographe, l'orthotypographie (usage des accents, majuscules, abréviations, manière d'écrire), la polysémie et la synonymie afin de proposer des points d'accès homogènes, de réduire l'ambiguïté du langage naturel, et de multiplier les points d'accès à partir des termes « équivalents ».

Pour rappel :

Les entités constituent un « langage contrôlé » :

- Chaque entité est représentée, décrite et insérée dans un graphe de connaissances, représentant un petit corpus de données. On passe ainsi d'un vocabulaire contrôlé à une encyclopédie plus ou moins développée selon les domaines.
- Chaque entité peut être interrogée, devenir une sorte de « mini-observatoire », ou encore être sélectionnée pour créer des conférences, des débats, des séminaires, etc.
- Les entités permettent de cartographier les études aréales et d'enrichir le « vocabulaire spécifique » de l'Inalco.
- Elles permettent d'indexer le contenu d'un document, tous formats confondus (textes, images, sons, audiovisuels), lors de son analyse, afin que l'utilisateur puisse accéder aux documents indexés ainsi qu'à leurs documents liés.

Au moment de la création d'une entité, des doutes peuvent apparaître :

- Il existe, par exemple, des entités qui n'existent pas en tant que telles dans les thésaurus BNF Rameau, IdREF, Loterre, qui peuvent ou non être présentes sur Wikipédia, Wikidata, etc. Il est alors important – lors de la création – de ne pas oublier de « fairiser » ces entités.
- Attention au « vocabulaire spécifique » relatif aux études aréales qui ne se trouvent pas dans les thésaurus.

Par exemple, la locution « études togolaises » n'existe pas en tant que telle, alors qu'on trouve fréquemment « études japonaises » ou « japanologie ». Il convient dès lors de créer cette locution sans oublier de compléter son graphe avec les langues, les peuples, les faits sociaux, les politiques, etc. associés.

Enfin, il s'agit de ne pas perdre de vue que **dénommer** au plus juste une entité va permettre **d'indexer** des documents (textes, images, audiovisuels, sons) et faciliter les **recherches** de l'utilisateur dans la base.

1. REMARQUES PRELIMINAIRES SUR LA CREATION DES ENTITES

Les remarques terminologiques sont issues du savoir-faire de la BNF et de la spécificité des dix domaines de connaissance LaCAS.

Nommer une entité

L'intitulé d'une entité – sous forme de mots ou de locutions en langage naturel - est choisi comme terme vedette (ou tête de vedette = élément initial du point d'entrée). Selon sa forme et/ou son usage, nous privilégions le terme le plus couramment attesté en français.

À cet intitulé peuvent s'ajouter des **termes équivalents** (ou encore termes rejetés, ou alternatifs – Cf. BNF). Ces derniers sont ajoutés – automatiquement via l'alignement -- ou manuellement si non aligné -. Ils apparaissent sous l'onglet « Autre information » et sous l'appellation « Expression équivalente ». C'est également sous cet onglet que l'on trouve les autres « noms » possibles de la même entité sous les appellations : « Locution », « Nom public », « Nom court », « Intitulé alternatif », « Variante orthographique », « Expression en langue originale », « Translittération », « Transcription », et « Expression historique ».

Les autres noms possibles sont à vérifier et à documenter en fonction de l'entité. **Documenter l'entité** revient à vérifier et compléter autant que faire ce peut l'ensemble des champs des onglets proposés. Il ne faut pas oublier que le moteur de recherche général de LaCAS est un moteur « full-text » permettant des recherches dans l'ensemble du site sur toutes les formes orthographiques possibles (essentiel du point de vue utilisateur).

Concernant les termes polysémiques :

Afin de désambiguïser les termes polysémiques, nous préconisons l'ajout à la suite de l'intitulé de l'entité d'un qualificatif entre parenthèses, toujours **en minuscule**.

Nous présentons en annexe une liste des qualificatifs les plus usités.

Par exemple : Afar (langue) - Afar (peuple) - Afar (région), ou encore Mékong (cours d'eau) et Mékong (région).

En cas de spécification particulière, nous ajoutons également à la suite de l'intitulé de l'entité un qualificatif entre parenthèses, toujours en minuscule.

Par exemple, pour les études de genre et l'ajout d'une spécificité, nous aurons études de genre (religion).

Concernant la synonymie :

Afin d'éviter la dispersion de l'information, nous privilégions la forme couramment usitée par la BNF.

Par exemple, l'emploi du mot « Séismes » – auquel on ajoutera une précision de lieu si besoin – est privilégié à son synonyme « Tremblements de terre » – qui devra apparaître dans les termes équivalents.

Remarques concernant la suppression et fusion d'entités – de notices

Attention, la fusion ne concerne que les doublons et non les concepts.

Lors de la fusion, les fiches doublent leur information réparties dans les onglets. Il convient donc de vérifier ses fiches et de les nettoyer si besoin après la fusion.

Les chapitres suivants apportent plus d'informations quant à la forme et aux principes. Nos observations sont réalisées sur des cas concrets et illustrer par des exemples. .

Quelques liens utiles

Catalogue BNF pour effectuer une vérification au préalable :

<https://catalogue.bnf.fr/recherche-sujets.do?pageRech=rsu>

Guide d'information Rameau à télécharger :

https://rameau.bnf.fr/docs_reference/guide_rameau

Règles typographiques générales :

<https://www.lacorreption.fr/post/regles-de-typographie-et-d-orthotypographie>

Grands principes rédactionnels – Etat français :

<https://www.gouvernement.fr/charte/charte-des-grands-principes-redactionnels/quelques-bonnes-pratiques-redactionnelles>

Règles typographiques, Le Robert en ligne :

<https://dictionnaire.lerobert.com/guide/titres-d-oeuvres-de-journaux-de-magazines-regles-typographiques>

<https://dictionnaire.lerobert.com/guide/qu-est-ce-que-la-typographie>

Bibliographie

Lexique des règles typographiques en usage à l'imprimerie nationale.
Editions Imprimerie nationale (beaucoup de réédition).

Compte tenu du travail d'indexation toujours en cours et de l'évolution régulière de la base de connaissance, le guide est mis à jour régulièrement.

2. CONSIGNES GÉNÉRALES À RESPECTER POUR ENTRER ET NOMMER UNE ENTITÉ

2.1. Intitulés

De façon générale

De façon générale, dans les titres, il est proscrit d'entrer ou d'ajouter : le / (slash), le ; (point virgule), la | (barre verticale), le _ (tiret bas) et les - (tirets).

Nous privilégions l'usage du mot ou de la locution en français, sauf pour les emprunts.

Nous privilégions également l'usage de la forme la plus couramment attestée (cf. supra voir synonymie « séisme ») idem dans le cas d'une locution (ex. Jeux olympiques d'été de 2024).

Voir les notations de datation (cf. infra). Toujours entre parenthèses sans espace et sans espace non plus en cas de tiret de date.

2.2. Noms communs

Les noms communs quantifiables doivent toujours être entrés au pluriel, sauf « quand la grammaire ou le sens l'interdisent »¹.

Ex. :. le nom "acteur" saisi au singulier, renvoie à la forme au pluriel "acteurs", il convient donc d'entrer « Acteurs » au pluriel et de rajouter si nécessaire la forme au singulier en terme rejeté (cad en « expression équivalente »).

¹ Cf. Guide Rameau 7^{ème} édition.

<u>Acteurs</u>	1030
Acteurs -- Agents Voir : <u>Agents théâtraux</u>	2
Acteurs -- Agents Voir : <u>Agents de cinéma</u>	3
Acteurs -- Direction artistique Voir : <u>Direction d'acteur</u>	14

2.3. Evènements historiques datés

La nomination d'événements historiques commence toujours par le nom de l'événement suivi de - *entre parenthèses et sans espace entre le tiret* - la datation :

Ex : Apartheid (1948-1991)

De même pour les siècles – Attention à la forme écrite des siècles, nous gardons la forme romaine suivie de l'abréviation « e » pour ième et du mot siècle.

Ex : Colonisation de l'époque moderne (XVe-XXe siècle)
Ici la datation (XVe-XXe siècle) = du XVe au XXe siècle

Attention pour la datation, le millésime n'est jamais abrégé :

Ex. Guerre mondiale (1939-1945) et non Guerre mondiale (1939-45)

L'abréviation est tolérée quand on parle de certaines dates historiques comme : « Les principes de 89 », par exemple.

Les cas des années

Années 1960 – signifie les années soixante – une période qui va de 1960 à 1969 et non 1960 à 1970.

Nous privilégions la forme chiffrée, cad : « Années 1960 », et en termes équivalents, nous ajoutons les déclinaisons suivantes : 1960 (décennie), Années 60, Sixties, années soixante,...

Exemple de cas particuliers, les noms de guerre

1/ Première Guerre mondiale - ou - Guerre mondiale (1914-1918)

Sur BNF Rameau, le terme vedette est **Guerre mondiale (1914-1918)**, en termes rejetés - c'est-à-dire en terme équivalent dans notre base – nous trouvons :

1ère guerre mondiale
1re guerre mondiale
Grande guerre (1914-1918)
Guerre de 1914-1918
Première guerre mondiale

Sur Loterre, nous trouvons aussi bien **Première Guerre mondiale**, que **Guerre mondiale, 1914-1918**

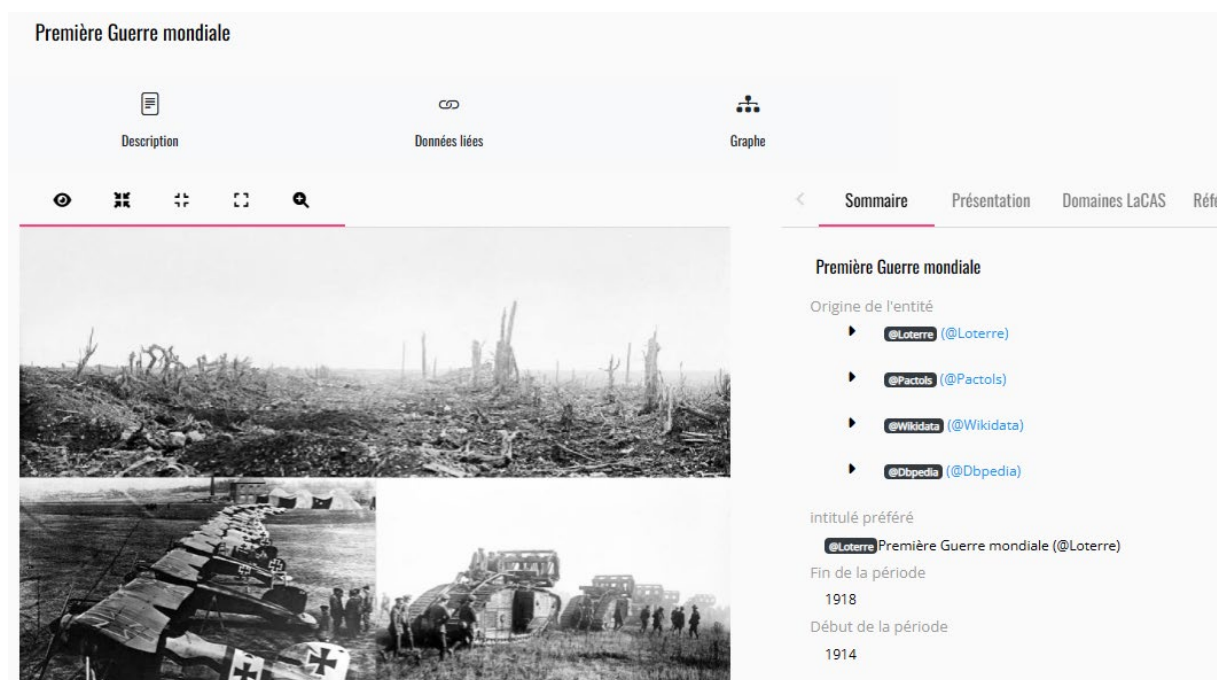
Nos remarques :

Ici, nous avons choisi – de prendre la forme la plus usitée dans les ouvrages de référence : **Première Guerre mondiale** comme terme vedette. Avec une majuscule à Première et à Guerre et une minuscule à mondiale.

Dans les termes équivalents, nous trouverons toutes les formes possibles. La datation sera également notée dans « Date de début » et « Date de fin » sur la page Sommaire et indiquée sous l'onglet « Domaine Lacas ».

Première Guerre mondiale

Description Données liées Graphe



2/ Concernant les autres entités incluant le mot « guerre » et dérivés

Pour mettre une majuscule au mot "guerre", il faut savoir dans quel contexte il est utilisé. En général, les noms de guerres sont constitués d'un nom générique (guerre, bataille, etc.) accompagné d'un nom spécifique qui caractérise ce nom générique, lui donne son unicité. On met une majuscule uniquement au nom caractéristique.

Ex. on écrit "la guerre de Cent Ans", "les guerres de Religion", "la guerre de Succession de Pologne", "la guerre des Six-Jours", "la bataille de Stalingrad", "la guerre du Golfe", "la bataille des Nations", "l'expédition des Mille", "la guerre des Deux-Roses", "la guerre de Sécession", etc.

Cf. <https://dictionnaire.lerobert.com/guide/guerre-et-bataille-majuscule-ou-minuscule>

En revanche, on écrit en minuscules l'ensemble de la dénomination lorsque le terme générique est déterminé par un ou plusieurs adjectifs, ou par une date.

Ex. on écrit "la quatrième croisade", "la guerre de 1914-1918", "la première guerre punique", "la dernière guerre mondiale", "la guerre russo-japonaise de 1904-1905", "la guerre froide".

Exception, on place une majuscule initiale au terme générique ainsi qu'à l'adjectif qui le précède dans les dénominations suivantes : "la Guerre folle", "la Grande Guerre", "la Première Guerre mondiale", "la Seconde Guerre mondiale".

Autres entités historiques :

Pour une locution spécifique non datée comme par ex. « Traite des esclaves » il ne faut pas ajouter de datation.

Attention à ne pas oublier de documenter le terme c'est-à-dire de toujours le dater dans le champ "historique" du Domaine LaCAS et de le localiser dans le champs « géographique ».

2.4. Noms géographiques

Les noms géographiques sont des noms propres particuliers. Il convient de se référer aux usages en cours :

👉 <https://dictionnaire.lerobert.com/guide/noms-geographiques-regles-typographiques>
(cf.rubrique noms géographiques)

Précisions dans les cas de polysémie :

Nous avons, **pour les fleuves et autres rivières**, privilégié la locution (**cours d'eau**) à mettre entre parenthèses à la suite de l'intitulé. Ex. Marne (cours d'eau) – ne pas oublier de localiser dans l'onglet Domaine LaCAS.

Pour les montagnes, on garde le nom « propre » de la montagne, suivi de l'information géographique. Ex. Altaï (**chaîne de montagnes**). Sans oublier de noter la localisation.

Attention aux cas multiples. Ex. Mékong (cours d'eau), Mékong (région) ou encore Oise (département) à dissocier de Oise (cours d'eau). (et voir Cf.infra).

A noter que l'on rencontre des entités – nom de fleuve ou de rivière par exemple) seules car sans problème de polysémie
Ex. Maroni ou encore Orénoque

Nous rencontrons également des noms de rivières précédés ou suivis de leur attribut en langue originale

Ex. Rio Tambo, Yangzi Jiang, Fleuve Sénégal

Ne pas oublier de localiser et documenter ces entités dans l'onglet « Domaine LaCAS » et de vérifier et compléter si besoin l'onglet « Autres Informations ».

Divisions et subdivisions administratives :

Il est très difficile de normaliser l'ensemble des références liées aux organisations, divisions et subdivisions administratives utilisées dans les pays et portant des noms différents (Ex. département, région, canton, land, comarca, comté, état, ...).

Afin de faire au plus simple et après consultation et comparaison de moult thesaurus en usage, nous avons opté pour le nom le plus usité (en usage) de l'entité suivi entre parenthèses des spécifications allant du plus particulier au plus générique.

Nous gardons les nominations administratives en usage dans les pays où se rapportent l'entité (cad, nous gardons le mot « land » pour les divisions en Autriche et en Allemagne, « canton » pour la Suisse, « barangay » pour les Philippines, ...

Pour suivre au plus près la BNF, nous privilégions le nom du pays après la virgule plutôt que son adjectif ou complément.

Ex. Salzbourg (land, Autriche) et non Salzbourg (land d'Autriche ou land autrichien)
Aisne (département) et non Aisne (département français).

Dans le cas, où il n'y a pas de problèmes de polysémie dans une entité géographique, nous pouvons omettre d'ajouter des spécificatifs supplémentaires.

Quelques exemples plus concrets :

Paris (ville, France) – Paris (collectivité à statut particulier, France), Paris (Illinois, Etats-Unis), Paris (Texas, Etats-Unis), ...

Mendoza (Mexique), Mendoza (Argentine) - ici on peut omettre l'ajout du mot « ville » si il n'y a qu'une seule ville dans le pays nommé (on doit donc toujours vérifier au préalable).

Mais nous aurons : Mendoza (Alava, Espagne) et Mendoza (Burgos, Espagne).
Et toujours Mendoza (cours d'eau) car l'entité désigne ici le fleuve.

Tecate (Basse-Californie, Mexique), Tecate (Oaxaca, Mexique)
Sachant que nous avons dans la base Basse-Californie (d'autres divisions administratives qu'il sera bon de noter dans le cas de synonymie).

Jackson (Mississippi, Etats-Unis), Jackson (Wyoming, Etats-Unis) Jackson (Michigan, Etats-Unis), Jackson (Tennessee, Etats-Unis) et Jackson (chaîne de montagnes)

Les spécifications de pays, ou de provinces, ... liées sont toujours à localiser et documenter dans le Domaine LaCAS.

Dans tous les cas, lors de l'alignement, il est nécessaire de vérifier dans l'onglet « Autres informations » que toutes les autres appellations de l'entité sont bien notées afin que n'importe quel utilisateur puisse retrouver facilement son entités.

On note également ici que les qualificatifs sont en minuscules, les noms « propres » gardent leur majuscule initiale.

Noms de divisions administratives modifiées dans le temps :

Exemple de changement de statut : Alsace
Jusqu'en 2016, l'Alsace était une région administrative composée de 2 départements, depuis elle est entrée dans la Région Grand Est, et en 2021, une « Collectivité européenne d'Alsace » a été créée lui conférant une certaine autonomie dans la région Grand Est...

Afin de simplifier, nous privilégions les indications comme sus expliquées auxquelles nous ajoutons la datation :

Alsace (région, France, 1956-2015)

Par contre, nous laissons comme entité la locution : « Collectivité européenne d'Alsace » et ajoutons dans le Domaine Lacas les références utiles.

Comté de Tyrol (1143-1919)

2.5. Noms propres

Nous avons retenu la forme courante suivante :

Nom de famille [virgule], Prénom – suivis des dates de la personne autant que faire ce peut).

Ex. Rodin, Auguste (1840-1917).

Attention, on ne met jamais mettre d'espace avant et après les parenthèses et avant et après le tiret dans les datations.

2.6 Noms d'institution

A des fins de cohérence avec les autres notices, il convient d'entrer :

- une seule fois le nom de l'institution sous sa forme développée, suivi entre parenthèses de l'acronyme (virgule) et du code.
Ex: Langage, langues et cultures d'Afrique (LLACAN, UMR 8235)
Ex : Centre d'études et de documentation économiques, juridiques et sociales (CEDEJ, UMIFRE-UAR 3123)

Attention les détails liés aux institutions doivent apparaître dans les onglets :

- **Sommaire** pour : l'adresse, l'acronyme, le statut de l'équipe, le pays, les affiliations, l'adresse postale (ensuite apparaissent les IdREF et fusionner avec).
- **Présentation** pour : description générale + ajouts détaillés si besoin

Il ne faut donc pas mentionner les noms des villes dans l'intitulé.

Attention aux sigles, nous privilégions le déroulé suivi du sigle entre parenthèses.

ANNEXE

Qualificatifs utilisés pour désambiguïser les entités

Note : La liste est régulièrement complétée et mise à jour

peuple	Usité pour les peuples vs langue
langue	Usité pour les langues vs peuple
cours d'eau	Usité pour les fleuves, rivières, ...
région	Usitée pour désambiguïser et dans la cas particulier d'une région géographique et non administrative (politique) Une région géographique peut s'étendre sur plusieurs pays.
ville	Usité pour désambiguïser, ex. ville vs région
Communauté autonome	Bien que pouvant être considéré comme une « division administrative » nous notons les communautés autonomes qui peuvent être composées de plusieurs régions.
chaines de montagne	Pour nommer l'ensemble de la chaines de montagne et non un massif en particulier
désert	Usitée pour désambiguïser
massif	Pour nommer un massif montagneux particulier
département, land, canton, wilaya, état, ...	Pour nommer et désambiguïser les organisations/divisions administratives des différents pays
religion	Usitée pour désambiguïser

A suivre...