



HAL
open science

Bonnes pratiques pour structurer un thésaurus

Emmanuelle Perrin

► **To cite this version:**

| Emmanuelle Perrin. Bonnes pratiques pour structurer un thésaurus. 2020. <hal-02987438>

HAL Id: hal-02987438

<https://hal.science/hal-02987438v1>

Submitted on 3 Jul 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License



Op
To

BONNES PRATIQUES

Bonnes pratiques pour structurer un thésaurus

PAR EMMANUELLE PERRIN · PUBLIÉ 28/09/2020 · MIS À JOUR 11/05/2022

[english version in progress]

[Les langages documentaires](#)

[La structure d'un thésaurus](#)

- [Les relations d'équivalence](#)
- [Les relations hiérarchiques](#)
 - [La relation générique \(genre/espèce\)](#)
 - [La relation partitive \(tout/partie\)](#)
 - [La relation d'instance \(classe/instance\)](#)
- [La polyhiérarchie](#)
- [Les termes de tête ou *Top Terms*](#)
- [Les relations associatives](#)

[Le modèle de données SKOS et sa mise en](#)

[oeuvre dans Opentheso](#)

[Diffuser son thésaurus](#)

[Bibliographie](#)

Les langages documentaires

Les langages documentaires sont des vocabulaires contrôlés ou normalisés, dont le but est de répondre à l'équivoque du langage naturel et aux problèmes que posent la polysémie et

l'homonymie. Ils servent à indexer et à décrire le contenu d'une ressource et visent à améliorer la recherche documentaire.

On distingue trois grands types de langages documentaires : les schémas de classification, les listes d'autorité ou de vedettes-matières et les thésaurus. Les classifications, à l'exemple de la Classification décimale de Dewey, propose une organisation hiérarchique des concepts. Les listes d'autorité ou vedettes sont des termes normalisés, qui doivent être obligatoirement utilisés dans le catalogage ou l'indexation, pour décrire le contenu des documents. On peut ici citer Rameau (Répertoire d'autorité-matière encyclopédique et alphabétique unifié), le langage d'indexation matière utilisé par la Bibliothèque nationale de France. Enfin un thésaurus structure un vocabulaire contrôlé en décrivant l'ensemble des relations sémantiques qui relient des concepts.

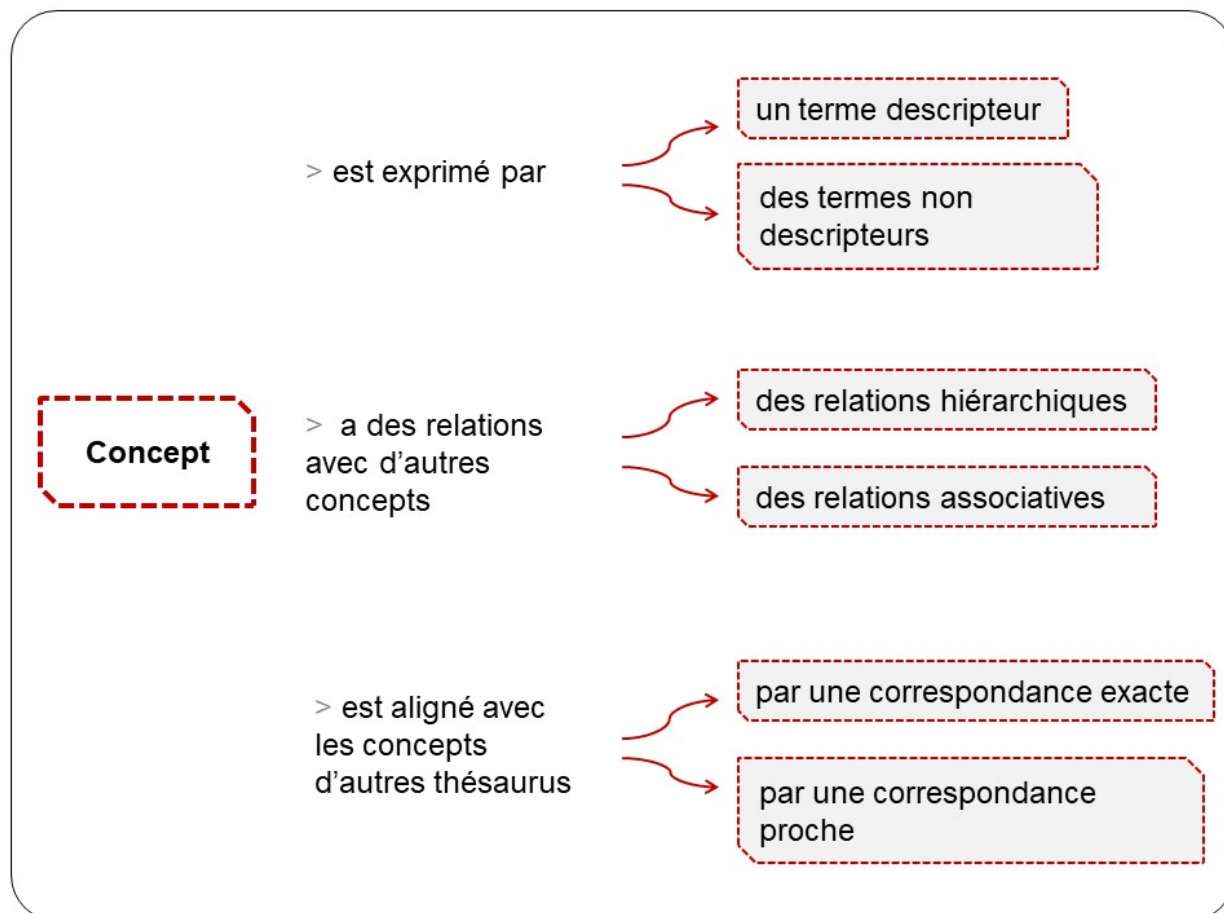
L'emploi de vocabulaires contrôlés est une recommandation de l'Agence nationale de la recherche pour l'évaluation de la qualité des données dans les plans de gestion des données. De même, selon les principes FAIR, « pour être FAIR, les données doivent être décrites à l'aide d'un vocabulaire contrôlé respectant lui-même les principes FAIR. Il doit être facile à trouver grâce à un identifiant pérenne et unique, documenté et lisible par les machines ».

La structure d'un thésaurus

La construction d'un thésaurus obéit à des normes ISO (ISO 25964-1:2011 et ISO 25964-2:2013).

La norme ISO 25964 distingue d'une part les concepts, comme "représentation mentale", et d'autre part, les termes, en tant expression linguistique d'un concept dans une langue donnée.

Un concept est exprimé par un terme préféré ou descripteur et des termes non-descripteurs. Il a des relations hiérarchiques et associatives avec d'autres concepts. Il est aligné avec les concepts d'autres thésaurus par des relations d'équivalences. Il peut être défini et documenté par différents types de notes.



La structure d'un thésaurus permet d'exprimer l'ensemble des relations sémantiques d'un concept.

▪ Les relations d'équivalence entre le terme préféré et ses synonymes

– Le terme préféré

Dans un système d'organisation des connaissances, le terme préféré (ou retenu, ou préférentiel) est utilisé pour se référer de manière non ambiguë à un concept. C'est le descripteur d'une indexation et le libellé qui apparaît à l'affichage.

⚠ Un thésaurus n'admet qu'un seul terme descripteur par concept et par langue.

⚠ Un même terme descripteur ne peut pas désigner plusieurs concepts. En cas d'homonymie, il convient de désambiguïser le champ d'application du terme en ajoutant un qualificatif entre parenthèses :

Bois

Bois (zoologie)

⚠ Un même terme ne peut être à la fois descripteur et non descripteur.

– Les synonymes

Le terme descripteur et les termes non descripteurs (ou rejetés, ou non préférentiels) qui représentent un même concept sont liés entre eux par une relation d'équivalence. Elle permet le contrôle des synonymes et de l'accès à l'information.

Selon les normes de construction d'un thésaurus, sont considérés comme synonymes :

- ◆ Noms complets et abréviations ou sigles : *imagerie par résonance magnétique* | *IRM*
- ◆ Noms courants et noms scientifiques : *araignées* | *Araneae*
- ◆ Noms scientifiques et noms de marque : *paracétamol* | *Efferalgan*
- ◆ Termes d'origines linguistiques différentes : *polychrome* | *multicolore*
- ◆ Variantes orthographiques : *clés* | *clefs*
- ◆ Appellations anciennes et modernes : *instituteurs* | *professeurs des écoles*

Les relations d'équivalence incluent également les quasi-synonymes. Il s'agit de termes trop spécifiques pour tous apparaître comme descripteurs selon le domaine traité par le thésaurus.

Végétarisme

EP : *lacto-ovo-végétarien, lacto-végétarien, ovo-végétarien, semi-végétarien, végétarien.*

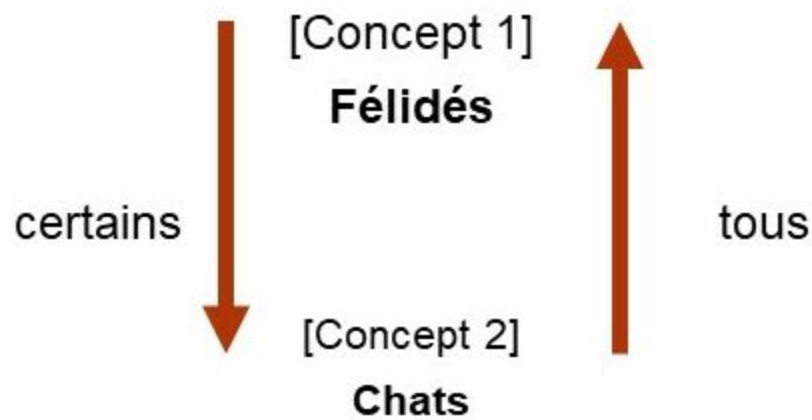
Les relations d'équivalence peuvent également englober les antonymes quand ils représentent deux aspects d'un même concept, par exemple *Alphabétisation* et *Analphabétisme*.

■ Les relations hiérarchiques

Une relation hiérarchique (ou générique-spécifique ou parents-enfants ou broader-narrower) doit relier, du plus général au plus précis, des concepts de même catégorie : entités, activités, agents, actions, propriétés.

Par exemple les termes *Musées* (entités), *Conservateurs* (agents) et *Muséographie* (action) n'appartiennent pas aux mêmes catégories et ne peuvent donc pas se trouver dans une relation hiérarchique.

Le test "certains/tous" permet de vérifier la validité de la relation hiérarchique :



Le test "certains/tous".

→ Certains félidés sont des chats ; tous les chats sont des félidés.

Les relations hiérarchiques générales peuvent être subdivisées en trois catégories :

– La relation générique

Sur le modèle de la relation genre-espèce, le concept représenté par le terme spécifique est un type ou une espèce particulière du concept représenté par le terme générique.

TGG Félidés

TSG Chats

TSG Lynx

TSG Panthera

TSG Puma

– La relation partitive

La relation hiérarchique tout-partie concernent :

- Les systèmes et organes du corps

TGP Appareil digestif

TSP Foie

TSP Pancréas

- Les lieux géographiques

TGP Canada

TSP Ontario

TSP Ottawa

TSP Toronto

- Les disciplines et leurs spécialités

TGP Médecine

TSP Cardiologie

TSP Chirurgie

TSP Dermatologie

- Les organisations hiérarchiques

TGP Corps d'armée

TSP Divisions

TSP Brigades

TSP Régiments

TSP Bataillons

– La relation d'instance

Cette relation caractérise le lien entre une catégorie ou une classe désignée par un nom commun et une instance de cette catégorie, représentée par un nom propre.

TGI Montagnes

TSI Alpes

TSI Pyrénées

▪ La polyhiérarchie

Chaque concept peut être relié à plusieurs autres concepts par des relations hiérarchiques.

TG Instruments de musique

TS Instruments à clavier

*TS **Piano***

TS Instruments à cordes

TS Instruments à cordes frappées

*TS **Piano***

TG Objet visuel

TS Motif géométrique

TS Motif géométrique cruciforme

*TS **Ankh***

TG Objets symboliques

TS Symbole

TS Ankh

■ Les termes de tête ou *Top Terms*

Un terme de tête ou *Top Term* est le terme préférentiel qui représente le concept de plus haut niveau dans la hiérarchie à laquelle il appartient. Il se situe donc à la “racine” du vocabulaire et constitue un point d’entrée dans l’une des branches d’un thésaurus.

■ Les relations associatives

Les relations associatives permettent de réunir des concepts qui ne sont pas liés hiérarchiquement mais sémantiquement. Pour associer deux concepts, la norme veut que l’un soit un élément nécessaire à la définition de l’autre. Le concept d’ “Approvisionnement en eau” est par exemple indispensable à la définition du concept de puits.

Les relations associatives sont réciproques. Elles permettent de décrire :

– La cause et l’effet :

Infections

TA Maladies infectieuses

– Un tout et sa partie :

Livres

TA Reliure

– Une action et un agent ou un instrument :

Jardinage

TA Jardiniers

– Une action et le produit de cette action :

Édition

TA Livres

– Une action et l’objet qui subit cette action :

Analyse de données

TA Données

– Une action et le lieu de cette action :

Enseignement primaire

TA Écoles primaires

– Une science et son objet :

Paléontologie

TA Fossiles

– Un objet ou une action et sa propriété :

Poisons

TA Toxicité

– Un objet ou une action et son rôle ou son but :

Citernes

TA Stockage de l'eau

– Un matériau brut et son produit :

Peaux

TA Cuir

– Un processus et son agent neutralisant :

Inflammation

TA Anti-inflammatoires

– Un objet et son origine :

Puits

TA Eaux souterraines

– Un objet ou un concept et une unité ou un mécanisme de mesure :

Température

TA Thermomètre

– Une profession et la personne qui exerce cette profession

Maçonnerie

TA Maçons

Les relations associatives permettent également d'associer des concepts "frères" qui partagent un même terme générique et dont la signification se chevauche (*navires et*

bateaux), des concepts contraires (*mariage et divorce*) et des concepts complémentaires (*enseignement et apprentissage*).

⚠ Deux concepts ne peuvent être liés à la fois par des relations hiérarchiques et par des relations associatives.

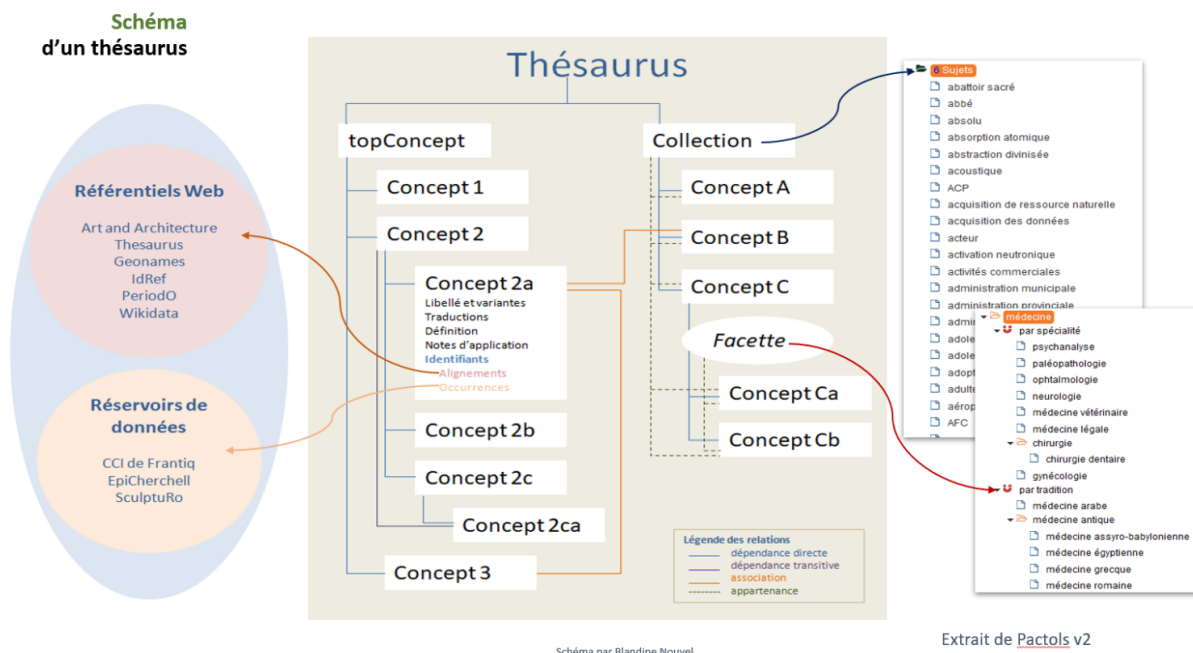
Type de relations	Sens de la relation	Terminologie en français	Abréviations	Terminologie en anglais	Abréviations
Relation d'équivalence	⇒ du descripteur vers le non-descripteur	Employé pour	EP	Used for	UF
Relation d'équivalence	⇒ du non-descripteur vers le descripteur	Employer	E ou EM	Use	
Relation hiérarchique	⇒ du terme spécifique au terme générique	Terme générique	TG	Broader Term	BT
Relation hiérarchique	⇒ du terme générique au terme spécifique	Terme spécifique	TS	Narrower Term	NT
Relation hiérarchique générique	⇒ du terme spécifique au terme générique	Terme générique (générique)	TGG	Broader term (generic)	BTG
Relation hiérarchique générique	⇒ du terme générique au terme spécifique	Terme spécifique (générique)	TSG	Narrower term (generic)	NTG
Relation hiérarchique partitive	⇒ du terme spécifique au terme générique	Terme générique (partitif)	TGP	Broader term (partitive)	BTP
Relation hiérarchique partitive	⇒ du terme générique au terme spécifique	Terme spécifique (partitif)	TSP	Narrower term (partitive)	NTP
Relation hiérarchique d'instance	⇒ du terme spécifique au terme générique	Terme générique (instance)	TGI	Broader term (instance)	BTI

Relation hiérarchique d'instance	⇒ du terme générique au terme spécifique	Terme spécifique (instance)	TSI	Narrower term (instance)	NTI
Relation hiérarchique	Concept le plus haut dans la hiérarchie	Terme de tête	TT	Top Term	TT
Relation associative	⇒ d'un descripteur à un autre descripteur	Terme associé ; Voir aussi	TA ; VA	Related Term	NT

Tableau récapitulatif des relations dans un thésaurus (d'après A. Serres, Introduction à l'indexation, <https://www.sites.univ-rennes2.fr/urfist/Supports/Indexation/Indexation4Langages.html>)

Le modèle de données SKOS et sa mise en œuvre dans Opentheso

Les thésaurus peuvent être décrits au format SKOS (*Simple Knowledge Organization System*), qui est une recommandation du W3C pour représenter des langages documentaires. Ce modèle de données permet de publier, d'échanger et de relier des vocabulaires structurés avec les technologies du Web sémantique.



Le modèle de données SKOS (schéma : Blandine Nouvel)

Les classes du modèle SKOS et leurs équivalents dans Opentheso :

Classes SKOS	Définition	Équivalent dans Opentheso
Libellés		
skos:prefLabel	Terme descripteur	Terme préféré
skos:altLabel	Terme non descripteur	Synonyme
skos:hiddenLabel	Le "libellé caché" permet d'inclure des écritures fautives des termes non descripteurs, intégrées à la recherche mais non visibles des utilisateurs	Synonyme caché
Relations sémantiques		
skos:broader	Concept générique	Concept générique
skos:narrower	Concept spécifique	Concept spécifique
skos:related	Concept associé	Concept associé
Annotations		
skos:note	Note	Note
skos:definition	Définition	Définition
skos:scopeNote	Note d'application, d'usage, d'emploi ou note explicative : brève explication sur les modalités d'emploi d'un terme descripteur.	Note de portée
skos:historyNote	Note historique : décrit un changement dans la forme ou la signification du concept	Note historique
skos:example	Exemple	Exemple
skos:editorialNote	Notes pour la gestion éditoriale du thésaurus (commentaires)	Note éditoriale
skos:changeNote	Notes pour la gestion éditoriale du thésaurus (documenter un changement)	Note de changement
Alignement ou équivalence avec des référentiels externes		
skos:exactMatch	Équivalence exacte	Équivalence exacte

Classes SKOS	Définition	Équivalent dans Opentheso
skos:closeMatch	Équivalence proche	Équivalence inexacte
skos:broadMatch	Lien d'alignement générique	Équivalence générique
skos:narrowMatch	Lien d'alignement spécifique	Équivalence spécifique
skos:relatedMatch	Lien d'alignement par une relation associative	Équivalence associative
Concept et collection de concepts		
skos:concept	Concept	Concept
skos:collection	Collection de concepts	Groupe
Notations		
skos:notation	Code identifiant un concept	Notation

Diffuser son thésaurus

1) Pour compléter la page d'accueil, cliquer sur « Modifier ».

2) Rédiger la présentation du thésaurus avec l'éditeur de texte et enregistrer.

Compléter la page d'accueil

Il est recommandé de rédiger une présentation de son thésaurus sur la page d'accueil. Il est utile en effet, pour la diffusion d'un thésaurus, d'en préciser les "métadonnées" — sujet, contexte et date de création, projet et équipe de recherche, institutions — et d'indiquer sous quelle licence de diffusion il est publié.

Citer cet article : « Bonnes pratiques pour structurer un thésaurus », par Emmanuelle Perrin. Publié sur *Opentheso* le 28/09/2020. Lien : <https://opentheso.hypotheses.org/67>.

Bibliographie

Aitchison, Jean, et Gilchrist, Alain, *Construire un thésaurus. Manuel pratique*, Paris, ADBS Éditions, 1992, 217 p.

Chichereau, Dominique, Contat, Odile, Dégez, Danièle et al., « Les normes de conception, gestion et maintenance de thésaurus. Évolutions récentes et perspectives », *Documentaliste-Sciences de l'Information*, 2007/1 (Vol. 44), p. 66-74 [[lire en ligne](#)].

Dalbin, Sylvie, Yakovleff, Nathalie, Zysman, Hélène, *ISO 25964-1- Thésaurus pour la recherche documentaire. Livre blanc*, Saint-Denis, Afnor, 2013, 37 p.

Essib, "Thésaurus", 2013 [[lire la notice en ligne](#)]

Glossaire des techniques de l'information, portail Joconde [[lire en ligne](#)].

Gouvernement du Québec, *Politique d'édition du thésaurus de l'activité gouvernementale* [[lire en ligne](#)].

Hudon, Michèle, « ISO 25964 : pour le développement, la gestion et l'interopérabilité des langages documentaires » *Documentation et bibliothèques*, vol. 58-3, 2012, p. 130–140 [[lire en ligne](#)].

Isaac, Antoine et Summers, Ed (ed.), *SKOS : Simple Knowledge Organization System Primer. W3C Working Group Note 18 August 2009* [[Lire en ligne](#)].

Keller, Loraine, *Encadrer la réingénierie d'un thesaurus*, mémoire pour obtenir le titre professionnel « Chef de projet en ingénierie documentaire », INTD, Conservatoire national des arts et métiers, Paris, 2013 [[lire en ligne](#)].

Serres, Alexandre, *Introduction à l'indexation*, URFIST Bretagne -Pays de Loire, 2003 [[lire en ligne](#)].

Alistair, Miles, Bechhofer, Sean, *SKOS : Système simple d'organisation de connaissances. Référence. Recommandation W3C du 18 août 2009*, trad. T. Francart [[Lire en ligne](#)].

Zeng, Marcia Lei, *Construction of Controlled Vocabularies, A Primer*, 2005 [[lire en ligne](#)]



Rechercher dans OpenEdition Search

Vous allez être redirigé vers OpenEdition Search

Dans tout OpenEdition

Dans Opentheso