

Description et indexation thématiques du corpus  
“l’histoire des inventions”

Cyril Grouin – [cyril@grouin.net](mailto:cyril@grouin.net)  
Thomas Szlamowicz – [szlam@free.fr](mailto:szlam@free.fr)

30 avril 2002

---

Sémiotique des médias  
Le genre du documentaire audiovisuel

Séminaire de DESS à l’Institut National des Langues  
et Civilisations Orientales (INaLCO)

Paris 2001–2002

# Table des matières

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Les taxèmes</b>	<b>4</b>
2.1	Liste canonique de taxèmes . . . . .	4
2.2	Tableau comparatif des taxèmes dans le corpus . . . . .	5
2.3	Annotation des taxèmes . . . . .	6
<b>3</b>	<b>Description des taxèmes</b>	<b>8</b>
3.1	Description taxinomique . . . . .	8
3.2	Description méréologique . . . . .	8
3.3	Représentation visuelle sous forme de scripts ou graphes conceptuels . . . . .	10
3.3.1	Graphe de base : les deux lectures possibles . . . . .	10
3.3.2	Sous-graphes : les différents taxèmes pour chaque thème	11
3.4	Enrichissement du dictionnaire de taxèmes . . . . .	16
<b>4</b>	<b>Extrait d'un thésaurus</b>	<b>18</b>
<b>5</b>	<b>Construction (d'un extrait) d'une ontologie</b>	<b>22</b>
<b>6</b>	<b>Établir une "bibliothèque de séquences AV"</b>	<b>27</b>
<b>7</b>	<b>Exemples d'indexation thématique de séquences AV</b>	<b>31</b>
<b>8</b>	<b>Description et indexation discursives de séquences AV</b>	<b>34</b>
<b>9</b>	<b>Description et indexation narratives de séquences AV</b>	<b>35</b>
<b>10</b>	<b>Thésaurus, lexiques, ontologies déjà existants</b>	<b>36</b>
10.1	Thésaurus des sujets de base du Gouvernement du Canada . . . . .	37
10.2	Thésaurus de Statistique Canada . . . . .	37
10.3	OECD Macrothesaurus . . . . .	38
10.4	Rodget's II Thesaurus . . . . .	38
10.5	Thésaurus de Péchoin : des idées aux mots, des mots aux idées...	39

# Chapitre 1

## Introduction

La première partie de notre travail consacrée à la description du corpus “l’histoire des inventions” a permis de mettre à jour les différentes séquences le composant. À partir de cette liste de séquences, nous allons dresser les taxèmes permettant une consultation ciblée de l’ensemble des documentaires, selon le type de lecture choisi. En effet, un corpus permet différentes lectures possibles, propres à chaque utilisateur.

Le but d’une description taxinomique consiste à indexer le contenu d’un document, en vue d’accéder directement à la séquence pertinente, et ainsi à l’information utile.

Notre documentaire admet – au moins – deux modes de lecture différents : une lecture fondée sur la diachronie, la seconde davantage axée sur les inventions. L’ontologie que nous nous proposons de construire, aura pour point de départ cette double vision. Ces deux approches autorisent une plus grande souplesse dans la recherche interne au documentaire, toutes les séquences se trouvant ainsi indexées sous ces deux entrées générales.

# Chapitre 2

## Les taxèmes

### 2.1 Liste canonique de taxèmes

C'est à partir de ces deux entrées générales, histoire et inventions, que vont découler tous nos taxèmes :

- Histoire : Préhistoire, Antiquité, époque Gallo-romaine, Moyen-Âge, Renaissance, XIX<sup>e</sup> siècle, XX<sup>e</sup> siècle ;
- Inventions : lieux d'invention, musées, maisons d'inventeurs, domaine farfelu, domaine technologique, domaine scientifique, domaine technique, spécialistes, inventeurs vivants, inventeurs décédés, exhibition d'inventions, démonstration d'inventions.

Chacun de ces taxèmes, est lié à un thème et un seul, ce qui assure une homogénéité à la structure taxinomique :

- Époques : Préhistoire, Antiquité, époque Gallo-romaine, Moyen-Âge, Renaissance, Xix<sup>e</sup> siècle, XX<sup>e</sup> siècle ;
- Lieux : lieux d'invention, musées, maisons d'inventeurs ;
- Domaines : domaine farfelu, domaine technologique, domaine scientifique, domaine technique ;
- Personnes : spécialistes, inventeurs vivants, inventeurs décédés ;
- Présentation d'objets : exhibition d'inventions, démonstration d'inventions.

## 2.2 Tableau comparatif des taxèmes dans le corpus

Chaque taxème est représentatif d’un thème plus général et permet d’indexer le contenu du corpus (hors génériques). Les 19 taxèmes retenus sont donc suffisamment représentatifs et satisfont aux différentes lectures envisageables en ce sens que chaque taxème est déclinable en plusieurs spécifiques.

<b>Taxème</b>	<b>Documentaire</b>
Préhistoire	1
Antiquité	1
Époque Gallo-romaine	1
Moyen-Âge	1, 2
Renaissance	1, 2, 3
XIX <sup>e</sup> siècle	1, 2, 3
XX <sup>e</sup> siècle	1, 2, 3, 4
Lieux d’invention	1, 2, 3, 4
Musées	1, 2, 3, 4
Maisons d’inventeurs	3, 4
Domaine farfelu	1, 4
Domaine technologique	2, 3, 4
Domaine scientifique	1, 2, 3
Domaine technique	1, 2, 3, 4
Spécialistes	1, 2, 3, 4
Inventeurs vivants	3, 4
Inventeurs décédés	2, 3
Exhibition d’inventions	1, 2, 3, 4
Démonstrations d’inventions	1, 2, 3, 4

## 2.3 Annotation des taxèmes

**Antiquité** : séquences dont la période est comprise entre -3000 et -500 avant JC. *Soie, métier à tisser* – #1 : 24.

**Démonstrations d’inventions** : séquences qui montrent les inventions dans leur pratique. *Poulie, palan* – #2 : 5.

**Domaine farfelu** : séquences pour lesquelles les inventions décrites relèvent de l’absurde, du farfelu. *Petites inventions inutiles* – #4 : 19.

**Domaine scientifique** : séquences pour lesquelles les inventions décrites découlent de travaux scientifiques (par opposition aux domaines techniques et technologiques). *Électricité appliquée à un mouvement rotatif (B.-Franklin)* – #3 : 6.

**Domaine technique** : séquences pour lesquelles les inventions décrites relèvent de méthodes de production traditionnelles, simples, et/ou artisanales. *Marteau, clou, pierre taillée* – #1 : 6.

**Domaine technologique** : séquences pour lesquelles les inventions décrites relèvent de techniques plus modernes. *Micro-informatique, Internet, jeux vidéos* – #3 : 19.

**Époque Gallo-romaine** : séquences dont la période est comprise entre -500 avant JC et 500 après JC. *Parures romaines, épingle à nourrice* – #1 : 23.

**Exhibition d’inventions** : séquences où les inventions sont simplement montrées. *Fer à cheval* – #1 : 11.

**Inventeurs décédés** : séquences qui présentent un inventeur connu décédé (par le biais de reconstitutions). *Gutenberg (reconstitution)* – #2 : 8.

**Inventeurs vivants** : séquences où les inventeurs sont présents. *Nakamatsu et sa machine à devenir intelligent* – #4 : 22.

**Lieux d’invention** : séquences qui présentent le lieu, le pays où des inventions ont été créées. *Égypte pour le cadran solaire* – #2 : 12.

**Maisons d’inventeurs** : séquences qui présentent la maison de l’inventeur. *Dans la maison d’un inventeur américain* – #4 : 23.

**Moyen-Âge** : séquences dont la période est comprise entre 500 et 1450 après JC. *Léonard de Vinci : grue, scie hydraulique* – #2 : 4.

**Musées** : séquences qui présentent un lieu où sont gardées et/ou exposées des inventions en grand nombre, ou des brevets. *Musée du Fer à Nancy* – #1 : 5.

**Préhistoire** : séquences dont la période est comprise avant -3000 avant JC. *Les hommes des cavernes* – #1 : 7.

**Renaissance** : séquences dont la période est comprise entre 1450 et 1800 après JC. *Les automates suisses* – #2 : 13.

**Spécialistes** : séquences qui font appel à des spécialistes pour la présentation des inventions (conservateurs, journalistes, gnous, professeurs). *Journaliste Benoît Chabert* – #2 : 21.

**XIX<sup>e</sup> siècle** : séquences dont la période est comprise entre 1801 et 1900. *Locomotive* – #2 : 16.

**XX<sup>e</sup> siècle** : séquences dont la période est comprise entre 1901 et 2000. *Aviation* – #2 : 21.

## Chapitre 3

# Description des taxèmes

### 3.1 Description taxinomique

Nous avons identifié précédemment une vingtaine de taxèmes de différents types :

- Taxèmes de type concret : lieu d’invention, musées, maisons d’inventeurs, démonstrations d’inventions, exhibitions d’inventions ;
- Taxèmes de type abstrait : Préhistoire, Antiquité, époque Gallo-romaine, Moyen-Âge, Renaissance, XIX<sup>e</sup> siècle, XX<sup>e</sup> siècle, domaine scientifique, domaine technique, domaine technologique ;
- Taxèmes de type rhétorique : spécialistes, inventeurs vivants, inventeurs décédés ;
- Taxème de type “cliché”, stéréotype : domaine farfelu.

### 3.2 Description méréologique

Plusieurs spécifieurs nous permettent de définir plus en détail le contenu et la portée de chaque taxème :

<b>Taxème</b>	<b>Spécifieur</b>
Préhistoire	Feu Outils pierres
Antiquité	Outils agricoles Habits, parures, maquillage Constructions
Époque Gallo-romaine	Habits Hygiène Outils fondamentaux Constructions

Moyen-Âge	Outils Tissage Constructions Imprimerie
Renaissance	Confort Développement scientifique Développement technique
XIX <sup>e</sup> siècle	Transports Révolution industrielle Restitution du son Électricité
XX <sup>e</sup> siècle	Aviation Révolution des mœurs Révolution technologique
Lieux d’invention	Europe : Allemagne, Angleterre, France, Italie, Suisse Asie : Chine, Japon Afrique : Égypte Amériques : États-Unis
Musées	Localisation : façade, salle d’exposition Implantation : nom, ville, pays
Maisons d’inventeurs	Localisation : façade, salle d’exposition Implantation : nom, ville, pays
Domaine farfelu	Utile Inutile
Domaine technologique	Transistor Internet Télévision
Domaine scientifique	Électricité Ondes radio Téléphone
Domaine technique	Nourriture Constructions Mesure du temps Imprimerie Transports
Spécialistes	Professeurs Conservateurs Journalistes Restaurateurs
Inventeurs vivants	Démonstrations
Inventeurs décédés	Reconstitution
Exhibition d’inventions	par un spécialiste ou inventeur par images d’archives
Démonstrations d’inventions	Publiques  Privées

### 3.3 Représentation visuelle sous forme de scripts ou graphes conceptuels

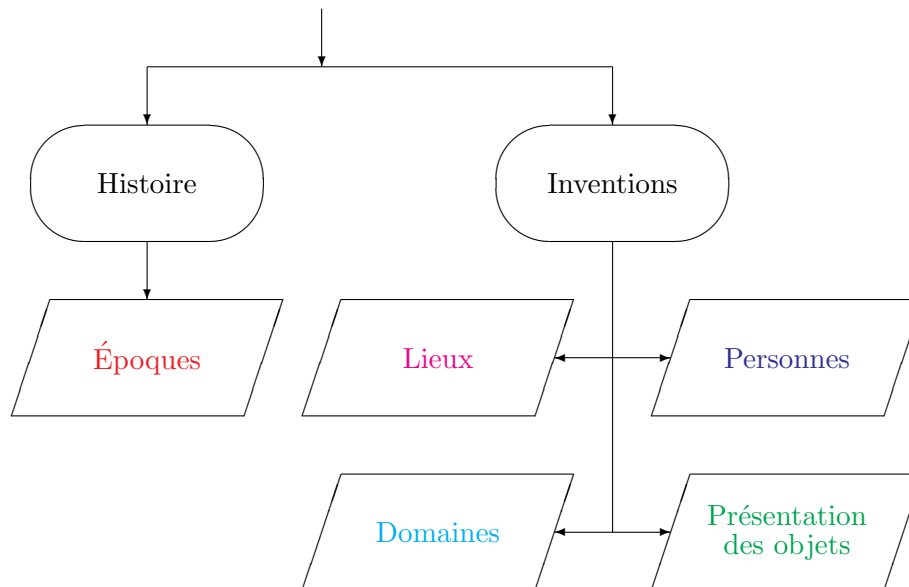
Nous avons adopté les conventions de représentation suivantes :

- Les deux types de lectures possibles du corpus sont représentés dans des cases de forme ovale ;
- Les thèmes principaux sont représentés dans des parallélépipèdes ;
- Les taxèmes sont représentés dans des cases de forme rectangulaire ;
- Les spécifieurs sont représentés dans des petites cases de forme ovale.

Concernant les couleurs utilisées pour le texte, il ne s’agit là que d’un repère visuel, signalant au lecteur à quelle partie du premier graphe se rattachent les sous-graphes, puisque nous n’avons pas pu réunir tous les éléments taxinomiques sur un seul graphe.

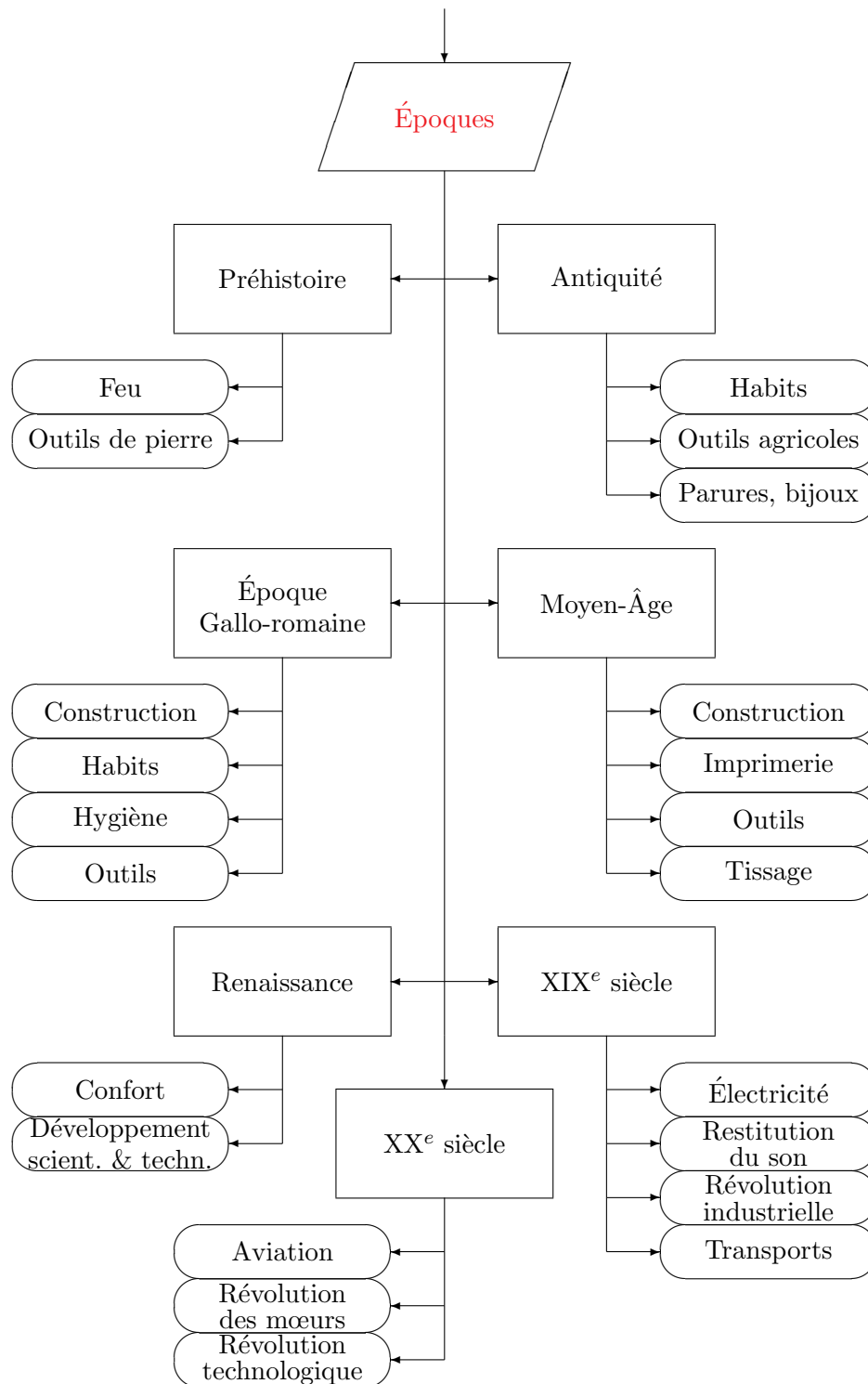
Ainsi, le deuxième sous-graphe (qui commence par une case rectangulaire dans laquelle est inscrite en rouge le thème “Époques”) est rattaché au premier graphe ; de fait, il est visible que les deux cases “Époques” se confondent dans les deux graphes.

#### 3.3.1 Graphe de base : les deux lectures possibles

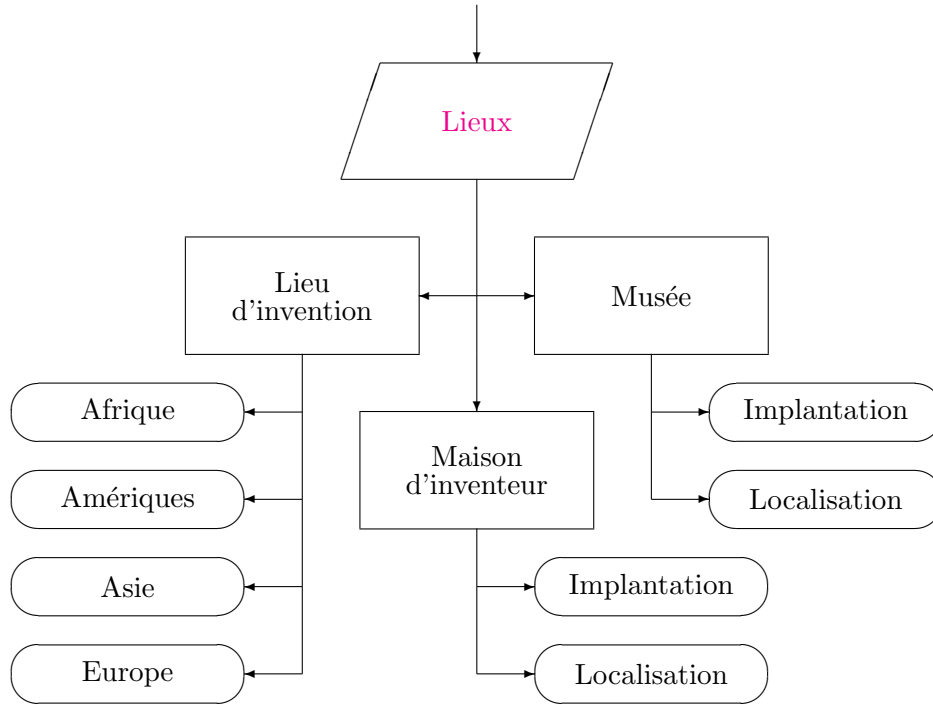


### 3.3.2 Sous-graphes : les différents taxèmes pour chaque thème

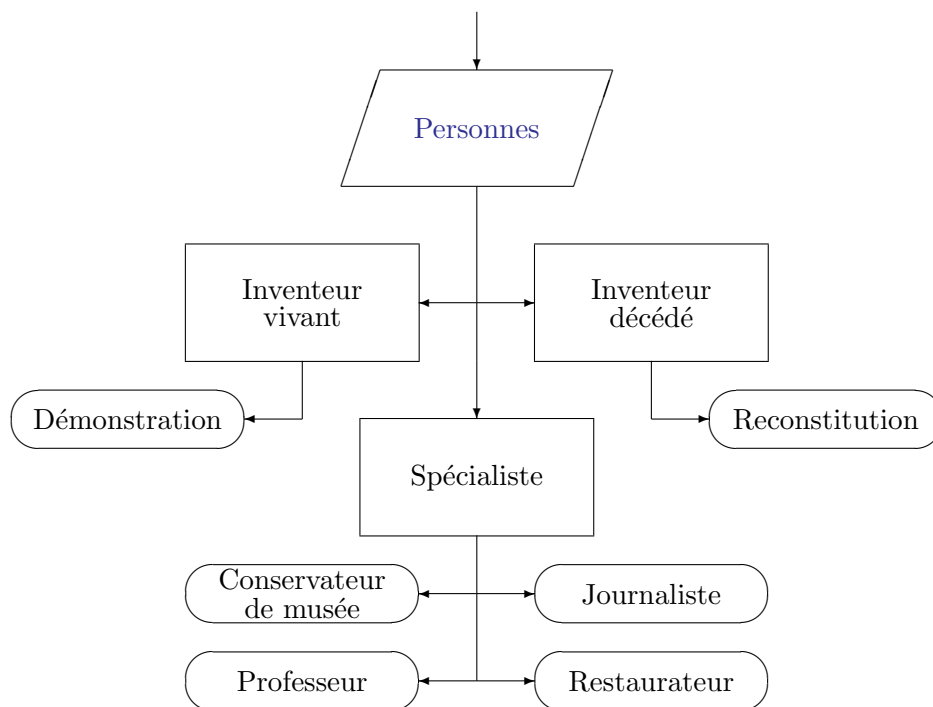
#### Thème “Époques”



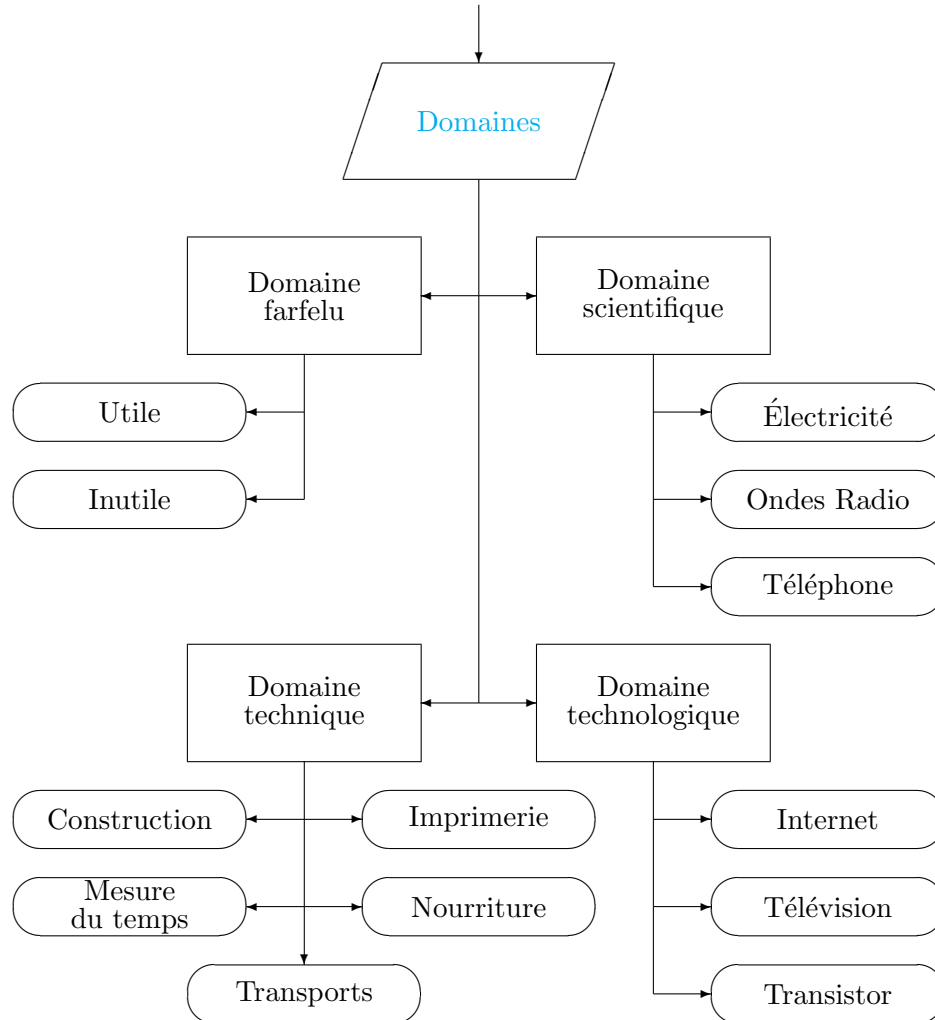
**Thème “Lieux”**



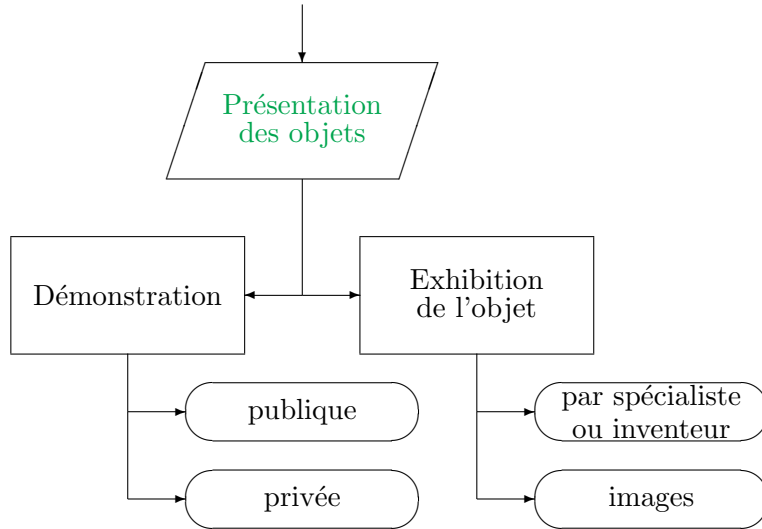
**Thème “Personnes”**



Thème “Domaines”



**Thème “Présentation des objets”**



### 3.4 Enrichissement du dictionnaire de taxèmes

**Antiquité** : séquences dont la période est comprise entre -3000 et -500 avant JC. *Soie, métier à tisser* – #1 : 24. Taxème spécifié par : Outils Agricoles – Habits – Parrures, bijoux, maquillage – Construction.

**Démonstrations d’inventions** : séquences qui montrent les inventions dans leur pratique. *Poulie, palan* – #2 : 5. Taxème spécifié par : Publiques – Privées.

**Domaine farfelu** : séquences pour lesquelles les inventions décrites relèvent de l’absurde, du farfelu. *Petites inventions inutiles* – #4 : 19. Taxème spécifié par : Utile – Inutile.

**Domaine scientifique** : séquences pour lesquelles les inventions décrites découlent de travaux scientifiques (par opposition aux domaines techniques et technologiques). *Électricité appliquée à un mouvement rotatif (B.-Franklin)* – #3 : 6. Taxème spécifié par : Électricité – Ondes radio – Téléphone.

**Domaine technique** : séquences pour lesquelles les inventions décrites relèvent de méthodes de production traditionnelles, simples, et/ou artisanales. *Marteau, clou, pierre taillée* – #1 : 6. Taxème spécifié par : Nourriture – Construction – Mesure du temps – Imprimerie – Transports.

**Domaine technologique** : séquences pour lesquelles les inventions décrites relèvent de techniques plus modernes. *Micro-informatique, Internet, jeux vidéos* – #3 : 19. Taxème spécifié par : Transistors – Internet – Télévision.

**Époque Gallo-romaine** : séquences dont la période est comprise entre -500 avant JC et 500 après JC. *Parures romaines, épingle à nourrice* – #1 : 23. Taxème spécifié par : Habits – Hygiène – Outils fondamentaux – Construction.

**Exhibition d’inventions** : séquences où les inventions sont simplement montrées. *Fer à cheval* – #1 : 11. Taxème spécifié par : Par un spécialiste ou un inventeur – Par images d’archives.

**Inventeurs décédés** : séquences qui présentent un inventeur connu décédé (par le biais de reconstitutions). *Gutenberg (reconstitution)* – #2 : 8. Taxème spécifié par : Reconstitutions.

**Inventeurs vivants** : séquences où les inventeurs sont présents. *Nakamatsu et sa machine à devenir intelligent* – #4 : 22. Taxème spécifié par : Démonstrations.

**Lieux d’invention** : séquences qui présentent le lieu, le pays où des inventions ont été créées. *Égypte pour le cadran solaire – #2 : 12*. Taxème spécifié par : Europe : Allemagne, Angleterre, France, Italie, Suisse – Asie : Chine, Japon – Afrique : Égypte – Amérique : États-Unis.

**Maisons d’inventeurs** : séquences qui présentent la maison de l’inventeur. *Dans la maison d’un inventeur américain – #4 : 23*. Taxème spécifié par : Localisation : Façade, Salle d’exposition – Implantation : nom, ville, pays.

**Moyen-Âge** : séquences dont la période est comprise entre 500 et 1450 après JC. *Léonard de Vinci : grue, scie hydraulique – #2 : 4*. Taxème spécifié par : Outils – Tissage – Construction – Imprimerie.

**Musées** : séquences qui présentent un lieu où sont gardées et/ou exposées des inventions en grand nombre, ou des brevets. *Musée du Fer à Nancy – #1 : 5*. Taxème spécifié par : Localisation : Façade, Salle d’exposition – Implantation : nom, ville, pays.

**Préhistoire** : séquences dont la période est comprise avant -3000 avant JC. *Les hommes des cavernes – #1 : 7*. Taxème spécifié par : Feu – Outils en pierre.

**Renaissance** : séquences dont la période est comprise entre 1450 et 1800 après JC. *Les automates suisses – #2 : 13*. Taxème spécifié par : Confort – Développement scientifique – Développement technique.

**Spécialistes** : séquences qui font appel à des spécialistes pour la présentation des inventions (conservateurs, journalistes, gnous, professeurs). *Journaliste Benoît Chabert – #2 : 21*. Taxème spécifié par : Professeurs – Journalistes – Restaurateurs – Conservateurs.

**XIX<sup>e</sup> siècle** : séquences dont la période est comprise entre 1801 et 1900. *Locomotive – #2 : 16*. Taxème spécifié par : Transports – Révolution Industrielle – Restitution du son – Électricité.

**XX<sup>e</sup> siècle** : séquences dont la période est comprise entre 1901 et 2000. *Aviation – #2 : 21*. Taxème spécifié par : Aviation – Révolution des mœurs – Révolution technologique.

## Chapitre 4

# Extrait d'un thésaurus

La construction de notre thésaurus repose sur deux étapes distinctes et successives :

- La première étape consiste à dresser une liste classée alphabétiquement, des concepts développés dans un corpus. Pour chacun de ces concepts, une ou plusieurs référence(s) renvoie(nt) au(x) passage(s) du corpus dans le(s)quel(s) le concept considéré est développé ;
- La deuxième étape vise à ordonner ces concepts selon une hiérarchie fondée sur les relations d'hyponymie, d'hyponymie et de synonymie. Nous pouvons ainsi pour une entrée donnée, fournir le concept, la catégorie plus générale à partir de laquelle il est dérivé (*ie.* l'hyponyme), et les termes “employés pour” (*ie.* les synonymes).

Ainsi, au terme de la première étape, nous obtenons un index , que nous organisons de manière hiérarchique dans la deuxième étape afin d'obtenir un thésaurus.

Notre indexation réalisée infra (chapitre 7) donne lieu à la liste suivante (extrait) :

Aviation : #2, 21.  
Confort : #1, 15.  
Conservateur de musée : #1, 5.  
Construction : #2, 2, 3, 4, 5.  
Développement scientifique : #3, 4.  
Développement technique : #2, 13.  
Électricité : #3, 5.  
Feu : #1, 7.  
Habits : #1, 23, 24.  
Hygiène : #1, 32.  
Imprimerie : #2, 7, 8.  
Internet : #3, 18.

Journaliste : #2, 16.  
 Mesure du temps : #2, 12.  
 Nourriture : #1, 20.  
 Ondes radio : #3, 16.  
 Outils : #1, 8.  
 Outils agricoles : #1, 7.  
 Outils fondamentaux : #1, 7.  
 Outils en pierre : #1, 7.  
 Parures, bijoux, maquillage : #1, 30.  
 Professeur : #1, 18.  
 Restaurateur : #1, 15.  
 Restitution du son : #3, 9.  
 Révolution Industrielle : #4, 8.  
 Révolution des mœurs : #1, 28.  
 Révolution technologique : #3, 18.  
 Téléphone : #3, 15.  
 Télévision : #4, 26.  
 Tissage : #1, 25.  
 Transistors : #3, 17.  
 Transports : #2, 16, 17.

Pour réaliser notre thesaurus, il nous reste à organiser hiérarchiquement cette liste, selon les trois relations sémantiques sus-mentionnées.

- Arts, Musique, Littérature
  - Arts
    - Artisanat
      - Tissage : #1, 25.
    - Ornements
      - Parures, bijoux, maquillage : #1, 30.
- Construction : #2, 2, 3, 4, 5.
- Économie et industrie
  - Infrastructure
    - Changements
      - Révolution Industrielle : #4, 8.
- Éducation et formation
  - Enseignement
    - Personne
      - Professeur : #1, 18, *employé pour* : Enseignant.
- Histoire et archéologie
  - Institution culturelle
    - Personne
      - Conservateur de musée : #1, 5.

- Liens et fonctions
  - Développement
    - Sciences et technologies
      - Développement scientifique : #3, 4.
      - Développement technique : #2, 13.
- Nature et environnement
  - Désastre naturel
    - Feu : #1, 7, *employé pour* : Incendie.
    - Électricité : #3, 5.
- Pensée Mœurs
  - Changements
    - Révolution des mœurs : #1, 28.
- Santé et sécurité
  - Santé
    - Hygiène : #1, 32.
    - Nourriture : #1, 20, *employé pour* : Aliment.
- Sciences et technologies
  - Changements
    - Révolution technologique : #3, 18.
  - Communication
    - Équipement de communication
      - Téléphone : #3, 15.
    - Médias
      - Télévision : #4, 26.
    - Transport
      - Ondes radio : #3, 16.
  - Informatique
    - Internet : #3, 18.
  - Lecture et écriture
    - Imprimerie : #2, 7, 8.
  - Outils : #1, 8.
    - Composition
      - Outils en pierre : #1, 7.
  - Son
    - Restitution du son : #3, 9.
    - Transistors : #3, 17.
  - Technologies agricoles
    - Outils agricoles : #1, 7, *employé pour* : Outils.
- Temps
  - Mesure du temps : #2, 12.
- Transports
  - Transport aérien
    - Aviation : #2, 21.

Nous tenons à mentionner que la réalisation de ce thesaurus a été longue et laborieuse (nous nous sommes inspirés des thesaurus existants pour former les catégories principales, *cf.* partie 10.1). Ce thesaurus comprend 26 concepts, classés sous 12 rubriques principales, divisées en 18 sous-rubriques. Il nous a fallu pas moins de 2 heures pour réaliser ledit thesaurus, qui ne représente qu’une petite partie de notre corpus audiovisuel.

## Chapitre 5

# Construction (d'un extrait) d'une ontologie

Nous postulons que ce qui fait la différence entre un thésaurus et une ontologie tient dans le point de vue adopté lors de la réalisation de l'ouvrage ; un thésaurus liste tous les termes d'un domaine ou d'une langue, et relie ces termes entre eux selon les acceptions sémantiques partagées. Quelle que soit la personne réalisant ledit thésaurus, ce dernier doit proposer une vision globale et commune à tous les locuteurs de la langue, et les différences que l'on ne manquerait pas de relever entre deux thésaurus ne sont pas d'avantage significatives que la non concordance de deux définitions d'un même terme dans deux dictionnaires différents.

Si le thésaurus apporte une vision partagée par tous les locuteurs d'une langue sur les liens sémantiques qu'entretiennent les termes de cette langue entre eux, l'ontologie correspond à une vision du monde propre à son auteur.

Dans ce travail d'analyse taxinomique du corpus "l'histoire des inventions", nous avons ainsi pu relever plusieurs thésaurus proposant différentes répartitions sémantiques pour le terme "invention" (p. 36 sq.). À l'opposé, aucune ontologie n'a pu être trouvée, c'est pourquoi nous proposons notre propre ontologie, correspondant à notre vision du domaine d'après le corpus audio-visuel mis à notre disposition.

Cette ontologie n'a aucun statut officiel et ne se veut représentative que de notre classification selon la lecture du corpus que nous avons adoptée. Nous avons ainsi choisi de considérer le corpus selon une approche diachronique, comme le résume le titre dudit corpus : "l'histoire des inventions".

Pour réaliser l'ontologie, nous avons utilisé l'éditeur d'ontologie développé à l'INA (v1.0 - 29.03.2002) par Antoine Isaac et Raphaël Troncy. Cet outil permet de générer une ontologie dans un fichier au format XML, utilisant la norme RDF. Nous proposons ci-dessous l'ontologie de manière "lisible", puis un extrait du fichier XML généré par DOE.

- AnyConcept
  - Domaines
    - DomaineFarfelu
      - Utile
      - Inutile
    - DomaineScientifique
      - Électricité
      - OndesRadio
      - Téléphone
    - DomaineTechnique
      - Constructions
      - Imprimerie
      - MesureTemps
      - Nourriture
      - Transports
    - DomaineTechnologique
      - Internet
      - Télévision
      - Transistor
  - Lieux
    - LieuInvention
      - Afrique
      - Amériques
      - Asie
      - Europe
    - MaisonInventeur
      - Implantation
      - Localisation
    - Musée
      - Implantation
      - Localisation
  - Personnes
    - InventeurVivant
      - Démonstration
    - InventeurDécédé
      - Reconstitution
    - Spécialiste
      - ConservateurMusée
      - Journaliste
      - Professeur
      - Restaurateur

- PrésentationObjets
  - DémonstrationObjet
    - Privée
    - Publique
  - ExhibitionObjet
    - ImagesArchives
    - InventeurSpécialiste

Voici un extrait de l’ontologie telle qu’elle a été générée par DOE, au format XML/RDF :

```

<ina:DifferentialPart>
<ina:Language name="French" code="fr"/>
<ina:Language name="English" code="en"/>
<ina:Language name="Dutch" code="nl"/>
<ina:DifferentialConceptList>
<ina:Notion key="id-1934921541019634562169979493441"
name="AnyConcept">
  <ina:Lexicalization language="Dutch">
  </ina:Lexicalization>
  <ina:Lexicalization language="French">
  </ina:Lexicalization>
  <ina:Lexicalization language="English">
  </ina:Lexicalization>
</ina:Notion>
<ina:Notion key="id-193492154101963456573413664512"
name="Domaines">
  <ina:subNotionOf ref="id-1934921541019634562169979493441"/>
  <ina:Lexicalization language="Dutch">
    <ina:PreferredTerm>Gebieden</ina:PreferredTerm>
  </ina:Lexicalization>
  <ina:Lexicalization language="French">
    <ina:PreferredTerm>Domaines</ina:PreferredTerm>
  </ina:Lexicalization>
  <ina:Lexicalization language="English">
    <ina:PreferredTerm>Field</ina:PreferredTerm>
  </ina:Lexicalization>
</ina:Notion>
<ina:Notion key="id-1934921541019634571062386910323"
name="DomaineFarfelu">
  <ina:subNotionOf ref="id-193492154101963456573413664512"/>
  <ina:Lexicalization language="Dutch">
    <ina:PreferredTerm>Vreemd gebied</ina:PreferredTerm>
    <ina:EncyclopedicDefinition>sequenties waarvoor de beschre-

```

```

    ven uitvindingen van het absurde, van vreemd vaststellen
    </ina:EncyclopedicDefinition>
  </ina:Lexicalization>
  <ina:Lexicalization language="French">
    <ina:PreferredTerm>Domaine farfelu</ina:PreferredTerm>
    <ina:EncyclopedicDefinition>séquences pour lesquelles les
    inventions décrites relèvent de l’absurde, du farfelu
    </ina:EncyclopedicDefinition>
  </ina:Lexicalization>
  <ina:Lexicalization language="English">
    <ina:PreferredTerm>Eccentric field</ina:PreferredTerm>
    <ina:EncyclopedicDefinition>sequences for which the descri-
    bed inventions concern the absurdity, of the eccentric
    </ina:EncyclopedicDefinition>
  </ina:Lexicalization>
</ina:Notion>
<ina:Notion key="id-1934921541019634576109927802680"
name="Utile">
  <ina:subNotionOf ref="id-1934921541019634571062386910323"/>
  <ina:Lexicalization language="Dutch">
    <ina:PreferredTerm>Nuttig</ina:PreferredTerm>
  </ina:Lexicalization>
  <ina:Lexicalization language="French">
    <ina:PreferredTerm>Utile</ina:PreferredTerm>
  </ina:Lexicalization>
  <ina:Lexicalization language="English">
    <ina:PreferredTerm>Useful</ina:PreferredTerm>
  </ina:Lexicalization>
</ina:Notion>
<ina:Notion key="id-1934921541019634576269761422791"
name="Inutile">
  <ina:subNotionOf ref="id-1934921541019634571062386910323"/>
  <ina:Lexicalization language="Dutch">
    <ina:PreferredTerm>Nutteloos</ina:PreferredTerm>
  </ina:Lexicalization>
  <ina:Lexicalization language="French">
    <ina:PreferredTerm>Inutile</ina:PreferredTerm>
  </ina:Lexicalization>
  <ina:Lexicalization language="English">
    <ina:PreferredTerm>Useless</ina:PreferredTerm>
  </ina:Lexicalization>
</ina:Notion>

```

Cette section correspond à l’extrait d’ontologie suivant (les concepts et les définitions en français ont été traduits en anglais et en néerlandais) :

AnyConcept

– Domaines – *Field* – *Gebieden*

    DomaineFarfelu – *Eccentric field* – *Vreemd gebied*

        séquences pour lesquelles les inventions décrites relèvent de l’absurde, du farfelu

*sequences for which the described inventions concern the absurdity, of the eccentric*

*sequenties waarvoor de beschreven uitvindingen van het ongerijmd, van vreemd vaststellen*

    Utile – *Useful* – *Nuttig*

    Inutile – *Useless* – *Nutteloos*

## Chapitre 6

# Établir une “bibliothèque de séquences AV”

Dans ce chapitre, nous avons réalisé une quinzaine de fiches descriptives de séquences audiovisuelles. Chaque fiche détaille la séquence en question : numéro d’identifiant, titre, type d’unité (*ie.* séquence ou sous-séquence), l’ensemble auquel appartient la séquence identifiée, la séquence qui la précède, la durée, et un résumé.

Ces fiches constituent notre “bibliothèque de séquences audiovisuelles” ; elles permettent de décrire le contenu de chaque séquence, mais également d’accéder directement à une séquence pertinente, selon le thème de recherche (tout ceci sera détaillé dans le chapitre 8, p. 34).

Identifieur	#1 – 7
Titre	De l’âge de pierre à l’âge du fer.
Type d’unité	Sous-séquence
Fait partie de	Grande séquence au Musée du Fer à Nancy
Précédée par	Générique
Durée	4’06”
Résumé	Présentation diachronique d’objets de la Préhistoire à l’époque Gallo-romaine.
Identifieur	#1 – 8
Titre	Fonte et marmite
Type d’unité	Sous-séquence
Fait partie de	Grande séquence au Musée du Fer à Nancy
Précédée par	“De l’âge de pierre à l’âge du fer”
Durée	0’26”
Résumé	Présentation de la fonte et la marmite par deux spécialistes.

Identifieur	#1 – 22
Titre	Réfréner la sexualité
Type d’unité	Sous-séquence
Fait partie de	Grande séquence sur les vêtements
Précédée par	“Gastronomie au XIX <sup>e</sup> ”
Durée	4’33”
Résumé	Présentation de divers objets ayant servis à contraindre le corps des hommes et des femmes.
Identifieur	#1 – 23
Titre	Vêtements antiques
Type d’unité	Sous-séquence
Fait partie de	Grande séquence sur les vêtements
Précédée par	“Réfréner la sexualité”
Durée	2’13”
Résumé	Sophistication et complication des vêtements au fil du temps.
Identifieur	#1 – 24
Titre	Soie, métier à tisser
Type d’unité	Sous-séquence
Fait partie de	Grande séquence sur les vêtements
Précédée par	“Vêtements antiques”
Durée	0’34”
Résumé	Intérêt de la soie et du métier à tisser dans la confection de vêtements.
Identifieur	#1 – 25
Titre	Lin & rouet
Type d’unité	Sous-séquence
Fait partie de	Grande séquence sur les vêtements
Précédée par	“Soie, métier à tisser”
Durée	1’02”
Résumé	Interview de paysannes sachant filer la laine.
Identifieur	#2 – 2
Titre	Construction de pyramides
Type d’unité	Sous-séquence
Fait partie de	Séquence sur la construction
Précédée par	Générique
Durée	0’35”
Résumé	Introduction à la séquence sur la construction.

Identifieur	#2 – 3
Titre	Inventions romaines
Type d’unité	Sous-séquence
Fait partie de	Séquence sur la construction
Précédée par	“Construction de pyramides”
Durée	2’08”
Résumé	Discipline des constructions romaines.
Identifieur	#2 – 4
Titre	Léonard de Vinci
Type d’unité	Sous-séquence
Fait partie de	Séquences sur la construction
Précédée par	“Inventions romaines”
Durée	3’09”
Résumé	Présentation de diverses maquettes du Maître.
Identifieur	#2 – 5
Titre	Poulie et palan
Type d’unité	Sous-séquence
Fait partie de	Séquence sur la construction
Précédée par	“Léonard de Vinci”
Durée	2’07”
Résumé	Facilitation de la construction.
Identifieur	#2 – 6
Titre	Boussole de Gengis Khan
Type d’unité	Petite séquence
Fait partie de	Transition
Précédée par	Séquence sur la construction
Durée	1’25”
Résumé	Comment Gengis Khan se dirigeait lors de ses conquêtes.
Identifieur	#2 – 7
Titre	Écriture en Asie
Type d’unité	Sous-séquence
Fait partie de	Séquence sur l’écriture
Précédée par	“Boussole de Gengis Khan”
Durée	1’04”
Résumé	Évolution des moyens d’impression en Asie.

Identifieur	#2 – 8
Titre	Gutenberg
Type d’unité	Sous-séquence
Fait partie de	Séquence sur l’écriture
Précédée par	“Écriture en Asie”
Durée	7’00”
Résumé	Gutenberg invente l’imprimerie
Identifieur	#3 – 4
Titre	Électricité de Benjamin Franklin
Type d’unité	Sous-séquence
Fait partie de	Grande séquence sur l’électricité
Précédée par	“Salle de l’Indépendance des États-Unis”
Durée	1’52”
Résumé	Présentation de la machine à produire de l’électricité de B.-Franklin.
Identifieur	#4 – 22
Titre	Nakamatsu, inventeur japonais
Type d’unité	Sous-séquence
Fait partie de	Séquence sur les Inventeurs
Précédée par	“Dépôt de brevets sur le mouvement perpétuel à l’INPI”
Durée	4’56”
Résumé	M. Nakamatsu nous présente ses inventions.

## Chapitre 7

# Exemples d'indexation thématique de séquences AV

Scènes indexées :

- Préhistoire (ÉPOQUE)
  - Feu : #1, 7.
  - Outils en pierre : #1, 7.
- Antiquité (ÉPOQUE)
  - Outils Agricoles : #1, 7.
  - Habits : #1, 24.
  - Parures, bijoux, maquillage : #1, 30.
  - Construction : #2, 2.
- Époque Gallo-Romaine (ÉPOQUE)
  - Habits : #1, 23.
  - Hygiène : #1, 32.
  - Outils fondamentaux : #1, 7.
  - Construction : #2, 3.
- Moyen-Âge (ÉPOQUE)
  - Outils : #1, 8.
  - Tissage : #1, 25.
  - Construction : #2, 5.
  - Imprimerie : #2, 7.
- Renaissance (ÉPOQUE)
  - Confort : #1, 15.
  - Développement scientifique : #3, 4.
  - Développement technique : #2, 13.
- XIX<sup>e</sup> siècle (ÉPOQUE)
  - Transports : #2, 16.
  - Révolution Industrielle : #4, 8.
  - Restitution du son : #3, 9.

- Électricité : #3, 4.
- XX<sup>e</sup> siècle (ÉPOQUE)
  - Aviation : #2, 21.
  - Révolution des mœurs : #1, 28.
  - Révolution technologique : #3, 18.
- Lieux d’inventions (LIEUX)
  - Afrique : #2, 2.
  - Égypte :
  - Amériques : #3, 10.
  - États-Unis :
  - Asie : #2, 6.
  - Chine :
  - Japon :
  - Europe : #1, 22.
  - Allemagne :
  - Angleterre :
  - France :
  - Italie :
  - Suisse :
- Musées (LIEUX)
  - Implantation : #4, 8.
  - Nom :
  - Ville :
  - Pays :
  - Localisation : #4, 8.
  - Façade :
  - Salle d’exposition :
- Maisons d’inventeurs (LIEUX)
  - Implantation : #4, 22.
  - Nom :
  - Ville :
  - Pays :
  - Localisation : #4, 22.
  - Façade :
  - Salle d’exposition :
- Inventeur vivant (PERSONNE)
  - Démonstration : #4, 23.
- Inventeur décédé (PERSONNE)
  - Reconstitution : #3, 4.
- Spécialiste (PERSONNE)
  - Conservateur de musée : #1, 5.
  - Journaliste : #2, 16.

Professeur : #1, 18.  
Restaurateur : #1, 15.

- Farfelu (DOMAINE)
  - Utile : #4, 15.
  - Inutile : #4, 19.
- Scientifique (DOMAINE)
  - Électricité : #3, 5.
  - Ondes radio : #3, 16.
  - Téléphone : #3, 15.
- Technique (DOMAINE)
  - Construction : #2, 4.
  - Imprimerie : #2, 8.
  - Mesure du temps : #2, 12.
  - Nourriture : #1, 20.
  - Transports : #2, 17.
- Technologique (DOMAINE)
  - Internet : #3, 18.
  - Télévision : #4, 26.
  - Transistors : #3, 17.
- Démonstration d’inventions (OBJETS)
  - Publique : #4, 17.
  - Privée : #4, 22.
- Exhibition d’inventions (OBJETS)
  - Par un spécialiste ou un inventeur : #1, 7.
  - Par images d’archives : #1, 12.

## Chapitre 8

# Description et indexation discursives de séquences AV

Les thèmes présents dans notre corpus audiovisuel n'ont pas la même répartition selon la séquence dans laquelle ils s'inscrivent. Pour une séquence donnée, un thème peut en être le thème directeur (*ie.* le leitmotiv) ; il peut être simplement un thème central, ou bien marginal. Un thème peut être le leitmotiv d'une séquence alors qu'il sera marginal dans une autre.

Ce chapitre vise à indexer de manière discursive le corpus audiovisuel, mettant ainsi en avant le poids de chaque thème dans chaque catégorie. La pondération de chaque thème par séquence est intéressante en ce sens qu'elle nous renseigne avec plus de précision sur le contenu de chaque séquence ; la recherche d'un concept se trouve ainsi affinée.

**NB :** *faute de temps, ce chapitre est vide.*  
*Les auteurs vous prient de bien vouloir les en excuser.*

## Chapitre 9

# Description et indexation narratives de séquences AV

*NB : faute de temps, ce chapitre est vide.  
Les auteurs vous prient de bien vouloir les en excuser.*

## Chapitre 10

# Thésaurus, lexiques, ontologies déjà existants

Les recherches effectuées sur le Web sont restées nulles quant à l’existence d’un Thésaurus relatif à “l’histoire des inventions”. Nous avons alors focalisé notre recherche sur le terme “invention”, celui-ci étant plus représentatif de notre corpus que le terme “histoire” qui pourrait, par ses ambiguïtés multiples, nous orienter vers de fausses pistes.

Les Thésaurus existants sur le Web sont pour la plupart, des commandes de Gouvernements d’États bilingues ; nous avons ainsi récolté deux Thésaurus émanant du Gouvernement canadien, l’un pour aider à définir correctement le contenu sémantique d’un site Web, l’autre étant plus statistique.

Nous avons également eu accès à des Thésaurus d’organisations internationales telles que celui de l’OCDE dont le but est de venir en aide aux Gouvernements confrontés à des problèmes liés à l’économie mondiale.

Enfin, deux Thésaurus se démarquent du lot par leur fonction première : il s’agit du Thésaurus de Rodget, un Thésaurus britannique qui existe depuis le XIX<sup>e</sup> siècle et considéré comme une institution dans ce pays, et le Thésaurus de Péchoin [Péchoin, 1992], un Thésaurus français, créé en 1992.

Nous avons utilisé les conventions suivantes pour la présentation des résultats :

- En premier lieu, l’entrée du Thésaurus (**Invention**), puis plusieurs sous catégories :
- **TG** désigne le terme générique (*ie.* la classe sous laquelle est classée l’entrée) ;
- **EP** indique que l’entrée est employée pour un autre terme (acceptation) ;
- **TS** désigne un terme spécifique (*ie.* un emploi particulier de l’entrée) ;
- Enfin, **TA** désigne les termes associés à l’entrée.

## 10.1 Thésaurus des sujets de base du Gouvernement du Canada

Ce Thésaurus a été créé en vue de correctement définir les descripteurs à inclure dans la métabalise "Controlled\_subject\_terms", lors de la création de sites Web. Dans cette optique, ce Thésaurus est en conformité avec la Normalisation des sites Internet, Standard 6.3.

Thésaurus accessible sur [http://fr.thesaurus.gc.ca/thesf/thes\\_f.html](http://fr.thesaurus.gc.ca/thesf/thes_f.html)

### Invention

- TG Sciences et technologies.
- TA Brevet, Innovation.

### Brevet

- TG Économie et industrie.
- TA Invention, Marque de commerce, Propriété intellectuelle, Recherche et Développement.

### Innovation

- TG Liens et fonctions.
- EP Changement.
- TS Innovation technologique.
- TA Gestion du savoir, Invention, Modernisation.

## 10.2 Thésaurus de Statistique Canada

Ce deuxième Thésaurus, également canadien, est similaire au précédent, tout en proposant toutefois, une classification légèrement différente.

URL : [http://www.statcan.ca/francais/search/thesaurus\\_f.htm](http://www.statcan.ca/francais/search/thesaurus_f.htm)

### Invention

- TA Brevet, Changement technologique, Innovations technologiques, Inventeurs.

### Brevet

- TA Invention, Marque de commerce, Propriété intellectuelle, Recherche et Développement.

### Changement technologique

- TA Indicateurs de l’activité scientifique et technique, Innovation, Innovations industrielles, Innovations techniques, Inventions, Technologies

**Innovations technologiques**

- TG Innovation, Technologie
- TA Changement technologique, Haute technologie, Innovations industrielles, Inventions

**Inventeurs**

- TG Professions
- TA Inventions

### 10.3 OECD Macrothesaurus

L’OECD<sup>1</sup> est une organisation internationale qui aide les Gouvernements à répondre aux défis économiques, sociaux, et de gouvernance, posés par une économie mondialisée. Cette organisation propose un Thésaurus assez sommaire.

URL : <http://info.uibk.ac.at/info/oecd-macroth/fr/1746.html>

**Inventions**

- TG Changement technologique
- TA Brevets
  - Innovations
  - Inventeurs
  - Recherche et Développement

### 10.4 Rodget’s II Thesaurus

En 1852 en Grande-Bretagne, Peter Mark Rodget fait éditer la première édition de son *Thésaurus des mots et des locutions anglaises, classés et mis en ordre pour faciliter l’expression des idées et aider à la composition littéraire*. Le Thésaurus rencontra un immense succès et connut près d’une trentaine de réimpressions du vivant de Rodget. Passé dans le domaine public, il est aujourd’hui dans le monde anglo-saxon un outil de référence aussi indispensable et courant qu’un dictionnaire pour un francophone [Péchoin, 1992].

Le célèbre Thésaurus de Rodget est disponible en version HTML à l’URL : <http://www.thesaurus.com/>

---

<sup>1</sup>Organisation Économique du Commerce et du Développement

**Invention**

NOUN: 1. Something invented: brainchild, contrivance, device.

See Machine, Make.

2. The power or ability to invent: creativeness, creativeness, ingenuity, inventiveness, originality.

See Ability, Make.

3. Any fictitious idea accepted as part of an ideology by an uncritical group; a received idea: creation, fantasy, fiction, figment, myth.

See Belief, Real.

Ce Thésaurus diffère des précédents en ce sens qu’il donne une petite définition de chaque emploi possible de l’entrée, plutôt que de donner le terme générique. Outre les termes associés, le Thésaurus spécifie également des renvois vers d’autres entrées.

## 10.5 Thésaurus de Péchoin : des idées aux mots, des mots aux idées...

En France, Daniel Péchoin a réalisé à partir de 1992 un Thésaurus pour le français, fondé sur le même principe fonctionnel que le Thésaurus de Rodget : “le *Thésaurus Larousse* est ordonné selon un plan en trois parties – le monde, l’homme, la société [...] Chaque partie est elle-même divisée en sous-parties”. Le Thésaurus comprend ainsi 873 articles, de l’article 1 Existence, à l’article 873 Jouets.

“Chacun des articles porte un titre indiquant la notion qu’il traite et se trouve lui-même divisé en paragraphes [...] qui regroupent les mots par famille de sens” [Péchoin, 1992].

Dans le cas qui nous intéresse, la notion d’invention est classée sous cinq articles différents (précédés d’un numéro d’ordre), ainsi organisés dans le Thésaurus :

- LE MONDE
  - LES CONCEPTS FONDAMENTAUX
  - IDENTITÉ
  - 32 Innovation
- L’HOMME
  - L’ESPRIT
  - L’INTELLIGENCE ET LA MÉMOIRE
  - 404 Imagination
  - LA CONNAISSANCE ET LA VÉRITÉ
  - 411 Découverte

- L’ACTION
  - LES OBJECTIFS
    - 556 Construction
- LA SOCIÉTÉ
  - LES ACTIVITÉS ÉCONOMIQUES
    - LE TRAVAIL ET LA PRODUCTION
      - 796 Production

Le mot “invention” apparaît sous les cinq articles ci-dessous. Comme dans le Thésaurus de Rodget, le Thésaurus de Péchoin propose des renvois (indiqués entre crochets) vers un autre article. Ceci permet de définir des interconnexions entre idées, et de rechercher l’expression la plus juste.

“À l’intérieur de chaque paragraphe, des tirets ou des points virgules signalent les regroupements de sens voisins, le tiret correspondant à une division forte, le point-virgule à une division faible. [...] Les caractères gras mettent en évidence les mots les plus usités” [Péchoin, 1992].

#### 32 INNOVATION

1 **Innovation** (*l’innovation*); création (*la création*). – Fondation, **invention**; changement, transformation 193; renouvellement, rénovation 558. – Exploration, recherche 412.

#### 404 IMAGINATION

3 **Création, invention**; affabulation, fabulation; confabulation [PSYCHOL.].

#### 411 DÉCOUVERTE

2 Déchiffrage, élucidation, **identification**, reconnaissance. – **Invention** 32.1; conception, intuition 424, vision; créativité, ingéniosité, inventivité; imagination 404, inspiration.

#### 556 CONSTRUCTION

3 **Fabrication**; confection [vieilli], création, élaboration, élucubration [vx], facture [MUS.], formation, production; mise au net, mise au point. – Constitution, établissement [vieilli], **fondation**, instauration, invention 411.

#### 796 PRODUCTION

3 Invention 32.1; établissement, fondation, **instauration**, institution.

# Bibliographie

[Péchoin, 1992] Péchoin, D. (1992). *Thésaurus, des idées aux mots, des mots aux idées*. In Extenso. Larousse, Paris.