



HAL
open science

Réflexions sur les rôles des documentalistes dans la numérisation

Eunjoo Eun-Joo Carre, Na

► **To cite this version:**

Eunjoo Eun-Joo Carre, Na. Réflexions sur les rôles des documentalistes dans la numérisation : le cas du “ Répertoire historique de l’administration coréenne ” de Maurice Courant . 10e rencontre annuelle du réseau DocAsie, Centre de Recherches sur la Corée, EHESS; IPRAUS de l’Ecole Nationale de l’Architecture Paris-Belleville; Institut d’études coréennes du Collège de France, Jun 2016, Paris, France. hal-01337488

HAL Id: hal-01337488

<https://hal.science/hal-01337488>

Submitted on 26 Jun 2016

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 Attribution - NonCommercial - ShareAlike 4.0 International License

Réflexions sur les rôles des documentalistes dans la numérisation : le cas du « Répertoire historique de l'administration coréenne » de Maurice Courant ¹

Eun-joo Carré-Na
(CRC-EHESS, CCJ-UMR 8173 CNRS)

Préambule

Cette communication a été présentée à la 10^e Rencontre annuelle du réseau DocAsie, à Paris, du 22 au 24 juin 2016, dont un des thèmes était « les fonds asiatiques à l'ère du numérique ». Le titre présenté dans le programme de cet événement était « quelques réflexions sur le développement d'applications numériques pour le « Répertoire historique de l'administration coréenne » de Maurice Courant », mais dans la présentation, je n'ai abordé qu'une question qui concernait la plupart des participants aux journées : « les rôles des documentalistes ».

1. Introduction

Depuis déjà quelques temps, nous entendons beaucoup parler de projets de « numérisation ». Et désormais la numérisation n'est plus une option. Après des débuts polémiques dans le monde professionnel de la gestion de documentation, la « révolution numérique », a bel et bien pris place dans notre domaine professionnel.

Ma présentation concerne le projet de « développement d'applications numériques pour le « Répertoire historique de l'administration coréenne » de Maurice Courant ». Ce projet n'est pas encore achevé, mais en cours d'opération. Donc, je fais cette présentation, en me limitant à la première phase du projet.

Un projet de numérisation peut être conçu pour un objectif scientifique. Mais jusqu'à présent, la plupart de projets de numérisation ont été réalisés dans le cadre de la gestion de documents, pour une bibliothèque ou de l'édition électronique, sans avoir forcément un projet scientifique (sur les documents en tant que tels). Il s'agit alors d'une numérisation pour conserver et valoriser le document.

¹ Ce travail fait partie du projet du « développement de l'application numérique de « Répertoire historique de l'administration coréenne » de Maurice Courant » subventionné par l'Academy of Korean Studies (AKS-2015-R28).

Je parle ici de ce deuxième cas : « un projet de numérisation dans le cadre de service de documentation ». Ce projet commence par la connaissance des documentalistes sur un document, sa valeur et son utilité et aussi par la connaissance des besoins de ses utilisateurs potentiels.

2. Choix du document : valeur et utilité

2.1. Le « Répertoire historique de l'administration coréenne » de Maurice Courant

Le document concerné par mon projet est le « Répertoire historique de l'administration coréenne » de Maurice Courant, que je vais appeler désormais « le Répertoire ». Ce document est une des références de base les plus utilisées chez des coréanologues, notamment chez les historiens.

Le manuscrit a été écrit par Maurice Courant en 1891 pour ses collègues diplomates français, pendant qu'il travaillait comme traducteur dans l'Ambassade de France en Corée.

Le manuscrit est conçu en 2 volumes : un répertoire principal et un index. Le « Répertoire » contient 349 pages et il est composé en 23 chapitres plus la préface de Maurice Courant et la table des matières, qui se développent de la page n°2 jusqu'à la page n°13.

La composition du contenu est comme le suit :

Préface de Maurice Courant : pp. 2-11

Table des matières : pp. 12-13

chapitres 1-4 : Maison royale

chapitres 5-10 : Les Ministères et les Administrations spéciales

chapitres 11-12 : Les Administrations locales

chapitre 13 : L'Organisation militaire

chapitre 14 : Les Examens civils et les Rangs officiels – La Noblesse

chapitre 15 : La Classe moyenne ou la classe des Interprètes

chapitre 16 : Les relations avec la Chine et le Japon

chapitres 17-18 : *Yamens* (Commis ; Valets)

chapitres 19-20 : Le peuple (classe honorable ; classe vile)

chapitres 21-22 : Le Confucianisme ; Le Bouddhisme

chapitre 23 : Les Administrations nouvelles

Le « Répertoire historique de l'administration coréenne » de Maurice Courant

TABLE DES MATIERES
I – La Maison Royale
II – Les Administrations de la Maison Royale
III – Les Administrations de la Cour
IV – Les Hautes Administrations générales
V – Les Ministères et les Administrations spéciales I – Fonctionnaires civils
VI - II – Cens
VII - III – Rites
VIII - IV – Armée
IX - V – Justice
X - VI – Travaux
XI – Les Administrations locales I – Seoul et les quatre Préfectures-fortereses
XII - II – les Provinces
XIII – L'Organisation militaire
XIV – Les Examens civils et les Rangs officiels – La Noblesse
XV – La Classe moyenne ou la classe des Interprètes
XVI – Les relations avec la Chine et le Japon
XVII – Les Commis des Yamens
XVIII – Les Valets des Yamens
XIX – Le Peuple : la classe honorable – Les Corporations
XX – Le peuple : classe vile – Les Esclaves
XXI – Le Confucianisme
XXII – Le Bouddhisme
XXIII – Les Administrations nouvelles

DocAsie 2016, carre-na.eun-joo@ehess.fr

Maurice Courant ne s'est pas contenté d'énumérer les titres d'officiers, mais il a mis des explications et des commentaires ainsi que l'évolution historique des systèmes d'administration et de divers termes spécifiques.

Marc Orange, « Book reviews » in *Cahiers d'Extrême-Asie* 1991
Volume 6 Numéro 1 pp. 315-316

L'ouvrage se présente en vingt-trois chapitres. Dans les seize premiers, Maurice Courant reprend la forme pyramidale et hiérarchisée de l'organisation administrative de l'époque : la maison royale, la cour et les administrations qui en règlent la vie, puis les ministères, au nombre de six, et les administrations à caractère plus local (capitale et provinces). Mais il a voulu aller plus loin que la sèche énumération de ces rouages administratifs et comme il le dit si bien dans sa préface : "... Je n'ai donc pas pu me borner au répertoire des organes du gouvernement coréen et, pour en faire voir le jeu, j'ai cherché à montrer la matière administrée." D'où ces chapitres consacrés aux différentes classes sociales ainsi qu'aux yamen (et à leur personnel), terme général désignant les bureaux administratifs locaux dont dépendait finalement la vie du peuple. Enfin, avant le dernier chapitre consacré aux administrations nouvelles (créées pour la plupart, sous l'influence des conseillers étrangers au service du gouvernement coréen ou pour répondre à de nouveaux besoins), deux chapitres sont consacrés au confucianisme et au bouddhisme, composantes non négligeables de la "matière administrée".

DocAsie 2016, carre-na.eun-joo@ehess.fr

On constate également qu'il a fait des efforts pour mieux comprendre la société coréenne, en décrivant ses classes sociales et ses deux principales pensées philosophique et religieuse : le Confucianisme et le Bouddhisme.

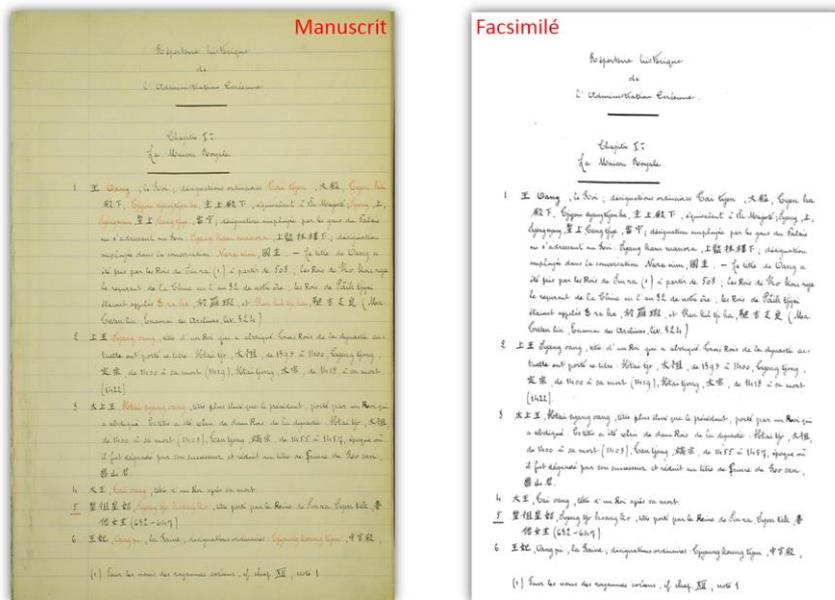
Par ce travail de recherche très riche, ce « Répertoire » fait partie des références importantes pour comprendre la société coréenne de fin du 19^e siècle et l'histoire de

l'administration coréenne. Surtout, il offre la traduction française des titres d'officier et de fonction ainsi que des termes spécifiques en caractères chinois.

Pour cette raison, ce document servait comme une référence indispensable pour les études coréennes, et il n'y a aucun doute que ce document mérite une valorisation numérique pour une plus large diffusion et aussi pour en faciliter d'accès auprès des usagers.

2.2. Evolution du document

Ce document, dont la valeur et l'utilité étant repérée très vite par beaucoup de coréanologues, a vécu différentes évolutions au fil du temps.



DocAsie 2016, carre-na.eun-joo@ehess.fr

2.2.1. En 1891, manuscrit en noir et rouge

En 1891 Maurice Courant a écrit le manuscrit à l'encre noire et rouge. Il l'a écrit en français, en utilisant des caractères chinois pour des titres d'officier, d'organisme ou d'autres termes spécifiques. Il a utilisé l'encre rouge pour transcrire la prononciation de ces termes. Il a adopté le système de transcription utilisé par les missionnaires étrangers de Paris. Maurice Courant a présenté ce manuscrit à un concours de traduction auquel il a gagné un prix, mais ce manuscrit n'a jamais été publié.

2.2.2. En 1986, publication de facsimilé en noir et blanc

En 1986, l'Institut d'études coréennes du Collège de France a publié le facsimilé du « Répertoire » en noir et blanc dans le N° 3 du Cahier d'études coréennes. Dans cette publication, a été ajoutée la préface de Daniel Bouchez, éditeur de cette publication. On a également ajouté dans la partie d'index la transcription du système McCune Reischauer faite par Min Yông-ûi. Le système de transcription de McCune Reischauer est un système de transcription officiellement utilisé dans le domaine scientifique des études coréennes.

The image displays three versions of a document: a manuscript, a facsimile, and a Word-pdf (HAL) version. The manuscript and facsimile show a list of Korean terms with their McCune Reischauer transcriptions. The Word-pdf version shows the same list in a digital format with a search bar and navigation options.

Manuscrit	Facsimilé	Word-pdf (HAL)
親世 357.115	親世 357.115	16 Chôngjong [T'iyeng t'jong]. 靖宗 敬宗 (1034-1046) employa les titres de : 院主 원주, Wônyu [Owen tyow], 宮主 궁주, Kungju [Kowng tyow], 院妃 원비, Wônbi [Owen pi], Dame.
親世 357.116	親世 357.116	17 Munjong [Moun t'jong]. 文宗 元宗 (1046-1083) garda les titres fixés par Hyônjong [Hyen t'jong] et y ajouta : 禮記 덕비, Tokpi [Tek pi], Dame du 3 ^e rang ; 賢妃 현비, Hyônbi [Hyen pi], Dame du 4 ^e rang. La dynastie Yi [Ri] a adopté la hiérarchie ci-dessous
親世 357.117	親世 357.117	18 嬪 빈, Pin, Dame du Palais (concubine royale) du 1 ^{er} rang ; 1a. Celles que le Roi favorise spécialement, reçoivent un nom honorifique en un caractère suivi du titre de Pin ; par une faveur plus grande encore, le Roi donne un nom en deux caractères suivis du titre de Kowng [Kowng]. 宮 中, 貴人 귀인, Kwin [Kou in], Dame du Palais du 2 ^e rang ; 1b ; 2 昭儀 소의, Sowî [Syo owi], Dame du Palais du 3 ^e rang ; 2a ; 淑儀 숙의, Sigwi [Syook owi], Dame du Palais du 4 ^e rang ; 2b ; 昭容 소용, Sowong [Syo yong], Dame du Palais du 5 ^e rang ; 3a ; 淑容 숙용, Sigwong [Syook yong], Dame du Palais du 6 ^e rang ; 3b ; 昭媛 소원, Sowôn [Syo owen], Dame du Palais du 7 ^e rang ; 4a ; 淑媛 숙원, Sigwôn [Syook owen], Dame du Palais du 8 ^e rang ; 4b ;

2.2.3. En 2007, à l'ère du numérique

Format word (2007)

En 2007, ce document entre à l'ère du numérique. Alain Delissen qui était directeur du CRC-EHESS, a fait saisir ce manuscrit sous format Word par Pierre-Emmanuel Roux qui était docteurant mais qui est actuellement maître de conférences à l'Université Paris Diderot. Dans cette version en word, Pierre-Emmanuel Roux a ajouté la transcription han'gûl, alphabet coréen.

Quand la saisie manuelle a été terminée, Alain Delissen m'a demandé de le mettre en ligne soit sur HAL soit sur le site-web du Centre Corée. Mais le site-web du centre basé sur Lodel, n'avait jamais, et n'a toujours pas la capacité technique pour

afficher un document de 360 pages. Et les deux premières versions de HAL non plus ne pouvaient pas publier ce manuscrit trop volumineux (2.3Mo). Il a donc fallu attendre la 3^e version de HAL. La 3^e version de HAL a enfin accepté d'afficher l'intégralité de ce document de 360 pages. Et en 2015 le résultat du projet de numérisation d'Alain Delissen est mis en ligne via HAL. Actuellement il est disponible au public à l'adresse suivante : <https://hal.archives-ouvertes.fr/hal-01149446> (HAL Id : hal-01149446, version 1).

Format image fixe (2012)

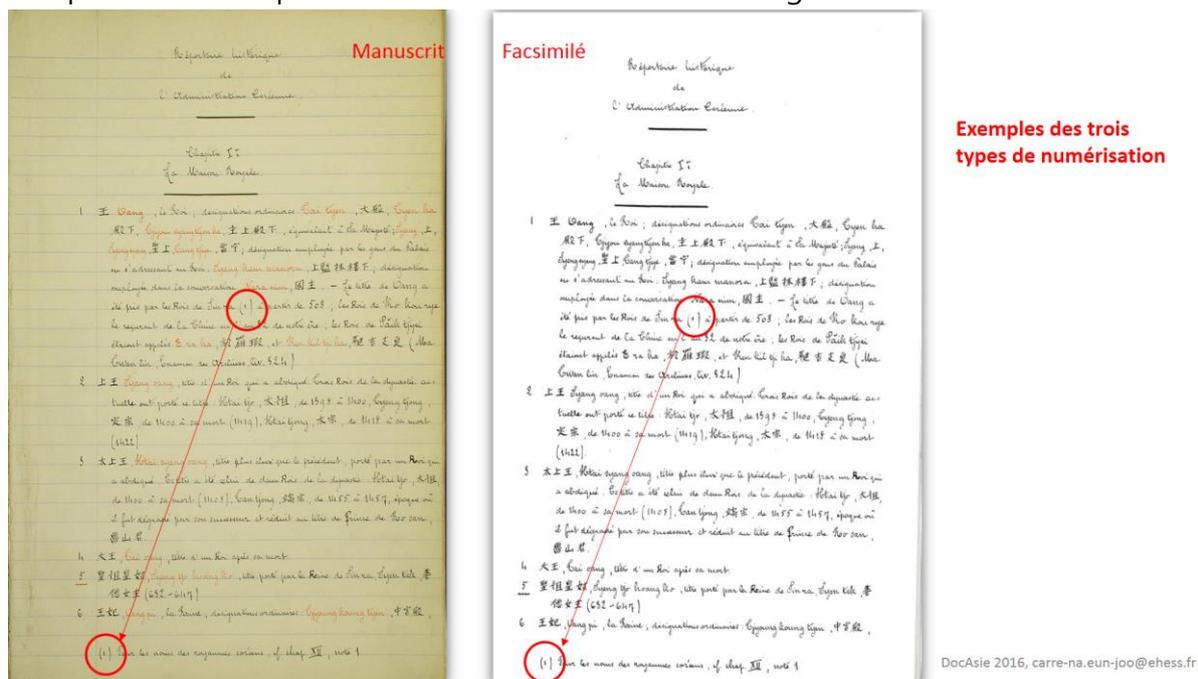
Entre 2010 et 2012, le manuscrit de Maurice Courant a été l'objet d'un autre projet de numérisation. Le document a été scanné et mis en image fixe dans le cadre du projet de « numérisation des anciens livres coréens » dirigé par la Bibliothèque Nationale de Corée en collaboration avec le Collège de France.

Le fait que le « Répertoire » de Maurice Courant a été l'objet de 2 projets de « numérisation » prouve son importance documentaire. Et surtout, nous constatons qu'il y a plusieurs types de numérisations.

3. Trois types de numérisation

Bien qu'on parle souvent de « la » numérisation, il faut être en conscient qu'il y a « des » numérisations et que chaque numérisation correspond à la pratique de différent type de numérisation qui demandent différents techniques et compétences.

Nous venons de parler déjà de 2 types de numérisation avec l'exemple du « Répertoire » tels que la numérisation sous format d'image fixe et celle de format word.



3.1. Pour conserver : en image fixe

Le premier type de numérisation consiste à scanner le texte original pour le mettre en image fixe. Ce type de numérisation est primordial pour la conservation de l'« état originel » du document, tout en mettant le document à la disposition aux usagers. Ce travail est nécessaire pour éventuelles vérifications par les chercheurs. La plupart de projets dit « numérisation » de bibliothèques ou de musées concernent ce type de la « mise en image ». Le projet de numérisation de la Bibliothèque Nationale de Corée en est un exemple. Le « Répertoire » est déjà mis en image, et les images du manuscrit original seront bientôt disponibles grâce au Collège de France via Salamandre.

3.2. Pour référencer : méta-données et textes

Le deuxième type de numérisation sert à fournir des références sur le document et/ou sur le contenu du document à des usagers et à des moteurs de recherche. On peut voir plusieurs types de techniques pratiquées : la description des métadonnées via un SIGB, l'insertion des tags (en xml), ou la saisie manuelle ou automatique (OCR) du texte. On peut utiliser un logiciel OCR avec un scanner pour donner plus d'informations sur le contenu du document. Quand il n'est pas possible de trouver un OCR adapté pour une raison particulière, par exemple dans le cas de documents multilingues, ou d'une langue rare, on peut saisir manuellement sous format word la totalité du document. C'était le cas du projet de « numérisation » réalisé en 2007 sur le « Répertoire » de Maurice Courant.

The image shows a side-by-side comparison of two digital representations of a text document. The left side, titled 'Word (Hal)', shows a list of entries with Korean text and French annotations. The right side, titled 'xml et html', shows the same text in a structured format with XML tags. A red circle highlights a specific entry in the 'Word (Hal)' column, and a red arrow points from it to a corresponding entry in the 'xml et html' column. A red box highlights a word in the 'xml et html' column. At the bottom, there is a red text box that says 'Exemples des trois types de numérisation'.

3.3. Pour explorer : applications de données et « database design »

Le troisième type de numérisation concerne l'exploration des données. Il ne s'agit pas d'afficher un document tel quel, mais de créer une possibilité d'exploration des données contenues dans le document concerné. Autrement dit, il s'agit de donner une nouvelle vie aux potentialités du document. Les pratiques techniques dépendent de la nature de données ainsi que de la nature de l'application à faire. Concernant des documents en texte ordinaire avec une typologie classique, on utilise souvent l'encodage xml avec des tags libres. L'encodage en xml demande une compétence pour comprendre le contenu du document. Cette compétence est en général suffisante pour traiter la plupart des documents textuels, parce qu'ils n'ont souvent qu'une structure textuelle classique et simple : chapitre, paragraphe et phrase. Mais quand il s'agit d'un document textuel qui a une structure atypique ou complexe, il faut une compétence d'analyse textuelle du document pour l'encoder. Le « Répertoire » de Maurice Courant appartient au deuxième cas. Donc, il fallait un travail d'analyse de texte pour son développement numérique.

3.4. Exemples des trois types de numérisation

<p>1 王 Wang, le Roi ; désignations ordinaires Roi Kyeon, 大殿, Oyen ha 殿下, Oyou syang tyeon ha, 主上殿下, équivalent à Sa Majesté; Oyang, 上, Oyangsang, 聖上 Sang tye, 當寧; désignation employée par les gens du Palais en s'adressant au Roi; Oyang kam manora, 上監抹樓下; désignation employée dans la conversation Nara nim, 國主. - Le titre de Wang a été pris par les Rois de Silla (Sin ra) à partir de 503; les Rois de Koguryō le reçurent de la Chine en l'an 32 de notre ère; les Rois de Paek tjejei étaient appelés E ra ha, 於羅環, et Ken kil tji ha, 健吉支更 (Ma Duan lin, Examen des Archives, liv. 324)</p> <p style="text-align: right;">Manuscrit</p>	<p>1 王 Wang, le Roi ; désignations ordinaires Roi Kyeon, 大殿, Oyen ha 殿下, Oyou syang tyeon ha, 主上殿下, équivalent à Sa Majesté; Oyang, 上, Oyangsang, 聖上 Sang tye, 當寧; désignation employée par les gens du Palais en s'adressant au Roi; Oyang kam manora, 上監抹樓下; désignation employée dans la conversation Nara nim, 國主. - Le titre de Wang a été pris par les Rois de Silla (Sin ra) à partir de 503; les Rois de Koguryō le reçurent de la Chine en l'an 32 de notre ère; les Rois de Paek tjejei étaient appelés E ra ha, 於羅環, et Ken kil tji ha, 健吉支更 (Ma Duan lin, Examen des Archives, liv. 324)</p> <p style="text-align: right;">facsimilé</p>
<p style="text-align: center;">Word (pdf HAL)</p> <p>1 王 왕, Wang [Oang], le Roi ; désignations ordinaires Taejōn [Tai tyeon], 大殿 대전, Chōnha [Tyeon ha], 殿下 전하, Chusang chōnha [Tjyou syang tyeon ha], 主上殿下 주상전하, équivalent à Sa Majesté; Sang [Syang], 上 상, Sōngsang [Syengsyang], 聖上 성상, Tangjō [Tang tje], 當寧 당저; désignation employée par les gens du Palais en s'adressant au Roi : Sanggam malluha (manura) [Syang kam manora], 上監抹樓下 상감말루하 (마누라); désignation employée dans la conversation Nara nim, 國主 국주. - Le titre de Wang [Oang] a été pris par les Rois de Silla [Sin ra] à partir de 503; les Rois de Koguryō [Ko kou rye] le reçurent de la Chine en l'an 32 de notre ère; les Rois de Paekche [Päik tjejei] étaient appelés Ōnaha [E ra ha], 於羅環 어나하, et Kōn'giljiha [Ken kil tji ha], 健吉支更 건길지하 (Ma Duanlin [Ma Twan lin], Examen des Archives, liv. 324).</p>	<p style="text-align: center;">xml-tei et html</p> <p>1 王 왕, Wang [Oang], le Roi ; désignations ordinaires Taejōn [Tai tyeon], 大殿 대전, Chōnha [Tyeon ha], 殿下 전하, Chusang chōnha [Tjyou syang tyeon ha], 主上殿下 주상전하, équivalent à Sa Majesté; Sang [Syang], 上 상, Sōngsang [Syengsyang], 聖上 성상, Tangjō [Tang tje], 當寧 당저; désignation employée par les gens du Palais en s'adressant au Roi : Sanggam malluha (manura) [Syang kam manora], 上監抹樓下 상감 말루하 (마누라); désignation employée dans la conversation Nara nim, 國主 국주.</p> <p>- Le titre de Wang [Oang] a été pris par les Rois de Silla [Sin ra] à partir de 503; les Rois de Koguryō [Ko kou rye] le reçurent de la Chine en l'an 32 de notre ère; les Rois de Paekche [Päik tjejei] étaient appelés Ōnaha [E ra ha], 於羅環 어나하, et Kōn'giljiha [Ken kil tji ha], 健吉支更 건길지하 (Ma Duanlin [Ma Twan lin], Examen des Archives, liv. 324).</p>

Prenons un exemple du « répertoire » pour comparer ces trois types de numérisation.

Tandis que les deux premiers types de numérisations affichent ce qui est produit par le travail, dans le troisième cas, le produit du travail de numérisation ne correspond pas à ce qui est affiché sur l'interface d'utilisation.

Dans les versions en image et en word, ce qui est affiché et vu est identique à ce qui est produit par le rédacteur. Mais dans le 4^e exemple, ce qui est affiché sur l'écran ressemble à la version en word, mais il s'agit du résultat d'un travail qui reste invisible pour l'utilisateur. Pour parvenir au résultat du 4^e exemple, il est nécessaire d'effectuer d'abord le travail en arrière-plan comme ces fichiers xml-tei et html.

```
<div type="theme" xml:id="cht1_RoiEtReine">
  <div xml:id="rhac.n-1">
    <ab-1</ab>
    <p>
      <roleName>
        <roleName xml:lang="zh">王</roleName>
        <roleName xml:lang="ko">왕</roleName>,
        <roleName xml:lang="ko-acc">왕</roleName>
        <roleName xml:lang="ko-sep">[Dang]</roleName>,
        <roleName xml:lang="fr">le Roi</roleName>
      </roleName>
      <seg>
        <roleName xml:lang="fr" n="1">designations ordinaires</roleName>
        <roleName xml:lang="ko-acc" n="1">Taejŏn</roleName>
        <roleName xml:lang="ko-sep" n="1">[Tai tyeŋ]</roleName>,
        <roleName xml:lang="zh" n="1">大興</roleName>
        <roleName xml:lang="ko" n="1">대흥</roleName>,
        <roleName xml:lang="ko-acc" n="2">Chŏnha</roleName>
        <roleName xml:lang="ko-sep" n="2">[Tyeŋ ha]</roleName>,
        <roleName xml:lang="zh" n="2">殿下</roleName>
        <roleName xml:lang="ko" n="2">전하</roleName>,
        <roleName xml:lang="ko-acc" n="3">Chusang chŏnha</roleName>
        <roleName xml:lang="ko-sep" n="3">[Tjyou syang tyeŋ ha]</roleName>
        <roleName xml:lang="zh" n="3">主上殿下</roleName>
        <roleName xml:lang="ko" n="3">주상전하</roleName>,
        équivalent à
        <roleName xml:lang="fr" n="2">Sa Majesté</roleName> ;
        <roleName xml:lang="ko-acc" n="4">Sang</roleName>
        <roleName xml:lang="ko-sep" n="4">[Syang]</roleName>,
        <roleName xml:lang="zh" n="4">上</roleName>
        <roleName xml:lang="ko" n="4">상</roleName>,
        <roleName xml:lang="ko-acc" n="5">Sŏngsang</roleName>
        <roleName xml:lang="ko-sep" n="5">[Syngssyng]</roleName>,
        <roleName xml:lang="zh" n="5">皇上</roleName>
        <roleName xml:lang="ko" n="5">성상</roleName>,
        <roleName xml:lang="ko-acc" n="6">Tanjŏ</roleName>
        <roleName xml:lang="ko-sep" n="6">[Tang tjye]</roleName>,
        <roleName xml:lang="zh" n="6">當堂</roleName>
        <roleName xml:lang="ko" n="6">당장</roleName>
      </seg>
    </p>
  </div>
</div>
```

Xml-tei

```
<seg>
  <roleName xml:lang="fr">designation employée par les gens du Palais en s'adressant au Roi</roleName> ;
  <roleName xml:lang="ko-acc">Sŏngsan mai luha (manura)</roleName>
  <roleName xml:lang="ko-sep">[Syng san sanora]</roleName>,
  <roleName xml:lang="zh">上皇陛下</roleName>
  <roleName xml:lang="ko">상강달루하 (단누리)</roleName></seg>
</seg>
<seg>
  <roleName xml:lang="fr">designation employée dans la conversation</roleName>
  <roleName xml:lang="ko-acc">Nara nira</roleName>,
  <roleName xml:lang="zh">禮主</roleName>
  <roleName xml:lang="ko">리주</roleName></seg></roleName>
</div>
<1> Le titre de Wang [Dang] a été pris par les Rois de
  <date>
    <date xml:lang="ko-acc">Silla</date>
    <date xml:lang="ko-sep">[Sin ra]</date>
  </date>
  <ref target="#001">1</ref>
  <date>à partir de 503</date> ;
  les Rois de
  <date>
    <date xml:lang="ko-acc">Koguryŏ</date>
    <date xml:lang="ko-sep">[Ko kou rye]</date></date>
  le règne de <geolName>la Chine</geolName> <date>en l' an 32 de notre ère</date> ;
  <roleName xml:lang="fr">les Rois de <date>Paikche [Paik tjeŋ]</date></roleName>
  étaient appelés
  <roleName xml:lang="ko-acc" n="1">ŏnha</roleName>
  <roleName xml:lang="ko-sep" n="1">[E ra ha]</roleName>,
  <roleName xml:lang="zh" n="1">於羅呼</roleName>
  <roleName xml:lang="ko" n="1">올러</roleName>, et
  <roleName xml:lang="ko-acc" n="2">Kŏn giljŏha</roleName>
  <roleName xml:lang="ko-sep" n="2">[Kŏn kil tji ha]</roleName>,
  <roleName xml:lang="zh" n="2">曠吉支夏</roleName>
  <roleName xml:lang="ko" n="2">관길지하</roleName> </roleName>
  </div>
  <bibi>[Me Duan lin [Me Ten lin], Examen des Archives, liv. 324</span></div>
</1></p>
```

DocAsie 2016, carre-na.eun-joo@ehess.fr

```
<div class="theme" id="cht1_RoiEtReine">
  <div>
    <p>
      <span class="roleName" lang="zh">王</span>
      <span class="roleName" lang="ko">왕</span>,
      <span class="roleName" lang="ko-acc">왕</span>
      <span class="roleName" lang="ko-sep">[Dang]</span>,
      <span class="roleName" lang="fr">le Roi</span>
    </p>
    <p>
      <span class="roleName" lang="fr">designations ordinaires</span>
      <span class="roleName" lang="ko-acc">Taejŏn</span>
      <span class="roleName" lang="ko-sep">[Tai tyeŋ]</span>,
      <span class="roleName" lang="zh">大興</span>
      <span class="roleName" lang="ko">대흥</span>,
      <span class="roleName" lang="ko-acc">Chŏnha</span>
      <span class="roleName" lang="ko-sep">[Tyeŋ ha]</span>,
      <span class="roleName" lang="zh">殿下</span>
      <span class="roleName" lang="ko">전하</span>,
      <span class="roleName" lang="ko-acc">Chusang chŏnha</span>
      <span class="roleName" lang="ko-sep">[Tjyou syang tyeŋ ha]</span>,
      <span class="roleName" lang="zh">主上殿下</span>
      <span class="roleName" lang="ko">주상전하</span>,
      équivalent à
      <span class="roleName" lang="fr">Sa Majesté</span> ;
      <span class="roleName" lang="ko-acc">Sang</span>
      <span class="roleName" lang="ko-sep">[Syang]</span>,
      <span class="roleName" lang="zh">上</span>
      <span class="roleName" lang="ko">상</span>,
      <span class="roleName" lang="ko-acc">Sŏngsang</span>
      <span class="roleName" lang="ko-sep">[Syngssyng]</span>,
      <span class="roleName" lang="zh">皇上</span>
      <span class="roleName" lang="ko">성상</span>,
      <span class="roleName" lang="ko-acc">Tanjŏ</span>
      <span class="roleName" lang="ko-sep">[Tang tjye]</span>,
      <span class="roleName" lang="zh">當堂</span>
      <span class="roleName" lang="ko">당장</span>
    </p>
    <p>
      <span class="roleName" lang="fr">designation employée par les gens du Palais en s'adressant au Roi</span> ;
      <span class="roleName" lang="ko-acc">Sŏngsan mai luha (manura)</span>
      <span class="roleName" lang="ko-sep">[Syng san sanora]</span>,
      <span class="roleName" lang="zh">上皇陛下</span>
      <span class="roleName" lang="ko">상강달루하 (단누리)</span>
    </p>
    <p>
      <span class="roleName" lang="fr">designation employée dans la conversation</span>
      <span class="roleName" lang="ko-acc">Nara nira</span>,
      <span class="roleName" lang="zh">禮主</span>
      <span class="roleName" lang="ko">리주</span>
    </p>
  </div>
</div>
```

Html (pour l'affichage en ligne)

```
<div class="1">
  <span class="date">
    <span lang="ko-acc" class="date">Silla</span>
    <span lang="ko-sep" class="date">[Sin ra]</span>
  </span>
  <a class="link_ref" href="#001" title="1">1</a>
  <span class="date">à partir de 503</span> ;
  les Rois de
  <span class="date">
    <span lang="ko-acc" class="date">Koguryŏ</span>
    <span lang="ko-sep" class="date">[Ko kou rye]</span>
  </span>
  le règne de
  <span class="geolName">la Chine</span>
  <span class="date">en l' an 32 de notre ère</span> ;
  les Rois de
  <span class="date">
    <span lang="ko-acc" class="date">ŏnha</span>
    <span lang="ko-sep" class="date">[E ra ha]</span>,
    <span lang="zh" class="date">於羅呼</span>
    <span lang="ko" class="date">올러</span>, et
    <span lang="ko-acc" class="date">Kŏn giljŏha</span>
    <span lang="ko-sep" class="date">[Kŏn kil tji ha]</span>,
    <span lang="zh" class="date">曠吉支夏</span>
    <span lang="ko" class="date">관길지하</span>
  </span>
  </div>
  <bibi>[Me Duan lin [Me Ten lin], Examen des Archives, liv. 324</span></div>
```

DocAsie 2016, carre-na.eun-joo@ehess.fr

Dans cette présentation, nous parlons de ce troisième type de numérisation : la numérisation pour « application ou exploration des données ».

4. Objectifs de travail : données à explorer, « Database design »

Je rappelle l'objectif du projet : « en sachant que nos chercheurs utilisent ce document comme une référence de base, je voudrais faire quelque-chose pour rendre plus facile et plus efficace l'utilisation de ce document, en repérant des données utiles ». Or, quelles sont les données explorables ou qui méritent exploration ? Comment les définir ? Mes collègues coréens appellent ce travail « data design ». Ici, j'ai mis une citation :

« Database design versus visualisation de l'information
 En vérité, lorsqu'on parle de design de données deux notions sont évoquées, qui ont trait à deux métiers différents. Une plus ancienne et plus technique fait référence au **processus de structuration de modèles logiques en vue de la création de bases de données**. Une autre, plus récente et moins répandue, englobe les **nouveaux métiers de l'analyse, de la scénarisation et de la représentation de données numériques** ayant comme but de rendre l'information compréhensible par le plus grand nombre de personnes. C'est cette dernière qui nous intéresse. »
 (Data Observer, *Notre métier, c'est aussi « designer » des données* (<http://data-observer.com/fr/design-des-donnees/>), consulté en mai 2016)

« analyser et scénariser le document pour mieux faire sortir sa valeur potentielle, en concevant un modèle logique de ses données à explorer » : ma propre définition

DocAsie 2016, carre-na.eun-joo@ehess.fr



J'ai emprunté cette définition que j'ai trouvée sur internet. Mais j'ai remodulé ses mots pour définir le « database design » à ma façon, parce que, au moins pour mon projet actuel, le travail concerne les deux parties : « analyser et scénariser le document pour mieux faire sortir sa valeur potentielle, en concevant un modèle logique de ses données à explorer ».

Le « database design » peut se faire avec un projet scientifique concret qui a un objectif particulier pour une application précise. Dans ce cas, on peut faire à partir d'un seul document, différents projets d'application numériques.

Mais un projet d'une application numérique peut se faire sans projet scientifique particulier pour certains documents. Surtout quand il s'agit d'un document qui a des caractères spécifiques, dont l'usage est explicitement défini lors de sa production.

Le « Répertoire », créé par un traducteur pour les diplomates français en Corée, est un bon exemple de ce type de document. Il n'est pas nécessaire d'une connaissance d'expert, pour repérer la potentialité d'exploration de ses données. Le « Répertoire » qui est une vraie boîte de concentration de données m'a causé plusieurs difficultés avec sa richesse dont on va parler plus tard.

En bref, j'ai voulu rester fidèle à la nature de ce document et à sa fonction de « Répertoire ». Et j'ai pris comme premier objectif de repérer et marquer les termes qui s'affichent en diverses formes linguistiques : caractère chinois, coréen et français, et les 2 types de transcription.

5. Travail préalable : droit d'auteur

Objet de droit d'auteur	Libre de droit
Préface et annotations qui contiennent de la créativité intellectuelle.	Manuscrit écrit en 1891, facsimilé, saisie du texte, transcriptions

DocAsie 2016, carre-na.eun-joo@ehess.fr

Il va de soi qu'il faut préalablement consulter pour le document à numériser des spécialistes du droit d'auteur. Dans le cas du « Répertoire » de Maurice Courant, la conclusion des deux experts (juridique et expert de numérisation de manuscrit) était identique :

- Le manuscrit écrit en 1891 est un document de libre droit.
- Ni le facsimilé ni la saisie manuelle du texte ne sont d'objet de droit d'auteur, sauf des ajouts comme la préface ou des annotations qui contiennent de la créativité intellectuelle.
- Les transcriptions ne sont pas d'objet du droit d'auteur.

Par conséquent, le présent projet du « développement d'application numérique du « Répertoire » » prend comme objet de travail seulement des parties libres de droit.

6. Choix des outils de travail

Type de numérisation	Outils de travail
Image fixe	Photographie, scanner (jpeg, png, gif, etc.)
Réréférences	Méta-données, SIGB, OCR, saisie manuelle de text (pdf, word, dublin core, rdf, etc.)
Applications numériques	Encodages (XML, TEI, etc.) éditeurs XML (oXygen, etc.)

DocAsie 2016, carre-na.eun-joo@ehess.fr

Chaque type de numérisation est autonome et il exige différents outils et compétences. Pour les deux premiers types de numérisation : la mise en image et le référencement, on utilise comme outil des appareils de photo et des scanners, ainsi que différents logiciels de SIGB ou l'OCR. Les outils du troisième type de numérisation pour des applications des données dépendent de l'objectif d'application. Mais quand il s'agit d'un document textuel, c'est souvent l'encodage en xml avec ou sans l'aide d'un éditeur xml.

6.1. Xml, TEI (Text Encoding Initiative) et oXygen

Les outils que j'ai choisis sont XML et son éditeur oXygen, ainsi que TEI. OXygen est l'éditeur xml le plus utilisé et recommandé en France. La licence pour l'usage académique coûte 108€. Mais on peut commencer par un essai gratuit. Avec oXygen on peut générer des schémas xml et convertir des fichiers xml en fichiers html. Le TEI (Text Encoding Initiative) est un "encodage normalisé" pour le document textuel.

Ici, j'ai pris quelques diapos de la présentation de TEI de Jean-Luc Benoit. Comme ce qui est indiqué dans les citations, l'encodage « sert à spécifier les caractéristiques d'un texte », mais comme c'est « une interprétation », il n'y a pas « un seul encodage complet et exhaustif ». Le TEI est un encodage standardisé basé sur des études de longue date des sciences de langage et d'informatique, et il propose divers modèles basés sur des prototypes textuels avec lesquels on peut décrire la plupart des genres de texte. Cela aide les documentalistes, qui n'ont pas toujours des connaissances suffisantes des langages informatiques, à trouver des schémas applicables à un type de document en cours de travail.

TEI : définition.

- Text Encoding Initiative.
- Encoding and Interchange.
- Groupe d'initiative pour un encodage normalisé.

BENOIT, Jean-Luc, Il y a une vie après MARC, 19 novembre 2007 : journée d'étude en hommage à Pierre-Yves Duchemin.

{ Digression : encodage ? 1 }

Un format d'échange s'exprime dans un encodage.

Pourquoi encoder ? Que veut-on encoder ?

- L'encodage exprime un aspect quelconque d'un document.
- tout encodage est donc une interprétation.
- aucun langage d'encodage ne peut donc prétendre être ni complet ni exhaustif.

{ Digression : encodage ? 2 }

Un langage d'encodage sert à...

- Spécifie les caractéristiques d'un texte.
- Rend explicite la/les structures aperçue/s dans un texte.
- Spécifie les méta-informations, renseignements contextuels etc.

A quoi sert la TEI ?

- **Faciliter la création, l'échange, et l'intégration des données textuelles informatisées.**
 - toute sorte de texte.
 - toutes les langues.
 - toute provenance temporelle ou culturelle.
- **La TEI doit servir également à aider...**
 - les *débutants*, cherchant des solutions bien connues et consensuelles.
 - les *experts*, cherchant à créer de nouvelles solutions.

TEI P5 Version 1.0

Les caractéristiques qui ont fait de la TEI le standard d'encodage préféré des 20 dernières années est conservé.

- L'accès libre et gratuit sur le Web.
- Une architecture modulaire qui peut être facilement être paramétrable.
- Un nombre très importants d'éléments capables de nuancer l'encodage de manière très fine et qui est capable de s'adapter à la presque totalité des textes.

DocAsie 2016, carre-na.eun-joo@ehess.fr

6.2. Compatibilités et tolérances linguistiques

Le choix des outils de travail doit se faire après l'examen des compatibilités. Il faut se demander si les outils sont compatibles avec différents environnements informatiques, Pc et Mac et aussi différents navigateurs. Et il faut voire aussi s'ils tolèrent différents types de caractères et langues, y compris des formes anciennes et des rares. Le « Répertoire » de mon projet contient des anciens caractères chinois avec quelques caractères rares. Quand j'ai inséré sa version word dans oXygen, tous les caractères ont été cassés, y compris le coréen qui est alphabet relativement simple à traiter dans l'informatique. Et j'ai dû convertir d'abord le fichier word en texte brut, puis l'insérer morceau par morceau, en vérifiant.

7. *Objet de travail : structure et composants*

Avant de commencer l'encodage, il faut d'abord analyser le document pour mieux comprendre ses caractéristiques textuelles, et aussi pour repérer et déterminer les données à explorer. C'est le début de « database design ».

7.1. Triple structure textuelle : typologique, thématique et numérative

Structures numérative et thématique

The diagram illustrates the triple structure of text through two pages of handwritten text. The top page shows entries 11, 12, and 13, with a box labeled 'Division thématique' pointing to a horizontal line separating the entries. The bottom page shows entries 14 and 15, with a box labeled 'Numéro souligné' pointing to the number 14 and a box labeled 'Numéro non-souligné' pointing to the number 15. The text is in French and includes Chinese characters and names.

DocAsie 2016, carre-na.eun-joo@ehess.fr

Au contraire d'un simple encodage en xml, le TEI exige la description de la structure hiérarchique du texte entier. Donc si on veut utiliser le TEI, il faut analyser la structure textuelle et déterminer sa hiérarchie structurale. Quant à la structure textuelle du « Répertoire », c'est plus complexe, car le « Répertoire » a une triple structure textuelle.

D'abord, il est construit d'après une « structure typologique » classique : chapitre, paragraphe et phrase. Ensuite chaque chapitre a des sub-divisiones thématiques. Je l'appelle une « structure thématique ». Maurice Courant a marqué cette structure par une ligne horizontale. De plus, en même temps que ces structures, le « Répertoire » construit tout au long du texte une « structure numérative ». La partie principale du document s'énumère de 1 à 1490. Chaque numéro contient un ou plusieurs titres de fonction administrative ou sociale, ou ceux d'organes gouvernementaux ou encore divers termes spécifiques. Mais il peut contenir aussi un ou plusieurs paragraphes d'explication sur les contextes socio-historiques du terme concerné. D'ailleurs, ces numéros sont classés en deux groupes : souligné et non-souligné.

Les structures typologique et numérative se chevauchent souvent sans cohérence. Cette structure textuelle atypique est un problème propre au « Répertoire » de Maurice Courant, et elle crée une très grande difficulté, surtout dans l'application de TEI qui exige la description de la structure hiérarchique du texte.

qui se suivent. Le 3^e exemple qu'on trouve dans les numéros de 13 à 15 contient un chevauchement multiple entre la structure de paragraphe et celle d'énumération. C'est-à-dire que n°14 et n°15 font respectivement partie des dernières parties du numéro précédent. Le n°13 « Hyŏnjong [Hyen tjong], 顯宗 현종 (1009-1031) fixa pour les Servantes les titres de : » et n°14 « et pour les Dames ceux de : ».

7.2. Composants linguistiques et contextuels

Composants linguistiques

Unité minimale

2 上王 Syang oang, titre d'un Roi qui a abdiqué.

2 上王 상왕, Sangwang [Syangoang], titre d'un Roi qui a abdiqué.

```
<div xml:id="rhac.n-2">
  <ab>2</ab>
  <p>
    <roleName>
      <roleName xml:lang="zh">上王</roleName>
      <roleName xml:lang="ko">상왕</roleName>,
      <roleName xml:lang="ko-mcc">Sangwang</roleName>
      <roleName xml:lang="ko-mep">[Syangoang]</roleName>,
      <roleName xml:lang="fr">titre d'un Roi qui a abdiqué</roleName>
    </roleName>
  </p>
</div>
```

DocAsie 2016, carre-na.eun-joo@ehess.fr

Dans le cas du « Répertoire », tous les termes en caractères chinois accompagnés par la traduction française faite par Maurice Courant constitueront des éléments les plus intéressants, et des données à explorer. Donc, dans mon projet les données à explorer constituent un ensemble multilingue qui contient 5 composants linguistiques : les caractères chinois classiques (voire anciens), le coréen, le français, et deux systèmes de transcription. Je définis comme unité minimale cet ensemble qui contient toujours ces 5 éléments linguistiques.

Composants contextuels

complémentaire

Contexte sociale :

- Relation
- Hiérarchique
- Fonctions

Contexte historique :

- Création
- Modification
- Suppression
- Date

Avec des composants linguistiques, on trouve aussi des composants contextuels. Ces composants contextuels peuvent se diviser encore en deux types différents : contexte social et contexte historique. Je définis comme contexte social ce qui décrit les relations sociale ou hiérarchique entre des éléments, ou encore les fonctions liées au terme en question.

Je définis comme contexte historique ce qui concerne la création, la modification ou la suppression du terme en question. Ce contexte historique est souvent accompagné par une expression de date ou de période.

7.3. Unité minimale

Avec ces composants, on peut définir une unité minimale et déduire sa forme.

7.3.1. Définition et forme

Définition	1 FR (+Ctxt sociale et/ou historique (= explication, précision))	=	1 ZH (+Ctxt sociale et/ou historique (= condition d'usage))
Forme	FR1(+ctxt)=ZH1(+ctxt)		FR1 = traduction Ctxt = contexte sociale ou historique

L'unité minimale contient un élément en caractère chinois (ZH1) et un autre en français qui est sa traduction (FR1). Cette unité minimale se présente avec un ou plusieurs contextes sociaux ou historiques (ctxt). Ces contextes peuvent se présenter soit explicitement soit implicitement. Il se peut qu'ils ne soient pas précisés.

Quand le contexte appartient à l'élément français (FR1), qui est la traduction de l'élément en caractère chinois (ZH1), il s'agit souvent d'une précision ou une explication concernant l'entité (ou le terme) en question. Quand le contexte appartient à une variété de l'élément en caractères chinois (ZH2), il s'agit souvent d'une condition d'usage du terme en question.

7.3.2. Exemples de l'unité minimale

Exemples de l'unité minimale

```

<term>
<roleName xml:lang="zh">魏妃</roleName>
<roleName xml:lang="ko">원비</roleName>
<roleName xml:lang="ko-mcc">Wŏnbi</roleName>
<roleName xml:lang="ko-mep">[Queen pŭ]</roleName>
<roleName xml:lang="fr">Dame</roleName>
</term>
EX. n° 42, n° 43 : contexte sociale

```

```

EX. n° 117 : contexte historique
<term>
<orgName xml:lang="ko-mcc">Chapkongsa</orgName>
<orgName xml:lang="ko-mep">[Tʃap kong sʌ]</orgName>
<orgName xml:lang="zh">龍工司</orgName>
<orgName xml:lang="ko">장공사</orgName>
</term>
<seg> sous
<persName>
<persName xml:lang="ko-mcc">Kyŏngdŏk</persName>
<persName xml:lang="ko-mep">[Kyeng tek]</persName>
<persName xml:lang="zh">景德王</persName>
<persName xml:lang="ko">경덕왕</persName> (742-765)</persName>
</seg>
<orgName xml:lang="fr">Bureau des Ouvrages divers</orgName>
</term>
EX. n° 223 : contexte historique
<orgName>
<orgName xml:lang="fr">Le Palais Kyŏngbok</orgName>
<orgName xml:lang="ko-mep">[Kyeng pok]</orgName>
<seg> (<geogName>Seoul</geogName>), bâti <date>en 1394</date>,
fut ruiné lors de la guerre avec les Japonais <date>en 1592</date>;
il a été reconstruit <date>en 1865</date> et est redevenu
<date>depuis 1884</date> </seg>
<orgName xml:lang="fr">la résidence royale</orgName>
</orgName>

```

N°42, 43, 56

N°117, 223

DocAsie 2016, carre-na.eun-joo@ehess.fr

Voici les exemples de plusieurs configurations de contextes dans une unité minimale.

N° 42, 43 et 56 sont des exemples contenant un contexte social, tandis que N° 117 et 223 ont des contextes historiques.

7.3.4. Conception d'encodage pour l'unité minimale

Une fois l'analyse faite, on peut concevoir un modèle d'encodage. J'ai décidé d'utiliser des tags sémantiques comme « roleName » ou « orgName » pour identifier chaque unité. Ensuite d'utiliser des attributs comme « @xml:lang » pour exprimer les caractéristiques linguistiques des éléments composants.

7.4. Variété de formes de l'unité

Les éléments composants peuvent se combiner de plusieurs manières. A partir de diverses combinaisons possibles de ces composants, on peut déduire certains types de composition et leurs variétés de forme.

Voici le tableau que j'ai conçu avec différentes variétés que j'ai repérées.

Variété de formes de l'unité

	Définition + forme		Traces linguistiques
Unité minimale	1 FR (+Ctxt sociale et/ou historique (= explication, précision))		= 1 ZH (+Ctxt sociale et/ou historique (= condition d'usage))
Forme	FR1(+ctxt)=ZH1(+ctxt)	FR1 = traduction Ctxt = contexte sociale ou historique	
Variété 1 (V1)	1 FR (+Ctxt sociale et/ou historique (= explication, précision))		= n ZH (+Ctxt sociale et/ou historique (= condition d'usage))
V 1-1	FR1=ZH1=ZH2	ZH2 = Synonyme	avec ou sans « appelé aussi », « nommé aussi »
V 1-2	FR1=ZH1=ZH2+Ctxt FR1+Ctxt=ZH1=ZH2	ZH2+Ctxt sociale (condition ou situation d'usage) FR1+Ctxt sociale (hiérarchie, fonction)	ZH2 : « nom littéraire », « désignation » ; FR1 : « dépend de »
V 1-3	FR1(+Ctxt ?)=ZH1=ZH2+Ctxt ?FR1+Ctxt=ZH1=ZH2	ZH2+Ctxt historique (évolution historique) FR1+Ctxt historique (précision)	« date », « roi », « créer », « supprimer », « séparer », « rassembler », « restituer »
Variété 2 (V2)	n FR (+Ctxt sociale et/ou historique (= explication, précision))		/ n ZH (+Ctxt sociale et/ou historique (= condition d'usage))
V 2-1	FR2=ZH2+Ctxt ZH1≠ZH2, FR1≠FR2	Ctxt = historique ZH2 = évolution historique	« date », « roi », « créer », « supprimer », « séparer », « rassembler », « restituer »
V 2-2	FR2=ZH2+Ctxt FR2-1 + FR2-2=FR2	Ctxt = sociale ou historique FR2 = segment divisé	

DocAsie 2016, carre-na.eun-joo@ehess.fr

7.4.1. Variété n°1 (V1) : 1 FR = n ZH

Le premier type de variété (V1) consiste dans l'unité qui contient un élément français (1FR) accompagné par plusieurs éléments en caractères chinois (n ZH).

Variété 1 (V1)	1 FR (+Ctxt sociale et/ou historique (= explication, précision))	=	n ZH (+Ctxt sociale et/ou historique (= condition d'usage))
-----------------------	---	----------	--

Cette variété n° 1 (V1) prend 3 formes différentes : celle qui n'affiche aucun contexte pour la forme variée de l'élément (ZH2) d'une façon explicite, celle qui contient un contexte social et celle qui contient un contexte historique.

7.4.1.1. *Forme n°1 de V1*

V 1-1	FR1=ZH1=ZH2	ZH2 = Synonyme	avec ou sans « appelé aussi », « nommé aussi »
-------	-------------	----------------	--

La première forme de la V1 (V1-1) concerne des synonymes, ou d'autres appellations. Elle peut avoir comme traces linguistiques « applé aussi », « nommé aussi », etc.. Les entités des éléments (FR1, ZH1 et ZH2) sont identiques.

Exemples de V 1-1 :
FR1=ZH1=ZH2

147
少年監典 소년감전, Sonyŏn kamjŏn [Syŏ nyen kam tyen], Bureau des Jeunes Eunuques (?), appelé aussi Choch'ŏnsŏng [Tyo htyen syeng] 鈞天省 조천성.

16
Chŏngjong [Tjyeng tjong], 靖宗 경종 (1034-1046) employa les titres de :
院主 원주, Wŏnju [Ouen tjyou], 宮主 궁주, Kungju [Koung tjyou], princesse ;
院妃 원비, Wŏnbi [Ouen pi], Dame.

N° 16
<roleName>
<roleName xml:lang="zh" n="1">院主</roleName>
<roleName xml:lang="ko" n="1">원주</roleName>
<roleName xml:lang="ko-mcc" n="1">Wŏnju</roleName>
<roleName xml:lang="ko-mep" n="1">[Ouen tjyou]</roleName>
<roleName xml:lang="zh" n="2">宮主</roleName>
<roleName xml:lang="ko" n="2">궁주</roleName>
<roleName xml:lang="ko-mcc" n="2">Kungju</roleName>
<roleName xml:lang="ko-mep" n="2">[Koung tjyou]</roleName>
<roleName xml:lang="fr">princesse</roleName>;
</roleName>

168
Sous la dynastie de Koryŏ ([Korye]), quelques fonctionnaires formaient la Maison, Yosok [Ryo syok], 德屬 邑司, de la Reine; ils portaient les titres suivants :
左右監事 좌우감사, Chwa-u ch'ŏmsa [Tja ou tchym sã], 左右司尹 좌우사음, Chwa-u sayun [Tja ou sã] [Page 31 youn], Superintendants de droite et de gauche ; on les appelait aussi 左右司監 좌우사감, Chwa-u sagŭm [Tja ou sã keum] ;
指諭 지유, Chiyu [Tji you], Chambellan ;
行宮 行宮, Hængsu [Hing syou], Premier Ecuyer ;
掌監 掌監, Kyŏlyŏng [Kyen ryong], Ecuyer ;
侍衛 수호, Suho [Syou ho], Chef des Gardes ;
侍衛軍 시위군, Siwŏn [Si oui goun], Garde, etc.

n° 147 :
<orgName>
<orgName xml:lang="zh" n="1">少年監典</orgName>
<orgName xml:lang="ko" n="1">소년감전</orgName>
<orgName xml:lang="ko-mcc" n="1">Sonyŏn kamjŏn</orgName>
<orgName xml:lang="ko-mep" n="1">[Syŏ nyen kam tyen]</orgName>
<orgName xml:lang="fr">Bureau des Jeunes Eunuques (?)</orgName>
<seg>appelé aussi
<terms>
<orgName xml:lang="ko-mcc" n="2">Choch'ŏnsŏng</orgName>
<orgName xml:lang="ko-mep" n="2">[Tyo htyen syeng]</orgName>
<orgName xml:lang="zh" n="2">鈞天省</orgName>
<orgName xml:lang="ko" n="2">조천성</orgName></term>
</seg>
</orgName>

n.168
<roleName>
<roleName xml:lang="zh" n="1">左右監事</roleName>
<roleName xml:lang="ko" n="1">좌우감사</roleName>
<roleName xml:lang="ko-mcc" n="1">Chwa-u ch'ŏmsa</roleName>
<roleName xml:lang="ko-mep" n="1">[Tja ou tchym sã]</roleName>
<roleName xml:lang="zh" n="2">左右司尹</roleName>
<roleName xml:lang="ko" n="2">좌우사음</roleName>
<roleName xml:lang="ko-mcc" n="2">Chwa-u sayun</roleName>
<roleName xml:lang="ko-mep" n="2">[Tja ou sã youn]</roleName>
<roleName xml:lang="fr">Superintendants de droite et de gauche</roleName>;
<seg>on les appelait aussi
<roleName>
<roleName xml:lang="zh" n="1">左右司監</roleName>
<roleName xml:lang="ko" n="1">좌우사감</roleName>
<roleName xml:lang="ko-mcc" n="1">Chwa-u sagŭm</roleName>
<roleName xml:lang="ko-mep" n="1">[Tja ou sã keum]</roleName>
</roleName>
</seg>
</roleName>

7.4.1.2. Forme n° 2 de V1

V 1-2	FR1=ZH1=ZH2+Ctxt	ZH2+Ctxt sociale	ZH2 : « nom littéraire »,
	FR1+Ctxt=ZH1=ZH2	(condition ou situation d'usage) FR1+Ctxt sociale (hiérarchie, fonction)	« désignation » ; FR1 : « dépend de »

La deuxième forme de la V1 (V1-2) concerne différentes appellations d'une même entité, qui dépendent de condition ou situation sociales diverses. Ces variétés ont souvent comme traces linguistiques des « désignations » ou des « noms littéraires ». Ce dernier est un terme utilisé par Maurice Courant avec sa propre définition mentionnée dans la note n° 2 dans le chapitre 3 (note n° 8 dans la version html) : « les appellations que je désigne comme noms littéraires, ne sont pas officielles : mais elles sont d'un usage fréquent dans les livres et dans le style épistolaire. »

Exemples de V 1-2 :
FR1=ZH1=ZH2+ctxt ; FR1+ctxt=ZH1=ZH2

FR1=ZH1=ZH2+ctxt

FR1+ctxt=ZH1=ZH2

« Les appellations que je désigne comme noms littéraires, ne sont pas officielles : mais elles sont d'un usage fréquent dans les livres et dans le style épistolaire. » (Maurice Courant)

n° 99

```

<orgName>
<orgName xml:lang="zh">尙衣院 </orgName>
<orgName xml:lang="ko">상의원 </orgName>,
<orgName xml:lang="ko-mcc">Sangüiwön </orgName>
<orgName xml:lang="ko-mep">[Syang eui ouen] </orgName>,
<seg type="nomLitteraire"> nom littéraire
<term>
<orgName xml:lang="zh">尙方 </orgName>
<orgName xml:lang="ko">상방 </orgName>,
<orgName xml:lang="ko-mcc">Sangbang </orgName>
<orgName xml:lang="ko-mep">[Syang pang] </orgName>
</term> </seg>,
<orgName xml:lang="fr">Garde-robe Royale </orgName>
</orgName>
</orgName>
n° 155
<orgName>
<orgName xml:lang="zh">內侍府 </orgName>
<orgName xml:lang="ko">내시부 </orgName>,
<orgName xml:lang="ko-mcc">Naesibu </orgName>
<orgName xml:lang="ko-mep">[Näi si pou] </orgName>,
<seg type="nomLitteraire">
nom littéraire
<term>
<orgName xml:lang="zh">內班院 </orgName>
<orgName xml:lang="ko">내반원 </orgName>,
<orgName xml:lang="ko-mcc">Naebanwön </orgName>
<orgName xml:lang="ko-mep">[Näi pan ouen] </orgName>
</term> </seg>,
<orgName xml:lang="fr">Administration des Eunuques </orgName>
</orgName>

```

99
尙衣院 상의원, Sangüiwön [Syang eui ouen], nom littéraire 尙方 상방, Sangbang [Syang pang], Garde-robe Royale, nommée, au commencement de la dynastie des Yi ([R]), Fabriques Royales (n° 127). La Garde-robe présente les vêtements officiels et bijoux pour le Roi, le Prince héritier, les Reines, Princesses et autres personnes du Palais.

136
符寶郎 符寶郎, Puborang [Pou pa rang], Secrétaire du Sceau, sous Üjong [Eul tjong], 觀宗 公 公 (1146-1170), nommé Inpurang [In pou rang] 印府郎 印府郎, sous Ch'ungnyö [Tchyoung nyel], 忠烈王 忠烈王 (1274-1308).

155
Taejo [Htai tjo], 太祖 太祖, en 1392, conserva cette administration, mais en modifia les titres : 內侍府 內侍府, Naesibu [Näi si pou], nom littéraire 內班院 內班院, Naebanwön [Näi pan ouen], Administration des Eunuques ;

DocAsie 2016, carre-na.eun-joo@ehess.fr

7.4.1.3. Forme n° 3 de V1

V 1-3	FR1(+Ctxt ?)=ZH1=ZH2+Ctxt ?FR1+Ctxt=ZH1=ZH2	ZH2+Ctxt historique (évolution historique) FR1+Ctxt historique (précision)	« date », « roi », « créer », « supprimer », « séparer », « rassembler », « restituer »
-------	--	---	---

La troisième forme de la V1 (V1-3) est le résultat de différences de contexte historique entraînant une pluralité de l'élément en caractères chinois (ZH2). Il s'agit de l'évolution historique d'une même entité. On trouve dans la plupart des cas des traces de « date », ainsi que : « nommé ... par ou sous un tel roi »

Exemples de V 1-3 :
FR1(+Ctxt) =ZH1=ZH2+Ctxt
FR1+Ctxt=ZH1=ZH2

189
平珍音典 平珍音典, P'yŏngjin'ŭmjŏn [Hŏyeng tjin eum tyen], appelé 掃宮, Sogung [So koug], par Kyŏngdŏk [Kyeng tek], 景德王 景德王 (742-765), Bureau du Balayage, du Palais de Silla (Sin ra).

#189
<orgName>
<orgName xml:lang="zh">平珍音典</orgName>
<orgName xml:lang="ko">平珍音典</orgName>
<orgName xml:lang="ko-mcc">P'yŏngjin'ŭmjŏn</orgName>
<orgName xml:lang="ko-mep">[Hŏyeng tjin eum tyen]</orgName>
</orgName>
<orgName>
<orgName xml:lang="zh">掃宮</orgName>
<orgName xml:lang="ko">소궁</orgName>
<orgName xml:lang="ko-mcc">Sogung</orgName>
<orgName xml:lang="ko-mep">[So koug]</orgName></orgName>
</orgName>
<persName>
<persName xml:lang="ko-mcc">Kyŏngdŏk</persName>
<persName xml:lang="ko-mep">[Kyeng tek]</persName>
<persName xml:lang="zh">景德王</persName>
<persName xml:lang="ko">景德王</persName> (742-765)</persName>
</persName>
</orgName>
<orgName xml:lang="fr">Bureau du Balayage, du Palais de Silla [Sin ra]</orgName>
</orgName>
<orgName>
<orgName xml:lang="zh">均田廳</orgName>
<orgName xml:lang="ko">당전청</orgName>
<orgName xml:lang="ko-mcc">Yangjŏnch'ŏng</orgName>
<orgName xml:lang="ko-mep">[Ryang tyeŏn htyeng]</orgName>
<orgName xml:lang="fr">Bureau du Cadastre</orgName>, fondé <date>en 1718</date>;
</orgName>
<seg>
<date>en 1663</date>, un Bureau semblable avait été nommé
</seg>
<orgName>
<orgName xml:lang="ko-mcc">Kyŏnjŏnch'ŏng</orgName>
<orgName xml:lang="ko-mep">[Kyouŏn tyeŏn htyeng]</orgName>
<orgName xml:lang="zh">均田廳</orgName>
<orgName xml:lang="ko">均田廳</orgName>
</orgName>
</seg>

406
均田廳 均田廳, Yangjŏnch'ŏng [Ryang tyeŏn htyeng], Bureau du Cadastre, fondé en 1718 ; en 1663, un Bureau semblable avait été nommé Kyŏnjŏnch'ŏng [Kyouŏn tyeŏn htyeng], 均田廳 均田廳.

#312
<name>
<name xml:lang="fr">Les Décrets Royaux</name>
</name>
<seg>
<date>au commencement de la dynastie de Koryŏ [Korye]</date>, étaient nommés
</seg>
<name>
<name xml:lang="ko-mcc">Cho</name>
<name xml:lang="ko-mep">[Tjyŏ]</name>
<name xml:lang="zh">趙</name>
<name xml:lang="ko">조</name></name></seg>
</seg>
<persName>
<persName xml:lang="ko-mcc">Sŏngjong</persName>
<persName xml:lang="ko-mep">[Syeng tjong]</persName>
<persName xml:lang="zh">肅宗</persName>
<persName xml:lang="ko">성종 (981-997)</persName>
</persName>
les appela
<name>
<name xml:lang="ko-mcc">Kyo</name>
<name xml:lang="zh">高</name>
<name xml:lang="ko">교</name></name></seg>
</seg>
<persName>
<persName xml:lang="ko-mcc">Chŏngjong</persName>
<persName xml:lang="ko-mep">[Tjyeng tjong]</persName>
<persName xml:lang="zh">淸宗</persName>
<persName xml:lang="ko">정종 (1034-1046)</persName>
</persName>
les appela
<name>
<name xml:lang="ko-mcc">Che</name>
<name xml:lang="ko-mep">[Tjyŏ]</name>
<name xml:lang="zh">淸</name>
<name xml:lang="ko">계</name>
</name>
</seg>

312
Les Décrets Royaux, au commencement de la dynastie de Koryŏ [Korye], étaient nommés Cho [Tjyŏ], 詔 宗 ; Sŏngjong [Syeng tjong], 肅宗 성종 (981-997) les appela Kyo, 教 高 ; Chŏngjong [Tjyeng tjong], 淸宗 정종 (1034-1046) les appela Che [Tjyŏ] 淸 制.

DocAsie 2016, carre-na.eun-joo@ehess.fr

7.4.2. Conception d'encodage pour la V1

J'ai décidé de traiter la première variété (V1) comme une seule unité. Quant à ses différentes formes, je les distingue soit en utilisant des numéros dans les attributs, soit en insérant un tag « seg » avec un attribut de « type ».

J'arrête ici la démonstration des exemples d'analyse de discours et de conception d'encodage. Parce que d'abord, ce travail n'est pas encore terminé, mais toujours en cours d'évolution, et surtout ce n'est pas l'objectif de cette présentation.

8. Recherche d'un modèle des données : « database design »

8.1. Tâtonnement avec diverses tentatives

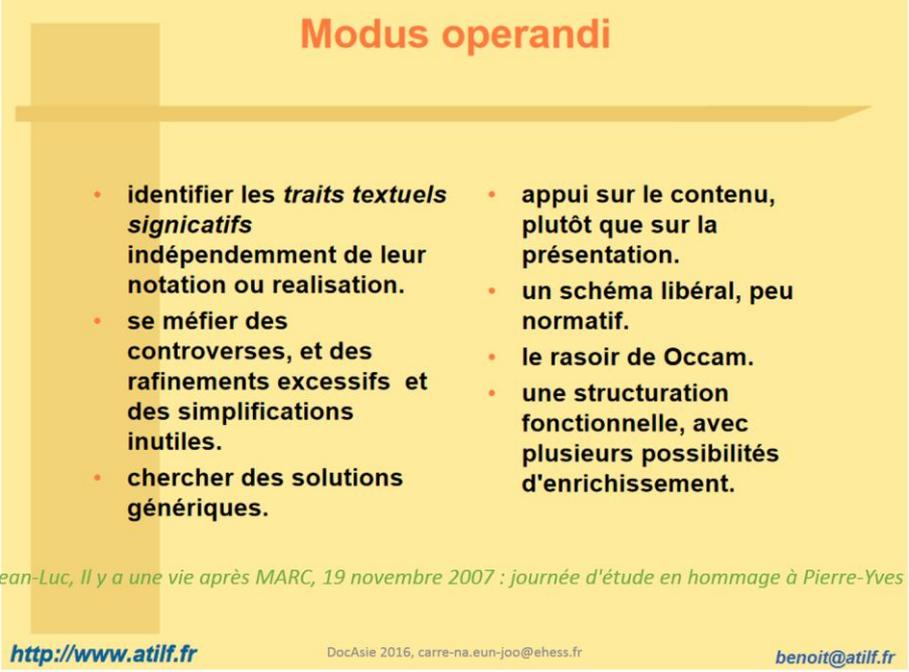
D'après le résultat de l'analyse du document, on cherche un modèle correspondant. Comme je n'ai pas assez de compétences techniques pour concevoir d'emblée un modèle adéquat pour ce document, j'ai d'abord essayé d'utiliser les modèles proposés par TEI.

Etant donnée que le « Répertoire » a une typologie textuelle très proche de celle d'un dictionnaire, pour ma première tentative, j'ai pris le modèle du « dictionnaire » de TEI. Ce modèle offre plusieurs tags pour décrire des variétés sémantiques et linguistiques,

mais la structure hiérarchique entre les tags est trop stricte et fermée pour l'appliquer au « Répertoire » de Maurice Courant qui a pris une grande liberté pour sa structure textuelle. Ma deuxième tentative fut de faire un mixage des tags du modèle du dictionnaire avec un modèle plus standard. C'est-à-dire qu'au lieu d'utiliser tout le modèle, j'emprunte quelques tags comme <pVar> pour exprimer la variété de prononciation. Ensuite, j'ai essayé d'appliquer les tags de <list><label> et <item> etc. La dernière tentative de la 1^{ère} phase de ce projet résida dans une symplification de codes, et consista à réduire au maximum le nombre de tags, en gardant seulement les plus universels comme <term><roleName><orgName> etc. Mais cela non plus ne constitua pas une version définitive. Pendant la préparation de cet exposé, j'ai encore modifié mes analyses et modèles, et par conséquent, mes tentatives et expériences continueront encore.

8.2. Quelques conseils

Ici, je présente des conseils de sagesse de Jean-Luc Benoit, qui peuvent être plus utile que mes pauvres tentatives.



Modus operandi

- identifier les *traits textuels signicatifs* indépendamment de leur notation ou réalisation.
- se méfier des controverses, et des raffinements excessifs et des simplifications inutiles.
- chercher des solutions génériques.
- appui sur le contenu, plutôt que sur la présentation.
- un schéma libéral, peu normatif.
- le rasoir de Occam.
- une structuration fonctionnelle, avec plusieurs possibilités d'enrichissement.

BENOIT, Jean-Luc, Il y a une vie après MARC, 19 novembre 2007 : journée d'étude en hommage à Pierre-Yves Duchemin.

<http://www.atilf.fr> DocAsie 2016, carre-na.eun-joo@ehess.fr benoit@atilf.fr

9. *Etat actuel du travail*

Bien que l'opération ne soit pas encore achevée, et que la recherche d'un meilleur modèle soit toujours en cours, j'ai quand-même encodé le document en intégralité. Et on a obtenu un certain résultat.

9.1. Etat de la database (données récupérées)

D'abord, on a pu récupérer des données. Les données récupérées par l'encodage se montent à environ 6 000 unités, et 30 000 éléments. Cela prouve bien que le « Répertoire » de Maurice Courant est une vraie boîte de concentration de données. Plus précisément, 2 400 titres de fonction administrative ou sociale, 1 500 titres d'organisation (organes gouvernementaux), 680 noms de personne, etc.

Etat de database (données récupérées)

6 000 unités = 30 000 éléments d'unité

Titre de fonction administrative ou social	2 400
Titre d'organisation	1 500
Nom de personne	680
Nom de lieu	340
date	920
Etc.	230

DocAsie 2016, carre-na.eun-joo@ehess.fr



9.2. Révision et correction de la version en word (HAL)

Cela a fait une occasion de réviser le travail de saisie. La saisie manuelle en format word, presque dans tous les cas, ne peut échapper à des erreurs : faute de frappe, manque et décalage de passages, répétition, etc. Pour moi, lors de ma 3^e tentative de simplification de codes, j'ai dû comparer la version word avec l'original, et j'ai pu trouver et corriger quelques erreurs ; mais surtout j'ai pu compléter les caractères chinois non-trouvés dans la version word. Pour l'instant, il y a eu en tout environ 120 corrections. Mais cette vérification n'a pas été effectuée sur l'intégralité du document, mais sur des extraits, donc, je pense qu'on aura quelques corrections de plus dans la 2^e phase du travail.

Visualisation par Html améliorée n°1

DocAsie 2016, carre-na.eun-joo@ehess.fr

9.4. Possibilité de diverses explorations et valorisation du document

Avec ce travail, le « Répertoire » offre désormais plusieurs possibilités d'exploration de ses données et d'extension vers d'autres bases externes qui existent déjà. La connection avec d'autres ressources extérieures comme l'Encyclopédie de l'histoire de Corée, donnera à ce document une valeur ajoutée. Et ce « Répertoire » français que les chercheurs coréens ne connaissent pas, pourra faire partie désormais des grandes références sur l'histoire de la Corée.

10. Conclusion

Eu guise de conclusion

Rôles des documentalistes dans la numérisation

Compétences	Opération de numérisation
Connaissance sur le document et les besoins	Sélection de documents à valoriser
Administratives	Enquêtes sur le droit d'auteur et les compatibilités
Techniques et de communication	Choix de type de numérisation et ses outils adaptés
Linguistiques, analytiques	Conception de base de données, « Database design »

DocAsie 2016, carre-na.eun-joo@ehess.fr

En guise de conclusion, je voudrais résumer les rôles des documentalistes dans la numérisation.

En tant que documentaliste, vous connaissez la valeur et les caractéristiques du document, ainsi que les besoins des utilisateurs potentiels, et vous êtes bien positionnés pour faire une sélection de documents à numériser.

Les documentalistes qui doivent toujours veiller à l'évolution des technologies concernant le domaine de gestion des informations et de documentation, ont la compétence pour bien choisir des outils adaptés aux objectifs d'un projet de numérisation.

Les documentalistes ayant les compétences administrative et linguistique nécessaires pour bien comprendre le document donné et l'analyser sont des acteurs principaux dans la conception de modèle de numérisation et/ou la réalisation de projet de numérisation.

Donc, les documentalistes, vous avez tout ce qu'il faut pour devenir « database designer ». Vous faites partie des acteurs principaux de l'ère du numérique.

Bibliographies

- Maurice Courant, *Répertoire historique de l'administration coréenne*, 1891
- Marc Orange, « Book reviews » in *Cahiers d'Extrême-Asie* 1991 Volume 6 Numéro 1 pp. 315-316
- Jean-Luc Benoit, *Il y a une vie après MARC*, 19 novembre 2007 : journée d'étude en hommage à Pierre-Yves Duchemin. (<http://www.enssib.fr/bibliotheque-numerique/documents/1335-tei-text-encoding-initiative.pdf>)
- Data Observer, *Notre métier, c'est aussi « designer » des données* (<http://data-observer.com/fr/design-des-donnees/>) consulté en mai 2016