Intelligibilité du numérique

N° 1 | 2020
Faire dialoguer les disciplines via l'indexation des connaissances : la recherche interdisciplinaire en débat

L’indexation automatique des archives audiovisuelles privées au prisme des Linked Open Data. Analyse sémiotique de l'architexte Famille

Lénaïk LEYOUDEC *Docteur*
TSH
COSTECH (EA 2223)
University of Technology of Compiègne

Édition électronique :
URL :
https://intelligibilite-numerique.numerev.com/index.php/articles/revue-1/2164-lindexation-automatique-des-archives-audiovisuelles-privees-au-prisme-des-linked-open-data-analyse-semiotique-de-l-architexte-famille
DOI : numerev_1694
Date de publication : 13/05/2020

Cette publication est sous licence **CC BY-NC-ND** (Attribution - No commercial - No derivatives).

Pour **citer cette publication** : LEYOUDEC, L. (2020) L’indexation automatique des archives audiovisuelles privées au prisme des Linked Open Data. Analyse sémiotique de l'architexte Famille. *Intelligibilité du numérique*, (1). https://doi.org/10.34745/numerev_1694

La démarche d’éditorialisation des documents est régie par des normes de modélisation de l’information (ontologies et bases de connaissances) permettant une interopérabilité entre les interfaces de programmation (API) du logiciel et des *linked open data*. Issue de bases de connaissances ouvertes, l’information contextuelle injectée dans le document cohabite avec l’information de nature mémorielle propre à l’archive. Or, en ajoutant des informations à un document patrimonial, on réinvente un document qui n’a jamais existé. Le clivage introduit ici entre des informations contextuelles mises en relation avec des informations de nature mémorielle au sein d’un même dispositif technique de gestion du patrimoine culturel incarne l’enjeu de cette étude. La conjonction de ces deux natures d’informations altère-t-elle la lisibilité culturelle du document ? Jusqu’à quel point ces agents automatisés – « robots » - sont-ils pertinents dans une démarche de valorisation du patrimoine culturel ?

The document editorial process is governed by information modeling standards (ontologies and knowledge bases) allowing interoperability between the programming interfaces (APIs) of the software and linked open data. Coming from open knowledge bases, the contextual information injected into the document coexists with information of a memorial nature specific to the archive. By adding information to a heritage document, you are reinventing a document that never existed. The cleavage introduced here between contextual information put in relation with information of a memory nature within the same technical device for the management of cultural heritage embodies the challenge of this study. Does the combination of these two types of information affect the cultural readability of the document? To what extent are these automated agents - "robots" - relevant in an approach to enhancing cultural heritage?

Mots-clés :
Linked Open Data, Sémiotique, Logiciel, Annotation, Patrimoine culturel, Indexation, Connaissance, Écrit d'écran, Linked Open Data, Semiotics, Software, Screen writing, Annotation, Cultural heritage, Knowledge, Indexing

---

1. Contexte

1.1. Statut des productions familiales

Le numérique est depuis deux décennies un support majeur pour l'ensemble des contenus. Les films de famille ou productions familiales ne font pas exception : hier analogiques, elles sont essentiellement produites en numériques aujourd'hui, les documents s'accumulant sur les disques durs, disques optiques et serveurs des particuliers. Or l'accès à un contenu numérique n'est possible qu'après un double procédé d'instrumentation et de manipulation. En effet, aux médias ne nécessitant pas de la part du lecteur d'instrumentation supplémentaire – tel que le livre –, s'opposent les médias technologiques soumis à une double manipulation : « à l'incontournable médiation culturelle, il faut donc ajouter une médiation technologique pour l'accès au contenu » (Mattelart, 1994). Le numérique, par définition, impose un décodage du contenu archivé afin d'en permettre son visionnage ou son utilisation. La production familiale numérisée ou nativement numérique en tant que telle est une ressource codée et l'outil informatique propose à l'utilisateur une reconstruction du contenu. Le média technologique est soumis avec le temps à deux problèmes qui impactent son utilisation : le fossé d'obsolescence correspondant à la préservation des formats associés et plus largement aux moyens technologiques d'accès au contenu archivé ; et le fossé d'intelligibilité (Bachimont, 2010, 22). Ce dernier incarne l'effet du temps sur la lisibilité culturelle du contenu. Décontextualisée, la production familiale perd son rôle de réceptacle de la mémoire collective car son intelligibilité doit, de fait, être restaurée.

1.2. Une plateforme qui valorise les productions familiales

Notre intérêt pour les productions familiales décontextualisées a trouvé un écho auprès d'un éditeur logiciel, Perfect Memory. Celui-ci édite Famille™, un système de gestion d'assets numériques (*Media Asset Management System*) dédié aux particuliers. En 2011, une version bêta de Famille® est publiée sous la forme d'une application Web (telefamille.perfect-memory.com), puis ses déclinaisons en applications Mobile et Télévision connectée en 2012. En 2014, l'application Web migre vers un nouveau nom de domaine (famille.pm).

Une convention industrielle de formation par la recherche (CIFRE) reliant l'auteur-doctorant, une unité de recherche (l'équipe d'accueil COSTECH de l'Université de technologie de Compiègne) et le partenaire industriel Perfect Memory est signée pour une durée de trente-six mois à partir de septembre 2013. Cette recherche doctorale en sciences de l'information et de la communication prend la forme d'une observation participante au sein de l'équipe de recherche et développement de l'éditeur logiciel.

1.3. Une plateforme connectée aux bases de données liées et ouvertes sur le Web

L'une des propriétés techniques majeures de la plateforme Famille™ est son interfaçage avec les bases de données liées et ouvertes sur le Web, les *Linked Open Data*.

Les *Linked Data* désignent des données reliées entre elles. Cette liaison est permise par la conjonction d'une publication sur le web et l'association à chaque donnée d'un identifiant unique : l'*Uniform Ressource Identifier* (URI). Par le biais de combinaisons « sujet-prédicat-objet » baptisées « triplets », ces données peuvent être mises en relation avec d'autres données tandis que leur relation peut être qualifiée telle qu'une équivalence. Le respect d'un standard commun - le RDF - au niveau du formatage des triplets valide l'intégration des données liées dans le réseau global baptisé « web des données » (Gandon, 2011).

L'*Open Data* (données ouvertes) désigne le mouvement d'ouverture des données des collectivités, des services publics et des entreprises qui émerge depuis la fin des années 2000. L'objectif est d'arriver à une meilleure transparence de ces acteurs en partant du principe que ces données représentent un bien commun selon l'acception de Elinor Olstrom (Olstrom, 2010) et devraient donc être accessibles par tous.

Afin d'être appréhendées comme ouvertes, ces données doivent respecter plusieurs critères : technique (fourniture des données dans un format non propriétaire), économique (liberté d'utilisation des données) et juridique - données sous licence ouverte - (Chignard, 2012). De fait, l'appellation *Linked Open Data* illustre la conjonction, à l'échelle des données, d'une structure (données liées) et d'un statut (données ouvertes). Ce contexte spécifique crée une interopérabilité de plusieurs sortes : technique (portée par le format RDF), juridique (absence de copyright) mais également économique (liberté d'accès) qui ouvre un large éventail de possibilités d'utilisation.

Historiquement, l'initiative *Linked Open Data* revient à Tim Berners-Lee, l'inventeur du *World Wide Web*. Celui-ci publie en 2006 quatre principes centrés autour de la notion d'URI, encadrant la publication des données en ligne et l'interconnexion des jeux de données (Berners-Lee, 2006). Mettant en application ces principes de référencement et de nomination des ressources, le projet *Linking Open Data* est créé en 2007. Il est initialement porté par des chercheurs universitaires en informatique ainsi que des développeurs évoluant dans diverses entreprises privées, le tout accompagné par le W3C, avant d'être rejoint par des institutions telles que la BBC, Thompson Reuters ou la Library of Congress (Heath et al., 2009). En investissant cet espace de publication de contenus, ces acteurs favorisent l'émergence d'un écosystème encyclopédique, utilisé par les communautés scientifiques et industrielles. En 2019, 1.239 bases de données sont en ligne, accessibles par des interfaces de programmation (API), qui permettent un export manuel (SPARQL End Point par exemple) ou automatique (à l'aide de requêtes SPARQL par exemple) au moyen d'agents automatisés.

Pour résumer, notre étude porte sur une plateforme logicielle qui manipule des productions familiales numérisées et pré-indexées par des institutions patrimoniales. La spécificité de cette plateforme est de mobiliser les bases de données liées et ouvertes sur le Web afin d'indexer automatiquement les documents à l'aide de notices encyclopédiques.

2. Problématique

La restauration de l'intelligibilité des productions familiales est l'enjeu qui nous anime. L'hypothèse que nous formulons est qu'en ajoutant des informations supplémentaires à la production familiale décontextualisée, on peut restaurer sa lisibilité culturelle pour tout un chacun. Selon le vocabulaire de l'ingénierie des connaissances, il s'agit d'associer une annotation à un objet éditorial - la production familiale en tant qu'objet logique ayant une représentation physique sous la forme d'une ressource numérique. L'annotation ne prend pas la forme d'une chaîne de caractères mais d'une entité sémantique avec des propriétés. Par exemple, « Paris » est une entité sémantique de type « Lieu » qui comporte plusieurs propriétés dont « Partie de » l'entité sémantique « Île-de-France ». Cette action, répétée de nombreuses fois par un utilisateur qui consulte le film ou par un robot qui analyse le film, est qualifiée d'éditorialisation : « processus consistant à enrôler des ressources pour les intégrer dans une nouvelle publication » (Bachimont, 2007b, 21). Si l'éditorialisation de la production familiale cherche, à terme, à reconstruire l'intelligibilité du document, elle donne lieu au premier abord à une recontextualisation, « reprise d'un objet culturel dans un nouveau contexte médiatique » (Treleani, 2013). Cette démarche peut s'appliquer depuis différents environnements techniques à l'instar de la couche système d'exploitation (édition des métadonnées de la ressource numérique directement depuis un explorateur de fichiers) ou depuis la couche logicielle. Dans notre étude, les processus appliqués aux productions familiales ont lieu par le biais d'un dispositif technique : le logiciel Famille™.

Le logiciel incorpore en son sein une fonctionnalité de consultation simultanée du film et des métadonnées associées à ce dernier. Les métadonnées historiques du film (provenant des documentalistes des institutions patrimoniales) sont élevées sémantiquement (passage de la chaîne de caractères à l'entité sémantique) puis un agent automatisé interroge les *Linked Open Data* et rapatrie dans l'interface utilisateur la notice correspondante. De même, d'autres agents automatiques viennent « gratter l'image » et le son et produisent d'autres métadonnées, elles-mêmes élevées sémantiquement. Enfin, l'utilisateur peut agir de son propre chef en produisant des annotations manuelles : identifier une personne ou un lieu à l'image par exemple. La saisie de la chaîne de caractères dans l'interface donne lieu au même processus de rapatriement de notices.

Après avoir clarifié les principes sous-jacents de notre démarche et décrit le périmètre de l'éditorialisation, formulons une question : est-il possible de reconstruire

l’intelligibilité d’un corpus de productions familiales en mobilisant les bases de données liées et ouvertes sur le Web ?

Issue de bases de connaissances ouvertes, l’information « encyclopédique » rapprochée du document cohabite avec un document patrimonial (au sens étymologique de document hérité de génération en génération). Le film enrichi par les annotations devient une nouvelle publication selon la définition de l’éditorialisation, or plusieurs périodes historiques coexistent dans cette publication hybride, inventée. Décrire la production familiale favorise-t-il la reconstruction de son intelligibilité ou l’empêche-t-il totalement ? Enfin, jusqu’à quel point ces agents automatisés – « robots » - sont-ils pertinents dans une démarche d’éditorialisation de productions familiales ?

3. Terrain

3.1. Identité du corpus multi-centré constitué

Notre corpus comporte quatre fonds associés à autant d’institutions. Il est construit à partir de plusieurs fonds préservés par différentes institutions patrimoniales parmi lesquels la Cinémathèque de Bretagne, l'Établissement de communication et de production audiovisuelle de la Défense (ECPAD), le Forum des images ainsi que l'Institut national de l'audiovisuel (INA). Au total, vingt-et-un films ont été choisis puis interrogés par le prisme de la sémantique différentielle, méthodologie décrite ci-dessous. Chaque film possède des métadonnées, produites majoritairement par les documentalistes des institutions. Aucun témoignage provenant des vidéastes-cinéastes amateurs demeure. À l’issue de l’analyse sémiotique, ces films sont versés sur la plateforme Famille™, enrichis et décrits à la fois par les agents automatisés et par l’utilisateur.

La Cinémathèque de Bretagne est une association Loi 1901 créée en 1986. Plus de 1.700 déposants ont jusqu’à aujourd’hui participé à l’enrichissement des fonds de la Cinémathèque de Bretagne, qui comptent plus de 27.510 films, vidéos et bandes son, 1.500 appareils de cinéma et près de 10.000 photogrammes et photographies numériques pour ne citer que la partie audiovisuelle du fonds. Outre le siège de l’association à Brest (Finistère), une antenne mobile « Est Bretagne » est placée à Rennes tandis que l’antenne Loire-Atlantique est située au sein des Archives départementales de Loire-Atlantique à Nantes. Cette dernière antenne accueille en particulier des fonds associés à ce département breton historique, parmi lesquels nous avons constitué notre fonds d’étude. Les cinq films étudiés préservés par la Cinémathèque de Bretagne partagent une même propriété : ils ont été réalisés par le nantais Eugène Dupont entre 1946 et 1947. Un corpus d’étude constitué de cinq films a pu être ainsi constitué.

L'établissement de communication et de production audiovisuelle de la Défense est depuis 2001 un établissement public à caractère administratif dont les missions sont la préservation et la commercialisation d'une collection de dix millions de clichés et près de 30.000 films liés à l'histoire militaire de la France. Elles accueillent également des archives versées par des particuliers et des associations patrimoniales, en lien avec la Défense et désireux de préserver la mémoire combattante. Les collections du « Fonds privés » accueille 94.000 photographies et 274 films, dont la quasi-totalité étant numérisés et disponibles à la consultation sur place au Fort d'Ivry. Parmi les films du fonds « FA », nous avons choisi cinq films dont deux d'entre eux réalisés par le même officier : le lieutenant-colonel Dominique Paulus.

Le Forum des images est une institution culturelle de la Ville de Paris dédiée au cinéma et à l'audiovisuel. Outre des cycles de projections, l'institution préserve près de 8.000 films ayant Paris pour sujet ou pour décor. Ces collections hétéroclites sont composées de courts-métrages, longs métrages, documentaires, entretiens filmés, actualités et cours de cinéma. L'ensemble de films amateurs réunis depuis la fin des années 1980 à l'initiative du pionnier de la télévision et directeur de l'institution Pierre Tchernia a éveillé notre curiosité. Tournés entre les années 1930 et les années 1980, les 336 films amateurs appartenant respectivement à 163 fonds ont été interrogés avant de faire le choix d'un fonds spécifique : le fonds Guillien. Fruit du dépôt par son propriétaire – Jean Guillien – en 1989 et 1990, le fonds accueille trois bobines 9,5 mm noir et blanc, soit au total dix films muets noir et blanc. À l'issue de notre étude, seuls deux des sept films du fonds (trois films de famille ont été rassemblés par les documentalistes pour n'en former qu'un) ont pu faire l'objet d'une analyse sémiotique.

En prolongement de ses missions liées au dépôt légal de l'audiovisuel, l'Institut national de l'audiovisuel a initié en juillet 2012 une collecte de films amateurs à plusieurs échelles : régionale d'abord avec l'Aquitaine en 2012 et la Lorraine en 2013, puis thématique avec le « Tour de France » la même année puis outre-Atlantique avec une collecte dédiée à la Martinique. Notre approche a consisté à interroger le premier pan de la collecte initiée par l'INA, dédiée à la région Aquitaine, lancée le 2 juillet 2012. Sur la plate-forme, qui constitue notre unique point d'entrée dans ce fonds, 2.373 vidéos sont accessibles dans le fonds Mémoires partagées, dont 218 liés à la collecte Aquitaine. Au total, neuf films de ce fonds ont été choisis et analysés. Compte tenu du caractère anonyme des films, les métadonnées fournies ne permettent pas de rassembler les films par réalisateur, chaque film est considéré comme indépendant et constitue un corpus d'étude à lui tout seul.

3.2. Présentation du *Linked Open Data* utilisé dans ce projet : Wikidata

Wikidata (www.wikidata.org) est une base de connaissances libre collaborative, elle est hébergée par la Wikimedia Foundation, qui édite Wikipédia. Porté par la section

allemande de Wikimedia, elle ouvre aux contributions le 31 octobre 2012. Le projet est financé par l'Allen Institute for Artificial Intelligence, la Gordon & Betty Moore Foundation et Google. En 2019, Wikidata comporte plus de cinquante-neuf millions de notices liées entre elles et pouvant être interprétées par des machines. La base de connaissances rend disponible ses données par des points d'accès (SPARQL End Point) à la fois aux humains et aux machines. Suite au développement d'un connecteur, une plateforme tierce peut ainsi interroger la base de connaissances de Wikidata et exporter n'importe quelle notice. Pour rappel, les données hébergées sur Wikidata sont en Licence Creative Commons Zero (CC0).

Dans notre expérimentation, l'action d'annoter temporellement une production familiale entraîne l'interrogation de la base de connaissances et le rapatriement dans l'interface de la notice liée au tag créé. En d'autres termes, quand l'utilisateur ajoute un tag temporel « Nantes » à un document du corpus depuis l'interface de Famille™, la machine interroge la base Wikidata et rapatrie dans l'interface l'éventuelle notice correspondante, ici la notice de l'entité « Lieu » « Nantes ».

Pour résumer, vingt-et-un films issus de quatre institutions patrimoniales sont collectés, ainsi que leurs métadonnées, puis versés dans la plateforme Famille™. Des processus d'annotation automatiques sont réalisées sur les documents, tandis que l'utilisateur peut ensuite par lui-même les annoter. Chaque action d'écriture (tag) réalisée sur la plateforme entraîne l'interrogation du *linked open data* et le rapatriement dans l'interface de notices liées au tag créé. En d'autres termes, des productions familiales sont enrichies manuellement et automatiquement par le biais d'une plateforme, interrogée dans cette étude.

4. Méthodologie

4.1. Le choix de la sémantique différentielle

Notre choix méthodologique a été influencé par ce postulat : « *la problématique interprétative dépasse les textes et peut s'étendre à d'autres objets culturels, comme les images (susceptibles des mêmes méthodologies : recueil de corpus, détermination des genres, indexation par des traits de l'expression)* » (Rastier, 1987, VI).

Dans un premier temps, la question de l'intelligibilité de la production familiale est mise en relation avec la notion de « compréhension ». Cette dernière est entendue comme « une interprétation : elle consiste à stipuler […] (i) quels traits sémantiques sont actualisés dans un texte, (ii) quelles sont les relations qui les structurent, et (iii) quels indices et/ou prescriptions permettent d'actualiser ces traits et d'établir ces relations, qui sont autant de chemins élémentaires pour des parcours interprétatifs » (Rastier et

als., 2001, 12). Ainsi, nous questionnons l'intelligibilité du document par une approche interprétative ou, plus particulièrement, un parcours interprétatif entendu comme « la suite d'opérations cognitives permettant d'assigner une signification à une séquence linguistique » (Rastier, 1987, 277). Ce parcours interprétatif lors du visionnage de l'archive a pour objectif d'extraire des traits sémantiques, qui sont des « moments stabilisés dans des parcours interprétatifs » (Rastier, 2009, VI). Le trait sémantique est une composante de la sémantique. Il intervient dans l'étude du champ lexical, déclinée en l'identification de proximités lexicales (synonymies, hyperonymie, métonymie, etc.) et la décomposition des mots en traits sémantiques (Ploux, 2011, 2). En guise d'exemple, « /transport/ est un trait sémantique partagé par les mots métro, train, autobus, etc.. /ferré/ est un trait qui différencie autobus de train » (Rastier, 1987).

Les traits sémantiques peuvent être regroupés en classes sémantiques, entendues comme un « ensemble d'éléments regroupés en vertu d'une ou de plusieurs caractéristiques communes. Les doubles barres obliques symbolisent une classe (//classe//) » (Hébert, 2009, 254). Les classes identifiées dans notre parcours interprétatif seront ensuite transcrites en autant de catégories d'annotation destinées au logiciel Famille™. Le trait sémantique est construit *ex nihilo* lors de la démarche interprétative, il émerge à partir du signifié d'un signe et est indiqué entre barres obliques. Lorsque deux traits sémantiques reviennent dans un même parcours interprétatif, une isotopie est identifiée. Christian Metz nous explique que « le film est un lieu où viennent se combiner différents éléments de signification co-présents » (Metz, 1971, 121).

Notre hypothèse réside dans le fait que le signifiant – le détail d'un vêtement mis à l'image par exemple – trouve son sens dans l'ensemble virtuel des éléments pouvant occuper sa place (axe paradigmatique), articulé avec son environnement qui participe de l'actualisation ou de la virtualisation des oppositions et différences identifiées (axe syntagmatique). Ce qui est propre à l'image, c'est le répertoire de formes qui, contrairement à la langue et au texte, ne mobilise pas un répertoire fini (phonèmes, alphabet) pour constituer des unités stabilités (les lexiques). De fait, les variations à l'échelle de l'image animée sont infinies. Une notion supplémentaire est nécessaire : la saillance perceptive. Autrement dit, nous nous appuyons sur ce qui se dégage perceptiblement lors de la consultation du document. Dès lors, notre parcours interprétatif s'articule autour de l'identification des saillances perceptives. Ces dernières sont ensuite sémantisées en fonction du jeu différentiel avec les autres saillances. Enfin, les saillances sont mobilisées pour interroger le signifiant dans sa construction technique et filmique, sous la forme d'hypothèses interprétatives.

François Rastier et les autres chercheurs liés à la sémiotique des cultures interrogent l'objet culturel selon un double processus : la constitution et l'interprétation. La constitution de l'objet culturel peut être décomposée en deux étapes : l'assignation, qui permet le passage du matériau aux formes et fonds de l'expression, puis la sémantisation qui marque l'établissement et contextualisation des signifiés (Rastier, 2011, 2, 9). Ces deux composantes forment la sémiotisation, ou *semiosis*. Une fois

constitué, l'objet culturel peut être interprété par un processus de textualisation qui « établit la détermination du global sur le local » (Rastier, 2011, 2). Le sens associé à l'objet peut ainsi s'établir dans le cadre de parcours interprétatifs par « recontextualisation maximale, dans des champs relationnels et des séries paradigmatiques que nous résumons sous le nom de corpus » (Rastier, 2011, 9). L'objet culturel rassemblé en corpus - ici un ensemble de films de famille - est ainsi interrogé par le biais d'un parcours interprétatif composé de plusieurs étapes.

Dans le cadre de notre parcours interprétatif il convient de fixer l'échelle minimale à laquelle nous allons interroger le film. Dans son entreprise interprétative, François Rastier met en avant la notion de « passage » comme lieu de la *sémiosis* locale. Il explique que « cette grandeur locale correspond indifféremment à un signe, à une phrase, ou par exemple à un paragraphe. Au plan du signifiant, le passage est un extrait, entre deux blancs s'il s'agit d'une chaîne minimale de caractères ; entre deux pauses ou ponctuations, s'il s'agit par exemple d'une période. Au plan du signifié, le passage est un fragment qui pointe vers ses contextes gauche et droit, proches et lointains » (Rastier, 2008, 6). Au niveau de l'objet film, le passage serait au plan du signifiant un extrait, l'espace entre deux interruptions de caméra : c'est-à-dire le plan filmique. Au niveau du signifié, le passage est un fragment qui regroupe différents marqueurs de sens, pointant vers les autres marqueurs de sens présents dans les extraits (plans) qui suivent et précèdent l'extrait étudié. Le plan filmique constitue ici un repère au niveau du signifiant, sans oublier que selon Rastier les significations ne sont pas uniformément isolables. De fait, on peut identifier des passages plus petits que le plan (notamment dans le contexte du plan séquence, figure du cinéma de fiction) ou plus grands (notamment dans les montages saccadés observés dans le cinéma de fiction d'action). Du point de vue du signifié, le passage comporte plusieurs effets de sens que l'on peut extraire. Notre parcours interprétatif s'applique ainsi à l'échelle du passage, porteur d'un ensemble de saillances perceptibles que l'on va agréger puis déconstruire.

Notre démarche interprétative appliquée au signifiant non-linguistique qu'est le film consiste en la rédaction d'une analyse plan par plan de chacun des films associé à notre corpus. Cette analyse comprend un paragraphe par passage, elle agrège les éléments techniques et thématiques : *time codes*, type de plan, formes perçues, détails techniques (image floue, éclairage, etc.). Ces éléments forment autant de saillances perceptives à partir desquelles nous élaborons des hypothèses interprétatives.

Le parcours interprétatif d'un élément du corpus est articulé autour de trois étapes : agrégation des saillances, sémiotisation des signes et actualisation des traits sémantiques.

La première étape consiste en l'agrégation des saillances perceptives qui émergent du visionnage. Ces saillances seront appréhendées comme autant de signes, mais de prime abord elles s'imposent au niveau de la perception. La sémiotisation devra alors leur conférer une valeur sémantique et les confirmer dans leur signifiance. Cette

première étape correspond donc à ce qui est perçu de prime abord.

Ensuite, les saillances sont mobilisées dans une opération de sémiotisation. La saillance est interrogée dans la double articulation des signifiants et signifiés, en la caractérisant comme un choix ou une valeur retenue parmi des choix possibles, que ce soit sur le plan du signifiant ou du signifié. Dans le cadre de notre parcours interprétatif, la saillance est ainsi valorisée sémiotiquement. Cette seconde étape correspond à ce qui résulte du parcours. Après le perçu, nous étudions le rendu.
Enfin, nous appliquons les codes de la sémantique interprétative à l'ensemble des signes : les traits sémantiques sont actualisés à partir des signifiés. Jusqu'ici nous demeurions à l'échelle micro-sémantique, à l'échelle de la saillance perçue au sein d'un passage. L'opération d'actualisation des traits se poursuit à l'échelle méso-sémantique (ensemble de trois passages) puis à l'échelle macro-sémantique (l'ensemble des passages du document et le paratexte). Cette opération permet d'identifier la récurrence de certains traits (formant des isotopies) et la présence de classes sémantiques regroupant plusieurs traits.

De la description brute à la classe sémantique, les trois étapes de notre parcours interprétatif ont pour enjeu de déconstruire la description brute afin de valoriser sémiotiquement la saillance, ce qui nous permet ensuite d'actualiser le trait sémantique. Ces étapes nous permettent de passer de la saillance au trait puis à la classe sémantique. Cette dernière incarne la matière première des catégories d'annotation du document constituant un axe des recommandations éditoriales et ergonomiques proposées pour l'architexte Famille™. En effet, la notion de classe sémantique repose ici sur l'hypothèse que les classes constituent la plus grande stabilité dans le système de l'expression, si bien qu'on les retrouvera d'un document à un autre et qu'elles préfigurent donc des catégories d'annotation et pas seulement d'interprétation. Autrement dit, l'actualisation répétée et constatée d'isotopies semblables permet de dégager des catégories stables qui seront autant de cadres interprétatifs à proposer à un utilisateur s'emparant d'un contenu pour l'annoter.

4.2. Ingénierie des connaissances

L'interprétation des traits sémantiques et leur regroupement en classes a pour fonction de faire apparaître des axes descriptifs majeurs des productions familiales considérées dans notre corpus. Ces axes forment autant de propriétés supplémentaires à prendre en compte dans notre approche d'éditorialisation du document numérique. À l'échelle de l'interface utilisateur du logiciel, il s'agit de types de tags supplémentaires, tandis qu'à l'échelle du *back office*, il convient d'ajouter ces propriétés au modèle de données existant.

L'une des spécificités du logiciel Famille™ est son architecture technique : une plateforme micro-services sémantique. Le terme « sémantique » ici signifie que la représentation formelle de l'information est sémantique, de fait pouvant être comprise

aussi bien par des machines que par des humains. Au lieu d'interroger une base de données relationnelles lors de chaque clic de l'utilisateur, le logiciel questionne une base de connaissances, régie par une ontologie informatique. Ce dernier terme est compris comme la « description formelle de propriétés, de catégories ou de types de relations existant entre des objets divers » (Douehi, 2011, 170). La discipline qui s'intéresse au mode d'existence des ontologies informatiques est l'ingénierie de la connaissance, champ de l'informatique qui « comprend d'une part l'ingénierie des représentations qui modélisent le sens des documents, et d'autre part, l'ingénierie des contenus qui modélise leur forme d'expression » (Bachimont, 2007a, 17). Dans notre étude, les productions familiales sont des contenus tandis que les annotations sémantiques associées aux productions sont des représentations. En d'autres termes, l'ontologie est une grammaire unifiée permettant de décrire à la fois les films et les tags : l'ontologie informatique. En apportant d'autres axes d'annotation, nous devons ajouter d'autres propriétés dans l'ontologie. Une fois ces modifications faites, une configuration de l'interface pourra permettre à l'utilisateur de taguer les films avec ces nouveaux axes. Une découverte du langage informatique sous-jacent à l'ontologie constitue une étape de notre recherche, tandis que les sciences informatiques composent ici avec la discipline sémiotique.

5. Résultats

La présentation de nos résultats est segmentée en trois temps : la présentation synthétique d'un échantillon de corpus, la discussion d'une partie des résultats de l'étude sémiotique et la description des étapes de l'éditorialisation d'un film du corpus.

5.1. Étude d'un segment de *Vues de Biarritz et de Bayonne* (inconnu, INA)

Afin de familiariser le lecteur avec l'approche interprétative appliquée aux vingt-et-un films du corpus, un film (*Vues de Biarritz et de Bayonne*) a été choisi comme échantillon représentatif. Un parcours interprétatif composé de trois plans consécutifs est discuté.

En niveau du signifiant, les trois passages sont des extraits et interviennent précisément aux times-codes suivants : 1'55''-2'00'', 2'01''-2'06'' et 2'07''-2'22''. Au niveau du signifié, les passages sont des fragments pointant vers des contextes latéraux : le premier passage (plan n°13) consiste en un plan d'ensemble mettant à l'image une rue passante entourée de bâtiments, tandis que le troisième passage (plan n°15) est un plan d'ensemble mettant à l'image le bord de mer et différents baigneurs.



Figure 1 : Vues de Biarritz et de Bayonne, extrait du plan n°13 (crédits INA)



Figure 2 : Vues de Biarritz et de Bayonne, extrait du plan n°14 (crédits INA)



Figure 3 : Vues de Biarritz et de Bayonne, extrait du plan n°15 (crédits INA)

Localement plusieurs saillances perceptives sont identifiables au sein du premier passage (plan n°13) : la présence d'une architecture spécifique, de fait pouvant être localisée. Le second passage (plan n°14) met à l'image des personnes vêtues avec distinction : l'étude du vêtement permet de positionner la localisation générale de la captation (Europe de l'Ouest) ainsi qu'une datation approximative du document par le biais des usages et modes vestimentaires. Enfin, le troisième passage (plan n°15) met à l'image un paysage à la fois naturel (plage) et humain (foule de baigneurs) suggérant un emplacement relatif de la captation (bord de mer, côte Atlantique) et une datation (postérieure aux années 1940 par l'étude du matériel, des installations balnéaires et du vêtement de plage).

De manière secondaire, l'interprétation croisée des trois passages permet de mettre en lumière des récurrences dans les saillances perceptives : l'architecture urbaine, le champ matériel peuplant l'image (voitures, objets du quotidien, matériel de plage) et enfin, le vêtement des personnes à l'image inscrivent la captation dans une temporalité spécifique (les années 1940-1950).

Dans un troisième temps, une approche macro-sémantique permet d'envisager le document par le prisme de l'intertexte : la conjonction des passages et des métadonnées sources produites par les documentalistes de l'INA.

« [Image amateur]. Balade dans Biarritz : le front de mer, la ville, la plage, le centre ville où des personnes distinguées se promènent. A Bayonne, une procession religieuse marche dans les rues de la ville. Vue de l'église Saint Michel au loin filmée depuis l'actuel Parc des Sports Saint Michel. Ambiance festive dans un quartier où des couples

dansent dans la rue au rythme d'un air joué par un groupe de musiciens (accordéon, batterie et guitare) » (INA, 2015).

Nous pouvons ainsi vérifier plusieurs hypothèses interprétatives. Tandis que le statut amateur des images est renforcé, la localisation précise des passages étudiés est précisée - Bayonne, sa plage et son centre-ville - de même que la datation de la captation est indiquée : 1950.

5.2. Bilan des traits et classes sémantiques

Les opérations suivantes sont la sémiotisation (identification des traits sémiotiques) puis la montée en généralité : l'identification des classes sémantiques s'appliquant au film étudié. À l'échelle du plan puis lors de la conjonction des trois plans, plusieurs traits ont été identifiés parmi lesquels : /voiture française d'après-guerre/, /complet-veston/, /maison balnéaire/, /ensemble tailleur/, /personne n°1 homme/, /personne n°2 femme/ ou /Bayonne/. Ces traits peuvent être rassemblés en trois classes sémantiques : //objet//, //lieu// et //personne//. Les étapes de la sémiotisation et la classification des traits en classes sont décrits en détail dans le mémoire de thèse (XXX, 2017), nous ne présentons ici qu'un aperçu des résultats afin d'introduire le socle méthodologique sémiotique de l'éditorialisation des productions familiales.

Pour résumer, plusieurs classes sémantiques ont été identifiées comme axes d'annotation pertinents pour décrire les productions familiales du corpus. Il convient désormais de configurer les paramètres d'annotation de la plateforme Famille™ afin d'intégrer ces axes, puis de verser l'échantillon du corpus dans la plateforme.

5.3. Éditorialisation du film dans Famille™

La dernière partie de notre étude du film *Vues de Biarritz et de Bayonne* consiste à verser la ressource numérique dans la plateforme, puis à l'éditorialiser. Cette étape expérimentale se fonde sur les axes d'annotation identifiés lors de l'étude sémiotique et cherche à vérifier l'hypothèse d'une possible recontextualisation de la production familiale *via* son éditorialisation manuelle et automatique.

Une fois la ressource versée dans la plateforme, la production est analysée par les modules connectés au logiciel tandis qu'un utilisateur crée ensuite plusieurs annotations afin de tester le rapatriement automatique des notices depuis Wikidata. Le film ne comportant pas de bande sonore, seules les métadonnées sources issues de l'INA ont été élevées sémantiquement par la machine. Le premier résultat que nous observons est, comme le montre la figure suivante, l'association de six tags au document. Outre les deux lieux qui sont mentionnés, le logiciel infère notamment la catégorie « Divertissement » à partir des mentions de danse et de musique, liées aux

derniers plans du film. Ni l'église Saint-Michel de Biarritz, ni le Parc des Sports Saint-Michel ne sont identifiés, car aucune notice correspondante n'est présente dans Wikidata. Les particularités géographiques régionales ne faisant pas l'objet d'une attention particulière dans cette plateforme collaborative, aucune notice ne peut être rapatriée automatiquement dans le logiciel. Une première limite est ainsi exprimée. L'utilisateur constatant une insuffisance de la base de connaissances devrait pouvoir créer lui-même une notice dans Famille™ afin de cristalliser la connaissance et permettre à d'autres utilisateurs d'y accéder. Une fonctionnalité supplémentaire est ainsi ajoutée au logiciel, permettant de résoudre cette limite.



Figure 4 : Tags créés automatiquement lors de l'import du film sur Famille™ (Crédits Perfect Memory, 2019)

Interrogeons désormais les tags produits automatiquement. Après vérification des notices sources sur wikidata.org, ces dernières comportent le site un nombre important de propriétés, à l'instar du label « Bayonne » dans chacune des langues de Wikidata. Une configuration a lieu afin que seules certaines propriétés de la fiche soient rapatriées dans le logiciel.

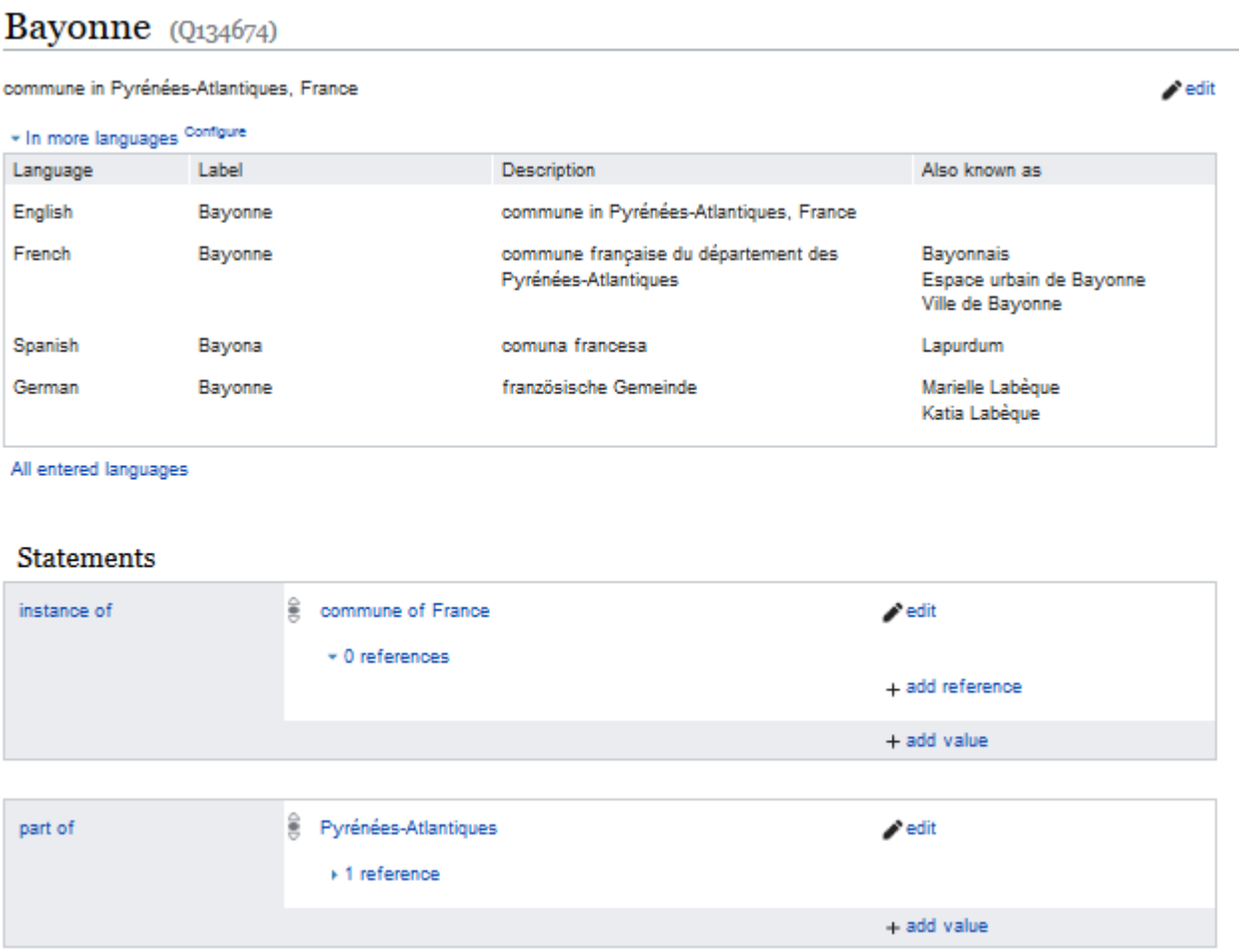


Figure 5 : Échantillon de la notice "Bayonne" sur wikidata.org (consulté le 09/09/19)

La dernière étape consiste en l'éditorialisation manuelle du film afin de créer un contexte de lecture du film enrichi des tags et notices. Les traits sémantiques identifiés ont chacun fait l'objet d'une annotation sous la forme d'un tag et certaines notices ont ainsi été rapatriée depuis Wikidata, à l'instar de la notice « Bayonne ». La couleur et la forme des tags sont différentes selon la classe sémantique, comme l'indique la figure suivante.

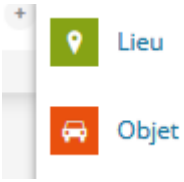


Figure 6 : Capture d'écran des classes sémantiques //Lieu// et //Objet// devenues des axes d'indexation (Crédits Perfect Memory, 2019)

L'utilisateur a ainsi la possibilité de créer des annotations de type « Lieu » dans le logiciel, associées à un time code précis du film consulté. Les propriétés des entités sémantiques « Lieu » sont décrites dans l'ontologie Perfect Memory ainsi :
http://www.perfect-memory.com/ontology/pmcore/1.0#Place

pmcore:Place rdf:type owl:NamedIndividual .

En suivant les recommandations du W3C, nous pouvons lire cet énoncé ainsi : toute annotation « Lieu » est une instance de l'ontologie « pmcore ». La catégorie d'instances « personne nommée » est décrite dans le langage « owl », elle est une instance de la classe de ressources « rdfs :Class »[1]. Enfin, l'annotation « Lieu » est une instance de la catégorie d'instances « personne nommée ». Une annotation ne pourra ainsi être acceptée que si elle correspond à la syntaxe de la catégorie « individu nommé », c'est-à-dire une chaîne de caractères interprétée comme une entité sémantique par le service. Si aucune entité sémantique des bases de connaissances ne correspond à la chaîne de caractère, la possibilité de créer une entité est proposée à l'utilisateur.

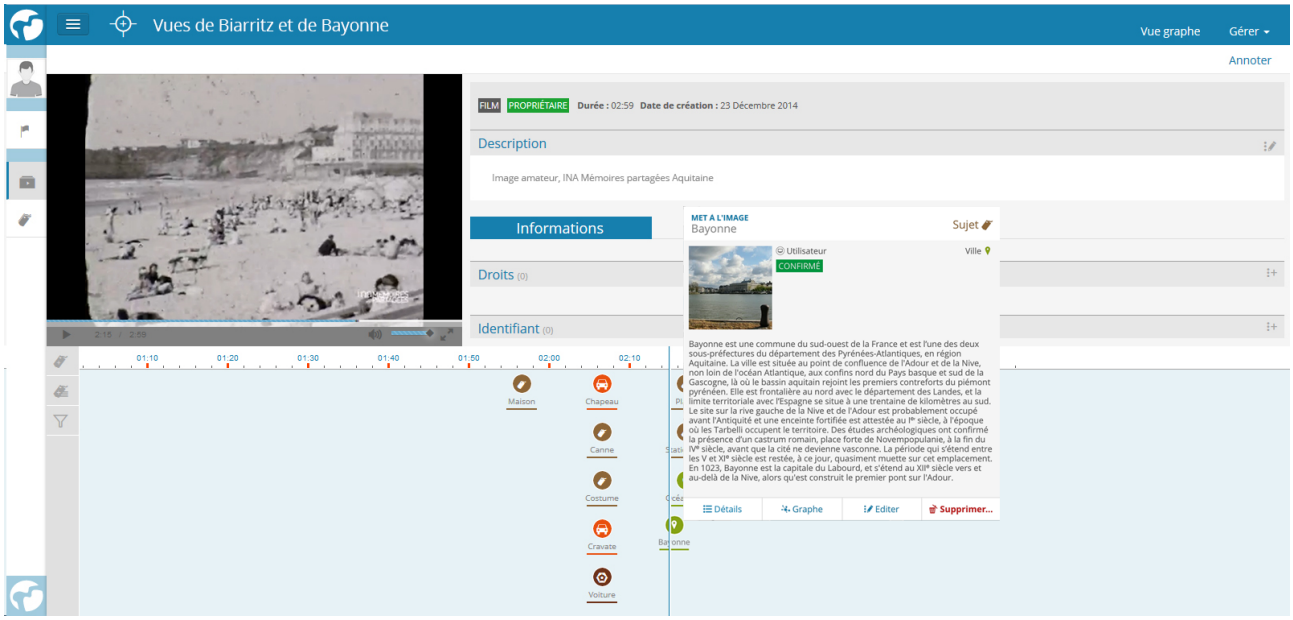


Figure 7 : Vues de Biarritz et de Bayonne enrichi dans Famille™ (Crédits INA, Perfect Memory, 2019)

À la suite de l'éditorialisation du film, une expérimentation peut être lancée : le visionnage du film enrichi depuis le logiciel Famille™. Outre le mode de lecture en plein écran, l'utilisateur a la possibilité, comme indiqué sur la dernière figure, de lancer la lecture du film tout en consultant les métadonnées associées à ce dernier. Celles-ci s'organisent en deux espaces : un bloc vertical déroulant qui comprend toutes les propriétés associées au film (« Description », « Droits », « Identifiants », « Durée », etc.) et un bloc horizontal – la ligne de temps – qui accueille les tags temporels et défile pendant le visionnage. Lors du survol d'un tag, un bloc apparaît et présente les propriétés du tag, les boutons d'actions liés au tag mais surtout un texte descriptif qui contient des informations sur l'entité sémantique taguée. Ce texte provient de la notice Wikidata associée à l'entité sémantique, qui comprend de nombreuses autres propriétés. Le choix ergonomique de ne présenter dans ce bloc que la description de l'entité répond au souci de ne pas noyer l'utilisateur sous l'information. L'impact de cette fonctionnalité sur l'expérience utilisateur n'est pas négligeable. Dans la même

vue, on peut consulter le film, découvrir les tags qui le décrivent mais aussi lire des informations supplémentaires sur les entités taguées. Cette fonctionnalité épargne à l'utilisateur la tâche d'aller se renseigner sur les tags lus qui ne trouvent pas d'écho dans son bagage, sans pour autant l'extraire de l'immersion propre au visionnage d'images en mouvement.

6. Discussion

De l'étude sémiotique des films a émergé une typologie de classes sémantiques, autant d'axes d'annotations du film à intégrer dans l'interface de Famille™. Nous avons fait le choix ici, en lien avec le format de l'article, de discuter d'un seul film, échantillon du corpus. Une mise en perspective des résultats obtenus par le biais de la convocation de vingt autres films - avec les trois paliers d'interprétation (micro-, méso- et macro-sémantique) – est présente dans le mémoire de thèse (Leyoudec, 2017). L'analyse des traits sémantiques actualisés dans l'ensemble du corpus permet d'identifier d'autres classes sémantiques, implémentées ensuite dans le logiciel. Notre hypothèse consiste à considérer ces classes comme des axes pertinents pour la recontextualisation de la production familiale. Notre typologie d'indexation des films prend ainsi sa source dans la sémantique différentielle, elle n'est pas inspirée par les conventions documentaires audiovisuelles professionnelles, stabilisées depuis les années 1990 (Briet, 1951, Bachimont, 1998). Plutôt qu'une taxonomie, notre typologie appartient à la catégorie des folksonomies (Broudoux, 2013 : 5), grammaire écrite par des non-professionnels. Cette typologie n'est pas construite pour être intégrée à la plateforme Famille™, elle est bâtie en amont. L'ontologie ou grammaire des ressources numériques manipulées dans la plateforme comporte de nombreuses propriétés. Un travail de rapprochement, par la création de liens d'équivalence (liens de type « same as ») est effectué pour rendre compatible la typologie sémiotique et l'ontologie existante. La singularité de notre approche d'éditorialisation se dégage ainsi à un premier niveau, en amont. La technique n'impacte pas la création de la typologie, elle prend sa source dans l'usage descriptif que l'analyse sémiotique accompagne.

En aval, notre approche bénéficie d'un interfaçage avec les bases de données liées et ouvertes sur le Web. Comparons la plateforme Famille™ à deux logiciels de gestion de collections patrimoniales, ActiMuseo (AA Partners) et Argus (Lucidea), installés chez de nombreuses institutions patrimoniales européennes. Ces logiciels, construits pour gérer des notices, proposent des fonctionnalités d'éditorialisation limitées : la taxonomie sous-jacente au logiciel est souvent plate (non-sémantique), tandis qu'annoter temporellement un contenu multimédia n'est pas possible.

Du côté des plateformes pensées pour les contenus multimédias, on peut citer Ligne de temps (IRI), dont l'ergonomie est inspirée des tables de montage audiovisuel. L'utilisateur peut sélectionner des extraits dans une ligne de temps et le décrire en l'annotant. Le potentiel pédagogique de ce type d'outil d'annotation collaborative est

considérable, nous soulignons à ce propos les travaux liés au projet Celluloid (Bourgatte et al., 2017). Une limite que nous observons à ces outils est la nature des connaissances versées dans le logiciel par l'utilisateur. Celles-ci prennent la forme de chaînes de caractères. De fait, seul un humain qui comprend la langue dans laquelle est écrite l'annotation peut interpréter la connaissance versée.

En interrogeant les bases de données liées et ouvertes sur le Web, autant de points d'accès à de nombreuses connaissances que nous souhaiterions mobiliser pour contextualiser les productions familiales, nous rencontrons un verrou technique quant à la mobilisation de ces données. Pour interroger les bases (Wikidata, DBpedia, etc.) l'utilisateur doit rédiger une requête dans un langage informatique – SPARQL (Corby et al., 2014) – dont la syntaxe nécessite un bagage technique considérable. La solution consistant à exporter l'ensemble de la base en local puis de l'intégrer au logiciel d'éditorialisation demande de nombreuses compétences techniques et entraîne un autre problème : les notices récupérées ne sont pas mises à jour, à moins de réexporter la base.

En comparant différentes plateformes avec Famille™, nous pouvons plus facilement identifier la singularité de la plateforme. L'utilisateur a la possibilité de consulter les productions familiales et de les taguer. Ces annotations étant sémantiques, elles permettent à la plateforme de rechercher si des notices contextuelles correspondantes existent sur le Web, puis de les rapatrier dans l'interface, le tout sans couper l'expérience utilisateur de consultation. Cette plateforme s'applique ainsi à un usage professionnel (le documentaliste qui décrit le document) qu'à un usage grand public (les familles qui témoignent d'événements mis à l'image dans les productions familiales).

7. Conclusion

Lors de l'expérimentation, un ensemble de notices ont été extraites automatiquement de la base de données liées et ouvertes sur le Web Wikidata vers le logiciel Famille™. L'éditorialisation des films par l'action conjointe des agents automatiques et de l'utilisateur final du logiciel a permis d'observer un segment de film éditorialisé. Les *Linked Open Data* généralistes semblent ainsi une source d'informations contextuelles pertinentes dans notre démarche d'éditorialisation. Les caractéristiques « ouvertes » et « liées » des données favorisent leur réutilisation dans le cadre de dispositif tiers comme Famille™. De même, le rapatriement automatique a été rendu possible par la configuration d'un connecteur entre Wikidata et Famille™.

L'annotation couplée avec le rapatriement de la notice permet de décrire un segment temporel du film. En appliquant cette démarche à l'ensemble du film, ce dernier est intelligible jusqu'à un certain pourcentage que nous ne pouvons exprimer aujourd'hui. Non seulement l'éditorialisation descriptive n'égale pas une démarche narrative (*via* le témoignage d'un membre de la famille à laquelle le film est lié) mais elle se heurte en

plus à des barrières culturelles : les bases de données liées et ouvertes sont généralistes donc ne contiennent pas pour le moment suffisamment de notices pour couvrir les spécificités (lieux, traditions, etc.) locales. Wikidata étant enrichi par son alter ego non-sémantique Wikipédia, l'augmentation des notices mises en ligne sur cette plateforme est une perspective pouvant répondre à la contrainte exprimée. En observant que plus de 25.000 articles sont créés chaque jour sur les différentes versions de Wikipédia, on peut estimer que l'éditorialisation « encyclopédique » dans Famille™ à l'aide des *Linked Open Data* sera invariablement plus précise au fur et à mesure de l'accroissement de la base de données.

Un film est enrichi de deux manières sur Famille™ : un utilisateur annote sémantiquement un film ou une combinaison d'agents automatiques produisent des métadonnées en interrogeant l'image et/ou la bande sonore, ces dernières étant ensuite élevées sémantiquement et des notices supplémentaires sont rapatriées. Il s'agit de ressources sémantiques comportant des propriétés explicitées dans l'ontologie informatique. Ces données riches car sémantiques constituent une matière première de choix pour d'autres modules dits d'« intelligence artificielle » qui ont besoin de données d'entraînement les plus riches possibles pour produire des résultats satisfaisants. Plus les métadonnées des productions familiales seront riches, plus « le robot » sera capable de décrire le film. De même, plus les métadonnées liées aux films seront nombreuses, plus facilement de nouvelles connaissances peuvent être déduites par des règles d'inférences compilées dans l'ontologie informatique. Contrairement à une base de données relationnelles dont la multiplication des tables peut entraîner une entropie, l'augmentation des métadonnées sémantisées se rapproche d'un système dit « néguentropique ». L'architecture technique sous-jacente de Famille™ désamorce ainsi un des écueils possibles d'une éditorialisation semi-automatique des productions familiales, sans pour autant apporter de réponse claire quant au calcul de la pertinence des métadonnées produites automatiquement par les modules d'« intelligence artificielle », lesquels feront l'objet d'une prochaine étude.

8. Bibliographie

Bachimont, B. (1998). Bibliothèques numériques audiovisuelles : des enjeux scientifiques et techniques. *Document numérique*, vol. 2, 219-242.

Bachimont, B. (2007a). *Ingénierie des connaissances et des contenus : le numérique entre ontologies et documents*. Paris : Lavoisier, 2007.

Bachimont, B. (2007b). Nouvelles tendances applicatives. De l'indexation à l'éditorialisation. In Gros, P. (dir.), *L'indexation multimédia*. Paris : Hermès, 2007.

Bachimont, B. (2010). La présence de l'archive : réinventer et justifier. *Intellectica*, n°53-54, 281-309.

Berners-lee, T. (2006). Linked Data. *W3C*. Repéré à http://www.w3.org/DesignIssues/LinkedData.html

Bourgatte, M., Tessier, L. (2017). Les outils d'annotation vidéo pour la recherche. In

Cavalié, É., Clavert, F., Legendre, O., Martin, D. (dir.), *Expérimenter les humanités numériques. Des outils individuels aux projets collectifs*, Montréal : Les Presses de l'Université de Montréal, 2017, ISBN: 978-2-7606-3802-0
Briet, S. (1951). *Qu'est-ce que la documentation ?*, Paris : Éditions documentaires, industrielles et techniques, 1951.
Broudoux, É. (2013). Quelles lectures du tagging : Modélisation, techniques et usages. *Document numérique*, vol. 16(1), 55-71. doi:10.3166/DN.16.1.55-71.
Chignard, S. (2012). *L'open data : comprendre l'ouverture des données publiques*. Limoges : Fyp Editions, 2012.
Corby, O., Faron Zucker, C., Gandon, F. (2014). SPARQL Template: A Transformation Language for RDF. Rapport de recherche RR-8514, INRIA, 2014, hal-00969068.
Doueihi, M. (2011). *Pour un humanisme numérique*. Paris : Seuil, 2011.
Gandon, F., Faron-Zucker, C., Corby, O. (2011). *Le Web sémantique : comment lier les données et les schémas sur le Web ?* Paris : Dunod, 2011.
Heath, T., Hepp, M., Bizer, C. (2009). Linked Data - The Story so far. *International Journal on Semantic Web and Information Systems,* « Special Issue on Linked Data ».
Hébert, L., (2007). *Dispositifs pour l'analyse des textes et des images*. Limoges : Presses universitaires de Limoges, 2007.
XXX., (2017), L'éditorialisation du film de famille. L'architexte Famille comme support d'un exercice de la mémoire, thèse de doctorat, Université de technologie de Compiègne.
Mattelart, A., (1994). *L'invention de la communication*. Paris : La Découverte, 1994.
Metz, C., (1971). *Langage et cinéma*. Paris : Larousse, 1971.
Ostrom, E., (2010). *La gouvernance des biens communs : Pour une nouvelle approche des ressources naturelles*, Paris : Commission Université Palais (collection « Planète enjeu »), 2010.
Ploux, S., (2011). Modèles et sémantique lexicale. In Kayser, D., Garbay, C. (dir.), *Informatique et sciences cognitives : influences ou confluence ?*, Paris : Maison des sciences de l'homme, 1-18.
Rastier, F., (1987). *Sémantique interprétative*. Paris : PUF, 1987, 3ème édition 2009.
Rastier, F., Cavazza, M., Abeillé, A., (2001). *Sémantique pour l'analyse. De la linguistique à l'informatique*. Paris : Masson, 2001.
Rastier, F., (2008). Passages et parcours dans l'intertexte. *Texto!*, vol. XIII n°4, 1-20.
Rastier, F., (2011). *La mesure et le grain. Sémantique de corpus*. Paris : Honoré Champion, 2011.
Treleani, M., (2013), Recontextualisation : ce que les médias numériques font aux documents audiovisuels. *Réseaux,* vol. 31, n˚177, 233-258.

9. Notes

[1] https://www.w3.org/TR/2004/REC-rdf-schema-20040210/ - ch_type ; https://www.w3.org/TR/2004/REC-rdf-schema-20040210/ - ch_class ; https://www.w3.org/TR/owl2-quick-reference/;

https://www.w3.org/TR/2012/REC-owl2-primer-20121211/ - Entity_Declarations