

Description et indexation des collections d'images en mouvement : résultats d'une enquête
Description and Indexing of Moving Images: Survey Results
Descripción y clasificación de las colecciones de imágenes en movimiento: resultados de un estudio

Michèle Hudon, James Turner and Yves Devin

Volume 47, Number 1, January–March 2001

URI: <https://id.erudit.org/iderudit/1032646ar>

DOI: <https://doi.org/10.7202/1032646ar>

[See table of contents](#)

Publisher(s)

Association pour l'avancement des sciences et des techniques de la documentation (ASTED)

ISSN

0315-2340 (print)

2291-8949 (digital)

[Explore this journal](#)

Cite this article

Hudon, M., Turner, J. & Devin, Y. (2001). Description et indexation des collections d'images en mouvement : résultats d'une enquête. *Documentation et bibliothèques*, 47(1), 5–12. <https://doi.org/10.7202/1032646ar>

Article abstract

During the course of a recently completed research project, the authors were interested in the vocabulary and structure of indexing languages used to describe the content of moving images of everyday objects and situations. Eleven organisations, managing fourteen collections of images, completed a questionnaire and were visited. The data leads to the conclusion that the collections are large and at least half are indexed with a more or less controlled vocabulary, often a thesaurus. Notwithstanding the similarities among the collections, the vocabularies have little in common. Moving image collections are often described and indexed according to in-house principles and techniques, with little standardisation and are rarely compatible.

Description et indexation des collections d'images en mouvement : résultats d'une enquête*

Michèle Hudon

Professeure adjointe, EBSI, Université de Montréal

James Turner

Professeur agrégé, EBSI, Université de Montréal

Yves Devin

Agent de recherche, EBSI, Université de Montréal

Dans le cadre d'un projet de recherche récemment complété, nous nous sommes intéressés au lexique et à la structure des langages documentaires utilisés pour représenter le contenu d'images en mouvement décrivant des objets et des situations de la vie courante. Onze organismes, gérant quatorze collections d'images ont répondu à un questionnaire et ont été visités. Les données recueillies ont permis de constater que les collections sont imposantes et qu'au moins la moitié d'entre elles sont indexées au niveau du plan à l'aide d'outils langagiers plus ou moins contrôlés, souvent un thésaurus. Mais malgré la similarité des collections, les divers lexiques montrent peu de recoupements. Les collections d'images en mouvement sont encore décrites et indexées selon des principes et des techniques établis localement, peu normalisés et rarement compatibles.

Description and Indexing of Moving Images: Survey Results*

During the course of a recently completed research project, the authors were interested in the vocabulary and structure of indexing languages used to describe the content of moving images of everyday objects and situations. Eleven organisations, managing fourteen collections of images, completed a questionnaire and were visited. The data leads to the conclusion that the collections are large and at least half are indexed with a more or less controlled vocabulary, often a thesaurus. Notwithstanding the similarities among the collections, the vocabularies have little in common. Moving image collections are often described and indexed according to in-house principles and techniques, with little standardisation and are rarely compatible.

Descripción y clasificación de las colecciones de imágenes en movimiento: resultados de un estudio*

En el contexto de un proyecto de investigación concluido recientemente, nos interesó el léxico y la estructura de los lenguajes documentales utilizados para describir el contenido de imágenes en movimiento que representan objetos y situaciones de la vida cotidiana. Se visitaron y respondieron a un cuestionario once organismos, que tienen a su cargo catorce colecciones de imágenes. Los datos obtenidos demuestran que las colecciones son imponentes y que al menos la mitad de ellas están catalogadas según un sistema que se vale de herramientas idiomáticas más o menos controladas, con frecuencia un diccionario de referencia. Pero a pesar de la similitud de las colecciones, los distintos léxicos tienen poca relación entre ellos. Las colecciones de imágenes en movimiento todavía se describen y clasifican según técnicas y principios establecidos localmente, poco normalizados y rara vez compatibles.

De tous temps, l'image a été utilisée pour représenter des concepts et des idées et pour communiquer des messages. La photographie, le cinéma, la télévision et plus récemment le développement fulgurant de la micro-informatique ont

conféré à l'image une importance de plus en plus considérable. Quel qu'en soit le format et le médium de présentation, l'image est un mode important de communication d'information dans la société actuelle. L'image joue un rôle crucial dans

des secteurs de l'activité humaine aussi diversifiés que la médecine, le journalisme, la publicité, l'éducation et le divertissement. Bien qu'une typologie officielle reste à établir, on sait déjà que les collections « visuelles » prennent différentes formes et

* Ce projet a été financé par le Steven I. Goldspiel Memorial Research Grant 1999 de la Special Libraries Association.
This project received funding from the Steven I. Goldspiel Memorial Research Grant 1999 from the Special Libraries Association
Este proyecto fue financiado por el Steven I. Goldspiel Memorial Research Grant 1999 de la Asociación de Bibliotecas Especiales (Special Libraries Association).

contiennent des éléments associés aux beaux-arts, au dessin, à la gravure, à la photographie, à l'infographie, etc.¹ Dans cette diversité, l'image en mouvement occupe une place importante. L'image en mouvement a été « *conçue par son créateur pour présenter une information en mouvement à la projection, souvent sur écran fixe* » (Turner 1998, 84). Les images en mouvement sont une mine d'or pour nombre d'organisations et d'individus et il importe de bien les décrire pour en faire ressortir la richesse et la complexité et pour en faciliter l'identification et le repérage. Il semble pourtant que les collections d'images en mouvement soient encore décrites et indexées selon des principes et des techniques établis localement, peu normalisés et trop rarement compatibles.

Notre projet de recherche, complété au cours de l'année 2000, portait sur le traitement documentaire dans les collections d'images en mouvement. L'objectif général de l'étude était d'en arriver à une meilleure compréhension des techniques et des outils lexicaux utilisés pour la représentation des contenus dans ces collections spécialisées. Nous nous sommes intéressés particulièrement au lexique et à la structure, ainsi qu'aux opérations de développement et de mise à jour des langages d'indexation. De notre objectif général découlaient en effet un certain nombre d'objectifs spécifiques : déterminer combien de termes, à l'exclusion des noms propres, sont inclus dans les outils lexicaux utilisés dans les collections d'images en mouvement nord-américaines ; estimer le rythme de croissance lexical de ces outils ; vérifier jusqu'à quel point les contenus lexicaux sont similaires ; évaluer l'opportunité de proposer un lexique unique pouvant être utilisé pour l'indexation et le repérage dans la majorité des collections d'images en mouvement représentant des objets et des situations de la vie courante.

Dans cet article, nous présentons d'abord le cadre théorique de notre étude. Nous traitons ensuite brièvement des aspects méthodologiques de notre projet de recherche avant de décrire les principaux résultats obtenus et d'en proposer une interprétation critique.

Problématique et cadre théorique

L'image et les collections d'images

Les images, fixes ou en mouvement, sont omniprésentes dans notre quotidien. Elles existent sur différents supports : du papier au numérique en passant par la pellicule. Elles sont par nature artistiques, historiques ou simplement représentatives de situations de la vie courante. Nous sommes exposés aux images dans toutes les sphères de notre vie personnelle, sociale et professionnelle. La société dans laquelle nous vivons donne à l'image une place de premier plan (Brown et al. 1996, 107 ; Armitage and Enser 1997, 287). Le prestige de l'image fait d'ailleurs en sorte que la représentation textuelle n'est plus nécessairement le paradigme dominant lorsqu'il s'agit d'informer ou d'instruire (Id., 287).

D'importantes collections d'images sont disponibles dans les bibliothèques publiques ou spécialisées, dans les centres de documentation et dans toute autre organisation pour laquelle ce type de document constitue une ressource incontournable. Grâce aux récents progrès de l'informatique et au développement du réseau Internet, nous assistons actuellement à une migration des supports ; des collections complètes d'images se voyant transférées d'un support matériel (la pellicule par exemple) vers un support électronique, autant pour en assurer la conservation que pour en élargir la diffusion.

La valeur même de ces collections ainsi que tous les efforts déployés pour développer des technologies encore plus performantes qui en permettent l'exploitation exigent la mise en place d'un système efficace de gestion. Il importe en effet que le repérage d'un document particulier au sein d'une collection donnée se fasse efficacement et rapidement. Sans la mise en place de nouvelles méthodes de stockage et de repérage de l'image, il se pourrait que nous devions bientôt faire face à des problèmes difficiles, sinon impossibles à résoudre (Turner 1998, 1).

La problématique de l'analyse et de la représentation du contenu de l'image est abordée selon deux approches distinctes qui n'ont entre elles que peu de caractéris-

tiques communes (Cawkell 1992, 180). La première approche repose en grande partie sur la technologie et les méthodes statistiques. Les spécialistes et chercheurs impliqués dans ce secteur sont davantage intéressés par la reconnaissance des formes et le traitement automatique de l'image. Quelques applications de ce courant sont liées, par exemple, au développement de thésaurus imagés comme celui de la NASA (Seloff 1990 ; Rorvig et al. 1999) ou encore aux thésaurus utilisés pour la gestion des collections d'images géographiques (Ramsey et al. 1999).

La seconde approche repose plutôt sur l'intervention humaine et s'inspire grandement des méthodes de travail développées au cours des ans pour la gestion des documents textuels par les professionnels de l'information documentaire. Il y est donc question de description, de classification et d'indexation. Un travail important d'observation et d'analyse a d'ailleurs déjà été réalisé en ce qui concerne les collections de documents visuels ; à titre d'exemples, voir les travaux de Shatford (1986, 1994), de Yee (1993), de Cawkell (1993 a, 1993 b) et de Rasmussen (1997).

Est-il possible et souhaitable d'appliquer aux collections de ressources visuelles les techniques d'analyse et de représentation des contenus mises au point pour les documents textuels ? À cette question, aucune réponse simple n'a encore été proposée. En ce qui concerne plus particulièrement les images en mouvement, il semble que les techniques déjà éprouvées pour stocker et retrouver les documents textuels ne soient pas toujours adéquates (Turner 1998, 21). Turner croit que, dans certaines situations, il est préférable d'indexer les images en mouvement plan par plan, plutôt que sur la base du document dans son entier (Id., 53, 75). En complément de l'image, le texte est utilisé pour créer des métadonnées qui serviront lors de la recherche. Cette information descriptive est d'autant plus importante qu'elle ne peut pas toujours être inférée de l'image elle-même (Turner, Hudon and Devin 2000).

1. Pour un essai de typologie et de représentation visuelle, voir Groupe départemental de recherche en information visuelle (GRIV) de l'École de bibliothéconomie et des sciences de l'information, Université de Montréal, 1998. *L'univers des collections visuelles = The world of visual collections*. Affiche recto-verso, 91 x 61 cm.

Compte tenu de l'expansion rapide des collections d'images, de la quantité d'information qu'elles contiennent et des sommes importantes nécessaires à leur développement et à leur exploitation, il devient essentiel que soient développées et implantées des méthodes communes de gestion de ces collections, autant pour retrouver rapidement le matériel dont nous avons besoin que pour partager les ressources informationnelles que nous possédons déjà. Une façon d'y arriver, par exemple, serait de procéder à l'indexation des ressources de diverses collections à l'aide d'un thésaurus commun.

Analyse et représentation du contenu de l'image en mouvement

À la différence de la majorité des textes, l'image peut être analysée et interprétée de diverses manières. Panofsky (1955) suggère que l'on peut extraire de l'image trois niveaux de signification. Le premier niveau, qualifié de préconographique, porte sur le sujet primaire ou naturel du document. Le second niveau, l'iconographique, s'intéresse au sujet secondaire ou conventionnel. Le troisième niveau, qualifié d'iconologique, s'attarde au sujet tertiaire ou au contenu symbolique. Dans ses travaux basés sur les notions avancées par Panofsky, Shatford (1986) met davantage l'accent sur les premier et second niveaux pour distinguer entre l'*ofness* et l'*aboutness* de l'image. L'*ofness* fait ici référence à la description pure et simple du contenu de la ressource visuelle (i.e. ce qu'on y voit) alors que l'*aboutness* s'attarde à l'interprétation ou à la signification du document (i.e. ce qu'on y lit).

Une image, fixe ou en mouvement, contient une grande variété d'informations et elle peut signifier différentes choses pour différentes personnes (Id., 42). Ce fait peut être problématique dans la perspective d'une normalisation de la description en vue du repérage, et c'est ce qui a amené l'Office national du film du Canada, par exemple, à indexer les *stockshots* au premier niveau de signification seulement (Turner 1990, 12). Il y a lieu de croire que la majorité des collections d'images en mouvement sont ainsi indexées et que les descripteurs nécessaires à la représentation des contenus nommeront des entités concrètes (par exemple, un chat et une

chaise) plutôt que des notions abstraites (par exemple, la quiétude ou le confort). Dans un tel contexte, l'utilisation d'un thésaurus présentera un intérêt certain. Le contrôle lexical et structurel offert par le thésaurus devrait contribuer à améliorer l'accès au contenu de la collection, à réduire le bruit et le silence au repérage, à augmenter le taux de précision et à mieux satisfaire l'utilisateur qui pourrait ainsi trouver ce qu'il cherche sans perte excessive de temps.

Le thésaurus est un outil dont le développement repose sur un ensemble de principes et de règles énoncés dans des normes internationales (Hudon 1994, 75-76). Les descriptions de l'outil thésaurus rendent compte à la fois de sa structure et de sa fonction :

Un thésaurus est une liste d'autorité organisée de descripteurs et de non-descripteurs obéissant à des règles terminologiques propres et reliés entre eux par des relations sémantiques (hiérarchiques, associatives ou d'équivalence). Cette liste sert à traduire en un langage artificiel dépourvu d'ambiguïté des notions exprimées en langage naturel (Association française de normalisation 1981).

Le thésaurus est un outil dynamique capable de s'adapter aux nouvelles réalités et aux nouveaux besoins dans les organisations où il est utilisé. Son lexique sera augmenté et bonifié pour mieux traduire les requêtes des utilisateurs. Le thésaurus est cependant un outil relativement exclusif, normalement développé pour décrire un domaine bien délimité, et reflétant les particularités propres à une collectivité d'utilisateurs (Van Slype 1987, 117). Contrairement aux schémas de classification documentaire et aux répertoires de vedettes-matières, le thésaurus traditionnel n'est jamais encyclopédique.

Quelques thésaurus ont été créés spécifiquement pour l'indexation des documents visuels (matériel artistique, photos, diapositives, films ou plans). Au nombre de ces derniers, le plus connu est le *Art and Architecture Thesaurus* (AAT)², dont la gestion est assurée par la Jean-Paul Getty Foundation. Le AAT propose une terminologie normalisée de 40 000 termes, couvrant l'art et l'architecture de l'Antiquité au temps présent. Le *Thesaurus for Graphic Materials* (TGM)³ de la Library of Congress offre pour sa part une série de descripteurs permettant l'indexation de maté-

riel graphique incluant les imprimés, les photographies, les dessins, les bandes dessinées, les affiches et les plans architecturaux. Au Canada, il faut mentionner le thésaurus de l'Office national du film du Canada (ONF), développé à partir de 1987 pour faciliter l'indexation des *stockshots* conservés à la cinémathèque des plans d'archives. S'écartant davantage des approches de représentation traditionnelles, quelques thésaurus visuels (*visual thesauri*) sont aussi disponibles. Ces thésaurus représentent les objets par l'image plutôt que par le texte. À l'aide de ces images-concepts, il est possible de repérer une série de ressources visuelles offrant une représentation de l'objet recherché. Cette technique permet d'accéder aux images sans avoir à recourir au texte (Rasmussen 1997, 182), un avantage certain en environnement multilingue. Le *NASA Visual Thesaurus*, développé pour faciliter l'accès aux archives (films) de la NASA (Johnson Space Center) est représentatif de cette catégorie d'outils.

Le développement d'un thésaurus est un travail complexe, réalisé en une séquence d'étapes nécessaires et d'opérations intellectuelles devant mener à un grand nombre de décisions. L'étape de l'élaboration du lexique en est une de première importance. Que les descripteurs proviennent de sources de référence, des images à indexer ou des questions des utilisateurs, ils doivent nommer clairement et avec suffisamment de précision les objets représentés dans les images à décrire.

Certaines sources, plus anecdotiques que formelles, suggèrent que le nombre de descripteurs nécessaires à l'indexation d'une collection d'images en mouvement est en fait limité, reflétant en cela un phénomène reconnu dans la langue naturelle, où le nombre de mots disponibles est de beaucoup supérieur au nombre de mots utilisés dans le discours et essentiels à la communication d'information⁴. Un nombre plus restreint qu'on pourrait le croire de descripteurs (noms communs seulement) serait donc suffisant pour décrire une majorité de catégories de personnes, d'objets et d'événements de la vie

2. <http://shiva.pub.getty.edu/aat_browser> (Page consultée le 3 février 2001).

3. <<http://lcweb.loc.gov/rr/print/tgm1>> et <<http://lcweb.loc.gov/rr/print/tgm2>> (Pages consultées le 3 février 2001).

4. Guiraud indique que 4000 mots couvrent 97,5 % de n'importe quel texte (1960, 93).

Tableau 1. Le contenu des collections

TYPE ET FORMAT	NOMBRE DE COLLECTIONS
<i>Film</i>	
8 mm	4
16 mm	8
35 mm	8
72 mm	1
Autres (film)	6
<i>Vidéo</i>	
3/4 po U-matic	11
1 po	8
2 po	5
Bétacam	13
Autres (vidéo)	10
<i>Autres formats</i>	
Disque optique, etc.	5

courante tels que représentés dans les ressources visuelles. Cette hypothèse est-elle conforme à la réalité? Si tel est le cas, un lexique commun pourrait peut-être suffire à la description de collections d'images en mouvement dans des organisations variées.

Méthodologie

Organisations participantes

Trente-trois organisations⁵ répondant à des critères de sélection préétablis, ont été identifiées comme participantes potentielles à ce projet (liste à l'Annexe 1). Les organisations sélectionnées géraient, depuis au moins cinq ans, des collections d'images en mouvement de nature autre qu'artistique. Chacune de ces organisations a reçu une trousse d'information sur le projet accompagnée d'une invitation à participer à l'étude. Au total, trois prises de contact préliminaires ont été tentées.

Vingt-deux réponses ont été obtenues (67%). Onze organisations (50%) ont accepté de participer au projet alors que neuf autres (41%) refusaient de s'impliquer plus avant. Les refus provenaient surtout du secteur privé pour qui la contribution à l'avancement de la recherche

Tableau 2. Volume des collections

	NOMBRE DE TITRES	NOMBRE D'HEURES
C1	4 962	n/d
C2	14 000	3 800
C3	n/d	n/d
C4	36 848	n/d
C5	11 755	750
C6	100 000	n/d
C7	n/d	n/d
C8	n/d	n/d
C9	50 000	17 500
C10	n/d	n/d
C11	18 500	17 848
C12	94 732	n/d
C13	5 600	n/d
C14	n/d	n/d

n'est pas nécessairement une priorité et parce que bon nombre de données y sont considérées comme leur propriété. Deux organisations (9%) initialement intéressées n'ont jamais donné de réponse définitive.

Chaque organisation participante disposait d'au moins une collection d'images en mouvement. L'ensemble des 11 organisations participantes gérait un total de 14 collections.

Questionnaire

Les organisations participantes ont répondu à un questionnaire conçu spécialement pour ce projet de recherche. Le questionnaire, disponible en français et en anglais, comprenait quatre sections distinctes: l'identification du milieu, la caractérisation des collections, le mode de gestion des collections, la caractérisation des outils langagiers utilisés pour l'indexation et le repérage.

Visite des milieux et entrevues

La visite des milieux a été une étape importante dans notre processus de cueillette des données. Une entrevue avec un répondant de l'organisation hôte a permis de vérifier et de compléter les données

obtenues par les réponses au questionnaire. Un canevas d'entrevue structuré autour de questions ouvertes a servi de cadre de référence. Tous nos répondants étaient associés de très près à la gestion des collections et tous possédaient une bonne connaissance de leur milieu, des collections et des instruments (bases de données, thésaurus, etc.) servant à la gestion de celles-ci. La visite des milieux a aussi permis de consulter et parfois même d'obtenir copie de différents documents susceptibles de contenir le type d'information dont nous avons besoin (par exemple, un thésaurus, une politique d'indexation, etc.).

Les organisations participantes se sont montrées particulièrement coopératives et prêtes à partager les données qui nous intéressaient lorsqu'elles étaient disponibles.

Résultats et discussion

Chacune de nos organisations a une structure de fonctionnement qui lui est propre et qui reflète la philosophie, les modes de gestion, les objectifs et les orientations de l'établissement. Les impératifs de temps et de production y sont omniprésents. À l'occasion, ces impératifs se traduisent concrètement par la duplication d'un matériel déjà disponible dans la collection, mais non retrouvé dans l'heure qui suit la demande d'une équipe de production. À ces impératifs s'ajoutent des exigences de rationalisation budgétaire qui impliquent de «faire plus avec moins». Cette combinaison apparaît de prime abord difficilement conciliable avec des investissements substantiels dans des outils de gestion documentaire complexes et en personnel spécialisé.

Identification des organisations

Les lieux où sont gérées les collections d'images en mouvement sont connus sous plusieurs dénominations, parfois même au sein d'une même organisa-

5. La plupart de ces organisations étaient représentées au congrès de l'Association of Moving Images Archivists (AMIA), tenu à Montréal en novembre 1999. Les premiers contacts informels ont été faits lors de ce congrès.

tion. La cinémathèque de *stockshots* est une dénomination populaire, utilisée par sept organisations. L'appellation centre d'archives est préférée en cinq endroits. Les autres dénominations suggérées sont: *Video Art Distributor* et *News Video Archive*.

Les collections

Sept collections sur quatorze (50%) sont des collections mixtes, couvrant à la fois des domaines généraux et des thèmes particuliers. Cette constatation ne surprendra pas lorsqu'on sait que le matériel conservé dans ces collections se compose de films, d'extraits de bandes de nouvelles et d'autres matériels d'intérêt public. Nos milieux sont liés de près à des réseaux de télévision ou de production cinématographique; pour des clientèles diverses, ils ont à couvrir une gamme très variée de sujets spécialisés et généraux. Trois collections (21%) sont de nature générale et deux seulement (14%) se qualifient de spécialisées. Il est intéressant de noter que deux collections ne peuvent apparemment pas être caractérisées aussi précisément. Douze collections sur quatorze (86%) ont moins de 50 ans d'âge.

Le Tableau 1 présente une vue d'ensemble du matériel disponible dans les diverses collections. Ce matériel est très diversifié, allant du film 8 mm au disque optique, en passant par le vidéo de type Bétacam. Cette diversité caractérise bien toute la pluralité propre à ce type de document qu'est l'image en mouvement. Les supports les plus largement représentés sont les films 16 mm et 35 mm et les vidéos de type U-Matic et Bétacam.

Les collections sont souvent imposantes. Les données du Tableau 2 ne donnent en fait qu'une image fragmentaire de l'importance des collections gérées par nos organisations témoins. Dans la plupart des cas, la seule information disponible est associée aux nombre de titres constituant la collection. Les données qui se rapportent aux valeurs en nombre d'heures de diffusion sont sommaires sinon inexistantes (n/d = non disponible), et aucune institution ne peut fournir de données sur le volume (c'est-à-dire le nombre de mètres linéaires de rangement).

Tableau 3. Niveaux de description et d'indexation

NIVEAU	NOMBRE DE COLLECTIONS
Titre	11
Séquence	5
Plan	8
Les trois niveaux	5
Autres	5

La description et l'indexation des documents

Compte tenu de l'importance des collections et du rythme de croissance de celles-ci, la contribution de l'informatique est depuis longtemps considérée essentielle. Les organisations visitées ont toutes créé des bases de données à structure complexe qui permettent un repérage plus ou moins efficace d'images représentant des entités ou des situations particulières.

Presque toutes les collections (11/14 ou 79%) sont cataloguées et indexées sur une base titre (ou document complet); ce pourcentage ne surprend pas si on pense à la facilité d'obtention et à l'importance de cet élément d'information. Quelques collections sont décrites et indexées plus en profondeur, au niveau de la séquence (5/14 ou 36%) ou du plan (8/14 ou 57%). Cinq organisations cataloguent et indexent aux trois niveaux (titre, séquence, plan). Cinq organisations disent aussi indexer à d'autres niveaux que ceux que nous leur avions proposés, au niveau d'une bobine complète par exemple (Tableau 3).

En majorité (10/14 ou 71%), les collections sont traitées au premier et au deuxième niveaux de signification définis par Panofsky. Il est surprenant de constater que cinq collections (5/14 ou 36%) sont apparemment indexées également au troisième niveau, soit au niveau de la symbolique: cette analyse qui implique une interprétation du matériel visionné est en effet plus courante dans le monde de l'œuvre d'art. Notons cependant qu'aucune organisation ne fait ce type d'interprétation sans recourir conjointement aux autres formes d'analyse.

Tableau 4. Langages documentaires utilisés

LANGAGE DOCUMENTAIRE	NOMBRE DE COLLECTIONS
Mots-clés	7
Classification	3
Thésaurus (commercial)	2
Thésaurus (maison)	5
Thésaurus (mixte)	1
Vedettes-matières	6
Autres	6

Les documents de quatre collections sont indexés en moyenne par cinq descripteurs ou moins. Le nombre maximum de descripteurs assignés peut atteindre 15 dans trois collections, et aucun maximum n'a été fixé dans deux autres. Dans tous les cas, le nombre maximum de descripteurs assignés dépend soit de la politique d'indexation, lorsqu'il y en a une, soit des capacités du système informatique en place.

Il est intéressant de souligner l'existence, dans les organisations participantes, d'un *continuum* dont une extrémité se caractérise par une absence totale de contrôle du processus d'indexation (tout est indexé ou rien n'est indexé) et l'autre par une indexation faite à l'aide d'un thésaurus développé spécifiquement pour décrire une collection particulière. Entre ces deux extrêmes, les autres modes de contrôle de l'indexation incluent l'utilisation des vedettes-matières de la Library of Congress (ou une adaptation de celles-ci), l'élaboration d'une liste de mots-clés ou d'une structure de classification simple et la combinaison de diverses techniques. Ce *continuum* vient par ailleurs faire ressortir la double tendance soulignée par Cawkell (1992) d'une utilisation maximale de l'informatique et du recours à des techniques déjà éprouvées dans le domaine des sciences de l'information. Il est très rare qu'une politique d'indexation formelle soit disponible et nous n'avons malheureusement pu en consulter aucune.

Vous trouverez au Tableau 4 les types de langages documentaires utilisés pour la représentation des contenus.

Le mot-clé, extrait de la langue naturelle et non contrôlé quant à sa forme et à

Tableau 5. Contenu des lexiques

OUTIL	NOMBRE TOTAL DE TERMES	NOMBRE DE NON- DESCRIPTEURS	NOMBRE DE NOMS PROPRES
Thésaurus 1	6 969	1 451	2 244
Thésaurus 2	344 500	8 850	220 000
Thésaurus 3	3 222	n/d	660
Thésaurus 4	1 163	42	744
Thésaurus 5	704	89	n/d
Thésaurus 6	3 680 (pour les lettres F, I, R)	1 346	1 204 (pers., coll., géographiques)

son sens, est utilisé dans sept collections (50%). Six collections (43%) sont indexées par des vedettes-matières. Un schéma de classification sert à l'organisation de trois collections (21%). Il faut préciser que les schémas de classification utilisés sont des classifications « maison », développées en fonction de besoins locaux. Nous avons pu constater lors de nos visites que six collections au moins sont également accessibles par le biais de recherches en plein texte, alors que deux gestionnaires de collection seulement avaient mentionné cette possibilité en réponse au questionnaire. Dans la catégorie « Autres » se retrouvent, notamment une liste de termes techniques spécifiques au domaine du cinéma et une liste de termes géographiques.

Une même organisation utilise généralement plus d'un instrument d'indexation pour une ou plusieurs de ses collections: la majorité des organisations ayant participé à cette étude disent utiliser entre deux et six outils langagiers différents pour la représentation des contenus de leurs collections.

Les thésaurus

L'utilisation d'un thésaurus comme outil de contrôle du vocabulaire est assez répandue. Six organisations sur un total de 11 (55%) font appel à un ou plusieurs outils qu'elles identifient comme étant un thésaurus. Deux collections sont indexées à l'aide d'un thésaurus commercial, une à l'aide d'un thésaurus mixte et cinq avec un thésaurus maison. Deux collections sont indexées à l'aide de deux outils de type thésaurus différents. Les six thésaurus que nous avons pu consulter nous ont été

présentés sous la forme traditionnelle d'une liste alphabétique de descripteurs. Une couverture spécialisée est relevée dans cinq outils (83%), alors qu'un seul est de nature résolument générale (17%). La terminologie est de niveau populaire dans trois thésaurus (50%) et davantage spécialisée dans les trois autres.

Plusieurs de nos questions avaient trait au contenu lexical de ces thésaurus (nombre total de termes, nombre total de non-descripteurs, nombre total de noms propres, etc.). Malheureusement, des données précises sur ce sujet ont été presque impossibles à obtenir des organisations participantes. Il s'avère que la plupart des thésaurus sont gérés par des logiciels informatiques propriétaires aux fonctions statistiques peu développées. Les chiffres du Tableau 5 sont donc pour la plupart des estimations à partir d'échantillons lexicaux prélevés dans les thésaurus auxquels nous avons eu accès. Nous les donnons à titre indicatif en soulignant la grande variation dans le nombre de termes constituant les lexiques.

D'abord, on notera l'importante proportion dans les lexiques, au moins 33% dans presque tous les outils, des noms propres de personnes, de collectivités et de lieux géographiques. On notera également la faible proportion de non-descripteurs dans tous les thésaurus. Il y a donc lieu de supposer que la réduction du vocabulaire par le contrôle des synonymes n'a pas été effectuée par les gestionnaires de ces lexiques; l'efficacité du thésaurus comme outil pouvant favoriser la qualité du repérage en est certes beaucoup diminuée.

Les termes qui constituent le lexique proviennent de différentes sources: des

Tableau 6. Structure relationnelle dans les thésaurus

TYPE DE RELATION	NOMBRE DE THÉSAURUS
Équivalence intralinguistique	4
Équivalence interlinguistique	1
Hierarchie	6
Association	6

ouvrages de référence généraux et spécialisés, des demandes faites par les usagers ainsi que des réseaux sémantiques déjà existants (dans d'autres thésaurus, par exemple).

La majorité des thésaurus en usage dans les organisations participantes repose sur une structure relationnelle explicite au sein de laquelle les descripteurs sont liés par des relations d'équivalence, de hiérarchie et d'autres types d'associations. Le Tableau 6 montre que les outils observés ont une structure relationnelle similaire, ce qui laisse croire que les normes de développement des thésaurus ont tout de même été prises en considération.

Le fait que quatre thésaurus seulement se préoccupent du contrôle des équivalences conceptuelles et terminologiques au sein d'une même langue suggère à nouveau que le contrôle sémantique n'y est que partiel et sans doute assez peu efficace. On notera qu'un seul des instruments observés tient compte de la relation d'équivalence interlinguistique (ou relation de synonymie entre deux langues). Cinq des outils consultés (83%) sont en effet des thésaurus de langue anglaise seulement; le sixième outil est un thésaurus bilingue dans lequel l'anglais et le français sont représentés.

L'efficacité des outils langagiers servant à l'indexation ne peut être maintenue que si leur contenu lexical et relationnel est à jour. Les réponses à une question précise concernant la fréquence et la régularité des opérations de mise à jour montrent que celle-ci est ponctuelle pour la moitié des thésaurus de notre échantillon (3/6), les changements requis étant immédiatement intégrés à l'outil. Dans les autres cas, la mise à jour est journalière (1/6), hebdomadaire (1/6) ou irrégulière

Tableau 7. Fréquence d'apparition (Fx) des termes dans sept outils lexicaux

	NOMBRE DE DESCRIPTEURS	POURCENTAGE
F1	1 680	73,0 %
F2	338	15,0 %
F3	134	6,0 %
F4	72	3,0 %
F5	47	2,0 %
F6	14	0,6 %
F7	7	0,3 %
TOTAL	2 292	100 %

(1/6). Dans la moitié des cas (3/6), une seule personne assume la responsabilité de la mise à jour et du contrôle de l'expansion du lexique. Deux personnes sont responsables d'un autre instrument et un nombre indéterminé d'individus (c'est-à-dire tous les utilisateurs) tiennent à jour les deux derniers thésaurus. Les organisations et les responsables ne disposent pas toujours de procédures et de directives formelles pour la mise à jour de ces outils lexicaux.

La mise à jour d'un thésaurus implique d'abord la création de nouveaux descripteurs. À la question concernant le nombre de descripteurs ajoutés sur une base annuelle, les réponses révèlent une situation pour le moins étonnante, puisque la moitié des outils s'enrichissent annuellement d'un maximum de 50 nouveaux descripteurs et l'autre moitié de plus de 300 nouveaux descripteurs. On peut s'interroger sur les causes de cette disparité dans des outils qui, en matière de contenu conceptuel, devraient être assez similaires. Les gestionnaires de thésaurus ne peuvent cependant pas dire avec certitude quelle proportion du lexique actuel avait été établi à la fin de la première, de la troisième ni de la cinquième année d'existence du thésaurus et à quel moment le taux de croissance des vocabulaires a ralenti et a finalement atteint son niveau actuel. Bien que nous sachions que le nombre de termes nécessaires à l'indexation d'une collection atteint toujours un sommet à partir duquel il n'y a plus beaucoup d'ajouts qui doivent être faits, les données obtenues ne nous ont pas permis de savoir où se situait ce sommet dans les collections d'images en mouvement.

Analyse lexicale

Idéalement, c'est le contenu intégral des différents outils lexicaux qui devrait être analysé. Dans les faits, compte tenu du très grand nombre de termes présents dans la plupart des outils langagiers observés, une telle démarche s'avère impossible. Les observations préliminaires qui suivent sont basées sur un échantillon constitué de tous les termes commençant par les lettres F, I, et R dans les sept outils lexicaux (thésaurus, listes de mots-clés et de vedettes-matières) dont nous avons pu obtenir copie. Les trois lettres de l'alphabet que nous avons retenues ont été choisies au hasard parmi la quinzaine de lettres qui servent d'initiale à un minimum de 900 et à un maximum de 5 000 mots dans la langue anglaise⁶.

Les chiffres, les noms propres de personnes et d'organisation et les titres (livres, chansons, films, etc.) ont été retirés des listes originales. Les termes qui restaient ont été combinés en une liste unique énumérant 2 292 termes distincts. De ce nombre, 1 858 (81 %) représentent des entités concrètes et 434 (19 %) des notions abstraites. Ces 434 descripteurs seraient plutôt utiles lors d'une indexation aux niveaux iconographique et symbolique.

Le Tableau 7 montre qu'une très large proportion des termes (1 680/2 292 ou 73 %) ne paraît que dans un seul outil, alors que sept termes seulement (0,3 %) sont présents dans les sept outils langagiers observés.

Cette observation a de quoi surprendre si on considère que le contenu des collections visuelles gérées par les organisations participantes est similaire. De plus, elle affaiblit notre hypothèse de départ suggérant que le nombre de termes nécessaires à la description des collections visuelles est limité et que ces termes seront les mêmes dans les collections d'images décrivant des objets et des situations de la vie quotidienne.

Cependant, avant de tirer des conclusions définitives, il nous reste encore à analyser de plus près le contenu des sept lexiques témoins que nous avons constitués. Il est plus que probable que les synonymes s'y retrouvent en très grand nombre et plus particulièrement dans la liste des 1 680 termes qui n'apparaissent qu'une seule fois dans le lexique intégré. La réduction du vocabulaire en fonction du

nombre de concepts différents représentés par les termes des sept langages d'indexation s'impose. De plus, il est également probable que des regroupements supplémentaires seront possibles lorsque les termes représentant des notions identiques ou similaires, mais à des niveaux hiérarchiques différents, auront également été identifiés.

Conclusion

Bien que nous n'ayons pu atteindre tous les objectifs que nous nous étions fixés, notre étude nous a permis de confirmer qu'il existe toujours une grande disparité dans les méthodes et les outils utilisés pour la représentation du contenu dans les collections d'images en mouvement.

Les données que nous avons pu recueillir confirment qu'en dépit de l'absence fréquente de personnel formé aux méthodes bibliothéconomiques de gestion documentaire, et malgré les pressions exercées dans un environnement très concurrentiel, les organisations participantes arrivent à repérer l'information voulue dans des délais raisonnables, mais ceci grâce surtout à la technologie qui permet de chercher de plus en plus rapidement. On ne sait pas cependant si ce qui est offert à l'utilisateur est toujours ce qu'il y avait de plus pertinent dans la base de données. Même dans les organismes publics où l'on gère des collections d'images en mouvement, les compressions budgétaires de la dernière décennie ont malheureusement entraîné un relâchement des contrôles au niveau des processus d'analyse et de la gestion des langages documentaires, avec les résultats anticipés quant à la qualité du repérage.

Tous les gestionnaires de collections qui nous ont offert leur collaboration se sont montrés intéressés au projet et aux hypothèses proposées. Tous ont également trouvé intéressante et réalisable l'idée d'un thésaurus unique pouvant servir de base à l'indexation de collections d'images en mouvement décrivant des

6. Les calculs ont été effectués dans les dictionnaires suivants: *Dictionnaire français-anglais*. 1990. Nouvelle édition enrichie. Paris: Larousse; *Harrap's New Shorter French and English Dictionary*. 1978. High Holborn, London: Harrap & Co.; *Robert-Collins dictionnaire français-anglais, anglais-français*. 1987. Nouvelle édition. Paris: Dictionnaire Le Robert.

objets et des situations de la vie quotidienne. Nous étudions présentement la possibilité de créer et de mettre à l'épreuve un tel thésaurus en nous inspirant entre autres des résultats de l'analyse lexicale des langages d'indexation auxquels nous avons maintenant accès.

Sources consultées

Armitage, L. H., and P. G. B. Enser. 1997. Analysis of user need in image archives. *Journal of Information Science* 23 (4): 287-299.

Association française de normalisation (AFNOR). 1981. *Règles d'établissement des thésaurus monolingues*. Z-47-100. Paris: AFNOR.

Brown, P. et al. 1996. The Democratic indexing of images. *The New Review of Hypermedia and Multimedia* 2: 107-120.

Cawkell, A. E. 1992. Selected aspects of image processing and management: Review and future prospects. *Journal of Information Science* 18: 179-192.

_____. 1993a. An Introduction to image processing and picture management. *Journal of Document and Text Management* 1 (1): 53-63.

_____. 1993b. Developments in indexing picture collections. *Information Services and Use* 13 (4): 381-388.

Guiraud, P. 1960. *Problèmes et méthodes de la statistique linguistique*. Paris: Presses universitaires de France. 145 p.

Hudon, M. 1994. *Le Thésaurus: conception, élaboration, gestion*. Clé en main. Montréal: ASTED. 220 p.

Hudon, M., J. M. Turner and Y. Devin. 2000. How many terms are enough? Stability and dynamism in vocabulary management for moving image collections. In *Dynamism and stability in knowledge organization: Proceedings of the Sixth International ISKO Conference*. Wurzberg, Germany: Ergon, p. 333-338.

Jackaniz, D. W. 1999. Reviews. *The American Archivist* 62 (1): 188-191.

Panofsky, E. 1955. *Meaning in the visual arts: Papers in and on art history*. Garden City, New York: Doubleday Anchor Press. 364 p.

Ramsey, M. C. et al. 1999. A Collection of visual thesauri for browsing large collections of geographic images. *Journal of the American Society for Information Science* 50 (9): 826-834.

Rasmussen, E. M. 1997. Indexing images. *Annual Review of Information Science and Technology* 32: 169-196.

Rorvig, M. E. et al. 1999. The NASA image collection visual thesaurus. *Journal of the American Society for Information Science* 50 (9): 794-798.

Seloff, G. A. 1990. Automated access to the NASA-JSC image archives. *Library Trends* 38 (4): 682-696.

Shatford, S. 1986. Analysing the subject of a picture: a theoretical approach. *Cataloging & Classification Quarterly* 6 (3): 39-62.

_____. 1994. Some issues in the indexing of images. *Journal of the American Society for Information Science* 45 (8): 583-588.

Turner, J.M. 1990. Representing and accessing information in the stockshot database at the National Film Board of Canada. *Canadian Journal of Information Science* 15 (4): 1-22.

_____. 1998. *Images en mouvement: Stockage-repérage-indexation*. Sainte-Foy: Presses de l'Université du Québec. 102 p.

_____. 1999. *A typology for visual collections*. Adresse URL: <<http://www.asis.org/Bulletin/Aug-99/turner.html>> (Page consultée le 3 février 2001).

Turner, J.M., M. Hudon and Y. Devin. 2000. Text as a tool for organizing moving image collections. In *Dimensions of a global information science: Proceedings of the 28th Annual Conference of the Canadian Association for Information Science (... , Edmonton, Alberta, 28-30 May 2000*. Adresse URL: <<http://www.slis.ualberta.ca/CAIS2000/turner.htm>> (Page consultée le 3 février 2001).

Van Slype, G. 1987. *Les langages d'indexation: conception, construction et utilisation dans les systèmes documentaires*. Paris: Editions d'Organisation. 277 p.

Yee, M. M. 1993. The Concept of work for moving image materials. *Cataloging & Classification Quarterly* 18 (2): 33-40.

ANNEXE 1. Liste des organisations invitées à participer au projet

- Academy of Motion Picture Arts & Sciences (Beverly Hills, CA)
- American Film Institute (Los Angeles, CA)
- Archives and Collection (Universal City, CA)
- Archives of Ontario (Toronto, ON)
- Canadian Broadcasting Corporation (Toronto, ON)
- 20th Century Fox (Beverly Hills, CA)
- Channel One (Los Angeles, CA)
- Chisholm Archives (Toronto, ON)
- Cinémathèque (Montréal, QC)
- CNN (Atlanta, GA)
- Film Reference Library (Toronto, ON)
- Fox News (New York, NY)
- International Image (Toronto, ON)
- Industrial Light and Magic (San Rafael, CA)
- Media Archive (San Francisco, CA)
- MGM Studios (Santa Monica, CA)
- Milestone Film and Video (New York, NY)
- MTV (New York, NY)
- NBC News Archives (New York, NY)
- New York Public Library (New York, NY)
- NT Audio (Santa Monica, CA)
- Office national du film (Montréal, QC)
- Prelinger Archives (San Francisco, CA)
- School of Cinema (Los Angeles, CA)
- Société Radio-Canada (Montréal, QC)
- The Image Bank (New York, NY)
- Thirteen/WNET (New York, NY)
- TV Ontario (Toronto, ON)
- UCLA Film & Television Archive (Los Angeles, CA)
- UCLA School of Cinema (Los Angeles, CA)
- V tape (Toronto, ON)
- Warner Bros. Hollywood (Hollywood, CA)
- WGBH (Boston, MA)