Using Audio Description for Indexing Moving Images

James M Turner* and Emmanuël Colinet**

*École de bibliothéconomie et des sciences de l'information, Université de Montréal, CP 6128, succursale Centre-ville, Montréal, QC H3C 3J7, <james.turner@umontreal.ca>

**Caisse de dépôt et placement du Québec, 1000, place Jean-Paul Riopelle, Montréal, QC H2Z 2B3, <ecolinet@lacaisse.com>

James M Turner is a professor at the École de bibliothéconomie et des sciences de l'information, Université de Montréal. He received his PhD in information science from the University of Toronto in 1994. His research activities are focussed on storage and retrieval of still and moving images, indexing images, metadata for digital images in a networked environment, and preservation of digital images. He teaches in the areas of multimedia information systems, managing visual and sound information and moving image archives, and preservation of digital information. More information about his professional activities is available from his Web site at http://mapageweb.umontreal.ca/turner/.

Emmanuël Colinet graduated in 2004 from the master's programme in information science at the École de bibliothéconomie et des sciences de l'information, Université de Montréal. He served as one of the research assistants on this three-year project. During his undergraduate studies and his graduate work, he developed expertise in managing digital information, structured documents, markup languages, and managing strategic information. He is now a researcher-analyst in the information service of the Caisse de dépôt et placement du Québec.

Turner, James M. and Emmanuël Colinet. **Using Audio Description for Indexing Moving Images.** *Knowledge Organization, 31(4).* 222-230. 9 refs.

ABSTRACT: This paper includes some of the results of a study that looks at three types of text for automatically deriving shot-level indexing to moving images. Audio description is a voice added to the sound track of moving pictures to provide information for the visually impaired. We analyse two one-hour parts of a television production broadcast as a mini-series in 1997. We compare our results with those of a previous study, which identifies some of the characteristics of audio description and the associated moving image. We found close correspondence among some aspects studied and for other aspects much less correspondence, but for reasons we are able to explain. In addition, in the process of conducting the current study we further developed our methodology and now feel that it is a mature method for analysing audio description text as a source for generating indexing to the associated moving image.

1. Introduction

As moving images gradually become recognised as valuable documentation in the broader world of information management, the question arises of whether access to their content is adequate. Description of the more complex nature of the production of moving image documents as compared to text documents is provided for by tools such as the *Anglo-American Cataloguing Rules* and the *Functional Requirements for Bibliographic Records*, but what of subject access to them?

We note that subject access to moving images is often provided by the use of tools created for books, such as subject headings and classification schemes. These cover completed productions such as films or

television series. However, deeper analysis of the content is rare, although films available in DVD format now typically include some form of access to the main divisions of the film, sometimes called "chapter headings." With few exceptions, there is no equivalent of a back-of-the-book index to help film researchers, artists, students, or other users find shots of specific persons, objects, or events in a moving image production. The two exceptions worthy of note are stockshot libraries and television newsrooms. Stockshot libraries catalogue and index individual shots, usually outtakes from other productions, because sale of the rights to use these shots in other productions can be very lucrative. Television newsrooms catalogue and index individual shots because they need to be able to retrieve the shots very quickly as news stories build from day to day or as suddenly-occurring events require them to find archived material for immediate broadcast.

The chief obstacle to providing access to the parts of a moving image production such as individual scenes, sequences, shots, or even frames is the cost of indexing. Stockshot libraries and television newsrooms justify the costs in terms of their needs, but how can we provide for other users? It has become clear that we either need to learn to automate indexing at these levels, or we will have no indexing at all. Computer science researchers work on the low-level approach, that is, statistical and algorithmic manipulation of pixels in the images to attempt to extract the semantic content. Information science researchers work on the high-level approach, that is, using humans to study the image and create searchable indexing terms. Also included in the high-level approach is the analysis of texts created for other purposes to derive useful indexing terms from these texts, either automatically or semi-automatically.

Much text is created in the process of making and editing moving image productions, and all such text is potentially useful as source material for deriving indexing to moving images. Here we report some of the results of a study which looks at three types of texts we believe are probably the richest sources for automatically deriving shot-level indexing to moving images. The three types of text are closed-captioning for the hearing-impaired, audio description (a technology for the vision-impaired), and shooting scripts broken down for the shoot. In this paper we look at some of the results related to the audio description part of the study.

Audio description consists of adding to existing productions such as movies and television programmes a voice that describes the picture for the blind and the vision-impaired. Much of the information vision-impaired people use to decode the action in moving images comes from the dialogues and narration, but the audio description text tries to provide other essential visual information that cannot be gotten from the existing voice tracks. Audio description elements are inserted into available sound spaces in the sound track. For purposes of analysis, we introduced the concept of "episodes" of audio description in our previous work in this area. What constitutes an episode?

> Ideally, an episode would be the description of a particular shot and the action that takes place therein. However, the correspondence between the various elements is not usually so straightforward, and some logical decision about where to divide the chunks needs to be made on a case-by-case basis (Turner 1998).

In general, we mean to say that an episode describes a discrete piece of information about the action.

The production we used to gather data for this study is entitled *Cher Olivier,* directed by André Melançon, produced for and distributed by Avanti Ciné Vidéo Inc. in 1997. "Based on the life of the great Quebec comic Olivier Guimond, this series recounts the highs and lows of his career and personal life" (QAV 2004). The production was broadcast as a mini-series in five one-hour parts on television in the spring of 1997 and later that year won eight *Gémeaux* awards (iFrance 2004). *Cher Olivier* is a French-language production and was the only one for which we were able to gather all the pieces for our study: the text of the closed-captioning, the text of the audio description, and the breakdown of the shooting script.

We analysed the first two one-hour parts of the mini-series, the only ones for which audio description was available. The series was broadcast in one-hour parts, but the cassettes we were provided do not include the commercials that were broadcast, so that the running time of each part of the production we analysed is actually 45 minutes. The audio description was made by a small describing company called Studio Vox, which was active for a few years in Montréal in the 1990s. In our previous study (Turner 1998), we looked at the first half-hour of three productions, a film (*Jurassic Park*), a television documentary (*This old pyramid*), and a drama scripted for television (episode 10 of Agatha Christie's *Poirot*).

The description for these productions was made by WGBH in Boston.

2. Background

Results obtained in our previous work provide a strong theoretical basis for the notion of using text created in the processes of pre-production, production, and post-production as source material for automatically deriving shot-level indexing to moving images. People who have the task of describing the content of images name the objects, persons, or events they see in the picture (Turner 1994a, 1994b, 1995) but few tools are available to manage this information (Hudon, Turner and Devin 2000). There is a rather direct correspondence between the information content of images and the words used to describe them, whether these words come from professional indexers, from associated texts such as descriptions in print catalogues or news databases, from descriptors provided by viewers, or from a number of other sources. Although people use a variety of terms to describe the same thing, the top terms emerge rather obviously, are consistent from one context to another, and provide the most obvious subject indexing for the shots because they are the terms named most often. There is already substantial use of "recycled" texts in existing information systems for managing moving images, and in the context of the ongoing move toward a digital production environment, we can imagine the day when most of the pieces of useful source text will be found in multi-module production databases that will accumulate textual information throughout the entire production cycle. Our research focusses on the general problem of how to exploit such a rich mine of useful indexing terms buried in large volumes of text by developing ways of automatically identifying these terms, generating the indexing, and matching it to individual shots, sequences, scenes, or other parts of a moving image production.

One of the findings of our 1998 study was that because audio description text is inserted into the production in available sound spaces, the image being described is not necessarily being shown at the same time the associated audio description text is being recited. In addition, sometimes an episode of description refers to a number of shots, often in rapid succession. We concluded that the best performance in deriving indexing automatically in this context would be to use both the shot and the sequence as units of documentation, but that the question would need further study in order to arrive at any firm conclusions about it (Turner 1998, 115).

Building on this work of developing appropriate theoretical approaches for shot-level indexing of moving images, in the part of the present study reported here we aim to consolidate the findings of our previous work on audio description by comparing the results of the two studies. Because we conducted the present study using a French-language production, we hope the results will also help us further our research goals of developing automated bilingual and multilingual shot-level indexing for moving images.

3. Methodology

Because complete details of the methodology used in this project are reported elsewhere (Turner and Colinet 2005), here we give only a brief overview of it, followed by the methodological details of the part of the study reported in this paper. In order to conduct our study, we first had to identify a production for which we could obtain the source texts that would provide the data for the three kinds of analysis we wished to conduct: subtitles, audio description, and a script broken down for the shoot. A Montréal studio, Avanti Ciné Vidéo, kindly provided us with a printout of its MovieMagic database for *Cher Olivier,* which constituted the shooting script broken down. The paper text of the subtitles was provided by another local company, the Centre national de sous-titrage PST. A third local company, Studio Vox, provided the text of the audio description, also in paper format, as well as videocassettes, which included the audio description sound track. Since only the first two parts of *Cher Olivier* were described for the vision-impaired, we limited our study to these.

A number of FileMaker Pro databases were structured and built, to cover various aspects of the two parts of *Cher Olivier* we studied. The databases relevant to the part of our study concerned with audio description cover:

1. the text of the audio description with the associated timecode;
2. the audio description parcelled into episodes;
3. the relationships between the episodes of audio description and the shots to which they refer (these are Excel files); and,
4. shot numbers and a textual description of the action of each shot

There is a database covering each of these aspects for each part of *Cher Olivier*, since the parts were analysed as separate productions.

The videos had been screened as part of our preliminary work, and we screened them again a number of times for various purposes as work on the project progressed. Having taken the audio description data from paper sources, this time we screened the videocassettes in order to harmonise the text of the final video edit with the content of our databases. We adopted the published video version as the official version of the text, and made corrections where needed in our databases to make them conform to the official version.

Another screening was used to identify individual shots, number the shots, write a description of each shot, and build the shot descriptions database to hold all this information. Once this was completed, we chunked the audio description text into episodes and recorded this information in another database, which had been created for this purpose (number 2 in the list above). We were now ready to match the episodes of audio description to the shots.

Building on the method we had developed in our previous study, we constructed spreadsheets using Excel to show the position of the episodes of audio description relative to the shots. This is a schematic representation (figure 1) of whether the audio description text is spoken at the same time the image to which it refers is shown, or whether it is spoken before or after the image is shown.

A	B	C	D	E	F	G	H
Shot number	nth next shot	2nd next shot	Next shot	This shot	Previous shot	2nd previous shot	nth previous shot
521							
522			172				
523							
524							
525				173			
526					173		
527							
528				174			
529				174			
530				174			
531		175					
532			175				
533				175			
534							
535							
536							
537			176				
538				176			
539							

Figure 1: *Detail of an Excel spreadsheet showing positions of the shots relative to the associated episodes of audio description. At shot 522, the audio describer recites episode 172 of the audio description, but this description corresponds to shot 523. At shot 525, the describer starts reciting episode 173, which describes this shot, but there is not enough time and the description continues while shot 526 is shown. Episode 174 covers shots 528-30 and is recited while these shots appear on screen. The describer begins reciting episode 175 at shot 531 and continues during shots 532 and 533. The episode refers to shot 533. Episode 176 is begun while shot 537 appears, and continues during shot 538, to which it refers.*

When the episode of audio description is not recited at the same time the image appears, we count how many shots away from the target shot the audio description episode occurs, and show this on the spreadsheet. The information contained in these spreadsheets (one for each part of *Cher Olivier*) is the basis for calculating the statistical information given in the results. This information expresses the various possibilities of the position of the episodes of audio description relative to the shots (the episode is recited at the same time the shot is shown, it is recited one shot before, two shots before, one shot after, two shots after, and so on).

We also analysed the types of information transmitted by the audio description text. For this stage,

we used the list of types we had developed in our previous study:

- Physical description of the characters
- Facial and corporal expressions
- Clothing
- Occupation or roles of the characters
- Information about the attitudes of characters
- Spatial relationships among the characters
- Movement of the characters
- Setting
- Temporal indicators
- Indicators of proportions
- Decors
- Lighting
- Action
- Appearance of titles
- Textual information included in the image

The last two types do not show in our analysis because there were no occurrences of either of them in the two parts of *Cher Olivier* we analysed. However, we added four new types in order to reflect the content of these productions:

- Description of sound
- Information about the creation and distribution of the production
- Information about a dedication of the production
- Weather

As part of the analysis we did for this study, we also counted the number of occurrences of each type of audio description, which we had not done in our previous work. This count makes a stab at beginning to quantify this type of information, and will contribute, we hope, to developing knowledge about how much information audio describers need to transmit.

4. Results

In our previous study, we analysed the first 27 minutes of each of three productions and in the present study we analyse two entire productions of 45 minutes each. In our previous study we gave the mean number of shots for the first 27 minutes and the mean number of episodes of audio description. In order to compare the data from the two studies, we give first the individual results from each production, and the means from each study (table 1). Then we give a simulation of these results, calculating first as if all productions lasted 27 minutes, then as if they all lasted 45 minutes (table 2).

	Number of shots	Number of episodes	Ratio
This old pyramid (27 mn)	173	53	1:3.26
Poirot (27 mn)	244	107	1:2.28
Jurassic Park (27 mn)	256	197	1:1.30
Cher Olivier part 1 (45 mn)	648	194	1:3.34
Cher Olivier part 2 (45 mn)	576	184	1:3.13
Mean for the first study (27 mn)	224	119	1:1.88
Mean for the second study (45 mn)	612	189	1:3.24

Table 1. *Number of shots and number of episodes of audio description.*

	Number of shots	Number of episodes	Ratio
This old pyramid (27 mn real)	173	53	1:3.26
Poirot (27 mn real)	244	107	1:2.28
Jurassic Park (27 mn real)	256	197	1:1.30
Cher Olivier part 1 (27 mn simulated)	389	116	1:3.35
Cher Olivier part 2 (27 mn simulated)	346	110	1:3.15
Mean for the first study (27 mn real)	224	119	1:1.88
Mean for the second study (27 mn simulated)	368	113	1:3.26
	Number of shots	**Number of episodes**	**Ratio**
This old pyramid (45 mn simulated)	288	88	1:3.27
Poirot (45 mn simulated)	407	178	1:2.29
Jurassic Park (45 mn simulated)	427	328	1:1.30
Cher Olivier part 1 (45 mn real)	648	194	1:3.34
Cher Olivier part 2 (45 mn real)	576	184	1:3.13
Mean for the first study (45 mn simulated)	373	198	1:1.88
Mean for the second study (45 mn real)	612	189	1:3.24

Table 2. *Number of shots and number of episodes of audio description for 27 mn. and for 45 mn.*

We can see from the ratios that either simulation could be used; we give both here for the sake of completeness. Independently of the number of shots, we note that the means for the simulated times of both studies are very close, 119 and 113 episodes respectively for 27 minutes, and 198 and 189 episodes respectively for 45 minutes.

In table 3 we give the characteristics of the audio description for the two productions we looked at in the present study. For the sake of clarity, in this table we present the data slightly differently from the way we did in our previous study. "Part 1" and "Part 2" refer to the corresponding productions of *Cher Olivier.*

	Part 1	Part 2	Mean
Shots accompanied by any audio description at all	54.3	41.0	47.7
Description corresponds to shot on screen	36.4	29.3	32.9
Some of the description corresponds to shot on screen	53.1	39.4	46.3
Description spoken entirely during corresponding shot	8.0	5.7	6.9
Description spoken entirely during corresponding sequence	14.8	9.5	12.2
Description refers to the next shot	11.0	11.5	11.3
Description refers to the previous shot	4.3	2.6	3.5
Description overlaps next shot	3.4	1.9	2.7
Description overlaps previous shot	4.9	5.0	5.0
Description refers to the second next shot	3.5	2.3	2.9
Description refers to the second previous shot	0.6	0.5	0.6
Description refers to the nth next shot	2.6	0.7	1.7
Description refers to the nth previous shot	1.1	1.4	1.3

Table 3. *Characteristics of the audio description. Figures represent the percentage of the shots exhibiting the characteristic.*

The results show that 54.3% of the shots in Part 1 and 41% of the shots in Part 2 are described, for a mean score of 47.7%. In our previous study (the first half-hour of three productions), the mean score was 48.3%. Next (table 4) we give the means for the characteristics measured in both studies.

	Previous study	Present study
Shots accompanied by any audio description at all	48.3	47.7
Some description corresponds to shot on screen	40.9	46.3
Description spoken entirely during the corresponding shot	25.8	6.9
Description refers to the next shot	2.3	11.3
Description anticipates next shot	8.8	2.7
Description refers to the previous shot	1.3	3.5
Description trails from the previous shot	5.9	5.0
Description refers to the second next shot	1.2	2.9
Description refers to the second previous shot	1.2	0.6

Table 4. *Mean scores for comparable characteristics from each of the two studies. Figures represent the percentage of the shots exhibiting the characteristic.*

In the previous study, we noted that the mean figure of 25.8% for the proportion of the shots during which the audio description was recited as the corresponding shot appeared on the screen was derived from figures covering a rather broad range, 11% for *This old pyramid,* 17% for the *Poirot* episode, and 49% for *Jurassic Park.* We further noted that when we added to this calculation the previous and next shots, the share rose to 44.1%, a figure approaching the total percentage of shots with any audio description at all (48.3%). Using the same procedure with the data from the present study, we arrive at 29.4%, a figure still far from the total percentage of shots with any audio description at all (47.7%).

However, because of our observation in the previous study to the effect that the unit of documentation should probably sometimes be the shot, and sometimes the sequence, in the present study we calculated the number of times the description was spoken entirely during the corresponding sequence. In Part 1 of *Cher Olivier* the figure was 14.8% of the time, and in Part 2 the figure was 9.5% of the time, resulting in a mean of 12.2%. Using this calculation might bring our figure up somewhat, but because we did not record the profile of each sequence, we can only speculate on this aspect.

We did notice that some parts of *Cher Olivier* are characterised by sequences of short shots that do not last long enough to include the entire episode of audio description, even if it could be recited at the

time the shot appears on screen. This observation is supported by the data simulation in table 2. The mean number of shots for the previous study is 373, while it is 612 for the present study, but the number of episodes of audio description is almost identical, 198 episodes for the previous study and 189 for the present study.

In our previous study we gave "a summary of the types of information transmitted by the audio description text, as observed during our analysis of the data" (Turner 1998, 114). In the present study we again analysed the data using the same categories, this time quantifying the data by making calculations on the occurrences of each type of information. In table 5 we give the data for Part 1 and Part 2 of *Cher Olivier*, as well as the mean, with figures representing percentages of the episodes of audio description exhibiting each type of information.

Type of information	Part 1	Part 2	Mean
Physical description	5.2	1.1	3.2
Facial and corporal expressions	13.4	21.7	17.6
Clothing	7.7	8.7	8.2
Occupation, roles of the characters	5.7	4.3	5.0
Information about attitudes of characters	32.5	32.1	32.3
Spatial relationships between characters	4.6	4.9	4.8
Movement of the characters	49.5	44.0	46.8
Setting	12.4	18.5	15.5
Temporal indicators	1.5	1.6	1.6
Indicators of proportions	0	0.5	0.3
Decor	6.7	3.8	5.3
Lighting	0.5	0	0.3
Action	28.9	37.5	33.2
Description of sound	0	0.5	0.3
Information about the creation and distribution of the film	1.0	1.1	1.1
Information about a dedication of the film	0.5	0.5	0.5
Weather	0.5	0	0.3

Table 5. *Types of information transmitted by the audio description text. Figures represent the percentage of the episodes of audio description transmitting the type of information. Types of information in italics represent new types not identified in the previous study Note: Results are rounded up if they are ≥0.05 and down if they are <0.05.*

We note that the last four types in our table represent very few episodes. However, we have included them for the sake of completeness, and to indicate the presence of new types of description not found in our previous study. "Description of sound" seems rather curious as a type, since audio description is created for "viewers" who cannot see and who rely entirely on sound to extract the meaning of the content of the production. In the present context (Part 2, episode 49), there are a number of layers of sound being heard at the same time. The action is taking place backstage, but the sound is coming from the action on stage, so the describer felt the need to sort this out for the visually-impaired "viewer." An equivalent in English of the description is " ... in the distance, music is coming from the show taking place on stage ...."

The other new type of description that merits explanation is "Weather." We felt the need to distinguish between seasons and weather. For example, the description may tell the vision-impaired audience that it is winter, but that does not necessarily mean that it is relatively cold, nor that there is snow on the ground. In the present case (Part 1, episode 166), the visually-impaired audience is told that "... it is raining cats and dogs ...."

5. Discussion

The number of episodes of audio description present in the two productions we studied, and how they relate to the corresponding shots, reinforces our observation from our earlier study to the effect that the unit of documentation sometimes needs to be the sequence, not the shot. In contrast, closed-captioning subtitles for the hearing impaired are pinpointed (with some exceptions) to the exact moment the dialogue is recited on screen. If the information system can handle requests based on the closed-captioning text, the database user who remembers even a snippet of dialogue can instantly and directly retrieve the associated shot.

Applying our research results from the two studies to a database management system, ideally the system would process a textual query in the following way: first it would check to see if the audio description is recited at the time the corresponding image is displayed. If so, it would simply retrieve the shot and offer to play it for the user. If not, the system would retrieve the sequence in which the shot occurs and offer to play the sequence. As a refinement, it might show a timeline representation of where the recital of the audio description occurs and where the corresponding shot occurs.

We are rather excited by our results concerning the similarity between the figures for the two studies

regarding the number of shots accompanied by any audio description at all (48.3% in the earlier study, and 47.7% in the present study). We now wonder, of course, whether this is a general pattern with described productions, but only further study could provide an answer to this question. The figures for shots in which some of the description corresponds to the shot appearing on screen are also quite close between the two studies, 40.9% of the time in the first study and 46.3% of the time in the second. These figures, when compared to those for the total number of shots accompanied by any description at all probably reflect the rigour of those responsible for inserting the episodes of audio description into the sound track of the production, in trying to place the description as close as possible to the action.

The discrepancy in the figures for audio description episodes being spoken entirely during the display of the corresponding shot is of great interest, 25.8% of the time in the first study but only 6.9% of the time in the second. We speculate that the type of production may explain this. *Cher Olivier* is biographical in nature, so perhaps there is more emphasis on the human aspects in the editing of the production. As we noted in the results section, some parts of *Cher Olivier* are characterised by sequences of short shots. Many of these are angle-reverse-angle shots where a dialogue occurs. We also noted that sometimes the shots do not last long enough for the entire episode of audio description to be recited entirely while the shot appears on screen. The fact that the mean number of episodes of audio description for the earlier study (198 simulated) very closely corresponds to that of the present study (189 real) supports the idea that our observation on the fast-paced cutting may be a useful explanation for this phenomenon.

We have not yet done any analysis comparing terms in English with terms in French. However, in the course of some previous work we did (Turner and Hudon 2002) we developed a methodology that we could use for doing so. It would undoubtedly be interesting to use the data from the present study to confirm or refute results from previous studies dealing with this question.

Finally, the results of our study further support the notion that audio description text is a useful source for deriving indexing to moving images. Further work is needed to establish the extent of the contribution of this type of text in relation to other types of text, but the results obtained in this study strengthen the idea that this research path is worth pursuing. As awareness improves of the need for providing better access to audiovisual productions for the blind and visually-impaired, more national governments will adopt policies requiring more productions to be described. This in turn will make more audio description text available for use as a source of indexing.

6. Conclusions

Some of the results we obtained lead us to think there may be patterns emerging that are common to all moving-image productions that are described for the visually impaired. As a result, we have taken a step closer to our goal of learning how to recycle the text of these descriptions for the purpose of automatically generating shot-level indexing to the moving image. In further analysis of the data we collected, we will have a better idea of just how much closer we are, but already the data presented in the present report are encouraging.

In the course of this project we further developed the methodology for analysis of audio description as a tool for indexing moving images and now feel that we have a solid basis for comparison of any number of productions. Any further studies undertaken in this area that use this methodology will contribute to consolidating the theoretical basis for using these valuable texts as a source for indexing. Combined with other text sources such as the closed captioning and the production script, to name only two, we are confident that with time we will be able to provide high-quality indexing to moving images, at the shot, sequence, and scene levels.

Acknowledgments

We are grateful to the Social Sciences and Humanities Research Council of Canada for financing this research. We thank Alexandre Delage, Rida Benjelloun, Marc Lemaire, and Patrick Beaulieu for their work on the initial stages of this project. We thank Avanti Ciné Video, the Centre national de sous-titrage PST, and Studio Vox for their generosity in sharing with us the material that made this research possible.

References

Hudon, Michèle, James M. Turner et Yves Devin. 2000. How many terms are enough? Stability and dynamism in vocabulary management for moving

image collections. *Proceedings of the 6th International ISKO Congress, Toronto, 10–13 July 2000, Toronto, Canada,* ed. Clare Beghtol, Lynne C. Howarth, Nancy J. Williamson. Würzburg: Ergon, 333–338.

i (france). 2004. Available at http://www.ifrance.com/teleromans/olivier_fiche.htm (document consulted 2004.11.24).

QAV. 2004. Québec Audiovisuel: QAV.ca – le portail de l'audiovisuel québécois. Available at http://www.qav.ca/FicheFilm/fichefilm-en.php?ID=459 (document consulted 2004.11.24).

Turner, James M. and Emmanuël Colinet. 2005. Using shooting scripts for indexing moving images. *The moving image: the journal of the Association of Moving Image Archivists* [in press 2005].

Turner, James M. and Michèle Hudon. 2002. Multilingual metadata for moving image databases: preliminary results. *L'avancement du savoir : élargir les horizons des sciences de l'information, Travaux du 30e congrès annuel de l'Association canadienne des scicnces de l'information,* éd. Lynne C. Howarth, Christopher Cronin, Anna T. Slawek. Toronto: Faculty of Information Studies, 34-45.

Turner, James M. 1998. Some characteristics of audio description and the corresponding moving image. *Information access in the global information economy: proceedings of the 61st ASIS Annual Meeting, Pittsburgh,, Pennsylvania, October 2429 1998,* vol. 35. Medford, NJ: Information Today, 108-117.

Turner, James M. 1995. Comparing user-assigned terms with indexer-assigned terms for storage and retrieval of moving images: research results. *Proceedings of the 58th ASIS Annual Meeting, Chicago, Illinois, October 9-12, 1995,* vol. 32, 9-12.

Turner, James. 1994a. Determining the subject content of still and moving image documents for storage and retrieval: an experimental investigation. PhD thesis, University of Toronto.

Turner, James. 1994b. Indexing film and video images for storage and retrieval. *Information services & use* 14: 225-236.