UniS

What Are They Talking About? Information Extraction from Film Dialogue

Videsh Lingabavan and Andrew Salway

School of Electronics and Physical Sciences

Department of Computing

CS-06-07

THE QUEEN'S ANNIVERSARY PRIZES

FOR HIGHER AND FURTHER EDUCATION

2002

UniS

University of Surrey

What Are They Talking About? Information Extraction from Film Dialogue

Abstract

The next generation of applications for retrieving and browsing digitised films will require the automatic generation of machine-processable representations of film content. An important aspect of film content is dialogue, but until now it has been neglected by researchers in the field of video content analysis. The aim of this research is to investigate what kinds of information are conveyed by film dialogue, how this information can be extracted automatically and how this information can be integrated with other aspects of film content to support future film retrieval and browsing systems. A corpus of subtitles from 60 films, comprising 516,299 words, with a vocabulary of 20,141 words, was analysed automatically to identify frequently recurring word clusters. Similar word clusters were manually grouped to propose major kinds of film dialogue event, e.g. Statement_of_Future_Intent and Statement_of_Lack_of_Knowledge. An information extraction system was developed to extract information about these kinds of events. It performed with Precision of 75% and Recall of 20%, with respect to a gold standard data set produced by 4 volunteers about a set of 6 evaluation films. A preliminary analysis of the extracted information from the corpus suggests that it could form the basis of inferences about higher-levels of film content, e.g. the current mental state of a character or predicting a character's future action. This research contributes some basic knowledge about film content with regards to the information provided in dialogue, and begins to show how this information may be extracted automatically. The work also gives empirically-grounded insights into the functions of film dialogue which may be of interest to film scholars. For further information about this concept refer to Lingabavan [1], which contains a detailed account of the work outlined, in this report.

1. Introduction

Films contain various information streams including visual content, music, ambient audio and dialogue, which are synthesised by humans to understand the story told by a film. The broad aim of this report is to identify how film dialogue one of these streams can be analysed to close the semantic gap between humans' and computers' understanding of film data. Our objectives are to:

- Review the role of film dialogue (Section 2)
- Gather a corpus of film dialogue for analysis (Section 3.1)
- Identify unusually frequent word clusters in the corpus and interpret these as being indicative of major kinds of film dialogue events (Sections 3.2 and 3.3)
- Develop an information extraction system using templates based around the unusually frequent word clusters (Section 4.1)
- Create a gold standard data set and evaluate the precision and recall of the information extraction system on a new set of subtitles (Section 4.2)

- Assess the potential uses for this data in future film retrieval and browsing systems (Section 4.3)

The goal of this work is to identify a link between film dialogue and film semantics. This is realised by producing a method for identifying exploitable word patterns that occur frequently in film dialogue.

2. Background

The fundamental importance of analysing and structuring video data in order to enable intuitive and effective retrieval and browsing is well understood Dimitrova et al. [2]. In the case of film video data, a number of models e.g. Corridoni et al. [3]; Roth [4]; Vendrig and Worring [5]; Salway and Xu [6] and techniques e.g. Sundaram and Chang [7]; Adams et al. [8]; Zhai et al [9]; Hanjalic and Xu [10]; Salway and Graham [11]; Chan and Jones [12]; Turetsky [13] and Dimitrova [2] have been proposed specifically for the structuring and automated analysis of film content.

However film dialogue has not been previously exploited as a source of information about film content. This is perhaps surprising given what an important role it plays in communicating stories in most films. If a film lacked dialogue this would greatly reduce the audience's ability to understand the story. In the silent movie era, the lack of sound was compensated for by more expressive action on-screen; furthermore text was displayed on-screen to clarify scenes which were difficult to understand.

It is perhaps important then to consider in more detail what functions dialogue has.

2.1 Functions of Film Dialogue

Film dialogue serves several purposes in communicating information to the audience. Kozloff [14] postulates nine functions of dialogue, which she organised into two categories. The first category, of six functions, revolves around dialogue used to communicate narrative.

1. Anchorage of the diegesis and characters
2. Communication of narrative causality
3. Enactment of narrative events
4. Character revelation
5. Adherence to the code of realism
6. Control of viewer evaluation and emotions

The second category is concerned with, "aesthetic effect, ideological persuasion, and commercial appeal" Kozloff [14]; it comprises of three functions:

7. Exploitation of the resources of language
8. Thematic messages/authorial commentary/allegory
9. Opportunities for "star turns"

Kozloff acknowledged that these nine functions are not a definitive list, but argued that they do cover the main purpose and usage of film dialogue. Here we outline the nine functions, adding our own examples.

2.1.1 Anchorage of the Diegesis and Characters

Anchorage of the diegesis and characters is dialogue used to communicate the story of the film or used to introduce characters. For example, films can introduce a character or place by using the line of dialogue; 'this is…'. This function of dialogue is commonly used in theatre where scenery is generic and dialogue is used to encourage the audience to imagine the backdrop. The technique is also used in films to allow for the introduction of new characters and for new places to be identified to the audience. For example Kozloff [14] outlines that the character Dorothy in *The Wizard of Oz* says, 'Toto, I don't think that we are in Kansas anymore,' indicating to the audience that the current location is no longer Kansas.

2.1.2 Communication of Narrative Causality

Communication of narrative causality is a function which allows a character to explain to the audience their actions. Dialogue is used to explain visual information when it is not totally clear from the visuals alone. For example, consider a character saying that they are chasing after someone for revenge, and explaining how they are going to catch up with their enemy.

2.1.3 Enactment of Narrative Events & Character Revelation

Enactment of narrative events is a function of dialogue where the "most common event is the disclosure of a secret or of crucial information, information vital to the plot, whose revelation poses some risk or jeopardy" Kozloff [14]. This function is best characterised by most of the antagonists from the *James Bond* series of movies. Where approaching the climax of a *Bond* film, James Bond would be captured by the antagonist who would explain his motives and current/future plans to James Bond. This allows the audience to understand the plot and current/future actions, whilst also informing James Bond how to defeat the antagonist.

The character revelation function of dialogue is used in a variety of ways to allow the audience an insight into a characters personality/psychology. This is achieved either directly by a character revealing a piece of information about themselves, or indirectly by another character testifying to another character's personality. For example, in *Con Air* when U.S. Marshal Vince Larkin says 'we got an ally on that plane' about the main character Cameron Poe, it reinforces Poe's personality and psychology.

2.1.4 Adherence to the Code of Realism

Adherence to the code of realism uses dialogue to establish a sense of realism within the film. For example a character walking into an office will greet other background characters. The dialogue from this scene may serve little purpose to the plot, however if scripted or improvised appropriately, it creates the illusion that what is being portrayed on-screen is realistic.

2.1.5 Control of Viewer Evaluation and Emotions

Dialogue may control the audience's emotions and opinions, e.g. by making a character seem innocent and the audience sympathise with them. In *Layer Cake,* the main character is a criminal and is shot at the end. However, despite this character's criminal past, the audience feels sorry for this character because the character talks about wanting to leave the life of crime but not being able to.

2.1.6 Dialogue to Aesthetic Effect, Ideological Persuasion, or Commercial Appeal

Dialogue in this category does little to express the plot in a film. Thematic messages, authorial commentary and allegory, are when a speech in dialogue holds a meaning and morale for the audience. This function of dialogue has been most exploited by propaganda films. According to Kozloff [14] this function of dialogue is criticised for seeming transparent and appearing to preach to the audience. For commercial reasons, star turns may impact on a film's dialogue as, for example Kozloff [14] points out, Robin Williams in *Mrs Doubtfire* shows off his range of impressions during a telephone call.

2.2 Speech Acts

Though it did not deal with film dialogue explicitly, Searle's work on speech acts may well be relevant in this context. Searle [15] classifies speech into categories that specify how it functions for the speaker. For instance if, 'how are you?' was asked, this represents a question where the speaker requires information, at least in the case when the question was genuine.

Searle [15] classifies general speech into 3 main categories:

Uttering words = performing utterance acts

Referring and predicating = performing propositional acts.

Stating, questioning, commanding, promising, etc. = performing illocutionary acts.

Searle also offers a fourth category - perlocutionary acts. A perlocutionary act is speech which has an effect on a hearer's action, thoughts or beliefs. So it is used to persuade someone by arguing or to scare somebody by talking to them.

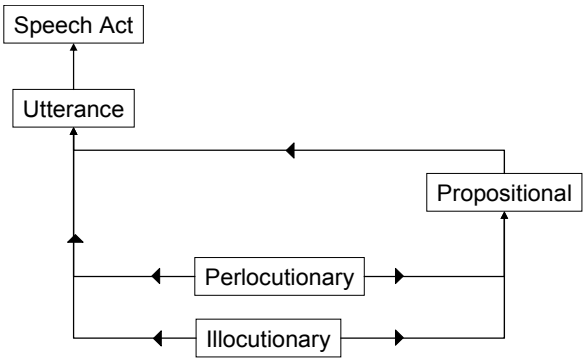

Figure 1: Hierarchy of speech act categories

Figure 1 shows how the different categories of speech acts link together. Searle states that a propositional act cannot exist by itself, rather it comprises of illocutionary and perlocutionary acts. An utterance comprises of either a perlocutionary, illocutionary or propositional act or any combination of these.

An utterance is using speech directly, uttering words, that is to say, exclaiming 'I promise' to someone does not require any more

explanation, rather it is just an utterance of the words. That goes to serve the speaker to just utter the words and mean exactly what they say.

The illocutionary act is similar to a perlocutionary act where the speaker intends the hearer to react, whether it is listening or performing an act. However an illocutionary act is only classified as successful if the audience recognises the speaker's desire to do a certain act. While a perlocutionary act is only successful if the hearer actually performs the speaker's desires, for example, if the hearer becomes scared after the speaker has talked to them. Thus essentially a perlocutionary act is linked to actions being performed by saying something, while an illocutionary act merely states what on the speaker's mind.

Searle offers sub-categories of the illocutionary acts which are:

Request
Assert, state (that), and affirm
Question
Thank (for)
Advise
Warn
Greet
Congratulate

Kozloff and Searle both offer evidence that dialogue at times, serves a deliberate purpose.

3. What Kinds of Information are Conveyed by Film Dialogue?

3.1 Film Dialogue Corpus

Subtitles from 60 films were gathered to form our film dialogue corpus; see Appendix 1 for a listing of the films. Ten films were selected from each of six film genres, according to their classification on www.imsdb.com <http://www.imsdb.com/>: Action, Comedy, Drama, Romance, Science Fiction and Thriller. The subtitles were obtained by using a program called *Subrip* [16]. This program allowed the subtitles to be extracted from a DVD into a text file, with lines of dialogue separated by different speaker and time code.

Film Statistics per Genre

Genre	Average Run Time	Average Vocab	Average Total Words of Dialogue	Average Words of Dialogue Per Min
Action	118	1317	7735	63
Comedy	105	1742	10226	97
Drama	113	1471	9483	83
Romance	110	1487	9422	86
Science Fiction	118	1405	7373	62
Thriller	108	1353	7391	68
Average	112	1462.5	8605	76.5

Table 1: Film Statistic per Genre

Table 1 shows an apparent variance between the amount of dialogue in different film genres. It seems that some genres rely more on visual impact, and less on dialogue, e.g. Action, Science Fiction and Thriller. Comparatively Comedy, Drama and Romance films seem to rely more on dialogue having higher word per minute rates.

Table 1 also highlights that comedy films seem to have the highest average words per minute. This may be due to comedy films needing to tell jokes via dialogue.

3.2 Identification of Word Clusters

The approach taken here is to identify unusually frequent word clusters in the film dialogue corpus, on the grounds that these should be indicative of main film dialogue events.

This approach is similar in spirit, though simpler in execution, than the method developed by Vassiliou [17] for the analysis of film screenplays and audio description scripts.

3.2.1 Method

The corpus analysis comprised of three main stages:

1. Identify keywords that occur unusually frequency in the film dialogue corpus compared to general language. This was done automatically using the SL/GL ratio Ahmad et al. [18] calculated using *System Quirk* [19].

2. Identify frequent word clusters that include the keywords. A cluster is a group of words which appear next to each other. This was done automatically using a program called *Antconc* [20].

3. The frequent word clusters were grouped together manually and, on the basis of the groupings, main film dialogue events were inferred.

3.2.2 Results

The 20 most unusually frequent words that occurred at least 100 times are listed in Table 2.

Top 20 Keywords

Rank	Keyword	Frequency
1	don't	3265
2	gonna	1086
3	didn't	736
4	hey	699
5	f**k	622
6	guy	557
7	s**t	533
8	uh	359
9	gotta	329
10	doesn't	300
11	wanna	300
12	hi	298
13	ain't	228
14	isn't	221
15	huh	214
16	a*s	208
17	wasn't	192
18	kid	190
19	haven't	183
20	wouldn't	165

Table 2: The top 20 keywords

Table 2: The 20 most unusually frequent words occurring at least 100 times in the film dialogue corpus. This list contains many slang words and swearing which is consistent with contemporary dialogue and film dialogue. It is interesting to note that 8 out of 11 verb forms in this list are in the negative; this may indicate that film dialogue contains a lot of negative statements.

Clusters including at least one of these 20 words were extracted automatically: Table 3 lists the top 25 clusters occurring at least 50 times in the corpus.

Top 25 Keyword Cluster

Freqency of Cluster	Cluster
1004	i don't
565	don't know
411	you don't
350	i don't know
262	I don't
258	don't you
226	the f**k
224	i didn't
181	don't want
172	don't have
155	don't think
142	gonna be
130	why don't
119	this guy
115	we don't
114	f**k you
108	what the f**k
107	you're gonna
106	why don't you
104	i gotta
102	don't worry
102	you gotta
96	i don't want
93	don't be
93	i don't think

Table 3: Top 25 keyword clusters

It is interesting to note from table 3 that only 11 of the keywords produced clusters which occur 50 times or more. The keyword 'don't,' produced the most clusters being 37 of the total 54 significant clusters.

From these clusters, we inferred two main film dialogue events.

1. Statement_of_Lack_of_Knowledge: based on the cluster 'I don't know.' This event represents when a character lacks knowledge about something.

2. Statement_of_Future_Intent: based on the clusters 'I'm gonna,' 'I gotta' and 'I wanna.' This event refers to a character openly stating their intentions to perform an action in the future.

These cluster and events were chosen for further analysis due to their frequency of their occurrence and belief they might yield potentially interesting findings.

4. Extraction of Information about Film Dialogue Events

To test the validity of these postulated events, an information extraction system was developed and a gold standard evaluation set was elicited.

For each event type, an information extraction template and associated heuristics were specified, based on word patterns seen from the previously extracted word clusters. The information extraction program, which was implemented in Perl, receives an input of subtitle files and outputs a file containing the extracted information formatted for loading into a database table.

4.1.1 'Lack_of_Knowledge'

The occurrences of the word cluster 'I don't know' were analysed to identify patterns, Figure 2. This led to the specification of an IE template, Figure 3 and a heuristic, Figure 4.

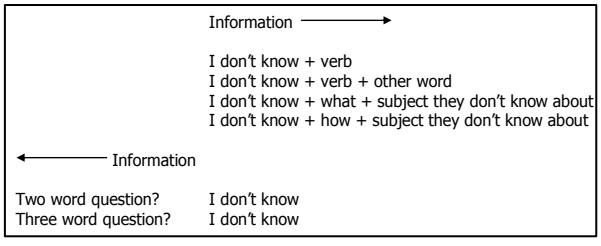

Figure 2: Information location about the event's occurrence

Figure 2 indentifes the word patterns which were used to identify the information to store from a cluster's occurance. Figure 2 represent a few of the most frequent cluster occurance, it does not include all the potential ways in which this cluster occurs in a pattern of words.

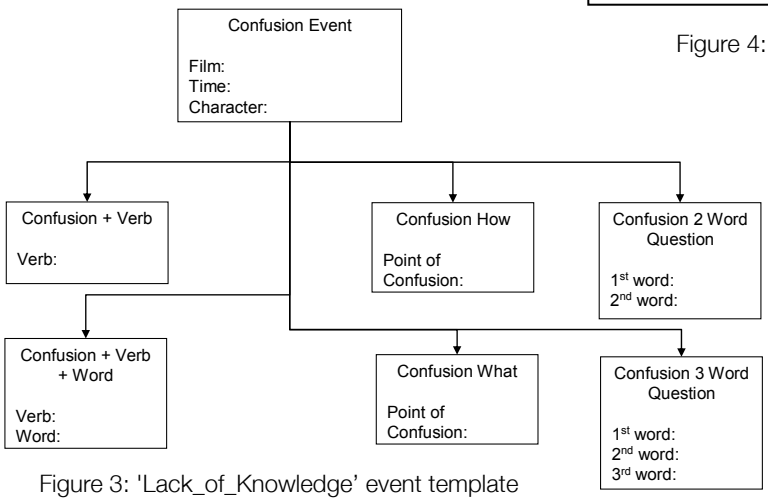

Figure 3: 'Lack_of_Knowledge' event template

Using the template outlined in figure 3 which was created according to the pattern identification in figure 2. A heuristic was created in order to be able to program an information extraction system to recognise the occurrences of the event.

```
For each line in the input file
Check if that line contains the cluster 'I
don't know'

If 'I don't know' exist
{
        If previous line contains a question
        mark
        {
                Is the question comprised of
                2 or 3 words?
                {
                        Store instance along
                        with question words
                }
        }
        If cluster occurs with only a verb
        {
                Store instance along with
                verb
        }
        If cluster occurs with only a verb
        and one word
        {
                Store instance along with
                verb and other word
        }
```

```
        If cluster occurs with how
        {
                Store instance along with
                words after how
        }
        If cluster occurs with what
        {
                Store instance along with
                words after what
        }
}
Next input line
```

Figure 4: 'Lack_of_Knowledge' event heuristic

4.1.2 'Future_Intent'

This event encompasses three sub-types, as depicted in the information extraction template in Figure 6, and the IE heuristic in Figure 7. From the cluster information it was determined that all the clusters forming this event all share the same word pattern figure 5.

Information →

Future Intent + verb
Future Intent + verb + other word
Future Intent + verb + 1st word + 2nd word

Figure 5: Information location about the event's occurrence

Figure 5 shows the main word patterns for the Future_Intent template. This event's word patterns were straight forward where the event occurs followed by a verb and a number of other words.

Due to the clear-cut word patterns the template for the future action event is simpler. As opposed to the Lack of knowledge event which generated many word patterns.

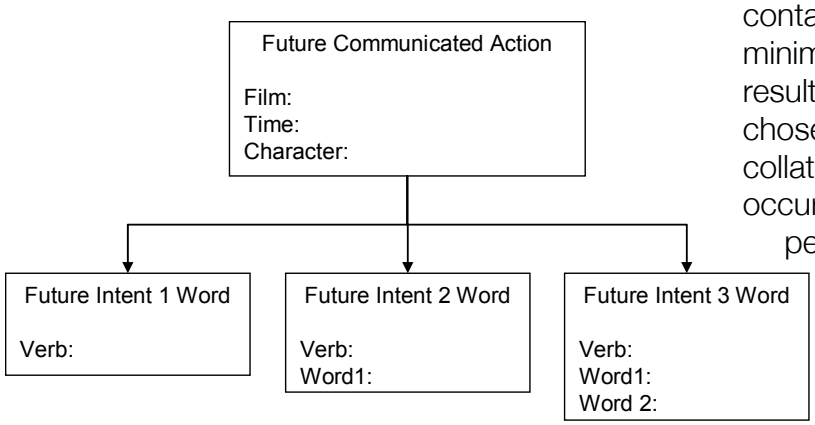

Figure 6: Future intention template

Thus the heuristic for this event is shown in figure 7.

```
For each line in the input file
Check if that line contains a future
intent cluster

Check if (I wanna or I gotta or I'm gonna)
ignore case of the letters
{
        Check if (after cluster there is
        only a verb)
        {
                Store that word as the
                future indented action
        }
        Check if (after cluster there is a
        verb followed by a word)
        {
                Store the future intent
                event
        }

        Check if (after cluster there are
        two words after a single verb)
        {
                Store the future intent
                event
        }
}
Next input line
```

Figure 7: 'Future_Intent' event heuristic

4.2 Evaluation

4 volunteers were asked to identify instances of the concepts of Lack_of_Knowledge and Future_Intent events in 6 scenes from 6 previously unanalysed films, totalling 51 minutes in length. These scenes contained the events identified by the research, but also contain a lot of other dialogue to minimise the effects of biasing the results from the scenes being chosen. The responses were collated by classifying an event occurrence when two or more people identify a particular word pattern to be an occurrence, this formed a gold standard data set. This was used to calculate Precision and Recall statistics for our information extraction system.

From the gold standard created the information extraction system was benchmarked according to Precision and Recall statistics using the set of previous unanalysed films, Table 4.

Set of Previous Unanalysed Films

Film Name	Scene Length	Genre
Batman Begins	04:16	Action
Coach Carter	04:13	Drama
Constantine	14:13	Thriller
Millers Crossing	06:51	Thriller
Wedding Crasher	08:15	Romance
Without a Paddle	13:38	Comedy
Total	51:28:00	

Table 4: Set of previously unanalysed films

Table 5 shows that the precision and recall for the information extraction system have been brought down by the poor performance of the Statement_of_Lack_of_Knowledge. The extraction of the Statement_of_Future_Intent event did however perform quite well, with the information extraction system managing to identify most of the events occurrences.

Information Extraction System Evaluation

Events	Testers Identified	System Identified	Correctly Identified Events	Incorrectly Identified events	Precision	Recall
Confusion	25	1	1	0	100.00%	4.00%
Future Intent	19	11	8	3	72.73%	42.11%
Total	44	12	9	3	75.00%	20.45%

Table 5: Results from Information Extraction System

The information extraction system shows that it is possible to automatically identify events from film dialogue, with the best evidence coming from the precision and recall for the future intent event. However, it is evident that not all event occurrences were identified. This was caused by the testers identifying word clusters which were not the ones identified by initial research as occurrences of events. The complicated nature of the Statement_of_Lack_of_Knowledge event caused further losses of event identification. Whilst initial results of this technique are promising, further tweaking of the information extraction system is required to improve overall precision and recall.

4.3 Making Inferences about Higher-Levels of Film Content?

The focus of the current work was on the discovery of common film dialogue events, and their validation through the development and evaluation of an information extraction system. However, at this point, we may speculate a little about how information about these events could be used as the basis for making inferences about higher-levels of film content.

It appears to us, though a proper statistical analysis still needs to be done, that the frequency of certain film dialogue events varies between different film genres. On the one hand, this observation could be useful to the development of film genre classification systems, on the other hand it may be of interest to film scholars interested in how genres work.

With regards to the Statement_of_Future_Intent event it may be possible to use information about instances of it in order to predict subsequent on-screen action, i.e. when a character says 'I'm gonna to XXX' then subsequent scenes depict them doing XXX. Care must be taken here because strictly speaking the dialogue is expressing what the character thinks they are going to do, which may not be what actually transpires. There may also be instances when the activity is not depicted, perhaps because it is enough for the audience to know it happens from the dialogue, or perhaps because the character was lying.

We also noted word patterns such as “I'm not gonna to...” and “I don't wanna...” which may or may not mean that the mentioned activity happens, but which does give us an insight into the character's mental state at that point. Also potentially useful for getting information about a character's mental state are the Statement_of_Lack_of_Knowledge events, though care must be taken to distinguish anguished “I don't know”'s from more casual ones. Our interest in learning more about character's mental states is motivated from our belief that they are a fundamental aspect of story telling in films, see Salway and Graham [11] and references therein.

5. Conclusions

This research contributes some basic knowledge about film content, with regards to the information provided in dialogue, and begins to show how this information may be extracted automatically. The work also gives empirically-grounded insights into the functions of film dialogue which may be of interest to film scholars. The main contribution, detailed more thoroughly in Lingabavan [1], is the discovery of word clusters in film dialogue that are unusually frequent in comparison with general language. We believe that groups of these clusters can be interpreted as expressing common film dialogue events, of which we elucidated two in this report. The two events that we postulated were validated to some extent (Statement_of_Lack_of_Knowledge less so) by the development and evaluation of an information extraction system: the heuristics of which followed directly from the patterns of word clusters. For the development of film content analysis applications the information extraction heuristics could be significantly refined, and extended, with an analysis of word patterns around the clusters already examined, and with the identification and analysis of further word clusters in a larger corpus. Our preliminary analysis of the information that we can extract from film dialogue, along with the premise that film dialogue plays a central role in filmic storytelling, encourages us that taken in conjunction with other data streams, information about film dialogue will enable the automatic analysis of high-level film content.

References

1. Lingabavan, V., (2006) "**Extracting Information from Film Dialogue to Access Video Data,**" MSc Dissertation, Dept. of Computing University of Surrey.

2. Dimitrova, N., Zhang, H-J., Shahraray, B., Sezan, I., Huang, T., & Zakhor, A., (2002) "**Applications of Video-Content Analysis and Retrieval**," *IEEE Multimedia*, Vol. 9(3) July, p 42-55.

3. Corridoni, J., Bimbo, A., Lucarella, D., & Wenxue, H., (1996) "**Multi-Perspective Navigation of Movies**," *Journal of Visual languages and Computing,* Vol. 7(4) December, p 445-466.

4. Roth, V., (1999), "**Content-based retrieval from digital video**," *special issue on Content-Based Image Indexing and Retrieval*, Vol. 17(7) May, p 531-540.

5. Vendrig, J., & Worring, M., (2002), "**Systematic Evaluation of Logical Story Unit Segmentation**," *IEEE Transactions on Multimedia*, Vol. 4(4) December, p 492-499.

6. Salway, A., & Xu, Y., (2005), "**Navigating Stories in Films,"** Dept. of Computing Technical Report CS-05-04, University of Surrey.

7. Sundaram, H., Chang, S.F., (2002), "**Computable Scenes and Structures in Films**," *IEEE Transactions on Multimedia*, Vol. 4(4) December, p 482- 491.

8. Adams, B., Dorai, C., & Venkatesh, S., (2002), "**Towards Automatic Extraction of Expressive Elements for Motion Pictures: Tempo**," *IEEE Transactions on Multimedia*, Vol. 4(4) December, p 472-481.

9. Zhai, Y., Rasheed, Z., & Shah, M., (2005), "**Semantic classification of movie scenes using finite state machines**," *IEE Proceedings - Vision, Image and Signal Processing*, Vol. 152(6) December, p 896-901.

10. Hanjalic, A., & Xu, L.Q., (2005), "**Affective video content representation and modeling**," *IEEE Transactions on Multimedia*, Vol. 7(1) February, p 143–154,

11. Salway, A., & Graham, M., (2003), "**Extracting Information about Emotions in Films**," *Proceedings of the 11th ACM Conference on Multimedia*, 4th-6th November, p 299-302.

12. Chan, C., & Jones, G., (2005) "**Affect-Based Indexing and Retrieval of Films**," *Proceedings of the 13th annual ACM international conference on Multimedia*, 6th-11th November, p 427-430.

13. Turetsky, R., & Dimitrova, N., (2004) "**Screenplay Alignment for Closed-System Speaker Identification and Analysis of Feature Films**," *In Proceedings IEEE ICME* Vol. 3 June, p 1659- 1662.

14. Kozloff, S., (2000) "**Overhearing Film Dialogue**", USA California: University of California Press.

15. Searle, J., (1976) "**Speech Acts an essay in the philosophy of language**", UK London: Cambridge University Press.

16. Subrip, (2006), "**Subrip**," [program], available from: http://zuggy.wz.cz/, Last Accessed: 16 June 2006.

17. Vassiliou (2006), "**Film Content Analysis: a text-based approach**," PhD Thesis, Dept. of Computing, University of Surrey.

18. Ahmad, K., Gillam, L., & Tostevin, L., (2000) "**Weirdness Indexing for Logical Document Extrapolation and Retrieval (WILDER)**," In (Eds.) E.M. Voorhees and D.K. Harman. The 8th Text Retrieval Conference (TREC-8). Washington: National Institute of Standards and Technology, 717-724.

19. System Quirk (2006), "**System Quirk**," [program], available from: http://www.computing.surrey.ac.uk/SystemQ/, Last Accessed: 16 June 2006

20. Anthony, L., (2006) "**AntConc**", [program], available from: http://www.antlab.sci.waseda.ac.jp/, Last Accessed: 16 June 2006

Appendix: List of Films used in Dialogue Research

Name	Year
A Guy Thing	2002
Addicted to Love	1997
Along Came A Spider	2001
Along Came Polly	2004
American History X	1998
Anger Management	2003
Back To The Future	1985
Basic	2003
Bridget Jones Diary	2001
Blade	1998
Blade Trinity	2004
Catch Me if You Can	2002
Changing Lanes	2002
Charlie and the Chocolate Factory	2005
Citizen Kane	1941
Collateral Damage	2002
Coyote Ugly	2000
Cube	1997
Daddy Day Care	2003
Equilibrium	2002
Evolution	2001
Four Feathers	1977
Four Weddings and a Funeral	1994
Ghost World	2001
Gothika	2003
Heavy Metal	1981
Heist	2001
High Fidelity	2000
Just Married	2003
Last Samurai	2003
Leon	1994
Mad Max 1	1979
Meet Joe Black	1998
Mimic	1997
Mission Impossible	1996
My Big Fat Greek Wedding	2002
Mystery Men	1999
Narc	2002
O Brother Where Art Thou	2000
Ocean Eleven	2001
Ocean Twelve	2004
Panic Room	2002
Pi	1998
Pulp Fiction	1994
Rear Window	1954
Reservoir Dogs	1992
Scarface	1983
Signs	2002
Star Wars Ep 3	2005
Star Wars Ep 4	1977
Swordfish	2001
Terminator 3	2003
The 6th Day	2000
The Beach	2000
The Cable Guy	1996
The Core	2003
The Fifth Element	1997
Top Gun	1986
When Harry Met Sally	1989
Zoolander	2001

Department of Computing
University of Surrey
Guildford, Surrey
GU2 7XH UK

Tel: +44 (0) 1483 683133
Fax: +44 (0) 1483 686051
E-mail: a.salway@surrey.ac.uk
www.surrey.ac.uk

UniS
University of Surrey