Analysing Film Content: A Text-Based Approach

Andrew Vassiliou

Submitted for the Degree of
Doctor of Philosophy
From the
University of Surrey

UniS

Department of Computing
School of Electronics & Physical Sciences

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

July 2006

Summary

The aim of this work is to bridge the semantic gap with respect to the analysis of film content. Our novel approach is to systematically exploit collateral texts for films, such as audio description scripts and screenplays. We ask three questions: first, what information do these texts provide about film content and how do they express it? Second, how can machine-processable representations of film content be extracted automatically in these texts? Third, how can these representations enable novel applications for analysing and accessing digital film data? To answer these questions we have analysed collocations in corpora of audio description scripts (AD) and screenplays (SC), developed and evaluated an information extraction system and outlined novel applications based on information extracted from AD and SC scripts.

We found that the language used in AD and SC contains idiosyncratic repeating word patterns, compared to general language. The existence of these idiosyncrasies means that the generation of information extraction templates and algorithms can be mainly automatic. We also found four types of event that are commonly described in audio description scripts and screenplays for Hollywood films: Focus_of_Attention, Change_of_Location, Non-verbal_Communication and Scene_Change events. We argue that information about these events will support novel applications for automatic film content analysis. These findings form our main contributions. Another contribution of this work is the extension and testing of an existing, mainly-automated method to generate templates and algorithms for information extraction; with no further modifications, these performed with around 55% precision and 35% recall. Also provided is a database containing information about four types of events in 193 films, which was extracted automatically. Taken as a whole, this work can be considered to contribute a new framework for analysing film content which synthesises elements of corpus linguistics, information extraction, narratology and film theory.

Key words: Film Content, Information Extraction, Film Corpora, Collocation, Semantic Gap, Text Analysis, Corpus Linguistics.

Email: a.vassiliou@eim.surrey.ac.uk; andro8472@hotmail.com

WWW: http://www.computing.surrey.ac.uk/

Acknowledgments

First and foremost I would like to thank Dr Andrew Salway for all his guidance and support, and the long meetings that were always useful, interesting and never long enough.

I would also like to thank Professor Khurshid Ahmad for the opportunity to do a PhD and for his help with all things computational/corpus linguistics and local grammars.

To Craig Bennett, Yew Cheng Loi, Ana Jakamovska and James Mountstephens: thanks for the welcome distractions and for allowing me to bounce ideas off you. A big thanks to the TIWO project group members, Yan Xu and Elia Tomadaki for their help and guidance. Also to Haitham Trabousli, who spent a great deal of time and patience explaining new concepts to me, a heart felt thank you. Paschalis Loucaides, Bouha Kamzi and James William Russell Green for their proof-reading efforts, thanks guys. To all the people who took the time to do the evaluation needed in this work: thank you! I would also like to thank Dr. David Pitt for his time, effort and many ideas.

A special thank you to Darren Johnson, Jamie Lakritz, Dominic Fernando, Vidsesh Lingabavan and Matthew Knight who implemented my work and showed me it has a use…

I would also like to warmly thank my parents, sister and family in general for their moral support and guidance, I am the person I am today because of them - Thank you!

Abbreviations

A list of abbreviations used throughout the thesis report.

AD *Audio Description*- A standard for describing what is happening in films to the visually impaired through a separate soundtrack complementary to the film's soundtrack.

SC *Screenplay*- The directing script of a Hollywood film appears in many forms, first, second, third, early, final drafts, post production and pre-production scripts etc.

FSA *Finite State Automaton* and in-plural *Finite State Automata*- a way of formally representing events and states. In our case they are used to represent restricted phrases of language that contain paths that can be followed to make a coherent phrase.

LSP *Language for Special Purpose*- A language used by experts in a specific domain that exhibits jargon and idiosyncrasies in that domain.

LG *Local Grammar*- A constrained set of words that could concurrently be used in a specific statement or context.

IE *Information Extraction* - A technology dedicated to the extraction of structured information from texts to fill pre-defined templates.

COL *Change of Location* Event- Developed as a template for IE of information pertaining to characters changing location from one area to another in a film.

FOA *Focus of Attention* Event - Developed as a template for IE of information pertaining to characters focussing their attention on other characters and on objects in a film.

NVC *Non-Verbal Communication* Event- Developed as a template for IE for when a character is non-verbally communicating with a body part.

ScCh *Scene Change* Event – Template that provides information of when a film's scene is changing.

1 Introduction

The amount of digitised film and video data available to us through online repositories of movies, DVDs and online video search engines is substantial. The challenge of automatically extracting meaningful video content that a user requires is still unresolved and thus there are minimal services provided to accommodate queries related to high-level semantics (e.g. character behaviours and interactions, emotions, plot points). The concept of a semantic gap, where there is a lack of interpretation between what a computer system can extract from video data and human users' interpretation of that same data, exists, that is non-trivial.

For visual data:

> "The semantic gap is the lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data have for a user in a given situation." Smeulders et al. [[89] pg 8]

The semantic gap is particularly pertinent for video content, where the information that can be extracted by existing audio-visual analysis techniques, from video data, and a user's interpretation of the video data content are not in agreement. This work aims to help bridge that semantic gap, for film content, through the analysis of texts that represent film: film scripts.

The digitisation of film data has made film more easily accessible, processable and easier to analyse than ever before. However, many issues still exist concerning the annotation, extraction and retrieval of semantic content from film data, which is still a non-trivial task. A user may want to query high-level concepts in film content, such as: emotional scenes (happy, sad) and atmospheres (suspense, dark) in a film, a specific action or dialogue scene, a scene in an ice hockey game where a specific player started a fight or an automatic summary of a sitcom they missed last week. The bridging of the gap between what the viewer *interprets* in film content and to what degree a computer system can match that viewer interpretation is of great concern in this work. Thus, there is a need for machine-processable representations of film content.

The objectives of this research are: to *investigate what information audio description scripts and screenplays provide about film content*, which is accomplished through the collocation analysis of certain highly frequent open class words in audio description script and screenplay representative corpora; *to explore how machine-processable representations of film content can be extracted from the texts*, achieved by the development and evaluation of an information extraction system

based on collocation analysis results; and *to consider how these representations enable novel applications for analysing and accessing digital film data*, where a database of film content information from audio description scripts and screenplays, gathered by the information extraction system, was used as a base to outline novel applications. This work goes on to describe in detail how we have considered the systematic exploitation of collateral texts to film (film scripts) as a novel approach to bridging the semantic gap for film content.

1.1 Films, Narrative and Collateral Texts

This section outlines key terminology and definitions that are used throughout the thesis report and serves to introduce representations of narrative and film content models and frameworks. A UML model of narrative is also introduced.

1.1.1 Key Definitions: *Film*

A film, otherwise known as a movie, moving picture, motion picture or 'flick' is defined as: "[a] sequence of photographs projected onto a screen with sufficient rapidity as to create the illusion of motion and continuity" [112] or "a form of entertainment that enacts a story by a sequence of images giving the illusion of continuous movement" [113]. A film provides us with the illusion of movement and sound and suspends our disbelief to provide an entertaining, immersive experience for the viewer. Film presents us with a story or narrative that is re-enacted through the interaction of characters. It can be argued that the cause and effect relationship in a film is governed largely by the characters' actions that cause events to change. Characters are said to be the agents of cause-effect relationships. [15]

Films conform to a film structure or film grammar. Films can be split into episodes, acts, scenes, events (events span many shots but may also only be one shot long) and shots; a shot being the smallest unit of a collection of frames [15], [63], [64]. Film can be considered to have a 'film grammar' which film directors commonly follow. "Like any other natural language this [film] grammar has several dialects but is more or less universal" [73]. Edit-effects such as fades, wipes, transitions between shots and scenes can be seen as *punctuation* in film grammar [22].

In the case of film it is made up of a *plot* and characters serve as the actors who act out a *story*. The story of a film can be considered the 'original' story a director/writer had in mind which considers only the important events that occur for a film's narrative to progress. "The term *plot* is used to describe everything visibly and audibly presented in the film before us," [[15] pg 92].

Film tells a story and has existents (characters, objects and locations) and events. In this work *Film events* describe events common to most films, e.g. action, dialogue, suspense and emotional

events. Its plot can be manifested in many ways and can be considered its discourse. When we talk of 'going to the movies' we almost always mean we are going to see a narrative film– a film that tells a story [15]. Hence, film can be considered a narrative and thus we can argue it conforms to narrative theory.

1.1.2 Key Definitions: *Narrative*

The Oxford Online Dictionary defines narrative as: "*An account of a series of events, facts, etc., given in order and with the establishing of connections between them; a narration, a story, and account.*"[115]. Narrative is described as a sequence of causally connected events, organised in space and time [18], [55].

It is argued that each narrative has two parts, the actual story and how it is communicated (discourse). Chatman, a leading *narratologist*, distinguishes between a story (*histoire*) and discourse (*discours*). "In simple terms, the story is the *what* in a narrative that is depicted and the discourse the *how*..." [[18] pg 19]. Story is considered "...the narrative in chronological order, the abstract order of events as they follow each other." [55] The story is "the set of events tied together which are communicated" or "what in effect happened" [Ibid]; in an event *something occurs*. A set of events can be considered *what happened* in a certain sequence and forms the idea of the *story*. *Events* are *actions* or things that just happen (*happenings*) and are associated with characters that either make the event happen, *agents*, or are impacted by an event, *patients*. In between events, characters, objects and settings exist in a *stasis* or some sort of permanent or temporary situation called a state. A representation can be seen in Figure 1.

Chatman describes discourse as the other necessary component of a narrative, "a discourse (*discours*) that is the expression, the means by which content is communicated." Discourse carries with it the notion of plot which is "how the reader becomes aware of what happened," that is basically, the "order of the appearance (of the events) in the work itself." Discourse is an abstract concept with many manifestations. "If discourse is the class of all expressions of story, in whatever medium possible be it (natural language, ballet, program music...[etc.]) it must be an abstract class containing only those features that are common to all actually manifested narratives." [Ibid pgs 27-28].[1]

[1] A UML model of narrative based on Chatman [18] and Herman's [45] work can be seen in Figure 9.

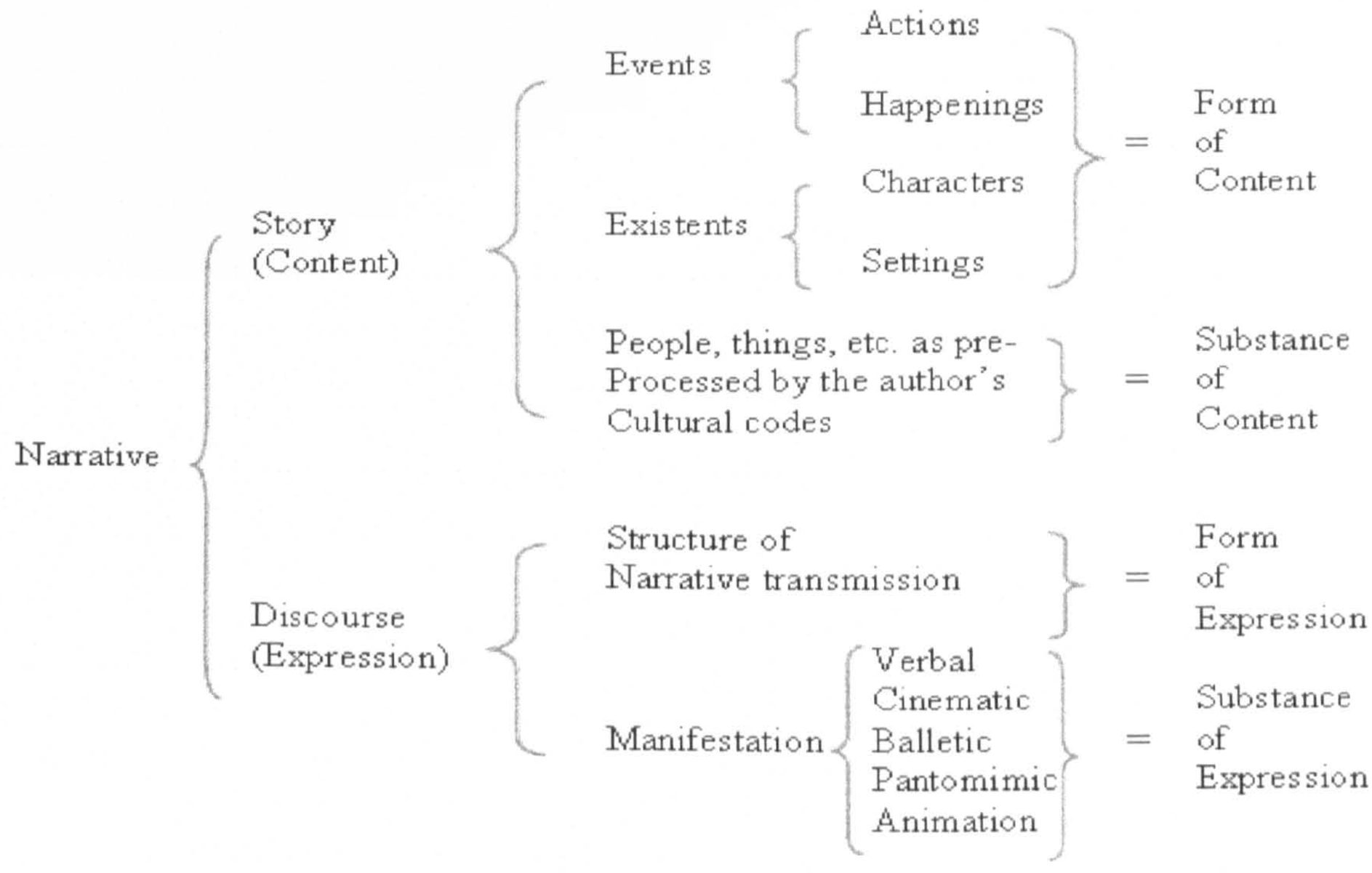


Figure 1 A proposed diagram by Chatman, representing narrative in terms of 'Story' and 'Discourse', and incorporating a new conceptual level of 'Content' and 'Expression' and their respective 'Form' and 'Substance'. Reproduced from Chatman [18], pg 26.

Thus a plot is, "explicitly presented events and added non-diegetic material", and story, "inferred events, back-story and associated story material."[15], [18]. As an example, many film *plots* can be written for one underlying *story* e.g. remakes of films, adaptation of films to theatre and vice-versa. NB Although in general language and in film theory the words *narrative* and *story* are used in slightly different ways, in this report they are used interchangeably.

Narrative theory tries to explain how these constituent elements form a narrative, where a story is more that the sum of the constituent narrative elements. It also describes how narrative is transmitted. Humans understand media at a high-level of content, that is, we understand the message or story it tells us. For instance a surveillance video of a man committing a crime would instantly be differentiated from a video of a cat chasing a mouse, through its context and meaning to us as human beings. Some types of video data, e.g. films, are made to entertain and others are made to educate, e.g. documentaries. Video data, in effect, is used for narration.

1.1.3 Key Definitions: *Collateral Texts*

Collateral texts are texts that transcribe, accompany or somehow describe the contents of multimedia artefact be that an image, a piece of audio or in the case of this work, video content. Srihari states that images do not appear [only] in isolation, but rather with accompanying text which is referred to as *collateral text* [93]. Collateral texts can be considered an archive of texts,

produced by experts, which elucidate these multimedia artefacts in depth [82]. This work deals with collateral texts associated with film, specifically *Audio Description* (AD) and *Screenplay* (SC) film scripts. Audio description refers to an additional narration track for blind and visually impaired viewers of visual media, which "weaves a scripted commentary around the soundtrack of [video data] exploiting pauses to explain on-screen action, describe characters, locations, costumes, body language and facial expressions"[2]. Examples of *audio description scripts* can be seen as follows:

Excerpt from "The English Patient" [164] RNIB Audio Description Script

01:05:20 They investigate his belongings, fingering photographs and paintings pushed between the pages of a book.

01:05:28 A simple piece of cloth is laid over his face with holes cut for his mouth and eyes. A dried palm leaf covers the cloth.

01:05:37 Swathed in blankets and strapped to a stretcher he is carried across the sand-dunes on the back of a swaying camel. Dimly he sees the outline of his rescuers through the fine mesh of the dried palm leaf which protects his eyes from the glare of the sun

01:05:55 He breathes with difficulty

01:06:00 An army Red Cross camp

Excerpt from "The English Patient" ITFC Audio Description Script

]]></ScriptText>

<AudioDescriptionScript>

<Block>18</Block> <ThreadValidity>FFFFFFFFFFFFFFF</ThreadValidity>

<OverallInTime>10:05:37.07</OverallInTime>

<OverallOutTime>10:05:56.24</OverallOutTime><ScriptText><![CDATA[<cue>(10.05.36.00)

<normal>camels' hooves kick up the sand as the shadows lengthen. The Bedouins lead their small caravan up the slope of a dune. Strapped to one camel is the pilot, bound to a rough stretcher, wrapped in cloths like a mummy.

</AudioDescriptionScript>

Excerpt from "High Fidelity" [162] BBC Audio Description script

Title number: 13

In Cue Time(HH MM SS Fr): 10:02:53:15

Out Cue Time(HH MM SS Fr): 10:03:03:01

Duration (MM SS Fr) : Outside, Laura calmly gets into her car. Rob shuts the window and slumps back into his leather armchair, staring blankly into space

Title number: 14

In Cue Time(HH MM SS Fr): 10:03:05:09

Out Cue Time(HH MM SS Fr): 10:03:07:02

---

[2] For more on Audio Description got to http://en.wikipedia.org/wiki/Audio_description [114] and the Audio Description section at [130] http://www.itfc.com/?pid=3&sub=1, accessed 16/05/06.

Duration (MM SS Fr) : he flicks off the music

Screenplays refer to film scripts written by *screenwriters*, which give directional advice to actors and crew on a film set. "A screenplay or script is a blueprint for producing a motion picture. It can be adapted from a previous work such as a novel, play or short story, or it may be an original work in and of itself."[116]. Screenplays come in different versions and drafts e.g. first, second, final drafts, post-production scripts and continuity scripts. In this work all types of film script are analysed, but collectively referred to as screenplays. A sample of a *screenplay* follows.

Excerpt from the "James Bond: The World is Not Enough" [156] Screenplay

GUN BARREL LOGO OPENS ON

GENEVA SWITZERLAND, an unnaturally clean city that melds old Europe with new money of both dubious and legitimate source.

INT. HALLWAY OFFICE BUILDING - DAY

An engraved brass plaque announcing the name of the "private banking institute" within.

INT. PENTHOUSE BANK OFFICE - GENEVA - DAY

JAMES BOND, dressed impeccably as ever, is being FRISKED by three THUGS in Armani suits. They remove A GUN from inside his jacket, a well-concealed knife, a metal case, laying them on the desk that separates Bond and...

LACHAISE, an extremely well-groomed gentleman. Behind him, three floor-to-ceiling windows lead out to a rooftop garden.

LACHAISE

Not the usual Swiss procedure, Mr. Bond,

but you understand, a man in my position...

BOND

Which is neutral, no doubt?

Archaise takes the joke a little tight-lipped. Gestures for Bond to sit

This work, while acknowledging that there are other collateral texts for films, such as plot summaries, subtitles and closed-caption text chooses to use audio description, because it was made available through the TIWO project [96], and screenplays; as these were publicly available for educational purposes and abundant.

1.1.4 Key Definitions: *Machine-Processable Representations*

Representations of video content that can be recognised and processed by computers automatically, without manual human action. One aim in this research is to explore and develop machine-processable representations of video data that capture some aspects of film content related to narrative and use these to allow information extraction of film content.

1.2 Problem Statement: Film Video Data and the Semantic Gap

When viewing a film at the cinema or on the small screen at home, a sense of escapism tends to engulf us. A very good film will grasp our attention, entertain us and help suspend disbelief enough to identify with what is happening on-screen. Everyone will form their own understanding of the film but there will be enough common elements to enable us to discuss the essence of the story with others. This *essence* of the story however cannot, as yet, be understood by a machine or computer system, but we would like it to be. We would like to be able to search through hours of feature film to find one specific scene of an elephant walking through the jungle in India at sunset, to be able to 'view' an automatic summary of an episode of our favourite TV show we missed, to be able to automatically segment scenes of home video or film content, to locate adult or unpleasant video clips to ensure our children do not watch them or be able to retrieve exactly the moment in a scene where a child wearing a red t-shirt scores a goal and is happy. However, this cannot currently be achieved, to that level of detail, and the reason for this is that there is still a semantic gap between what humans want and what computers can provide in terms of interpreting film content.

As mentioned, the semantic gap for film content refers to the lack of coincidence/correlation between what a human views, understands or wants returned from a query, with respect to film content, and what a machine can interpret automatically from a film (audio-visual analysis, text analysis). Thus, the problem of specifying and automatically extracting machine-processable representations of film content to bridge the semantic gap arises. This section explores the difficulties of analysing film content and models that attempt to represent film content and narrative data.

1.2.1 Why is the Analysis of Film Content Difficult?

Narrative can be defined as "a chain of events in cause-effect relationships occurring in time and space" [15]. A good narrative allows for the suspension of disbelief, presents believable characters that we can relate to or loathe and provides engaging storylines. Chatman, in his book "Story and Discourse" (1978) [18], asks what are the necessary components of narrative?

Due to narrative's multiple components, modelling narrative is no trivial task. For instance, narrative can be, and has been, modelled as just a story and its components or a series of plot units [58], but such approaches have, as yet, been too time consuming to manually index and have not captured the *full* essence of the narrative. Also the problem when modelling the concept of *discourse* for narrative is that it is an abstract concept and has many manifestations. Discourse carries with it the notion of plot which is "how the reader becomes aware of what happened". Thus, the issue of how to model narrative remains.

A film is a complicated multi-layered, multi-faceted entity. It tells a story and has existents (characters, objects and locations) and events. Its plot can be manifested in many ways and can be considered its discourse. Therefore, film can be considered a narrative and thus we can argue it conforms to narrative theory. A film is made up of a *plot* and characters serve as the actors who act out a *story*. The story of the film considers *all* events in a film even; Inferred events, back-story and associated story material [18]. The story of a film includes cultural codes and if we consider narrative theory, consists of *events* and *existents*. Events are either acted out by 'actors', i.e. some character does something, or events just happen. These 'happenings' occur through some unseen force or through divine intervention, such as a storm 'happens', an accident or a miracle. "Events… are either *acts* or *actions* in which an existent is the agent of the event, or *happenings*, where the existent is the patient."[18] There is a cause and effect chain that occurs through a film, where events are interrelated and cause other events to follow. It can be argued that the cause and effect relationship in a film is governed largely by the characters' actions in films that cause events to change. Characters' states, whether they are physical, emotional, or cognitive (e.g. making decisions, being puzzled), can be considered the driving force of event changes. Characters are said to be the agents of cause-effect relationships [15]. Thus, the characters' (and objects and locations that they interact with) are integral to a film's story and are collectively called existents.

Film can also be conceived as having layers, e.g. structural and temporal. Structural layers refer to the concept of film structure, as discussed by Metz [63], [64]; films can be split into episodes, acts, scenes, events[3] and shots; a shot being the smallest unit of a collection of frames. Film can be considered to have a 'film grammar', where edit-effects, such as: fades, wipes, transitions between shots/scenes, can be seen as *punctuation* in film grammar.

The temporal layer refers to the length of the film and the concept of *story time* vs. *plot time* (see Figure 2). Since "[t]he story is simply the chronological order of events," [[55] pg 20], story time refers to the set of all events in a film both ones explicitly presented and those the viewer infers. "The term *plot* is used to describe everything visibly and audibly presented in the film before us," [[15] pg 92], "explicitly presented events and added non-diegetic material" Chatman [18]. Thus,

plot time corresponds to the full length of the film. Plot time may be non-linear as in "Memento" [166] and "Pulp Fiction" [173]. Plot time represents a film's discourse and story time represents a film's story.

Figure 2 A diagram depicting the difference between story and plot time in a film taken from Bordwell and Thompson [15] page 93. NB Non-diegetic material refers to material not essential to the plot.

A film has many manifestations in the form of different cinema releases (International vs. domestic), different DVD releases (extended versions, director's cuts, alternate endings), film scripts (continuity scripts, various screenplay draughts and audio description scripts) and other textual manifestations (plot summaries, book adaptations). This makes modelling and choosing plot units (main story units) a difficult *manual* choice.

Thus it can be said that digitised film data is a difficult thing to represent. Even elements of film content that can be extracted automatically from low-level visual features (keyframes), audio features (sound intensity) and textual features (extracting text internal to the video such as sport game scores and subtitles) are not easily extracted with high confidence. Capturing more 'abstract' film elements automatically such as camera movements and effects, dialogue and special effect sequences are also difficult to extract. Films' different modalities lead to many issues when trying to extract semantic information from films and make the crossing of the semantic gap very difficult.

1.2.2 Formalising Film Content

One of the ways in which we can bridge the semantic gap is by creating systems dealing with knowledge at a human and machine level [100]. Film content must be 'understood' by a computer system (be machine processable), not necessarily in the human sense of narrative understanding (which may never happen: "Since movies are the result of an artistic process sometimes causing confusion on purpose, the necessity for human judgement will always remain to some extent" [[103] Pg 493]), but in a representational sense. It is necessary to represent film content, automatically, for a computer/machine to allow us to analyse films and be able to access film content and still retain the film as a whole, or the essence of the film. It is necessary to encode a system's knowledge-base formally (e.g. in an ontology) [[100] pg. 2] and then convert into a representation. Representations of film content that are machine processable may be a step in the

[3] Events span many shots but may also only be one shot long.

right direction for bridging the semantic gap. What follows is research that has attempted to represent film content or narrative with some degree of success.

Corridoni et al. [22] models components of film and organises film information into a formal schema, which is independent of the specific content of a film but mirrors the structure of the film as it is made. Film elements are described in an object-oriented conceptual model in terms of objects, attributes and semantic relationships among objects, this is illustrated in Figure 3. The work here demonstrates that a film and its elements can be modelled and represented based on film grammar and film theory techniques. The conceptual schema treats the elements of a film 'uniformly' in respect to film hierarchy, film edit-effects, film objects and other elements. Once the structure is modelled and manually annotated, this allows users to view a movie from different perspectives, effectively choosing what to watch through 'graph operators of perspective and filters'. This is a step forward for content representation and retrieval for movies in the semantic content domain as the whole film can be represented at multiple levels (low-level features such as lighting, spatial relationships of objects, images, mid-level features such as edit-effects and shots and high-level semantic features such as scenes). Representing a film, and its elements, in this way presents many retrieval possibilities for querying the content of films.

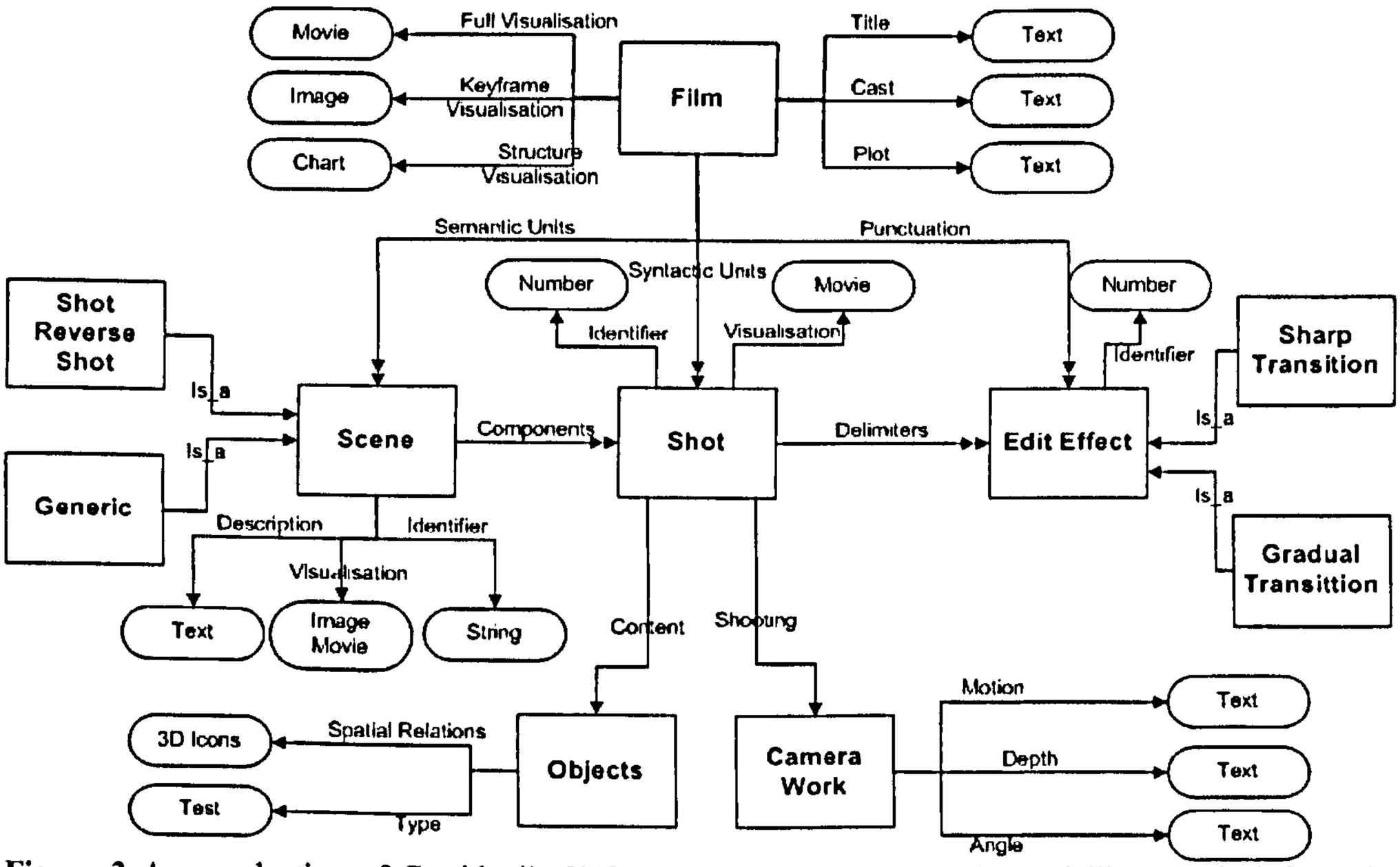


Figure 3 A reproduction of Corridoni's [22] Conceptual Schema Graph, modelling movie informative structure.

Hanjalic et al. [40] present a strategy for segmenting movies into Logical Story Units (LSU). The LSUs are based on events, not single shots, as the natural retrieval unit and correspond to a single event or groups of complex interlinking events both referred to as 'meaningful segments' of film and in [40] are referred to as *episodes* (see Figure 4). In [40] approximations of film episodes are

obtained which they call LSUs. Vendrig and Worring [103] present a Logical Story Unit definition based on film theory. Visual Similarity and dissimilarity of shots is compared to ascertain whether or not a boundary for a Logical Story Unit exists. A temporal distance measurement is taken for a time window (pre defined) to ascertain whether the shots are similar over this time frame. This work provides us with a possible unit for film content, the Logical Story Unit. It is speculated by [103] and [40] that films can be automatically divided into LSUs.

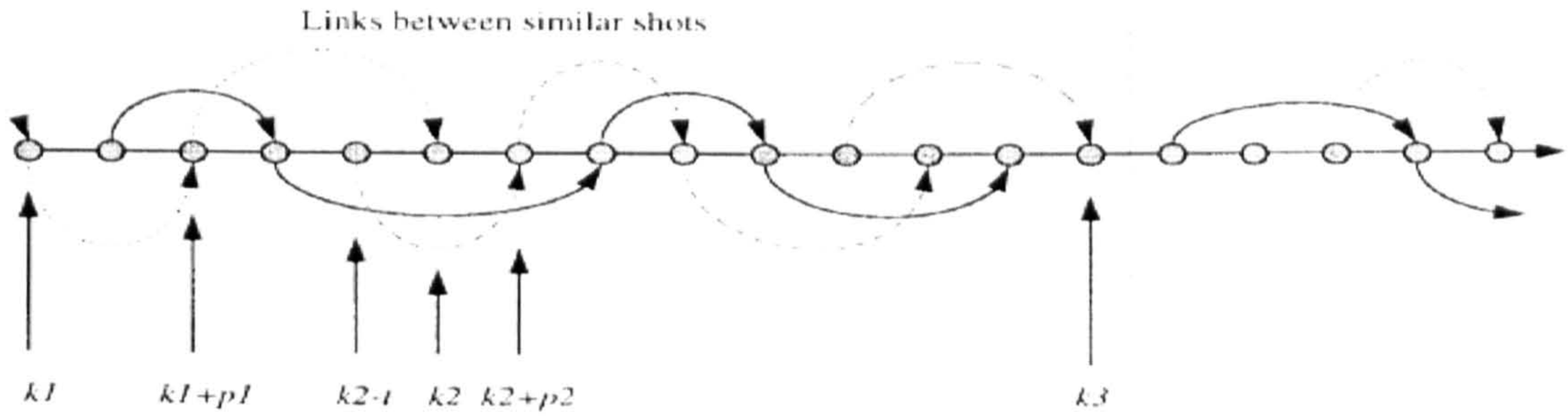


Figure 4 An illustration Logical Semantic Units characterised by overlapping links connecting similar shots. Taken from [40] pg 583.

Another way to represent film content is through a semantic network[4]: graph with nodes that represent physical or conceptual objects and arcs that describe the relationship between the nodes, resulting in something like a data flow diagram. Roth [76] presents a readily explorable framework for content-based video retrieval based on semantic networks. Roth refers to the entities visible in a video as objects. The objects or film elements can be represented as nodes on semantic nets with meaningful links. Each frame, or keyframe, has what are called sensitive regions of 'importance' to the frame or even to the shot as a whole. The semantic links can be built up manually. A system for film content could query on cause and effect as well as individual film elements, which are demonstrated partially in the prototype system (see [76]). However, this approach although providing a semantic network of useful information for a film with the possibility of growing into an ontology[5] (or ontologies) with respect to the film's content has, at present, to be manually populated. This, apart from being laborious and time intensive, relies on human judgement of frames, shots and film elements for each and every node of the film as well as providing relevant links between nodes. Thus, presently to automatically populate a semantic network for film information is beyond the scope of the work.

---

[4] **Semantic Network**: a graph with nodes that represent physical or conceptual objects and arcs that describe the relationship between the nodes [126].
[5] The concept of 'ontology' for video content is covered in Chapter 2, section 2.1.15.

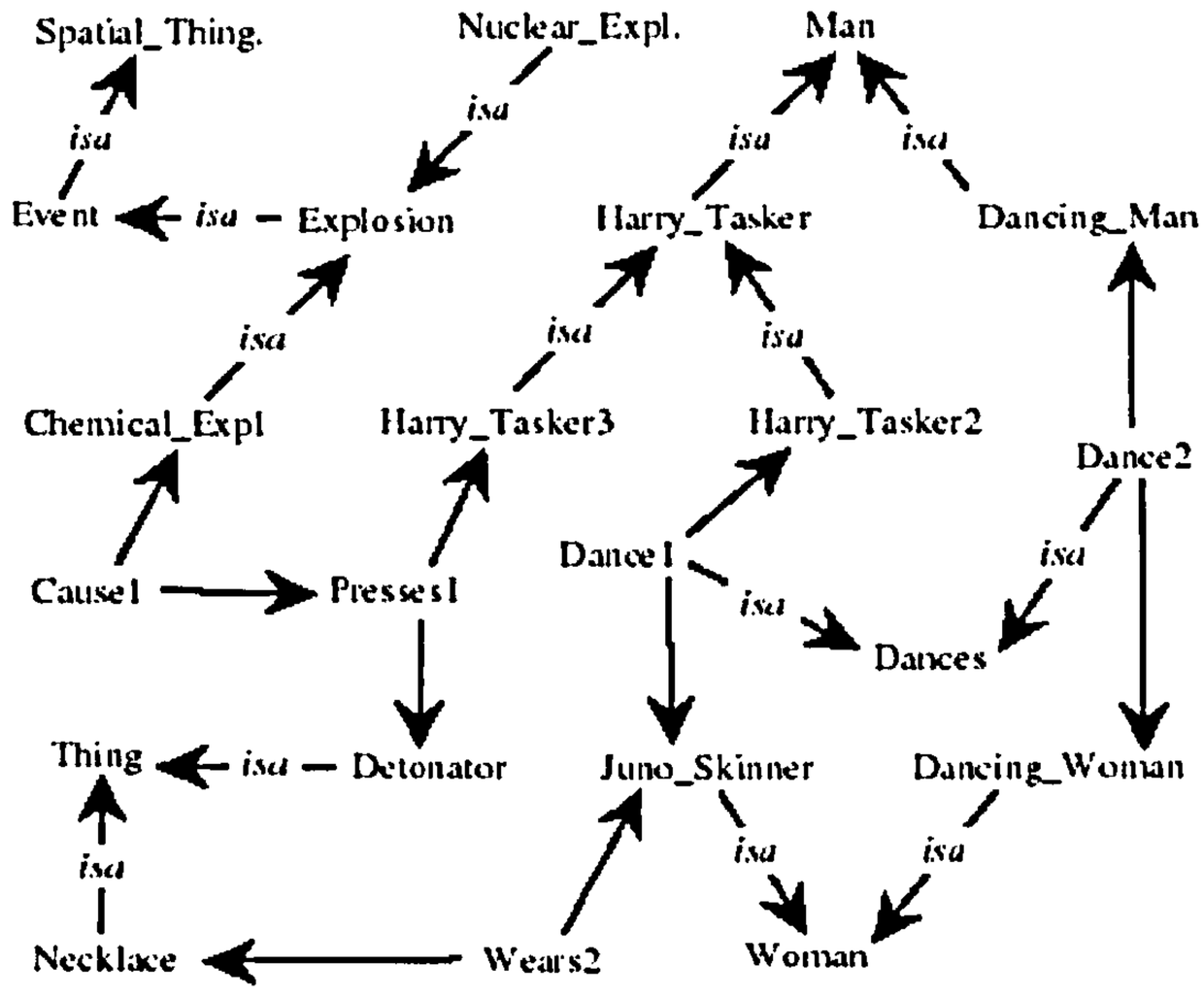


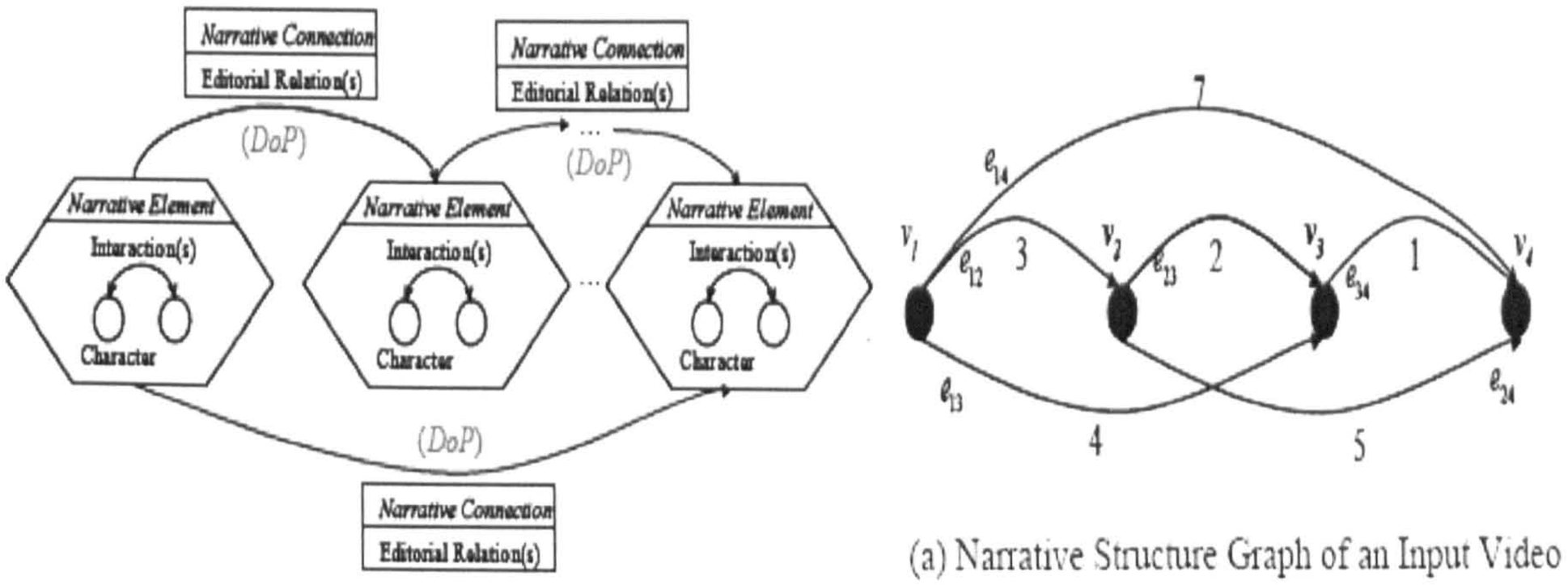
Figure 5 An excerpt of the semantic network representation of the scenes from the motion picture True Lies [158]. Taken from [76] page 537.

Allen and Acheson [7] argue that the best level of analysis for videos is not shot and scene boundaries but plot segments. They adapt Lehnert's [58] notation for plot units, events, mental states and their associated links. Allen and Acheson focus on the structure of stories through events and reactions of characters. This work provides the opportunity to model a story in terms of its important events through plot units. A high concentration of events and mental states can be considered a *plot cluster* and may signify an important event. Other smaller clusters are deleted leaving only the essential plot of the story. The system, Story-Threads, allows plot clusters to be labelled and represented through a story's 'keyframe' which may be an image or video shot and linked with each other to represent the story sequence and the cause-effect relationships between events. This approach captures the essential events of the plot and can be used to summarise the key events in a story. However, it has many limitations. It relies heavily on human judgement to annotate each event and the mental state of characters. This is time intensive and an automatic method is needed to extract such events. Also people's opinions capturing an event may differ and the plot focus excludes other aspects of the story. There are no descriptions of relationships amongst characters and no way to show characters' reactions, assumptions or awareness. Finally, there is difficulty representing non-linearity in stories. This kind of approach has been extended by Xu [107] who has developed a system that represents films in terms of plot units (events/mental states) and devised a querying system that allows the navigation of a film based on characters, their mental states and key film events.

Nack and Parkes [66], [67] describe the AUTEUR system architecture, which can be used to generate comedy videos from existing, manually described, video shots. In terms of film content, Nack and Parkes describe video at a shot level and each shot description is hierarchical,

descending from general features to specific details. The knowledge base in the AUTEUR system contains conceptual structures representing events and actions along with representations of visual features underpinned by a network of semantic fields. The structure of a scene is represented by an event, which is described by its name, number of actors, intentional actions, main actions of actors involved and a link to a higher element within the story structure. AUTEUR also contains hierarchically organised conceptual knowledge about objects, locations, directions and moods. This work provides an insight into how complex a 'joke' is to be understood by a computer system but demonstrates the possibility of a system, with human guided initial conditions and narrative theory guidelines, capable of producing humorous video.

Jung et al. [53] also have a representation of film content. They treat film as a formal system and develop a formal representation of narrative (film). Jung et al. proposes a narrative abstraction model which considers 'narrative elements' basic units, narrative elements such as dialogue and action scenes in films, which can be automatically extracted (see Figure 6 for example). They use facial recognition and shot and scene detection through edit-effects, spacial and temporal changes and rhythm to help identify action and dialogue scenes.



(a) Narrative Structure Graph of an Input Video

Figure 6 Depicts the framework of the narrative abstraction model (left) where *DoP* is a function that determines the progress of a story between two video segments and (right) a narrative structure graph of an input story oriented video where the duration of each vertex v_1-v_4 is measured in minutes. Taken from [53].

1.2.3 Modelling Narrative and Film Content

Reviewing the literature concerning narrative and film theory, and research into how narrative is represented, gave us an insight into the structure and theories of narrative and, by extension, film. The definitions above highlight the complexity of narrative and its constituent elements, such as plot, events, existents, actions, states, happenings and characters. In this work these narrative elements are considered, for film as *film elements*. As discussed, the modelling of the structure and elements of narrative and film is a non-trivial task. After all "a narrative cannot be reduced to the sum of its sentences" [55], there are the issues of coherence and cohesion of the story to

consider as well as deciding which are the important (kernel) events versus the extraneous, unnecessary events (satellite). Human interpretation is also a problem as humans do not always agree on what constitutes an important event and we all conjure up our own image or story essence when viewing a film.

Chatman's analysis of story and discourse accentuates the examination of narrative on *other* levels such as human inferencing, importance of cultural codes, temporal sequencing and structuring of events, narrative's semiotic structure and the importance of form and substance. Ultimately, Chatman's "Story and Discourse" manages to bring together these different elements to form one 'big picture' of narrative. This is why we believe Chatman's book was useful to model and represent narrative concisely. The 'big picture' of narrative seems especially useful for modelling narrative; it tackles diverse areas of narrative and tries to represent them simply. Largely through the aforementioned definitions, and based on Chatman's work [18], a model for narrative structure was developed. Through the Unified Modelling Language (UML) it was possible to build up a class diagram of the 'big picture of narrative'. Each narratologists' key chapters (Chatman [18], Herman [45] and Lacey [55]) definitions of narrative element were considered for the model. Initially we started bottom-up with 'events' (see Figure 8) and grew the model up to accommodate narrative as a whole as defined by the narratologists (Figure 9). The diagrams depicting the different stages of modelling narrative structure and theory, going from Figure 8 (existents and events) to Figure 9 can be seen on the **CD:** Additional Thesis Material; that accompanies this report. Figure 7 depicts the key symbols and representations used in the modelling of narrative for Figure 9 in UML.

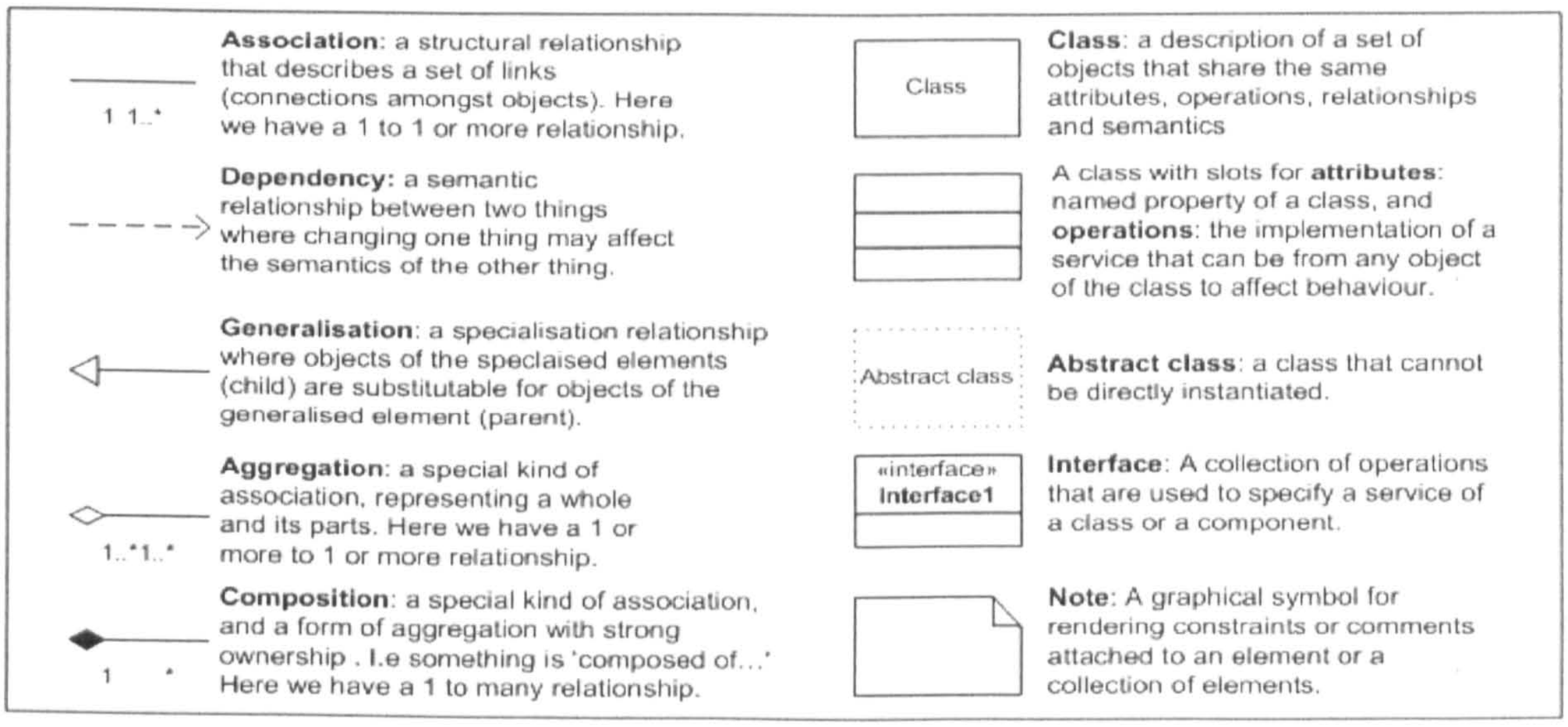


Figure 7 A key of terms and symbols used in the drawing of UML graphs.

Systems exist that deal with processing video data in many forms but what of the content of that data? We wish to be able to process, index, retrieve and navigate video content. Narrative theory provides an interesting mechanism for abstracting the content in video data, if we say that every

piece of video tells some sort of *story* then we can say that that piece of video in question conforms to the narrative theory. It is our belief that this representation of narrative, although not complete, provides enough information to represent narrative structure and possibly be instantiated automatically for video content. For this instantiation to occur however we had to be able to automatically extract narrative elements for video content. The video content we are examining in this work is film. Films are abundant, widely viewed and extensively studied. However, what interests us most is the rich description, selection, abundance and availability of films' collateral text data: film scripts. Thus, we examine film content in films by exploiting collateral texts that describe film: audio description scripts and screenplays.

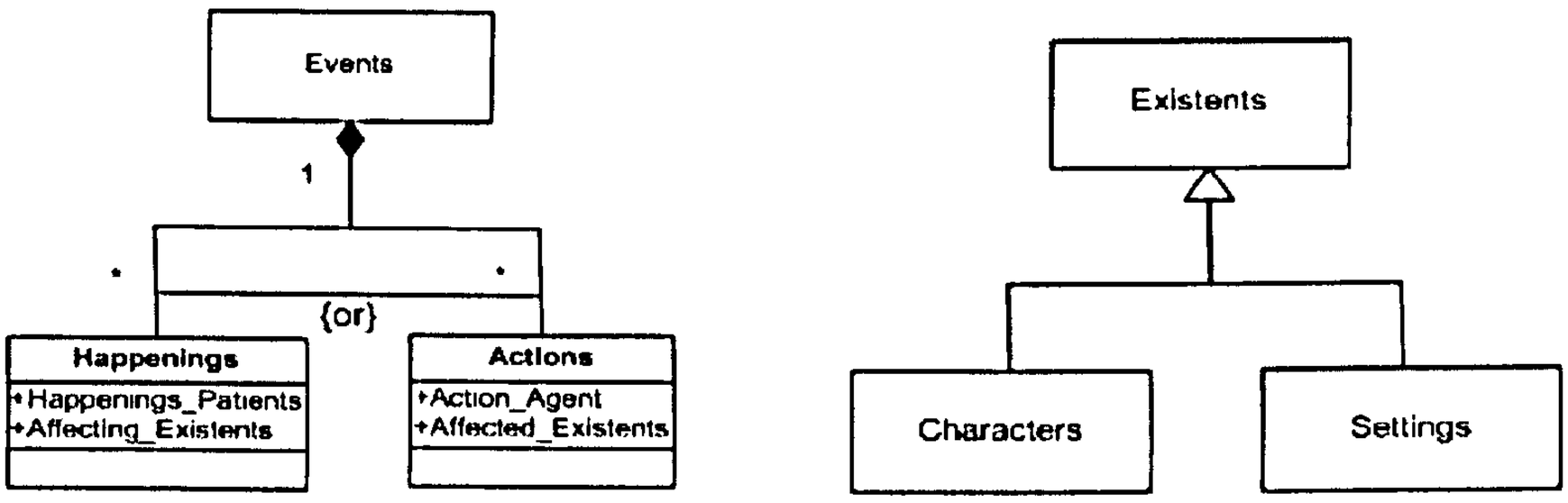


Figure 8 An illustration depicting possible relationships between 'Events', 'Actions' and 'Happenings'. This shows that one event is composed of one or more action or one event is composed of one or more happening. The attributes refer to the agents/patients involved in actions or happenings and, an illustration depicting possible relationships for 'Existents', 'Characters' and 'Settings'.

1.2.4 Why is it Important to Bridge the Semantic Gap?

Bridging the semantic gap will allow systems to have an understanding of film content or a story in a film. It would allow a machine-processable version of a film, providing the opportunity to develop applications to deal with film content which are currently not possible. For example: more concise querying of film content, (e.g. all the scenes when a cat is walking along a fence on a moonlit night), higher-level semantic querying (characters' behaviours, goals, motives, emotions and moods), short video summaries of a film etc. This however is a challenging subject. We must first gain an understanding of film content and the concept of a film's story. We must then translate that somehow into a machine-processable representation of a film, or machine-processable elements, events and objects of a film that are interconnected through their respective relationships (as can be seen in Figure 9).

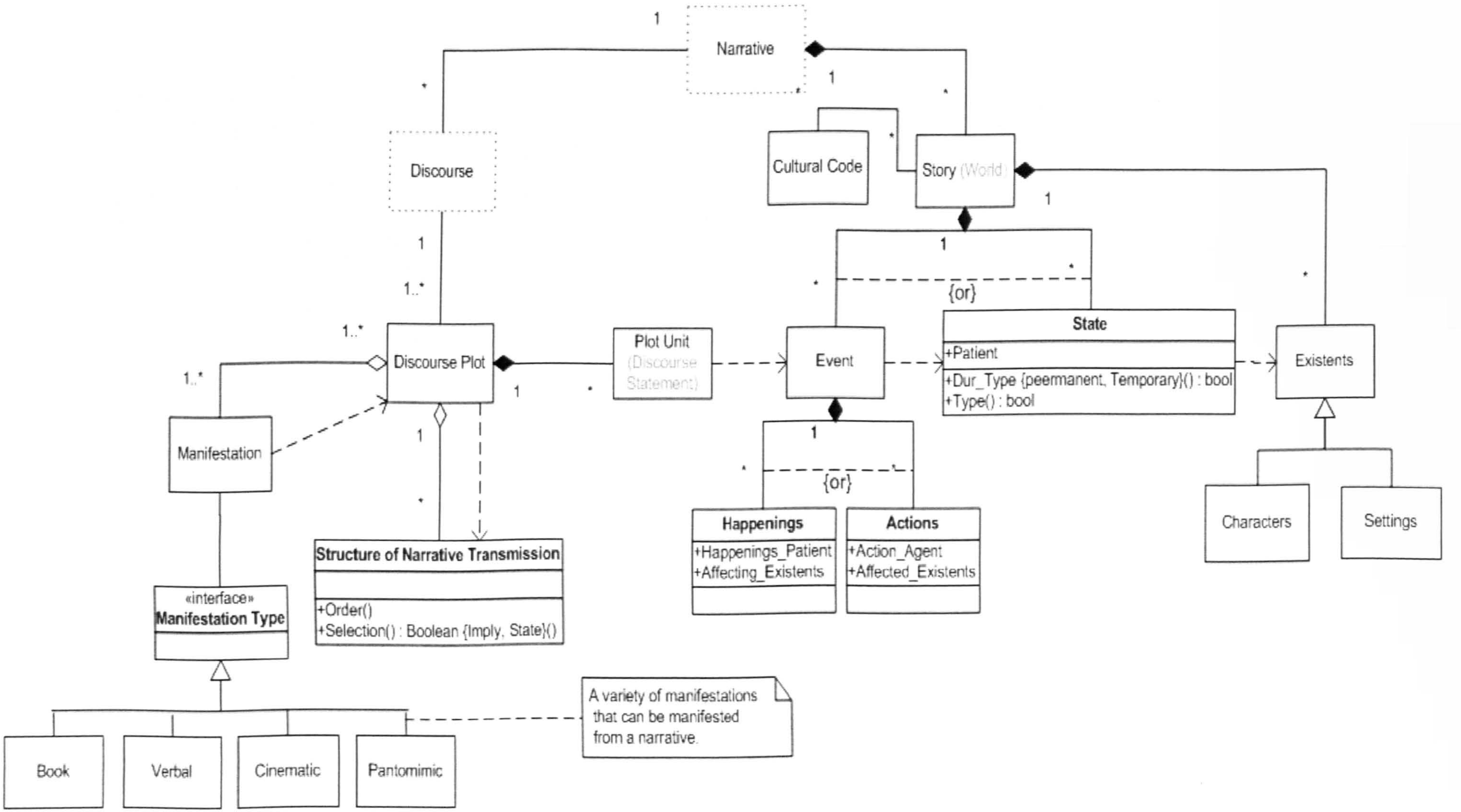


Figure 9 A UML diagram depicting a possible model for the 'Big Picture' of Narrative with respect to 'Story' and 'Discourse'. The diagram illustrates an overall hierarchical top-down structure for narrative with key emphasis on Story and Discourse.

As seen in Section 1.1.2, although attempts have been made to model and represent film and narrative in general, they are only successful to a degree and have limitations. Mostly the limitations are to do with manual instantiations of the models, i.e. there is no way of instantiating the representations reliably and automatically. Bridging the semantic gap may solve that issue.

1.3 Thesis Overview

The aim of this work is to bridge the semantic gap with respect to the analysis of film content. Our novel approach is to systematically exploit collateral texts for films. We believe that the systematic study of collateral texts, in our case audio description scripts and screenplays, can provide information about film content. Moreover, we can locate frequently recurring text patterns that will aid the extraction of high-level semantic elements of film, help automatically classify elements of film and provide evidence that screenwriters have idiosyncrasies in their screenplay writing, which may show us that collateral film texts hold a language for special purpose. We believe that our analysis of collateral texts, and the recurring patterns isolated, can be used towards bridging the film content semantic gap.

The work provides basic knowledge about films which can be utilised in an information extraction capacity. This is interesting to narratologists as we are categorising events through empirically gathered data. We suggest new building blocks for narrative, specifically film, and gaining an insight into how stories work. The four events found here give focus to the *coincidence* of what and how humans write for films, through scripts, (their interpretation) and what a machine can extract. This work also provides the possibility of new applications with film content, narrative content and any video content that tells a story. Video content analysts will be interested at the possibility of video ontologies in this work. New statistical information for script writers is also offered. We have provided writers with the most common elements in script writing, common textual structures and the most common formatting techniques used. From this information an automatic writing Thesaurus that is specific to language for film is possible for audio describers and screenplay writers.

Through the synthesis of corpus linguistics, information extraction, narrative theory and film theory it can be argued that we have developed a framework for representing film content and by extension narrative content. This may be of interest to the video content analysis community as a narrative framework could be used to model any video content data that tells a story and has collateral texts and a database of information could be automatically instantiated. For example television series, documentaries and court room video transcripts. This may also be useful to narratologists (even Herman talks about corpus-based narratology [45]); we have possibly

provided a framework for narrative elements that can be instantiated automatically and may possibly be used to instantiate a model such as in Figure 9.

Through this study we have aided the bridging of the semantic gap allowing, possibly, high-level semantic concepts to be extracted from video data and instantiated automatically in a database that could be interpreted by humans and a computer system in as equal measures as possible.

Chapter 2 reviews the extent to which film content can be automatically analysed and identifies emerging ideas in the field of video content analysis that are relevant to improving automatic film content analysis. Chapter 2 addresses two questions: how far across the semantic gap is the multimedia community with respect to film video data? And what are the important and relevant recent developments in the field of video content analysis? It comprises a review of approaches to video content analysis, and applications for accessing video data, with a focus on techniques for analysing film content on the basis of audiovisual features, collateral texts and multimodal fusion. Chapter 2 also assesses how far current technologies are able to bridge the semantic gap for video data and film content and identifies approaches for new research and development.

Chapter 3 describes the steps taken in investigating corpora of collateral texts that describe film using *collocation* analysis. The chapter seeks to determine basic knowledge about collateral texts for films that can be used to extract machine-processable representations of film content from these texts. Chapter 3 asks two questions: what information about film content do these texts contain? Are the ways in which this information is expressed sufficiently regular to allow for reliable information extraction, and the automatic generation of information extraction templates and algorithms? In order to answer both questions, a representative corpus of audio description scripts and a representative corpus of screenplays were analysed to identify unusually frequent open class words and collocations involving these words. Formal descriptions of frequent collocations, in the form of Finite State Automata (FSA), are interpreted to suggest four common types of event that are commonly described in audio descriptions and screenplays.

Based on evidence gathered from this chapter we claim that the language used in audio description and screenplays contains idiosyncrasies and repeating word patterns, specifically an unusually high occurrence of certain open class words and certain collocations involving these words, compared to general language. The existence of these idiosyncrasies means that the generation of information extraction templates and algorithms can be mainly automatic.

We believe that this chapter contributes an extended and tested, existing, mainly-automated method to generate templates and algorithms for information extraction of film content. The corpus linguistics method applied here has never been tested on corpora of film scripts before. This is interesting to the corpus linguistics community as it validates their techniques on a new data set. Enough statistical evidence is gathered to provide evidence of Languages for Special

Purpose (LSPs) and Local Grammars. Providing evidence for a specialist language for film will be of interest to corpus linguists and script writers.

Chapter 4 explores whether events in film can be automatically extracted from film scripts based on the frequently occurring collocation phrases or local grammars. Chapter 4 explores how the basic knowledge from Chapter 3 can be applied to support novel applications that rely on machine-processable representations of film content. Continuing the question from Chapter 3, Chapter 4 asks: are the ways in which this information is expressed sufficiently regular to allow for reliable information extraction and the automatic generation of Information Extraction templates and algorithms? Chapter 4 also enquires: what kinds of novel applications may be enabled by machine-processable representations of film content-based on the four events identified in Chapter 3? To answer these questions, an information extraction system is developed. Algorithms, based solely on the four templates identified in Chapter 3 are designed and implemented using a Natural Language Processing text analysis system. To evaluate the system, a Gold Standard data set is gathered from an hour of film clips from five Hollywood films. The Gold Standard data set is used to evaluate the system by examining its performance, through precision and recall, against the segments of audio description (AD) scripts and screenplays (SC) for the five film clips. The system is used to populate a database of information for the four event types for 193 AD and SC film scripts. Based on the results of the database, potential novel applications are outlined for the querying, retrieval, browsing, display and manipulation of film content.

Results from Chapter 4 support our claim that there are four types of event that are commonly described in AD scripts and SC for Hollywood films: Focus of Attention (FOA), Change of Location (COL), Non-verbal communication (NVC) and Scene Change events (ScCh); and that information about these events will support novel applications for the automatic analysis of film content. Chapter 4 contributes a database containing information about four types of events in 193 films, which was extracted by the system automatically and a Gold Standard data set of the events, for an hour of film scenes.

Chapter 5 summarises the thesis and examines how this work can aid the bridging of the semantic gap. The contributions and claims of the thesis are presented and discussed and opportunities for future work and future applications are outlined.

2 The Analysis of Video Content

Automatically analysing Video Content is a non-trivial task. The past 15 years have seen many proposed methods to automatically browse, search, retrieve, process, segment, abstract, summarise and manipulate video content. The idea of a semantic gap emerges, with respect to video content, where there is a lack of coincidence between automatically extracted video information provided and the interpretation that data has for users, in a given situation. The prospect of representing video content, to allow a machine to understand the content such as a human being, still eludes the research community and signs of commercial applications to do so are currently not evident.

This chapter seeks to establish the extent to which film content can currently be analysed automatically, and to identify emerging ideas in the field of video content analysis that are relevant to improving the automatic analysis of film content. It addresses two questions: how far across the semantic gap are we with respect to film video data? And what are the important, relevant recent developments in the field of video content analysis? It comprises a review of approaches to video content analysis, and applications for accessing video data, with a focus on techniques for analysing film content on the basis of audiovisual features, collateral texts and multimodal fusion.

Thus, Section 2.1 looks at analysis techniques and standards for general kinds of video data. Section 2.2 then concentrates on the analysis of film content. In conclusion, Section 2.3 assesses how far current technologies are able to bridge the semantic gap and identifies approaches for new research and development.

2.1 Approaches to Video Content Analysis

The last 6 years have seen the size of home-use PC hard disks rise from 10 GB to 1 Terabyte. Why? Because of the sheer volume of multimedia data that can be easily accessed and stored on home PCs. It is now possible to buy your favourite TV series, in crystal clear quality, download and store it on your home PC, create and store hours of home videos on digital cameras, store volumes of photos and store thousands of music albums and videos. However, easily and efficiently browsing, organising, retrieving, processing or describing the *content* of this media is not a trivial task and nor is it, arguably, commercially developed; "we still have limited applications to describe, organise and manage video data."[28]. There seems a need for applications to aid the organising, summarisation and retrieval of the sheer volume of video data and, more specifically, it's content.

The multimedia community and industry realise this. Research into dealing with video content has, arguably, progressed significantly over the last 15 years. However these progressions have yet to impact the marketplace.[6] Currently video content description and indexing is manually generated, making it time consuming and costly and to some extent impossible. Therefore automatic classification, indexing and description generation, or automatic 'understanding' of video content, is necessary [28], [90].

2.1.1 Video Content Analysis Systems and Standards

This section examines video content analysis techniques of different types of video (different media types and genres: sports, CCTV footage, microbiological videos). It also explores the MPEG 7 standard, the TRECVID conference and various systems that analyse video content.

2.1.1.1 Analysis of Visual and Audio Features, Collateral Texts and Multimodal Fusion

The analysis of visual features in video is extensive. Stemming from pattern recognition [10] and the analysis of images, there are many techniques available to analyse visual features in video, e.g. motion vectors (and other motion features) [1], [2], [62], [73], [74], [108], [109], [111] background subtraction [33], [62], colour histograms [104], comparisons of changes in colour, shape, lighting etc. per video frame [1], [2], [19], [73], [94], [104]. However, bridging the semantic gap and extracting high-level semantic features from video using visual features is a non-trivial task. It is an ill-determined problem where video features under constrain the problem and do no take account the context of what is being extracted well. Attempts to bridge the

[6] With the possible exception of automatic digital video editing programs, e.g. www.muvee.com [151].

semantic gap in terms of recognising high-level concepts have been attempted for all sorts of video data, such as recognising suspicious events in CCTV footage automatically [62].

In analysing video content, the analysis of audio features seems a less commonly studied modality than visual features. However, there have been studies of audio features in sports videos [41], [111], speech detection [108], [109], emotional intensity [16], [42] and other high-level feature extraction. Audio features are commonly studied in sports videos as the change of intensity of the crowds' reaction in sporting events often indicate interesting, exciting or dramatic events.

The use of collateral texts in video has also been extensive. Collateral texts have been examined both as part of the video data (subtitles, basketball game scores [111]) and collaterally (accompanying news feeds [see Informedia-II], closed-caption texts [49], screenplays [101], [102], continuity scripts [75]). Collateral texts and transcriptions have been used to index video data and in video retrieval for E-Learning Videos (lecture videos) [110]. High-level concepts such as emotions [77], [102] have been extracted using collateral text data.

The complexity of analysing video data lies partially in its multimodal, temporal nature and partially in interpreting low-level features and mapping them to high-level concepts. Video data encompasses different types of visual, audio and textual features, changing over time, which are 'fused' together. The question arises: can they be analysed separately and still capture the video data temporal narrative nature and intended meaning? The gestalt[7] view of the whole being greater than the sum of its parts rings true here. Snoek and Worring [90] suggest that "Effective indexing [of video content], requires a multimodal approach in which either the most appropriate modality is selected or the different modalities are used in a collaborative fashion." Thus, methods to analyse and fuse the different modalities of video data in order to analyse video content must also be addressed. [8]

2.1.1.2 Evaluating Video Content Extraction Systems: TRECVID

In terms of evaluating a video content analysis system, the TRECVID conference offers "a large test collection, uniform scoring procedures, and a forum for organizations interested in comparing their results"[117]. The TREC conference series is sponsored by the National Institute of Standards and Technology (NIST) with additional support from other US government agencies. TREC's goal is to promote progress in content-based retrieval from digital video via open, metrics-based evaluation. TRECVID (based on 2006 guidelines) has a series of four tasks for

---

[7] **Gestalt**: A physical, biological, psychological, or symbolic configuration or pattern of elements so unified as a whole that its properties cannot be derived from a simple summation of its parts [112].
[8] It may be interesting to note that [111] uses all three features in its analysis of sports videos.

participants to enter into and they must complete at least one task to attend the workshop. The tasks are:

1. Shot Boundary Determination: Identify the shot boundaries with their location and type (cut or gradual) in the given video clip(s).

2. High-Level Feature Extraction: Participants are asked to identify pre-defined High-Level Features such as shots containing: sports actions, office spaces, human faces, a car.

3. Search (interactive, manually-assisted, and/or fully automatic): A high-level task including at least query based retrieval and browsing. It models an analyst looking for persons, objects, locations events etc.

4. Rushes (raw material used to produce a video) Exploitation (exploratory): participants in this task will develop and demonstrate at least a basic toolkit for support of exploratory search on rushes data.

To bridge the semantic gap for video data, the ability to analyse low-level video features, raw data or collateral media for video data, to extract high-level semantic concepts is a necessary and non-trivial task; as the need for TRECVID to assess progress in this field demonstrates.

2.1.1.3 Standards for Representing Video Content and Narrative: MPEG-7

MPEG-7 became an ISO/IEC standard in December 2001 and was developed by MPEG (Moving Picture Experts Group) for multimedia content description [118]. MPEG-7 provides standardised tools for describing multimedia data. It uses XML tags to store metadata about multimedia, such as attaching time code to particular events, the lyrics to a song or descriptions of a film scene or medical image. "MPEG-7 provides a standardised description of various types of multimedia information, as well as descriptions of user preferences and usage history pertaining to multimedia information"[28].

MPEG-7 can be used to support the interpretation of the information. MPEG-7 has tools available for describing abstract concepts, the semantics of multimedia content and even captures elements of narrative "MPEG-7 refers to the participants, background context and all other information that makes up a single narrative as a 'narrative world'" [[59] pg 127]. The MPEG-7 semantic entity tools describe entities such as narrative worlds, objects, events, states places and times and a narrative world is represented using the semantic description scheme [[59] pg 130]. The fact that MPEG-7 contains tools to represent a 'narrative world', and narrative elements, demonstrates to us the need for narrative elements when describing complex or rich multimedia data.

2.1.1.4 Commercial and Prototype Systems that Retrieve Video Content

Commercial systems exist which analyse video data. Google Video [119] and Yahoo Video [120] both offer the ability to search for video clips on the Internet. Both, like their image search engines, search the accompanying textual description of the video for relevant information.

Google Video takes it a step further by allowing American TV series' video content to be searched using the closed-captions of the TV series. Blinkx [121] allows a user to search video content for over 60 media companies. It claims to use "visual analysis and speech recognition to better understand rich media content".

All of these systems are interesting and provide good steps forward with respect to searching and browsing video content but are not advanced enough to allow 'accurate' searches for multiple high-level semantic concepts in video. Many: "show me the scenes where…" type questions to do with character behaviours, story plots (turning points in a film's plot) or specific scene types (love, action, suspense etc.) could not be retrieved accurately with these systems partially because the annotation of the video content is not adequate enough to capture the semantic concepts. The Virage system [122] offers a variety of retrieval methods and claims to be able to input and understand all rich media content and allow users to search extensive video 'assets' with pinpoint accuracy. However there is no indication to what extent Virage can handle semantic concepts in its query system.

Advanced prototype systems that model, or analyse video content, have been developed. VideoQ [123] was developed at Columbia University and is a content-based video search engine. It uses novel search techniques that allow users to search video content-based on a rich set of visual features and spatio-temporal relationships [10]. Visual features include colour, texture, shape and motion. Informedia-II [124] is a fully fledged digital video library provides content search and retrieval of current and past TV and radio news and documentary broadcasts. Speech recognition, visual feature analysis and natural language processing are used to automatically transcribe, segment and index the video. They are able to summarise news stories automatically.

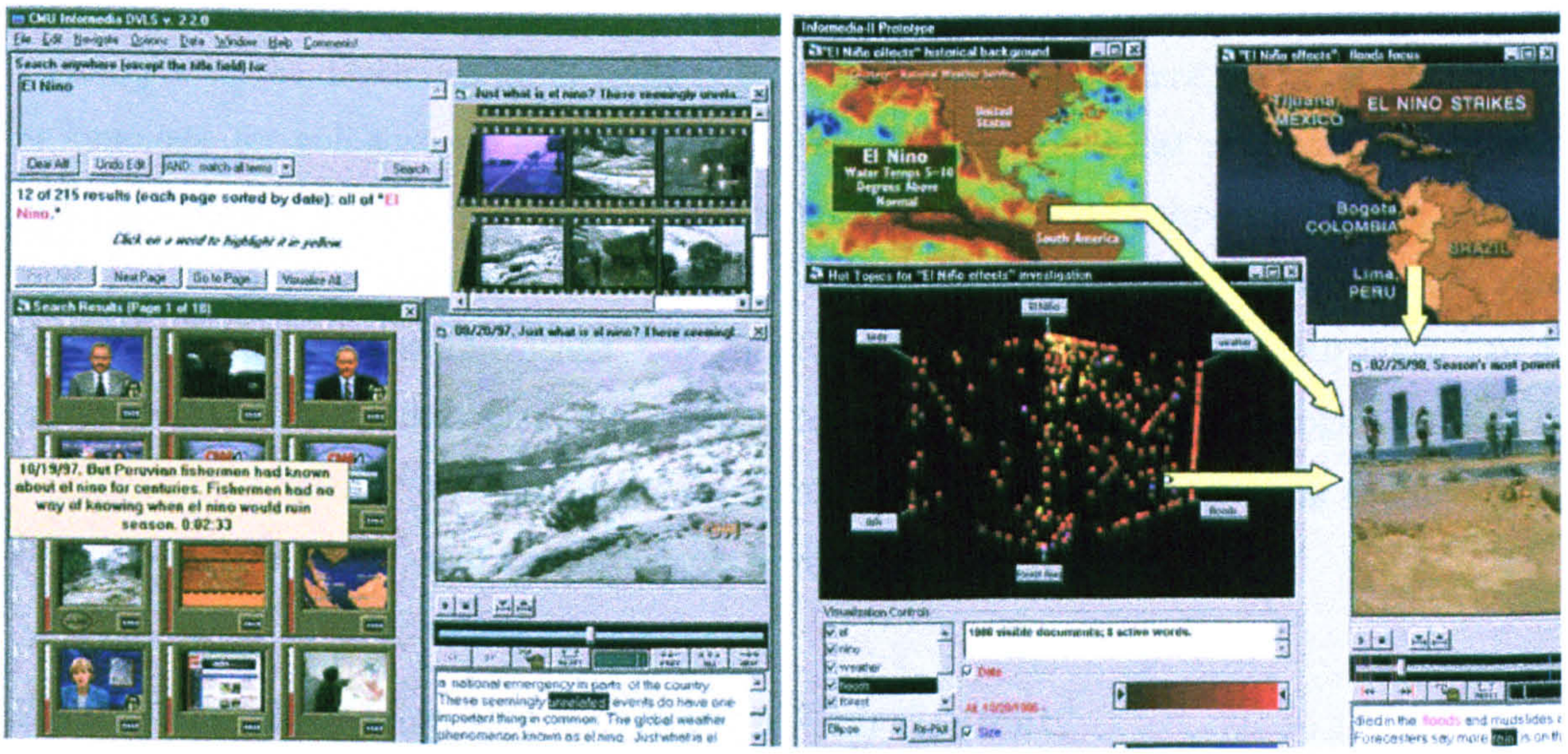


Figure 10 *Left:* INFORMEDIA-II IDVL interface showing 12 documents returned for "El Niño" query along with different multimedia abstractions from certain documents. *Right:* Additional views provided by the Informedia-II interface for an inquiry into "El Niño effects". Taken from [124].

The Fischlar [125] system, developed at Dublin City University, Ireland, is a web-based video recorder (Fischlar TV), which includes camera shot boundary detection, keyframe extraction, closed-caption analysis, video streaming and XML architecture for video retrieval. It is an automated video archive of daily TV news (Fischlar News Stories), where a user can search news stories by keywords, browse news stories by date/year and browse full details of a story and video and Library (Fischlar Nursing) which allows a user to search nursing related video programs.

Most of these systems are not commercial systems and are prototype systems; they are specialist systems available from universities. They do however deal with video content and specifically with the organizing, retrieval and processing of video content automatically.

2.1.1.5 Video Ontologies

An *Ontology* is a formal, explicit specification of a domain [50]. It is the hierarchical structuring of knowledge about things by subcategorising them according to their essential (or at least relevant and/or cognitive) qualities [126] and is a controlled vocabulary that describes objects and the relations between them in a formal way. An ontology is a description of the concepts and relationships of interest in a given domain. Ontologies, provide formal descriptions/conceptualisations of a domain. Our domain is video content and there exist ontologies to represent video. In Nevatia et al. [68], an ontology for a video event representation is examined leading to a language to describe an ontology of events: Video Event Representation Language (VERL) with a mark up language to annotate the events. In Dasiopoulou et al. [27] an ontology is described for visual components of video to allow semantic video analysis and annotation of video. Jaimes et al. [50] use an ontology of multimedia objects and *modal keywords* (keywords that represent perceptual concepts in a given modality) to 'understand' videos. Ontologies can be developed for different types of video data, for instance, Bao et al. [11] present an ontology for colonoscopy videos. There have also been ontologies expressed for narrative [100] and movies/films [35]. In [35] the author describes a method to populate a movie ontology from IMDB which is accomplished with high precision and recall. Dasiopoulou et al. [27] argue that the use of domain knowledge is perhaps the only way higher-level semantics can be incorporated into techniques that capture semantics by automatic parsing. Ontologies can be used to capture the higher-level semantic and incorporate them into automatic parsing and thus it can be argued that ontologies make steps towards bridging the semantic gap.

2.1.2 Discussion: How can we Analyse Film Content?

When analysing film content many researchers have chosen to focus on one or two modalities (audio and visual). Studying all modalities at once, or a fusion of them, does not seem often

attempted and appears a non-trivial task. The ACM Multimedia Conference in San Francisco in 2003 focussed on the issue that research into multimedia content should focus on multimodal sources of multimedia rather than a single modality when dealing with automatically extracting information from multimedia; after all multimedia uses visual, textual and auditory channels to express itself. Hence video content is intrinsically multimodal [90]. However, so far dealing with one modality as a source of eliciting high-level concepts from video content is preferred and has offered adequate results. Systems such as Blinkx, VideoQ, Fishclar and Informedia-II identify high-level semantic concepts from video content through the analysis of low-level features. Blinkx and Informedia-II also analyse low-level audio features of speech, and Informedia-II, Google and Yahoo video and Fischlar also use text analysis/natural language processing to elicit high-level semantic content from video data. The fact that the multimedia community has a method for evaluation of such systems, TRECVID, is a positive step towards encouraging and evaluating researchers' efforts to bridge the semantic gap. Also, developing standards with frameworks for representing multimedia and more precisely 'narrative worlds' in the form of MPEG-7 presents the opportunity for universal annotation of all video content and the opportunity for systems to be created that can annotate, analyse and retrieve video data automatically.

All of the aforementioned systems as well as techniques for modelling video content (section 1.2.2) represent and analyse video content in one way or another. There are elements of all systems that can be applied directly to analysing film content specifically. All the models mentioned in 1.2.2 are possible representations of film content; some are manual and others can be automatically instantiated. The same applies to the commercial and prototype systems presented; however, the question becomes what elements of these systems can be combined to elicit high-level concepts from film data *automatically*. Also the question of which modality is superlative and most constructive to analyse? Is the *fusion* of these modalities the best option when trying to automatically extract high-level concepts from film content?

The next section reviews research that has attempted to bridge the semantic gap with respect to extracting high-level film concepts from film data using low-level features of film and different modalities of film.

2.2 Analysis of Film Content

Film content, as opposed to generic video content, implies 'story-oriented' video (for example, films or movies, TV series and animations etc.) [53]. These story-oriented videos comprise rich sets of events, characters and intricate – often non-linear – plots, open to human interpretation and follow aspects of film theory, grammar and structure. This makes the analysis and extraction of their content a non-trivial task. This section explores attempts to analyse the content in such story-oriented videos in the computing community.

2.2.1 Film/Genre Classification

In [73], Rasheed et al. present a framework to classify films into genres based on computable visual features (such as shot length, colour variance motion content and lighting). They study over 100 movie previews: trailers, TV spots etc, and categorise them into four main categories: comedy, action, drama and horror, and combinations of these. They examine the average shot length, colour variance, motion content and lighting key (how much lighting per frame) which is represented as a 4-D feature space. Different degrees of variance in these, over certain thresholds, allow for the detection of shots and then mean-shift based clustering over the 4-D feature space is used to index the trailers. They manage to classify 84/101 films into the genres specified by them. The use of low-level features for classification of genres using mean shift classification seems successful but thus far this method has only been applied to movie trailers which are shorter and are created to relay the essence of the film to the viewer. It is not certain whether this method of classification would work if applied to the whole of the film.

Zhai et al. address the problem of classifying scenes in films into categories and propose a framework that utilises finite state machines; low-level and mid-level features [71], [108], [109]. This work analysed low-level features in the form of MPEG motion vectors and audio energy and intensity and mid-level features in the form of facial and body recognition, which has allowed them to identify conversations and speakers and classify conversation, action and suspense scenes. Film scenes were classified by Finite State Machines accepting relevant criteria, i.e. the criteria for a conversation scene are low motion activity, medium audio energy and multiple speakers (>2). They manage to classify conversation and non-conversation scenes for 50 movie clips in [108] with above 90% precision and recall. In [109] they experimented on over 80 movie clips and managed to classify 35 conversation scenes, 16 suspense scenes and 33 action scenes all with above, on average, 93% precision and recall.

This work presents another method to classify scenes with low-level features of the video however, unlike in [73], audio features are also utilised here. The mid-level features, face and body recognition, although currently not developed enough, serve the method well. The method presented in [71], [108], [109] could not however be used for animations or special effects sequences without recalibration as it seems the face and body recognition although not specified only works on humans. Also the method classifies only a small number of film scenes, although, arguably, these scenes may be integral to a film's plot and thus considered the 'important' scenes.

2.2.2 Film Theory Techniques, Shot/Scene Segmentation & Film Abstraction

In [95], Tavanapong and Zhou focus on scene segmentation for films. They strictly define a *scene* for a narrative film only (as opposed to any other type of video), which is less subjective, to

provide more *familiar* scenes to viewers when browsing and searching. A shot clustering technique, ShotWeave, is presented that extracts visual features from two predetermined areas of a keyframe of a film (called *colour keys*) as opposed to visual features from the entire keyframe. The regions are carefully selected to reduce *noise* (noise often confuses other existing techniques) and maintain viewers' thought in the presence of shot breaks. Then related shots are grouped together, using a clustering algorithm, by comparing keyframe similarity. In their experimental study two full-length feature films were segmented into scenes and ShotWeave was compared to two other systems. ShotWeave outperformed the other systems. Tavanapong and Zhou have provided a stricter definition of a film narrative scene, which makes for better shot clustering to extract scenes in films and a technique to group related shots into scenes of a film. By contextualising the study of scene clustering to that of film the authors may have provided a step to extracting film 'structural' units from video. However, their system is based on visual features only and may benefit from a combination of audio and textual features.

Sundaram and Chang present a computational scene model and derive novel algorithms for computing audio and visual scenes and within-scene structures in films [94]. They develop notions of video and audio computable scenes (v-scenes and a-scenes) by considering camera placement, lighting continuity and audio features (based on the psychology of audition). They extend an existing memory model making it causal and finite for film [[94], pg 484] to be able to determine 'coherence' of shots in the film. From the implementation of the computational scene model, scene change points can be detected which are considered scene boundaries. In [94] the scene segmentation algorithms were tested on 3 hours of films. There was a computable scene (c-scene) detection of 94%. The structure detection (dialogue structures) algorithm gave 91% precision and 100% recall. What is interesting to note here is that they discuss the *coherence* of shots as opposed to treating them as discrete elements for study. This lends to the idea of shots and scenes being interlinked throughout the film with relationships in terms of story and plot and not just collections of images. We speculate that dialogue scenes will have cause and effect links that advance the plot and capturing information about their coherence may be useful in tracing such links.

Adams et al. [1], [2] present a method to compute a novel measure of movie *tempo*. They attempt to show that Tempo is a useful high-level semantic construct in its own right and a component of the rhythm, tone and mood of a film. Adams et al. define Tempo as carrying with it important notions of time and speed: "the rate of performance or delivery". They analyse low-level features of video data such as shot length, motion, editing (shot boundaries or cuts) and lighting. Pace, which is calculated based on changes of motion and shot length over time, is used to detect dramatic sections of a film. Edges of the function that defines pace are identified as indicators of events using Deriche's recursive filtering algorithm (see Figure 11). In [2] four films were

analysed and 20 more are claimed to be analysed, giving varying results with few false positives and negatives. In Adams et al. [3] a novel method of detecting act boundaries in films using tempo analysis is presented and a probabilistic framework utilises low-level features to derive semantic narrative structures of a film (acts). Adams et al. define what an *act* is and use tempo peaks as in [2] to approximate act boundaries to find where there are dramatic events (Figure 11). Adams et al. [3] is another application of Tempo to find more semantic structures in a film. Bayesian formulation is used to draw the identified factors together. 25 films were experimented on with a low error in act boundary detection found (~3.5%).

In this work the tempo/pace function, based on its definition with respect to film grammar, examined how directors' manipulate tempo in film (and to a lesser degree *motion*). Adams et al. discuss how information about low-level features and their changes can be combined to give information about the speed, pace or tempo of a film. Adams et al. developed a method, using Tempo, to identify dramatic story sections or events in a film and story structures in the form of acts. It seems that Tempo may be able to provide information about a film's content, and about where 'dramatic' scenes are in film and about their intensity.

In the same spirit as [2], Chen et al. [19] aims to extract the most semantically important story units and segments of an action movie based on film tempo analysis. They calculate Tempo differently in [19]; by the combination of the motion activity intensity, shot length and *audio* effects. The peak high tempo shots are found, story boundaries are cut and the algorithm is applied again to get a film segment. Experimentation was performed on three action films and managed to successfully detect various action scenes in the film. Human judgement evaluated the system based on relevance of scenes recognised in the film.

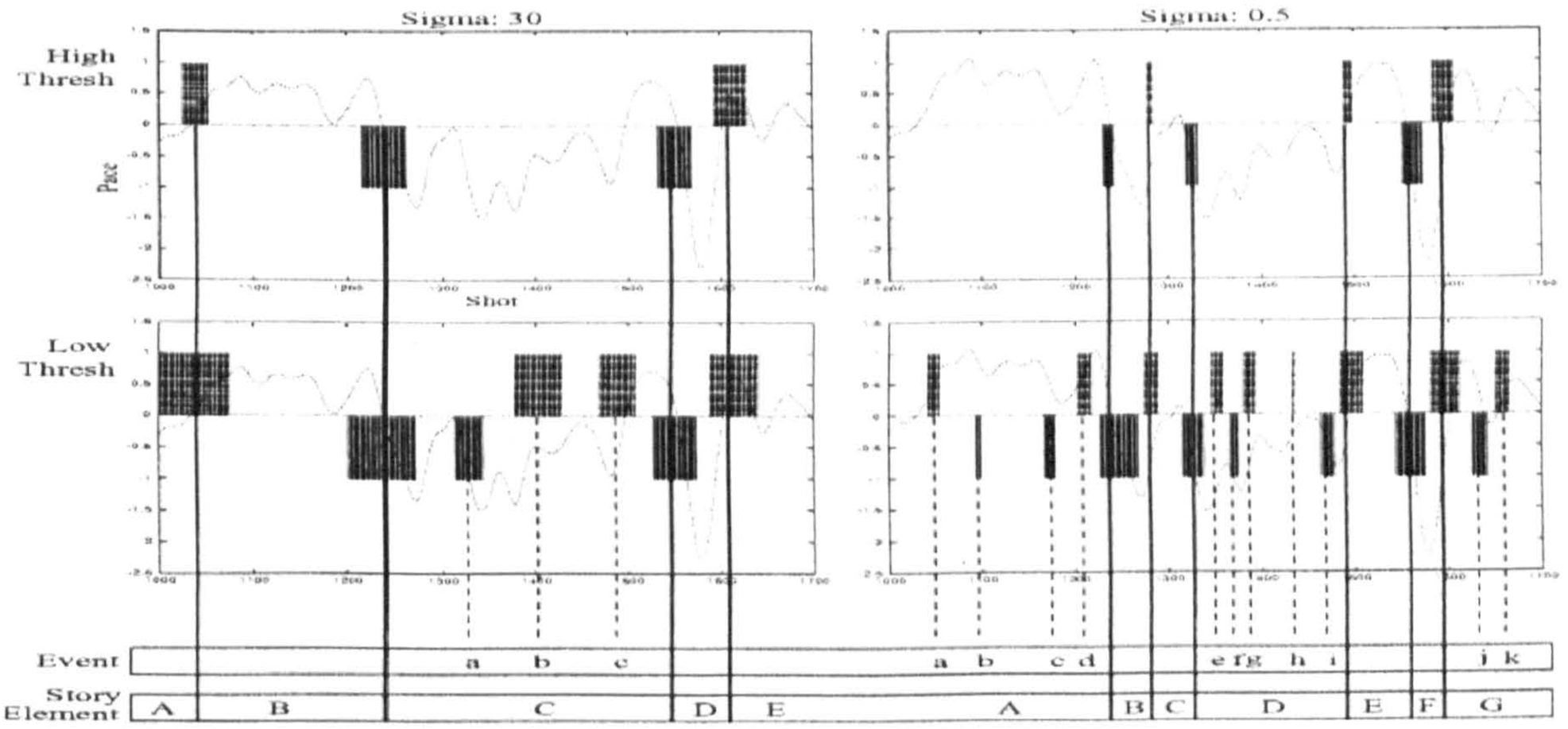


Figure 11 Results of edge detection on Pace flow and corresponding story sections and events from *Titanic* [157]. Taken from Adams et al. [2] pg 476.

Shiraham et al. [86] follow characters' movement in a film to find meaningful film segments involving characters, and annotate movie sections based on the rhythm of characters' appearance and disappearance in shots. Shiraham et al. explore the notion of following what characters *do* within a film to provide interesting/important events or even map the plot arc of the story.

Wei et al.'s [104] work utilises pace, however this differs from tempo [2], [19] as it uses colour histograms to ascertain a general mood for either scenes in a film and what types of moods or atmospheres occur in the film as a whole. Wei et al. investigate the colour characterisation of films at global and local (group of shots) levels for mood analysis and provide a method to capture the dominant colour ratio and pace of the film. They experimented on 15 films and reported approximately 80% accuracy for 'mood tone' association.

Truong and Venkatesh [99] present an algorithm for the extraction of flashing lights from films in order to identify certain dramatic effects in films. Here, Truong and Venkatesh state that they can identify dramatic effects that are intensified by the use of flashing lights such as supernatural power, terror, excitement and crisis. As part of a bigger system, this work may be a small step towards mapping low-level features to high-level concepts, such as characters' emotion, allowing us to help classify scenes and detect emotional intensity.

Aner-Wolf [9] seeks to generate high-level semantic descriptions for film scenes based on the extraction of low-level illumination features and by using film rules regarding scene lighting. This work presents a method to generate semantic descriptions for film scenes, automatically, based on the changes of illumination over time for a film scene. This method could be used, as part of a system, to identify changes in time i.e. day to night or changes in an area's illumination (e.g. turning on a light, opening a door to let light in.)

The idea of *video abstraction* is explored by Hwan Oh et al. [49] where video abstraction refers to a short representation of the original video. A video summary and video skim of video data present a summary of the video or a shorter representation of it respectively. A *video summary* is a set of keyframes from the original video which are detected automatically by comparing colour histogram changes per frame; motion-based selection based on the rate of change of 'optical' flow vector spaces of keyframes are then calculated and cluster-based keyframe selection ensues. *Video Skimming* refers to a collection of image sequences along with related audios from the original video and possesses a higher-level of semantic meaning than a video summary does. A video skim either *highlights* the most interesting parts of a video (found by: scene boundary detection, dialogue extraction, high motion scene extraction and average colour scenes) or can produce a summary sequence which renders an 'impression' of a video (a mixture of compressing audio and a Model-based method: selecting important scenes, and a dialogue detector from closed-caption

text data). The video abstraction methods have been applied to five feature films with relative success but this method has not been implemented.

2.2.4 Story-Oriented Video Analysis

Jung et al. [53] present a method and system to create video abstracts or summaries of TV dramas through a narrative abstraction model that deals with 'dramatic incidents' (events) or narrative elements in TV series. The narrative abstraction model is developed which treats 'film' as a type of formal system: a film system, where scenes in a video are related to each other to provide information or make a story understandable. The narrative abstraction model considers a 'story-oriented' video (video that has a narrative) a sequence of dramatic incidents or scenes. The basic unit considered here is that of a 'narrative element', which is either a dialogue or an action scene. A narrative connection must be found between the possible pairs of narrative elements in the model. In the model the progress of the story can be determined through three aspects: the intensity of edit-effects, the intensity of character interaction between an incident and the importance of the characters in a dramatic incident.

The proposed narrative abstraction model is represented as a weighted acyclic graph called a Narrative Structure Graph (NSG). The video abstraction process happens in three stages. Firstly, pre-processing is conducted, where face recognition, shot detection and scene detection occur. Secondly, the input video is modelled through the narrative abstraction model by distilling narrative elements from scenes, identifying each narrative connection and calculating *DOP* values for each connection. Finally an NSG is constructed from the model. Abstraction occurs by constructing subgraphs until the duration of the subgraph reaches the desired duration (which can be chosen by the user).

In [53], two episodes from two different TV shows were experimented on. The shorter the target duration of the abstraction, the higher precision and recall reached. Precision and recall were on average >70%.

Jung et al. have presented a way of representing story-oriented narratives through narrative elements which can be automatically extracted. They use facial recognition and shot/scene detection through edit-effects, spacial and temporal changes and rhythm to help identify action and dialogue scenes. This work is unique as it presents us with a formal model for narrative, algorithms that can automatically analyse a video and detect scenes in the narrative and a system that provides summaries with a flexible duration decided by the user. Though it can be argued that other works presented here have a model, an algorithm and a working system, this work touches on all three adequately. Jung et al. deal with the issue of modelling video content in terms of narrative and film theory and presents a formal description of the model, describes algorithms that

can analyse the video automatically and detect characters, shots and scenes and presents a system to summarise the video automatically with some success.

2.2.5 Audio Visual Features Mapping to Affect and Emotion

Hanjalic et al. [42] attempt to model affective video content in films. They define the affective content of a video clip as: the intensity and type of feeling or emotion (which are both referred to as affect). They differentiate between the cognitive level and affective level of video content perception. The cognitive-level algorithm aims to extract information that describes facts and the affect level deals with feelings or emotions. Affect has three basic underlying dimensions: Valence (type of affect), Control (dominance) and Arousal (intensity of affect). Hanjalic et al. map the affective video content onto the 2-D emotion space by using the models that link the arousal and valence dimensions to low-level features extracted from video data.

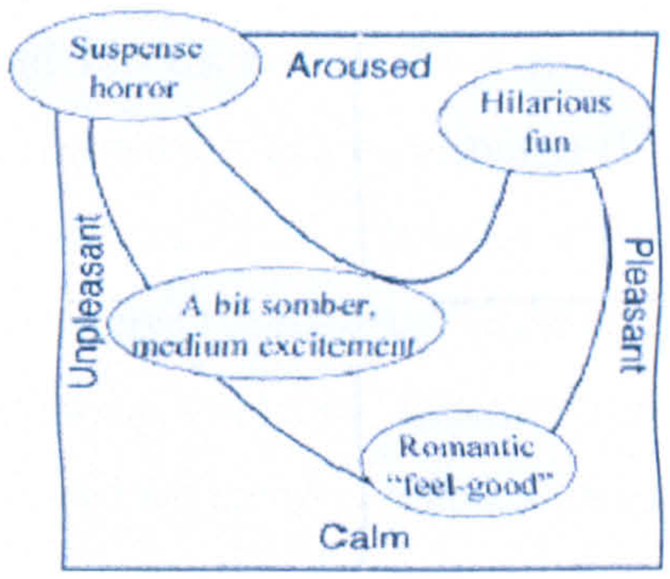


Figure 12 Indexes, in the form of labels, are assigned a priori to different regions of the 2-D emotion space. The affect curve, which represents affect over time, is a combination of the arousal and valence change with time curves. The arousal model considers: audio (change in sound energy), motion (motion activity in each frame) and rhythm (time varying shot lengths).

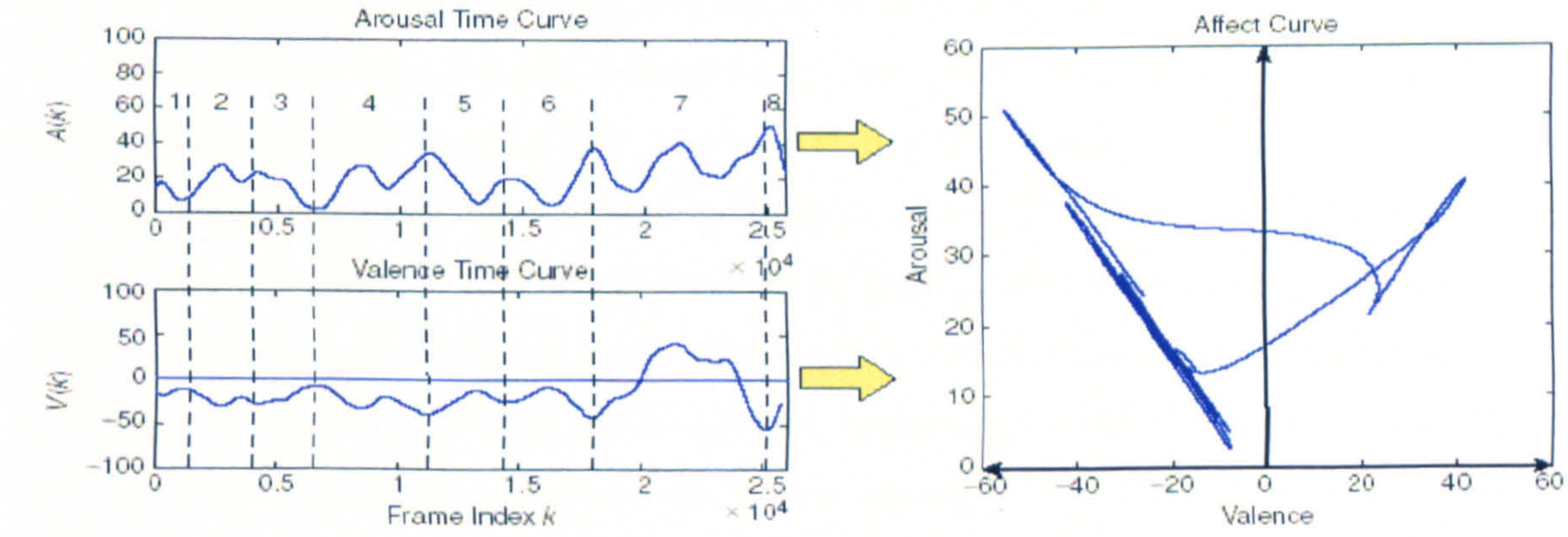


Figure 13 Arousal and Valence curve obtained for an excerpt from the movie "Saving Private Ryan" [172]. Valence curve obtained on the basis of the pitch-average component. Taken from [42][43] pg 137.

Hanjalic et al. developed a model of affect that was applied to films of different genres. Chan and Jones [16] go a step further to Hanjalic et al. and attempt to categorise the emotions/affect retrieved by the different arousal curves by mapping the audio content to a set of keywords with predetermined emotional interpretations. Kang [54] use colour and motion intensity and Hidden Markov Models to categorise affective content. Xu et al. [106] analyse the soundtracks of horror and comedy films in terms of change in sound energy and achieve >90% precision and recall in terms of correctly identifying comedy or horror moments.

This is an example of extracting higher-level meanings, in this case the mood or affect of film scenes, from low-level features of film (audio and motion changes). This is a novel approach to dealing with emotions of the characters in the films although there is some question as to whether it is the emotions of the films' characters', the emotions the director *wants* us to feel or the *expected* emotion of the viewer that is being extracted. In [43] Hanjalic considers "how we [the viewer] feel about the content we see or hear...we are interested in obtaining the information about the feelings, emotions and moods evoked by a speech, audio or video clip" [pg 90]. Hanjalic et al. [42] state that the expected affective response (feeling/emotion) can be considered objective as it results from the actions of the film director, or reflects the more-or-less unanimous response of a general audience to a given stimulus. The resulting valence/arousal models can be considered a solid basis for obtaining a reliable affective video content representation and, if nothing else, visual representations of the mood or emotion changes in films.

2.2.6 Information Extraction from Collateral Texts Mapping to Emotions

Salway & Graham [77] and Vassiliou et al. [102] deal with extracting information about emotions in a film from collateral film text data, i.e. audio description scripts and screenplays of films. These are early attempts to elicit meaningful information about films as a whole from film collateral data. Lists of emotion words are mapped to 22 base emotion types as outlined by Ortony [69]. These emotion types are then represented against time to get a visualisation of the types of emotion that occur over time in a film. Vassiliou et al. [102] attempt to formalise this in terms of emotional state changes trying to capture the possible cause-effect links occurring in films, based on emotion types and their relations as outlined by Ortony. Both papers attempt to visualise the types of emotion occurring across a film's timescale. Vassiliou et al. devise a comparison metric to compare the emotional information for different manifestations of the same film i.e. a film may have a screenplay and two audio description film scripts. [77] & [102] both experimented on gathering emotions from 45 audio description scripts and Vassiliou et al. also looked at 70 screenplays.

These papers deal with extracting information about emotions in a film from collateral film text data, i.e. audio description scripts and screenplays of films. Other research has extracted other film elements and information about high-level concepts from film scripts. Ronfard and Tran-Thuong [75] present a method to align a film's video segments to a continuity script through matching dialogue and subtitles and a possible framework to extract structural units of film, such as shots, scenes, actors and dialogues, from the script. They format the continuity script (updated throughout shooting of the movie and includes the breakdown of scenes and shots) into XML. Subtitles and candidate shot cues are then extracted for the respective film. The alignment between the script and the detected video segments is performed by matching a temporally sorted string of shots and dialogue lines from the continuity string with the shots and subtitles from the film. Alignment was conducted on "The Wizard of Oz" [161] continuity script and film and matched ~82% of script shots and ~88% of detected video shots

Though a method to extract structural elements of a film from film scripts is appealing to our research these scripts are specialised (as they provide a breakdown of the film) and are not readily available. Also the method relies on the analysis of the video itself and access to the subtitles. The issues of different versions of the film (DVD, international versions, directors' cuts etc.) are being ignored here and this method would require some manual verification for omitted or extra shots before alignment cold be done. The work is in its early stages.

Turetsky and Dimitrova [101] attempt to derive high-level semantic information (dialogue, character identification) from films using the films' screenplay. They align the screenplay of a film with its correct video time code using the film's time stamped subtitles. First, the screenplay is parsed to obtain the location, time and description of a scene, the lines of dialogue and speaker and the action direction for the actors. Second, the time stamped subtitles and metadata are found and pre-processed. Then the screenplay is aligned using dynamic programming to find the 'best path' across a similarity matrix. Alignment was conducted on four films with on average about 90% overall accuracy and some dialogue speaker ID was also conducted. The idea of aligning any script to a film, through dialogue and being able to automatically distinguish between dialogue and action scenes is very pertinent to our work. Also being able to distinguish which character is talking or in a scene is also relevant and useful. The fact that this is collateral text film data giving information about film elements and a film's content is interesting to us. However, this work is not developed and relies on time stamped data existing already for each film's dialogue.

2.3 Discussion: How Far are we across the Semantic Gap?

Current research into bridging the semantic gap for video content utilises mostly low-level audio-visual features for analysis. Text and collateral text analysis and fusion of all these analyses

(multimodal analysis) are also being utilised to bridge the semantic gap. However, the research has not successfully or satisfactorily bridged the semantic gap enough to make the human and machine understanding of film coincide. It is still difficult to process video data in any form due to its physical size and the complexity of a video or film's story and *essence*. We still need a better understanding of film content to allow the machine-processable representation of film, be that in the form of better modelling and frameworks of film that can be automatically instantiated or ontologies of film content that will aid said automatic instantiations. Thus, we are *not far enough* across the semantic gap.

Solutions to the issue of accessing video content have been proposed in the last two decades that involve providing information about events, characters, objects and scenes, generating overviews, summaries and abstracts of films, sports, news, CCTV surveillance, meetings and educational video material. Low-level audio and visual features in films have been utilised to categorise films and scenes in films, elicit information on characters' emotions, characters' movements in a film, moods and atmospheres in a film and types of events (action, dialogue, suspense etc. and affective content). Automatic segmentation methods have also provided opportunities to automatically locate scenes, events and shots in a film for further analysis. Textual features and collateral texts have been utilised to extract information about film content such as character's emotions and categories of event and to produce film summaries.

Film and narrative in general has been modelled 'bottom up' in terms of semantic nets, plot units, multi-perspective networks and top down: Logical Story Units (see Section 1.2.2). Frameworks have been proposed for the structuring of video content: AUTEUR, Semantic Nets, Narrative Abstraction Models, A Multi-Perspective Network, and Story-Browse. Even the MPEG-7 standard allows the annotation and modelling of narrative data in 'narrative worlds'. Working and prototype systems that browse, search and retrieve video and film data are available commercially. There is also a body to evaluate such systems in the form of TRECVID.

All the research mentioned here has in some way provided steps to bridge the semantic gap for video content where there is a lack of coincidence between automatically extracted video data provided by a computer system and the interpretation that data has for users' in a given situation. [89]. However, there is still progress to be made with respect to crossing the semantic gap, these proposals still leave issues unanswered and there has been no decision made in terms of which is the best modality to utilise when extracting high-level semantic concepts from film data.

For extracting 'user specific' high-level semantic units from films the computer must have a better 'understanding' of what the user is looking for: "a deeper understanding of the information at a semantic level is required" [17]. The concept of an ontology for film content could be considered as a method for representing an 'understanding' of film data. Frameworks that have

modelled film and stories in general have had restricted interfaces in terms of queries (as in [22], [66], [76]) but they were manually instantiated whereas on the other end of the scale Google and Yahoo searches are unrestricted but they often give inaccurate and non-specific answers. When dealing with Affective Content [42] it is the users' emotions and feelings, or cues to these, that are being captured, this may provide additional search information for users wanting to make 'mood driven' queries, e.g. find me the *sad* scenes in a film. A better understanding of what a user is looking for embedded in a representational system (e.g. MPEG-7 annotation of film content) could allow for more accurate user-specific results to a film content query.

A major issue still facing the automatic representation of film content is that there is still a need for manual instantiation of models, "Since movies are the result of an artistic process sometimes causing confusion on purpose, the necessity for human judgement will always remain to some extent." [103]. The manual indexing of film content can be an expensive and time consuming process. Therefore automatic classification of video content is necessary [28], [90].

The question of which modality is best when eliciting high-level meaning from film content is still in debate. Section 2.1.1.1 touches briefly on research that has used different types of modality to analyse film content. There was no clear evidence to support one modality over another. As mentioned earlier the ACM Multimedia Conference in San Francisco in 2003 discussed the issue that research into multimedia content should focus on multimodal sources of multimedia rather than a single modality when dealing with automatically extracting information from multimedia. Even though using more than one modality in the analysis of film content has not proven better, researchers still believe two or more modalities of analysis are better than one. After all "the content of a video is intrinsically multimodal." [[90], pg. 6].

Though the question of which modality or combination thereof is best suited to extracting high-level semantic units remains, we believe that text, specifically collateral film texts, still has merit for automatic analysis of film content. Adams and Dorai [2] refer to text [and image] analysis techniques as 'mature' and that existing interrogative techniques (text and image) "fall short in mining meaning from the unique modes open to the video medium" [ibid pg 472]. Chrystal et al. [25] however, believe, and to some extent have shown through their research, that text analysis is still more accurate when the text modality, specifically collateral texts in the form of news transcripts and closed-captions, are available with respect to retrieving video content, than just using low-level visual features of video for analysis. As Wilks [105] states the text world is vast and growing exponentially, and we should not be seduced by multimedia into thinking that text and how to deal with it, how to extract its content, is going to go away.

Thus, the current field of video content analysis is not far enough across the semantic gap to allow the coincidence of what a user views in a film and what a machine understands of the same data.

Though there have been important developments to understand, access and process video content and film content and models to represent such data, the research reviewed here is not developed enough as yet (though some is more developed than others) and this is largely due to the complex nature of semantic video content and, specifically, film and the multimodal nature of video.

We believe that the systematic study of collateral texts, in our case texts that describe films in the form of audio description scripts and screenplays, can provide information about film content. We also believe they exhibit frequently recurring textual patterns that will aid the extraction of high-level semantic film content, help automatically classify elements of film and provide evidence that audio describers and screenwriters have idiosyncrasies in their writing, showing us that collateral film texts hold a language for special purpose. We believe that our analysis of collateral texts and the patterns which we will isolate can be put to use to bridge the film content semantic gap.

3 Collocation Analysis of Film Corpora

This chapter seeks to discover basic knowledge about collateral texts for films that can be used to extract machine-processable representations of film content from these texts. It asks two questions: what information about film content do these texts contain? Are the ways in which this information is expressed sufficiently regular to allow for reliable information extraction, and the automatic generation of Information Extraction (IE) templates and algorithms? The first question is important in order to establish the potential for using these texts to cross the semantic gap. The second question is important because automated information extraction relies on information about entities and events being expressed with some regularity.

In order to answer both questions, a representative corpus of audio description scripts and a representative corpus of screenplays were analysed to identify unusually frequent open class words and collocations involving these words. Formal descriptions of frequent collocations, in the form of Finite State Automata (FSA), can be interpreted to suggest four common types of event that are commonly described in audio descriptions and screenplays. At the same time, these formal descriptions, along with the unusually high occurrence of certain open class words, suggest a high degree of regularity in how information about the four event types is expressed. These idiosyncrasies can be described as local grammars and taken as evidence that audio describers and screenwriters use a language for special purpose (LSP).

Section 3.1 introduces the idea of collocation and expands on why collocation is important to the work. Section 3.2 presents an overview of an existing, extended, mainly-automated method to generate templates and algorithms for information extraction. Section 3.3 explains the seven main stages of the method and presents results for each stage based on the analysis of a corpus of audio description scripts (73 films with 714,681 words) and a corpus of screenplays (125 films with 3,211,640 words). Section 3.4 interprets the collocation data in two ways: first, identifying the kinds of information about film content that these texts contain, specifically four types of event; second, arguing that the audio describers and screenwriters use an LSP, in particular certain idiosyncratic local grammars. Section 3.5 considers the implications of these results for designing templates to conduct information extraction to help cross the semantic gap for film content.

3.1 Theory and Techniques Chosen

Lacey states that: "Arguably what makes narrative a key concept in media studies is its usefulness in looking at texts as a whole, particularly demonstrating similarities between texts that appear completely different [55]." Since we consider films as telling a story or narrative, it is these 'similarities [and regularities] between [and within] texts that appear completely different' that we search for in the analysis of film script corpora. To answer the question: what information about film content do these texts contain?, we could search for regularities of individual words across the corpora and treat each corpus as a 'bag of words', searching for frequencies of individual words using *corpus linguistics*. Instead however we opt to examine the combinations of words and in what context they co-occur using *collocation analysis* through corpus linguistics. Collocation analysis is statistically grounded and collocations can be expressed formally as Finite State Automata (FSA) allowing any regularities in the corpora to be expressed.

These formal expressions of collocations allow us to answer the question: Are the ways in which this information is expressed sufficiently regular to allow for reliable information extraction, and the automatic generation of Information Extraction (IE) templates and algorithms? There seems to be sufficient regularity in the expressions and collocation phrases found in the corpora to provide evidence of local grammars in the corpora and a language for special purpose being used by the writers of the film scripts. The collocation FSA and local grammar formal expressions are regular enough to allow templates of common events in films to be established which may allow for reliable information extraction of film content.

This section defines the theories and techniques chosen to examine the corpora of audio description scripts and screenplays to investigate whether regularities exist and how they can be expressed formally.

3.1.1 Why do Collocation Analysis?

Collocation analysis is a technique from *corpus linguistics* and is defined as a sequence of words, which co-occur more often than would be expected by chance, and that correspond to arbitrary word usages and words that a *nucleate* or target word commonly co-occurs with [13]. There are indications that collocations are pervasive in English; they are common in all types of writing, including both technical and non-technical genres [127]. Generally, collocations are word pairs and may be neighbours or may co-occur with other interspersing words. They are the way in which words are used together regularly, e.g. which prepositions are used with particular verbs, or which verbs and nouns are used together. As an example, in English the verb *perform* is used with *operation*, but not with *discussion*: "The doctor performed the operation." Smadja states that they are domain-dependent and that domain specific collocations are numerous. Technical jargons are

often totally unintelligible for the layman and contain a large number of technical terms. "Linguistically mastering a domain, such as the domain of sailing, thus requires more than a glossary; it requires knowledge of domain-dependent collocations."[[88] Pg.147] Collocations have particular statistical distributions. This means that, for example, the probability that any two adjacent words in a sample will be "red herring" is considerably larger than the probability of "red" times the probability of "herring." The words cannot be considered as independent variables. This fact has been taken advantage of to develop statistical techniques for retrieving and identifying collocations from large text corpora [88].

We wish to investigate what information audio description scripts (AD) and Screenplays (SC) provide about film content. Our domain is *film*, thus we must consider what Smadja's statement that mastering a domain requires domain-dependent collocations. Also, we wish to explore what *form* the film content information can take. The form of a *collocation* seems a robust, data driven, statistically grounded way to express any film content information found through the analysis of AD and SC corpora. Thus we wish to explore the statistically significant, arbitrary and recurrent co-occurring word combinations that exist in the film corpora in order to elicit information about film content.

3.1.2 Why Choose Corpus Linguistics?

Corpus linguistics is the study of language as expressed in samples (corpora) or "real world" text [128]. The essential characteristics of corpus linguistics are that it is empirical and analyses the *patterns of use* in texts, i.e. the lexical patterns of general language used in texts. It utilises a large principled collection of natural texts from a disciplined domain (such as biochemistry, physics), known as a 'corpus', as the basis for analysis. It depends on both quantitative and qualitative analytical techniques and makes use of computing for analysis, using both automatic and interactive techniques [[13] pg. 4]. The analysis of a 'representative corpus' can provide information about language use, in particular a corpus-based approach allows the identification and analysis of complex *association patterns*: the systematic ways in which linguistic features are used in association with other linguistics and non-linguistic features [[13] pg. 5].

Corpus linguistics has also been used in the field of Information Extraction (IE) from text in many domains. The FASTUS system for instance [46] extracts information from natural language texts for entry into a database. It describes five systematic stages for the extraction of 'event structures' (a series of phrases and patterns corresponding to an event or theme) which are merged, where appropriate, to form the templates for database entries. The system can be theoretically applied to any corpus of texts to elicit 'event structures'. Poibeau [70] describes a corpus-based approach to IE. An IE system is described, intended to extract structured information from general texts. They

propose an integrated framework to semi-automatically acquire resources to feed an IE system [70] [71]. Their goal is to acquire knowledge and structured information from a corpus.

Thus, corpus linguistics is a well established method for analysing large corpora, is empirical and concentrates on the statistical and logical modelling of natural language in corpora. Corpus linguistics explores *patterns of language use* in representative corpora to identify association patterns. Corpus linguistics has also been utilised for IE to populate databases using 'event structures' derived from corpora for a specific domain as templates. In our case we wish to learn about the patterns of language use in AD scripts and SC, thus we employ a representational corpus for both and techniques and methods from corpus linguistics, specifically collocation analysis. We also wish to populate a database of film content information about films and believe that 'event structures', or templates, from the AD and SC corpora can be elicited using collocation analysis, to capture information about film structure and narrative elements such as story (existents and events in film).

3.1.3 Why Introduce the Idea of LSP and Local Grammar?

A Language for Special Purpose (LSP) refers to a language used by experts in a particular subject field: "…within specialist communities the spoken and written communications of experts lead to the development of a special language with an idiosyncratic vocabulary and grammar; in particular special languages tend to have a profusion of terms with which to articulate the important concepts of a domain, and a restricted grammar to minimise ambiguity and ensure precise communication of ideas", from [78] quoting [5].

Audio description scripts are created by trained professionals and screenplays are written by experienced screenwriters, therefore the film scripts have a set of conventions that are followed when produced, possibly providing idiosyncrasies in the texts. We wish to exploit these 'communicative needs of language users (AD and SC writers)' and we believe these idiosyncrasies will provide strong evidence towards showing that a language for special purpose exists in these specialist texts. We believe that frequent language used in the film scripts will correspond to common, important film content and that there exists a language for special purpose for both audio description and screenplay film scripts that differs from general language. Our approach is to use corpus analysis techniques to identify idiosyncratic linguistic features in a collection of audio description scripts and screenplays.

Linguistic phenomena such as technical jargon, idioms or clichés lead to common syntactic constraints that can be accurately described locally. Local grammars are rules that govern the simultaneous choice of a set of words used in a specialist context [98]. "Local grammars can be used to represent such linguistic phenomena [and]…can be compactly represented by a finite

[state] automaton" [[65] pg. 84]. Gross states "Local grammars are finite-state grammars or finite-state automata that represent sets of utterances of a natural language" [[37] pg. 229]. Thus the collocations that could be extracted from the film corpora for film content can take the form of Finite State Automata (FSA). We believe that frequent, repeating phrases (collocations) exist in film scripts which differ from general language and could be considered local grammars as they may refer to a restricted set of phrases concerned with a type of film content.

A local grammar is a constrained set of words that could concurrently be used in a specific statement or context. Finding local grammars in any text domain is not a trivial task. Gross describes a method for constructing local grammars around a keyword or a semantic unit [37]. It was suggested that Gross's work could be extended, in that a corpus of texts could be used systematically to find local grammars [98]. Our work expands the method presented in [31] for systematically constructing local grammars and applies it to the domain of texts that describe film: AD and SC scripts.

3.2 Overall Method

To investigate what information audio description scripts and screenplays provide about film an existing method was expanded and adapted to accommodate film scripts. The method was first developed at the Department of Computing at Surrey University and first appeared in FINGRID [31] which was a project analysing financial information texts and involved the analysis of textual data about financial markets to gain insight into perceptions about the market and market sentiment (Time series and news data)[9]. The corpus linguistic method was used to elicit frequently recurring phrases and words to populate a database and thesaurus of terms about the financial markets and has since appeared in Ahmad et al. 2005 [6], which extracts Local Grammars and 'sentiment' in financial texts. The method described in [31] has also been adapted by Almas et al. [8] for the analysis of Arabic and English news articles. In our work the author has adapted the method to apply to the film script corpora and extended the method by adding sections that semi-automatically expand the collocation analysis, generalise and abstract the collocation results and possibly generate representations of local grammars.

Sections **I.-VI.** give a brief overview of the method and tools used throughout the method are introduced. Figure 14 provides a visual breakdown of the steps involved in the method.

[9] FINGRID [31] at: http://www.esrc.ac.uk/ESRCInfoCentre/

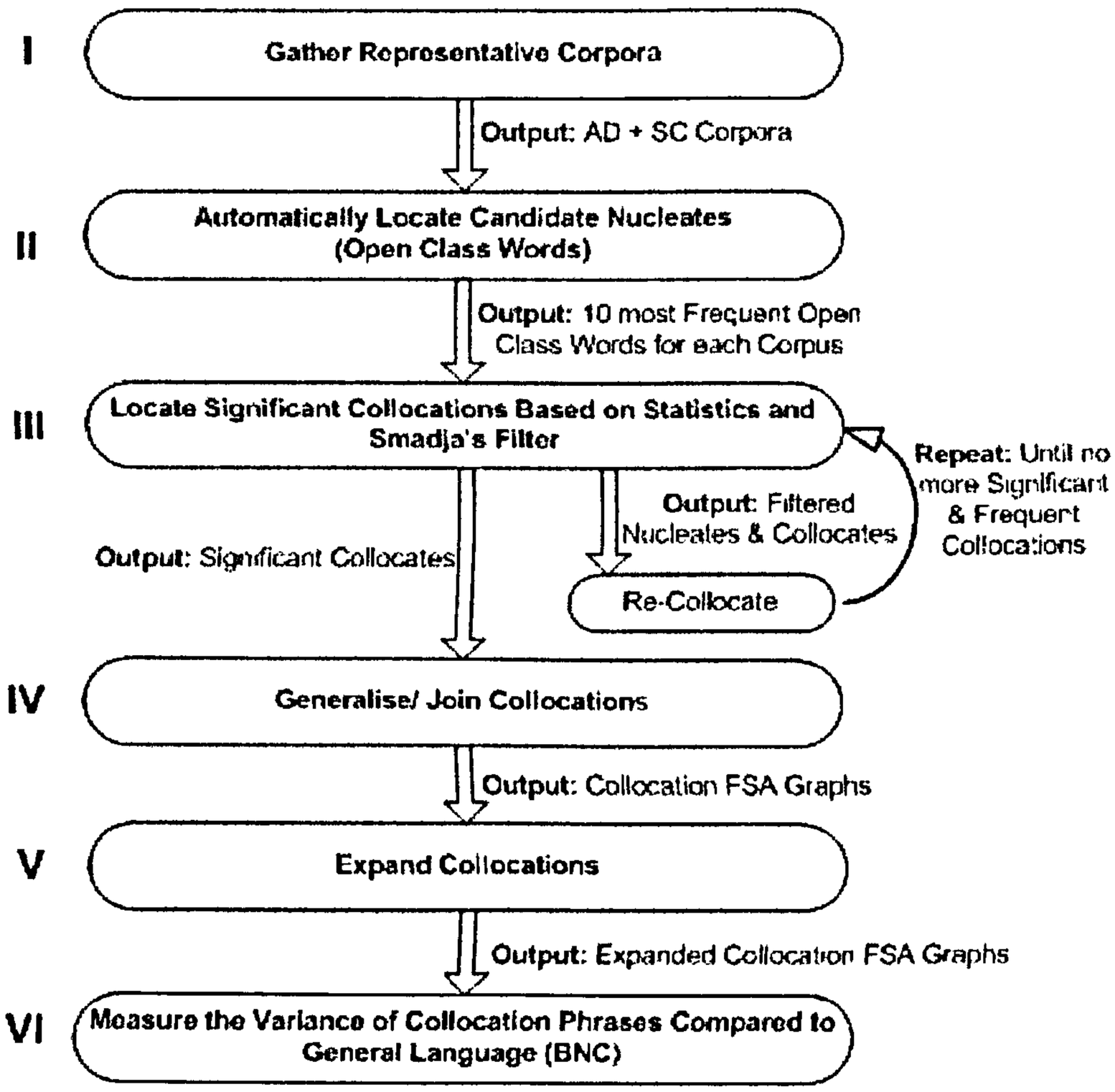


Figure 14 The six stages of Corpus Linguistic analysis employed to investigate what information audio description scripts and screenplays provide about film content and the form that information can take.

3.2.1 Step I. Corpus Gathering

Corpus linguistics involves the study of representative corpora, thus accordingly we had to gather representative corpora for films. Many texts that describe or transcribe films exist to choose from. Plot summaries give an overview of films' plot and have been analysed using corpus linguistics [97]. However they are lacking in key information about film content as they are merely short summaries of films. Reviews of films tend to describe the film as a whole and not the films' content. Subtitles are a rich source of film content but are mostly concerned with dialogue and not enough description about what is happening in the film.[10] The TIWO [96] project presented us with a film collateral text: the Audio Description (AD) script that was written by trained professionals (BBC [129], ITFC [130] and RNIB [131]) and described where possible, in detail, what was happening in any given scene of a film. AD is conveyed via another soundtrack that takes advantage of the audio breaks, recorded over the original soundtrack of a film to communicate what is happening in a film, to the visually impaired. The fact that it provides a rich

[10] Subtitles for the hearing impaired give more details but still do not provide a rich description of what is occurring visually.

description and that it is scripted makes it a reliable and quite accurate source of film content. AD has the bonus of being time-coded allowing precise information extraction of film content.

Thus, an AD corpus was compiled by colleagues in the TIWO project: BBC, ITFC, RNIB British organisations that produce AD scripts. This consisted of a total of 73 films with 714,681 words. To try and ensure a representative corpus we consulted two audio description experts at ITFC [130] and established nine categories of films in terms of how the experts thought audio description would vary. Hence, we believe this to be representative of the AD script language and by extension representative of the film language.

Another collateral text that described what was happening in films was film scripts or screenplays[11]. Screenplays are written by screenwriters for Hollywood movies and, over the 100 or so years of cinema, have adopted a certain format that they adhere to. S*creenplays* include many incarnations: Early draft, first, second and third drafts and Final drafts (the film at different stages of production or story edits), the shooting script (the script used while filming), post production script (the script released with all changes after film has been completed) and transcripts (where people have transcribed the film). We believe that the screenplays to be a rich source of film content, and film structure as well, as they are designed for directional purposes and contain most elements of the film (e.g. locations, characters, dialogues, descriptions of events and scenes, shots, scenes and acts). Also we believe that screenplays are complementary to audio descriptions as audio descriptions exist separately (separate soundtrack) to films and explain more of what is *visually* occurring in a film.

A representative corpus of screenplays was gathered of 125 films with 3,211,640 words, from various movie enthusiast Internet sites[12]. To try and ensure a representative corpus we kept in mind a balance between gathering a variety of films, film genres and types of script, i.e. some final drafts, first drafts, film transcripts, post production scripts etc. We also gathered films to complement the AD corpus.

Both corpora were sorted into film genres according to the Internet Movie Database [136] and basic information about the films, such as length, was also recorded. A test corpus was also set aside for each of the corpora: 38 AD and 75 SC scripts. Details of the film titles, film information and genre classification for both corpora can be seen in the film scripts section at the end of the report.

Having gathered the representative corpora for films we investigated the collocations that existed in the corpora to study what information about film content they contain.

---

[11] Throughout this report we use ***Screenplays*** to represent all types of Hollywood film script.

3.2.2 Identifying Significant Collocations

II. Locate Candidate Nucleates

Firstly we find the *frequency* of the words in each corpus and then contrast the frequency of the words in each corpus to the frequency of a general language corpus, in this case the British National Corpus (BNC)[13]. This gives us the *relative frequency* of each word in each of the corpora. Then the relative frequency of each word, in each corpus, is compared to the relative frequency of that word in the BNC, giving us a *weirdness*[14] ratio for each word. The *average frequency* and *standard deviation* of frequency of all words in each corpus and the average weirdness and standard deviation of weirdness of all words in each corpus are calculated. Finally we calculate the *z-score*s of frequency and weirdness for each corpus. Based on the z-scores, we filter out the open class words (from closed class) by the *Relative Frequency z-score* being > 1 and the *weirdness z-score* being greater than the **median** of the *weirdness z-score* values.

III. Locate Significant Collocates

The first ten open class words from each corpus are taken as *nucleate* words, i.e. words to be collocated.[15] The frequency of the nucleate word co-occurring with any other words in the corpus is then computed and five words to the left and five to the right of the nucleate words are examined. Statistics of frequency, *k-score*, *U-score* and *p-strength* are calculated and Smadja's [88] tuple {U-score=10, k-score=1, p-strength=1[16]} is used to filter out significant collocates for all nucleate words. The resultant *nucleate* and *collocate* phrases, which have been filtered as significant collocations, are treated as new *nucleate phrases* and *re-collocated* using Smadja's tuple to filter out significant collocates. The re-collocation process is continued until there are no more frequent, significant collocations i.e. above a frequency of five. The collocation phrases and statistical information are recorded at each stage of the process.

IV. Generalise/Join Collocations

Once a set of collocations have been found for each open class word we generalise the collocation phrases to remove redundancies and present a more abstract or simplified expression for the open class word and its frequent collocates. Based on an algorithm by Fargues [30] we are able to *join*

---

[12] www.script-o-rama.com/ [132]; www.simplyscripts.com/ [133]; Internet Movie Script Database: www.imsdb.com/ [134] and Daily Script: www.dailyscript.com/movie.html [135] Last Accessed 17/06/06

[13] The BNC is a general English language sample of over 100,000,000 words.

[14] **Weirdness**: comparison of the relative frequencies of two corpora, one of which being the mother corpus

[15] Also called *node* or *target* words

[16] p-strength = 1 is for general language a p-strength of 1.5 was found to reveal more frequent collocations

the collocation phrases where possible and *simplify* to remove any redundancies (*maximum overlaps*). The generalisation of the collocations results in a Finite State Automaton (FSA) for each nucleate open class word.

V. Expand Collocations

The nucleate word of each resulting generalised collocation is replaced with a wildcard and searched for in the corpora. Alternative nucleates occurring with frequency of above 10% of the overall number of instances are selected. These alternative nucleates are recorded and a predominance test is conducted to see whether the original nucleate word is the most predominant or frequent above all the alternative nucleates.

VI. Measure Variance of Collocations with General Language (BNC)

To compare the collocations to a 'ground truth' set of general language we compare the collocations to the British National Corpus. Each collocation is taken in turn and searched for in the BNC. The number of instances of each collocation (frequency) in the AD or SC corpus is compared to that in the BNC. Then, the relative frequency of each collocation (including BNC) is calculated. Then *t-scores* and *mutual information* (MI) statistics are calculated for all instances of the words in the corpora as a basis for comparison to measure the variance of each collocation to that of statistics of the same words in the BNC.

3.2.3 Tools used for Corpus Linguistics and Statistical Calculations

SystemQuirk is a *Language Engineering Workbench*, developed by the University of Surrey: Department of Computing [137], available for free. It is a package of integrated tools for building and managing term bases and makes text analysis techniques available. It allows a corpus of texts to be analysed and compared to the BNC. ***Kontext*** is a module of SystemQuirk that handles text analysis. It is language independent and can generate word lists, indexes and calculate weirdness values against the BNC. ***ColloQuator*** and ***COLLOCATOR*** are prototype systems in SystemQuirk that calculate statistics of the strength of co-occurrence of nucleate words to other words in a corpus.

Unitex, Founded at LADL (Laboratoire d'Automatique Documentaire et Linguistique), under the direction of its Director, Maurice Gross [138], is a corpus processing system, based on automata-oriented technology. It allows collocation, concordance and frequency analysis of corpora. One of

its major functions is pattern matching with regular expressions and recursive transition networks. In this work it is predominately used for concordances and representing FSA.

3.3 Detailed Method, Results and Examples

This section provides an in-depth look at how the method was implemented, examples of the method in use and results at each stage. The rest of this section describes, in detail, a step by step breakdown of the method and presents any definitions needed to understand that method.

3.3.1 II. Identify Candidate Nucleates

This section identifies the candidate nucleates from both corpora (AD and SC) based on the weirdness of the frequency of the words in the corpora compared to the BNC. Figure 15 shows the steps of the method for this stage.

I. INPUT $CORPUS_{GL}$ /* a general language corpus comprising N_{GL} individual words*/
INPUT $CORPUS_{FL}$ /* a corpus of specialist film texts comprising N_{FL} individual words*/

II. **CONTRAST** the distribution of words in $CORPUS_{GL}$ and $CORPUS_{FL}$
COMPUTE Frequency $n_{FL}(w)$ /*of all words, *w* in $CORPUS_{FL}$*/
Frequency $n_{GL}(w)$ /*of all words, *w* in $CORPUS_{GL}$*/
Relative frequency $f_{FL}(w) = n_{FL}(w)/N_{FL}$
Relative frequency $f_{GL}(w) = n_{GL}(w)/N_{GL}$
$WEIRD(w) = f_{FL}(w)/f_{GL}(w)$ /* weirdness ratio for each word *W**/
$avg_f := (\sum f_{FL}(w))/ N_{FL}$ /* average frequency of all words, *w* in $CORPUS_{GL}$*/
$\sigma_{frequency} := (\sum (f_{FL}(w) - avg_f)^2/ (N_{FL}*(N_{FL}-1))$ /*stdev of frequency of all words, *w* in $CORPUS_{GL}$*/
$avg_{weird} := (\sum WEIRD(w))/N_{FL}$ /* average weirdness ratio for each word *W**/
$\sigma_{weird} := (\sum (WEIRD(w) - avg_{weird})^2/ (N_{FL}*(N_{FL}-1))$ /* stdev weirdness ratio for each word *W**/
$z_{frequency}(w) := (f_{FL}(w) - avg_f)/ \sigma_{frequency}$ /* Z-score based on frequency of *w* in $CORPUS_{FL}$*/
$z_{weird}(w) := (WEIRD(w) - avg_{weird})/ \sigma_{weird}$ /*Z-score based on weirdness of *w* in $CORPUS_{FL}$*/

SELECT Open Class words

a. **LET** S^{OPEN} and S^{CLOSED} be the sets of open and closed class words
b. **CATEGORIZE** Words
   i. IF $WEIRD(w) >> 1$ THEN $w \in S^{OPEN}$
   ii. IF $WEIRD(w) \approx 1$ THEN $w \in S^{CLOSED}$
   iii. IF $z_{frequency}(w) > \tau_{frequency}$ & $z_{weird}(w) > \tau_{weird}$ THEN $w \in S^{OPEN}$
c. **COMPUTE** h_0 the median $\forall$ $z_{weird}(w)$, h /* $h_0 = 0.005$ for this work */
d. **ADD** h_0 Until change in number of S^{OPEN} words >20% number of previous words /*i.e. at least 20% drop in number of S^{OPEN} words */
e. **RECORD** S^{OPEN} Words when > 20% drop in number of words
f. **SELECT** CANDIDATE NUCLEATE Words
   i. 10 most frequent words from S^{OPEN} Words

Figure 15 Pseudo code of the method for automatically choosing open class words from corpora.

3.3.1.1 Identify Candidate Nucleates: Method, Definitions and Tools

After the corpora had been gathered (step I.), the following statistics were calculated for each word in each corpus:

Frequency (absolute): instances of a word that exists in a corpus thus disclosing how frequent that word is in a corpus.

***Relative frequency*:** frequency of each word over the total number of words in the corpus. It allows relative comparison to other corpora.

Weirdness: comparison of the relative frequencies of two corpora, one of which being the mother corpus. In our case the relative frequencies are divided by the British National Corpus relative frequency values [57].

$$Weirdness = \frac{W_s / t_s}{W_g / t_g} \quad [\alpha]$$

Where: W_s & W_g= frequency of word in specialist language/general language corpus respectively

t_s& t_g = total count of words specialist language/general language corpus respectively [4]

Z-scores: derived by subtracting the sample mean from an individual (raw) score and then dividing the difference by the sample standard deviation.

$$z = \frac{X - \bar{X}}{s} \quad [\beta]$$

The dimensionless quantity *z* represents the number of standard deviations between the raw score and the mean; it is negative when the raw score is below the mean, positive when above. [139]

The ***average relative frequency*** for words in the corpora and the ***standard deviation*** of the *frequency* of all the corpora's words are calculated as well as for the *weirdness* of the corpora's words. ***z-score*** values for Relative Frequency and weirdness for both corpora are calculated.

Based on the z-scores, we filter out the open class words (from the closed class) by the *Relative Frequency z-score* being greater than 1 and the *weirdness z-score* being greater than the **median** of the *weirdness z-score* values $\{h_0\}$. The value above the median *h* was altered by adding $+0.005$[17] at every iteration, giving a different set of open class values each time. The iteration that involved at least a 20% drop in the number of candidate words was taken as the ***open class words*** for that corpus (see Table 5 & 6 for examples). In both of our corpora it was found that 0.01 above the median produced at least a 20% drop in the number of open class words which may be an indication of a usable *h* value threshold. The top ten most frequent open class words for each corpus are then considered as candidate nucleates. Any redundancies and possible errors are examined and removed if necessary. A final list of candidate nucleates, from both corpora is presented in Table 1.

3.3.1.2 Identify Candidate Nucleates: Results

Sixteen open class (see Table 1) words were chosen as the candidate nucleate words to be analysed for collocations in both corpora.

Table 1 Tables depicting the open class words for both corpora, for the 100 most frequent terms in each corpus and the *candidate nucleates:* top ten most frequent open class words, selected from both corpora, after removing redundancies and errors.

Audio Description Candidate Nucleates	*looks, door, turns, head, towards, eyes, room, takes, walks, behind*
Screenplay Candidate Nucleates	*int, looks, day, ext, night, room, door, around, away, head*
Candidate Nucleates	**looks, door, turns, away, head, towards, eyes, room, takes, around, walks, behind, int, ext, day, night**

3.3.1.3 Identify Candidate Nucleates: Detailed Results and Discussion

Frequency Analysis: Audio Description and Screenplay Corpus Results

In this analysis we concentrated mostly on the top 100 most frequent words in the corpora as they usually constitute ~45% of the amount of words in the whole corpus. In the Audio Description corpus the first 100 most frequent terms account for 43.4% (about 175,000 terms), of the total corpus, with the first 50 terms making up 37.5% of the corpus alone. Table 2 is a *percentile* table which shows the first 100 terms, split into ten terms per row, along with the percentage of the respective ten terms in the entire corpus (created using System Quirk). See Ahmad et al. [4].

Table 2 Shows the 100 most frequent words in percentiles with the open class words highlighted for the Audio Description corpus.

AD

Frequent Words	% of Corpus
the,a,and,of,to,his,in,he,on,her	22.4
at,she,up,with,it,him,is,as,out,into	6.6
down,back,looks,from,they,over,by,door,you,through	3.9
off,then,i,man,them,turns,away,head,one,an	2.5
are,for,towards,eyes,default,hand,their,face,around,room	2.0
who,takes,two,walks,behind,car,sits,that,across,hands	1.6
white,stands,tom,other,men,open,john,side,pulls,smiles	1.3
stares,goes,look,round,onto,puts,steps,front,watches,another	1.2
along,all,water,again,opens,table,black,this,but,inside	1.1
window,runs,stops,has,way,me,outside,its,woman,bed	1.0
Total Percentage of Corpus	**43.4 %**

[17] A small number (0.005) is needed to bring about a 'controlled' change in the number of candidate nucleate words. The number was chosen after trial and error.

It was interesting to note that the corpus contained frequent use of words that appear to refer to: character names {john, tom}, characters staring or looking {looks, stares, watches} and character actions {turns, walks, sits, takes, stands, pulls, smiles, steps, opens, runs, stops}. These terms all reflected the descriptive nature of audio description and were used mainly in descriptive phrases to describe scenes and shots to the visually impaired viewer throughout the corpus.

In the case of the Screenplay corpus the first 100 most frequent terms account for approximately 44.2% (about 870,000 terms), of the total corpus, with the first 50 terms making up 36.5% of the corpus alone, Table 3.

Table 3 Shows the 100 most frequent words in percentiles with the open class words highlighted for the Screenplay corpus. (Italicised words are *possible* open class words).

SC

Frequent Words	% of Corpus
the,a,to,and,of,you,I,in,it,he	20.4
is,his,on,at,with,that,as,we,up,her	6.6
out,him,for,she,they,from,what,are,this,into	4.2
int,back,me,but,down,one,have,be,all,can	2.9
not,there,your,an,looks,then,my,like,just,over	2.4
by,don,day,do,them,off,ext,was,now,know	2.1
through,night,get,room,see,door,man,about,here,two	1.8
if,who,so,has,around,their,will,right,look,go	1.5
away,head,turns,got,eyes,hand,time,where,other,some	1.2
face,when,going,how,cut,come,think,car,way,been	1.1
Total Percentage of Corpus	**44.2 %**

This corpus contained specific words that were not commonly used in everyday language {ext, int}. They referred to whether a scene is an internal (INT) and external (EXT) one. The words {day, night} were also very frequent, describing whether a scene is a daytime or nocturnal scene. There were also frequent terms that possibly referred to characters staring or looking {looks, see} and characters' actions {go, turns, come, think}.

Automatically Choosing Open Class Words through Z-scores

Only the 100 most frequent words in the corpora were of interest, here as they made up over 40% of the corpora. Thus, they were taken as 'samples' to extract the open class words from the two corpora (see Table 2 and Table 3). Table 4 shows the metric: {weirdness>1, Relative Frequency Z-score >1, weirdness Z-Score > (median of weirdness Z-Score + 0.01)}. In the 23 most frequent words the word 'looks' is flagged as an open class word according to the metric.

Table 4 Shows the 23 most frequent words in the AD corpus and their respective frequencies, relative frequencies, weirdness, and the relative frequency and weirdness z-scores. 'looks' has been flagged as an open class word by the open class metric (Relative Frequency Z-score>1, Weirdness Z-score > h_1).

Word	Frequency	Relative Freq.	Weirdness	Rank	Relative Freq. Z-score	Weirdness Z-score	Flag
the	28335	0.04	0.59	1	99.34	-0.107	0
a	12094	0.02	0.72	2	42.36	-0.107	0
and	10598	0.01	0.51	3	37.11	-0.107	0
of	6688	0.01	0.29	4	23.39	-0.108	0
to	6185	0.01	0.31	5	21.63	-0.108	0
his	6184	0.01	1.83	6	21.62	-0.106	0
in	5836	0.01	0.40	7	20.40	-0.107	0
he	5225	0.01	0.99	8	18.26	-0.107	0
back	1799	0.00	2.26	22	6.24	-0.106	0
looks	1797	0.00	20.03	23	6.23	-0.091	1

Audio Description and Screenplay Corpus Automatically Extracted Open Class Words

For the audio description corpus the value h_0 was -0.10629, h_1 onwards were calculated by adding +0.005 (chosen arbitrarily). The results of this comparison can be seen in Table 5. The set of open class words at h_2= -0.10529 is chosen as there is a steady reduction of open class words after that i.e. for h_3-h_6 the number of open class words only drops by one word per iteration. Thus it was decided that the open class words retrieved *before* this steady change would be taken as the set of open class words. h_2 was 0.01 above the median h_0 for the weirdness z-score. There were 32 open class words at h_2 for the audio description corpus.

Table 5 Open Class words, at varying levels of weirdness z-score and Relative frequency z-score, for the AD corpus. The row in Italics is chosen as the main set of open class words for the AD corpus at h_2.

	Weirdness Z-Score	Open Class Words (in first 100 words)
	h_1: -0.10579	down, looks, door, turns, away, head, towards, eyes, default, hand, face, room, takes, walks, behind, car, sits, across, hands, white, stands, tom, pulls, smiles, stares, goes, onto, puts, steps, front, watches, along, opens, table, inside, window, runs, stops, bed
Chosen Open Class words →	**h_2: -0.10529**	*looks, door, turns, head, towards, eyes, default, room, takes, walks, behind, sits, across, hands, stands, tom, pulls, smiles, stares, goes, onto, puts, steps, front, watches, along, opens, inside, window, runs, stops, bed*
	h_3: -0.10479	looks, door, turns, towards, default, takes, walks, sits, hands, stands, tom, pulls, smiles, stares, goes, onto, puts, steps, front, watches, opens, inside, window, runs, stops

For the Screenplay corpus the value h_0 was -0.10368, h_1 onwards were calculated by adding +0.005. The open class words that were flagged from altering h can be seen in Table 6. The set of open class words at h_2= -0.10268, as with the audio description corpus, h_2 was 0.01 above the median h_0 for the weirdness z-score. There were 17 open class words retrieved; less than for the audio description corpus.

Table 6 Open Class words, at varying levels of weirdness z-score and Relative frequency z-score, for the SC corpus. The row in Italics is chosen as the main set of open class words for the SC corpus at h_2.

Weirdness Z-Score	Open Class Words (in first 100 words)
h_1: -0.10318	up, out, him, int, back, down, looks, don, day, off, ext, night, room, door, man, around, look, away, head, turns, eyes, hand, face, cut, car
h_2: -0.10268	*int, looks, don, day, ext, night, room, door, around, away, head, turns, eyes, hand, face, cut, car*
h_3: -0.10218	int, looks, don, ext, night, room, door, turns, eyes, cut

Chosen Open Class words →

The words 'don' (SC) and 'default' (AD) were ignored as 'don' referred to the word 'don't' and was incorrectly picked up by the software and 'default' is a word used as a *tag* in audio description scripts and thus not part of the language of the script but part of the formatting.

3.3.2 III. Identify Significant Collocations

This section identifies significant collocation phrases from both corpora (AD and SC) based on the candidate nucleate words in both corpora.

Figure 16 shows the steps of the method for this stage.

III. **FIND COLLOCATION** patterns for each S^{OPEN} CANDIDATE NUCLEATES
  a. FOR EACH WORD S^{OPEN} CANDIDATE NUCLEATES IN BOTH CORPORA
    i. **COMPUTE** $n(W, W^{\#})$/*frequency of a word W co-occurring with $W^{\#}$ -where $W^{\#}$ is any other word in the corpus- interspersed by any k words; where $-m \leq k \leq m$, and m=5 (Smadja 1991)*/
    ii. **EXTRACT** Significant collocates $W+W^{\#}$ and/or $W^{\#}+W$ by based on *z-scores* and other moments of $n(W, W^{\#})$ where $k = \pm 1$
    iii. **REPEAT** Steps i & ii with $W\,W^{\#}$ or $W^{\#}\,W$ as the new term to collocate
    iv. **REPEAT** Steps iii until there are no more significant collocates to extract.
    v. **EXTRACT** Sentence or Phrase for W^N $\{X_{1m}\}$ /*where m=aa…zz*/
    vi. **EXTRACT** Significant collocates $W+W^{\#}$ and/or $W^{\#}+W$ where $k = \pm 2\text{-}5$
    vii. **REPEAT** Steps iii-iv until there are no more significant collocates to extract.
    viii. **EXTRACT** Sentence or Phrase for W^N $\{X_{2m\text{-}5m}\}$
  FILTER COLLOCATION PHRASES $\{X_N\}$ against Smadja Inequalities
  b. FOR EACH WORD S^{OPEN} CANDIDATE NUCLEATES IN BOTH CORPORA
    i. **COMPARE** All $\{X_N\}$ (U-score, K-score, P-strength)} against {U-Score=10, K-score=1 , P-strength=1} /*(Smadja 1991)*/
    ii. IF $X_N(U, K, P) \geq (10,1,1)$ THEN ACCEPT $\{X_N\}$
    iii. IF $X_N(U, K, P) < (10,1,1)$ THEN DECLINE $\{X_N\}$
    iv. **EXTRACT** $\{Y_N\}$ = ACCEPTED $\{X_N\}$

Figure 16 Pseudo code of the method for systematically identifying significant collocations.

3.3.2.1 Identify Significant Collocations: Method, Definitions and Tools

Collocation: an arbitrary and recurrent word combination. Collocates are word pairs and may be neighbours or may co-occur with other interspersing words [88].

Re-Collocation: the process of taking a nucleate word, and its collocate, as a *new nucleate phrase* and collocating again. This process can continue as long as there are statistically significant collocations remaining.

U-score: subtracts the minimum of the range and divides by the range. It creates a standard uniform random variable. In this case it refers to [ϵ_2].

k-score: refers to the *strength* of a word score and is closely linked to z-score. It represents the number of standard deviation above the average of the frequency of the word pair w and w_i and is defined as [88]:

$$k_i = \frac{freq_i - \bar{f}}{\sigma} \qquad [\gamma]$$

Due to the varied nature of collocations Smadja [88] describes that up to five left and five right 'neighbours' of a high frequency word appear to be significant. He defines a peak or lexical relation containing a high frequency word w as a tuple (w_i (collocate term), *distance*, *strength*, *spread*, j (p_i^j)) that verify certain inequalities:

$$strength = \frac{freq_i - \overline{f}}{\sigma} \geq k_0 \qquad [\delta]$$

$$spread \geq U_0 \qquad [\epsilon_1]$$

$$U_i = \sum_{j=1}^{10} (p_i^j - \overline{p}_i)^2 \qquad [\epsilon_2]$$

$$p_j^i \geq \overline{p}_1 + (k_1 \times \sqrt{U_i}) \qquad [\zeta]$$

Where: Equation [δ] is used to eliminate low frequency collocates. The $freq_i$ is the frequency of collocation w_i with the nucleate term w; $\overline{f}$ is the average frequency, σ the standard deviation and k_0 is the strength threshold. This threshold usually has a value of one for the task of language generation (according to [88]).

Equation [ϵ_1] requires that the histogram of the ten relative frequencies of the appearance of w_i within five words of w to have at least one spike. The histograms are rejected if the variance threshold U_0<10 according to Smadja's research. The variance is equated using [ϵ_2] where p_j^i and $\overline{p}_i$ are the frequency of one collocate at certain distance from w and their average respectively.

Equation [ζ] finds the significant relative positions of two words. This inequality eliminates columns whereas inequalities [δ], [ϵ_1] and [ϵ_2] select rows. It states the frequency threshold of one collocate at certain distance from *w* be at least one standard deviation above the average frequency of one row collocates (k_1=1). However the value of k_1 is task dependent and can be adjusted to suit the task at hand [88].

For each nucleate these statistics were calculated for every neighbouring word that occurred ± 5 positions to the left and right of the nucleate. COLLOCATOR [137], was able to automatically calculate these statistics for the corpora allowing the inequalities to be easily used to eliminate certain collocate terms w_i that did not strongly collocate with the node terms *w* and was also used as a basis to further investigate other terms.

COLLOCATOR was used to calculate the most frequent collocations and re-collocations of the candidate nucleates for both corpora with respect to the ten neighbouring words (5 to the left and right) that co-occurred; i.e. the candidate nucleates were examined for *both* corpora in terms of statistical information of the strength of co-occurrence of other words to the candidate nucleates. The collocations and re-collocations were recorded along with their relevant U-scores, k-scores and p-strength (k_x). The second phase involved flagging which frequent phrases from the collocation analysis results were *accepted* according to Smadja's inequalities {[δ], [ϵ] & [ζ]}= {U-score>10, k-score>1, p-strength>1}. (Section 3.3.2.3 shows an in depth example of the method.) Downward and upward collocating was conducted to see whether there were any words that collocated strongly with each other, most of the nucleate words seemed to collocate with more frequent words (upwards collocation).

3.3.2.2 Identify Significant Collocations: Results

Collocation Analysis Results: Audio Description Corpus

In the audio description corpus 44 frequently collocating phrases were accepted according to the tuple {U_0=10, k_1=1, p_0=1} outlined by Smadja, with 21 frequent phrases being rejected. Diagrams of the frequent phrases found by ColloQuator can be seen in Appendix A. Table 7 shows the 44 collocate phrases of the open class candidate words. Since both corpora were collocated the same open class nucleate words it must be noted that the words 'day', 'night', 'int' and 'ext', which are highly frequent in the SC corpus, did not produce any significant collocates for the AD corpus.

Table 7 The 44 collocate phrases of the AD corpus' 10 most frequent open class words greater than the tuple $\{U_0=10, k_1=1, p_0=1\}$. ▲ and ▼ denote upward and downward collocates respectively.

AD Corpus Collocation Phrases			
looks up	▲	she shakes her head	▲
looks up at	▲	towards the	▲
looks up at the	▲	back towards the	▲
and looks up at the	▲	his eyes	▲
he looks up at the	▲	closes his eyes	▲
she looks up at the	▲	her eyes	▲
looks at	▲	closes her eyes	▲
looks at the	▲	the room	▲
and looks at the	▲	of the room	▲
the door	▲	out of the room	▲
opens the door	▲	takes a	▲
and opens the door	▲	he takes a	▲
turns to	▲	and takes a	▲
she turns to	▲	she takes a	▲
he turns to	▲	around the	▲
away from	▲	looks around the	▲
away from the	▲	and walks	▲
his head	▲	and walks away	▲
shakes his head	▲	behind him	▲
he shakes his head	▲	door behind him	▲
her head	▲	the door behind him	▲
shakes her head	▲	behind him and	▲

It is interesting to note that the majority of the phrases in Table 7 are *actions*.

Table 8 shows the 36 collocate phrases of the open class nucleate candidate words in the ± two and five positions, most words collocate with more frequent words (upward collocation) though there are some downward collocating words (indicated in grey in Table 8).

Table 8 The phrases from the neighbouring word position collocation analysis (±2 to 5 positions) of the AD corpus' 10 most frequent open class words > $\{U_0=10, k_1=1, p_0=1\}$. ▲ and ▼ denote upward and downward collocates respectively.

AD Corpus collocation phrases ±2-5 position			
and looks around	▲	looks down the	▲
away at the	▲	looks through the	▼
away from the	▲	looks over at the	▼
and turns away	▲	looks round at the	▼
and walks away	▼	looks up at the	▲
the door behind	▲	the living room	▼
looks around at	▼	into a room	▲
looks back at	▲	into the room	▲
looks back at the	▲	in his room	▲
looks down at	▲	in the room	▲
looks down at her	▲	walks along the	▼
looks down at the	▲	walks down the	▲
looks over at	▼	walks into the	▲
looks round at	▲	walks through the	▲
looks up at	▲	walks up to	▲
looks around the	▼	walks over to	▲
looks at the	▲	and walks over to	▲
and looks at the	▲	walks over to the	▲

Collocation Analysis Results: Screenplay Corpus

In the audio description corpus 58 frequently collocating phrases were accepted according to the tuple $\{U_0=10, k_1=1, p_0=1\}$ outlined by Smadja, with 11 frequent phrases being rejected. Diagrams of the frequent phrases found by COLLOCATOR can be seen in Appendix B. Table 9 shows the 58 collocation phrases of the open class words. Notice that unlike the AD corpus, this table contains the open class words 'day', 'night', 'int' and 'ext', which are frequent in the Screenplay corpus. In the case of the 'day' and 'night' collocation phrases they collocate with less frequent words, i.e. downward collocation.

Table 9 The 58 collocate phrases of the SC corpus' 10 most frequent open class words greater than the tuple $\{U_0=10, k_1=1, p_0=1\}$. ▲ and ▼ denote upward and downward collocates respectively.

SC Corpus Collocation Phrases			
looks at	▲	the room	▲
looks at the	▲	of the room	▲
he looks at the	▲	out of the room	▲
and looks at the	▲	takes a	▲
she looks at the	▲	he takes a	▲
the door	▲	around the	▲
opens the door	▲	around the room	▲
he opens the door	▲	looks around the room	▲
she opens the door	▲	look around the room	▲
opens the door and	▲	around the corner	▲
turns to	▲	and walks	▲
turns to the	▲	turns and walks	▲
he turns to the	▲	he turns and walks	▲
away from	▲	turns and walks away	▲
away from the	▲	behind him	▲
away from the curb	▲	door behind him	▲
backs away from the	▲	the door behind him	▲
his head	▲	to int	▲
shakes his head	▲	cut to int	▲
her head	▲	to ext	▲
shakes her head	▲	cut to ext	▲
she shakes her head	▲	room day	▼
towards the	▲	living room day	▼
towards the door	▲	drawing room day	▼
his eyes	▲	motel room day	▼
closes his eyes	▲	room night	▼
he closes his eyes	▲	living room night	▼
her eyes	▲	shower room night	▼
closes her eyes	▲		
she closes her eyes	▲		

Table 10 shows the 80 collocate phrases of the open class nucleate candidate words in the screenplay corpus in the ± two and five positions, most words collocate with more frequent words (upward collocation) though there are some downward collocating words (indicated in grey in Table 10).

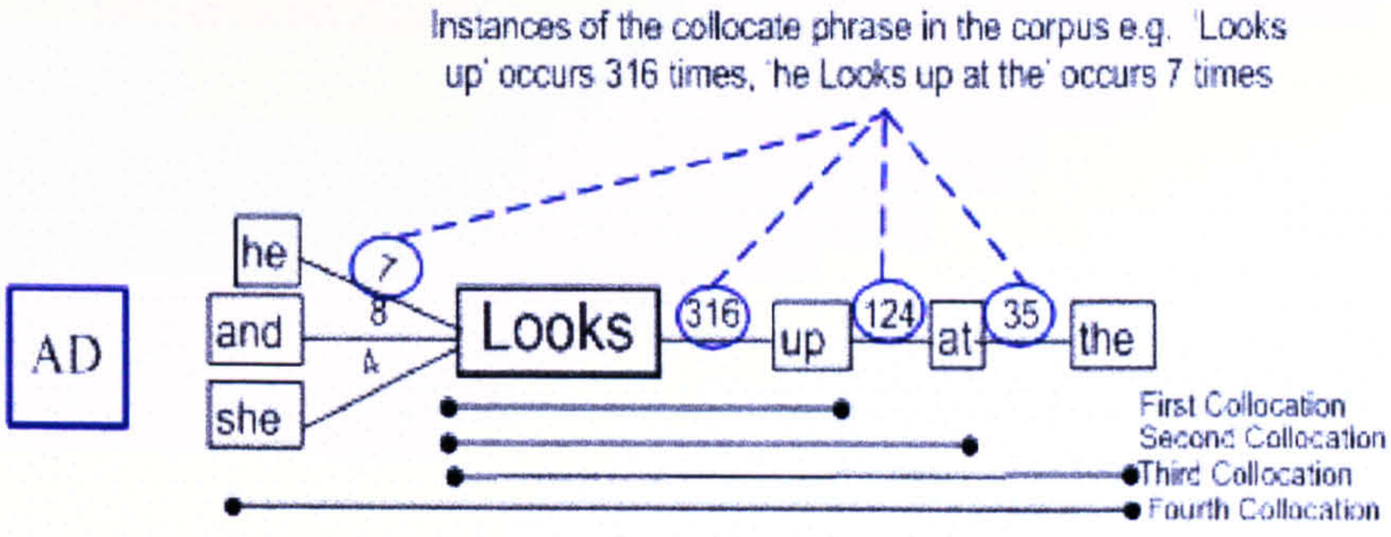
Table 10 The phrases from the neighbouring word position collocation analysis (±2 to 5 positions) of the SC corpus' 10 most frequent open class words > {U_0=10, k_1=1, p_0=1}. ▲ and ▼ denote upward and downward collocates respectively.

SC Corpus collocation phrases ±2-5 position					
he looks around	▲	looks around at	▲	looks into his	▲
he turns around	▼	looks around at the	▲	looks over his	▲
around a corner	▲	he looks around	▲	looks over his shoulder	▲
around the corner	▲	looks back at	▲	he looks at his	▲
around his neck	▲	looks back at the	▲	across the room	▲
away and the	▲	looks down at	▲	across the room to	▲
away as the	▲	looks down at his	▲	the control room	▼
away at the	▲	looks down at the	▲	the living room	▼
away by the	▲	he looks down at the	▲	the throne room	▼
away from the	▲	looks over at	▼	turns from the	▲
away in the	▲	looks over at the	▼	turns off the	▲
away into the	▲	looks up at	▲	turns to the	▲
and backs away	▼	looks up at her	▲	turns toward the	▼
and turns away	▼	looks up at him	▲	walks across the	▲
and walks away	▼	looks up at the	▲	walks around the	▲
to walk away	▼	looks at the	▲	walks down the	▲
he turns away	▼	he looks at the	▲	walks into the	▲
he walks away	▼	she looks at the	▲	walks out the	▲
to a door	▲	looks into the	▲	walks through the	▲
to his door	▲	looks like the	▼	walks to the	▲
to the door	▲	looks out the	▲	walks toward the	▲
to the door and	▲	looks through the	▼	walks up the	▲
goes to the door	▲	looks to the	▲	he walks to	▲
moves to the door	▲	he looks to the	▲	she walks to	▲
looks at him	▲	looks at his	▲	as he walks	▲
she looks at him	▲	he looks at his	▲	as she walks	▲
just looks at him	▲	looks at his watch	▲		

3.3.2.3 Identify Significant Collocations: Detailed Results and Discussion

This section provides a more in depth version of the method and provides a running example of the analysis of the nucleate word 'looks' and its collocations.

After inputting the nucleate candidates ColloQator provided the most frequent collocations/re-collocations according to Smadja's inequalities. These were visualised as graphs as can be seen in Figure 17 and are all available in Appendix C and on the accompanying **CD** along with their relevant U-scores, k-scores and p-strength (k_x).



U-Scores	Std. Dev.	K-Scores	P_1	
6554.49	139.36	24.66	1.86	
1291.04	35.93	14.61	2.85	
28.00	5.29	6.47	2.98	ar
49.80	7.06	4.74	2.79	h
35.44	5.95	5.78	2.84	sh

Figure 17 A visualization of the 3 re-collocations of the word 'looks' in the audio description corpus, accompanied by the U-scores, K-scores and p-strengths of the respective re-collocations. See Appendix A and accompanying **CD**: Additional Thesis Material.

N.B. It must be noted that the only the most 'significant' collocate phrases were recorded in this research. 'Significant' refering to there being more than *two* instances of the phrase, the phrase being *highest* in the rankings in terms of values such as U-score and frequency of the collocate and that no more re-collocations can be made. For instance in the case of AD 'looks up at the', seen in Figure 17, 'and/he/she looks up at the' were returned as the three highest rank phrases after three re-collocations with '8', '7' and '4' instances of the phrases respectively. Table 11 also gives an idea of the how COLLOCATOR ranks terms.

The re-collocations of the nucleate open class words, such as those seen in Figure 17, are presented by COLLOCATOR as a table of values. An example of results from COLLOCATOR is shown in Table 11 which excludes standard deviation and average number of instances of a collocate in all positions. "Frequency f" refers to the frequency of the nucleate term, "frequency W" refers to the frequency of the collocating "word W." The columns marked -5 to 5 represent the collocates' 10 neighbouring positions to the nucleate word (in this case 'looks'). "Max" represents the highest instance of a collocate word in any position for that specific collocation, e.g. for 'look at' the maximum highest instance is '1237' in the '1' position (shaded **dark grey**). The U-scores and k-scores are also presented. The P_1 column has been manually added to illustrate the strength of a collocate in a specific position, COLLOCATOR automatically determines this number but it is not shown in the results as it is not accessible; The three coloured areas in Table 11 represent three stages of filtration as outlined by Smadja [88]. The first stage is the elimination of low frequency collocates by examining the k-scores {**Yellow**}. Taking the strength threshold k_0 as above 1 we can eliminate any collocates with k-scores below that. The second stage requires examining that there is at least one spike of the histogram of the ten relative frequencies {**Green**}. This is done by looking at the U-score and the histograms are rejected if the variance threshold U_0<10, in the case of our example the values are much higher than 10. The third stage eliminates columns instead of rows {**Blue**} and finds the significant relative positions of two words {**Lavender**}. Inequality [ζ] is used to locate terms to collocate further.

Table 11 Results of a collocation of the word 'looks' in the Screenplay corpus. Values shaded grey correspond to high instances of a collocate word in a certain position around the nucleate word.

Nucleate	Freq f	Word W	Freq W	-5	-4	-3	-2	-1	1	2	3	4	5	P_1	Max	U-score	K-score
looks	4831	at	2147	27	57	55	5	0	1237	595	47	71	53	2.69	1237	144032.01	37.49
looks	4831	he	1227	56	47	49	30	771	2	52	60	86	74	2.98	771	47171.41	21.38
looks	4831	up	900	30	25	21	41	13	715	17	6	16	16	3.00	715	43487.80	15.65
looks	4831	the	2903	235	299	518	197	9	13	666	481	227	258		666	40227.81	50.74
looks	4831	around	407	7	7	12	12	0	354	7	3	4	1		354	10921.21	7.01
looks	4831	she	533	23	31	17	15	357	0	12	23	34	21		357	10331.41	9.22
looks	4831	like	396	16	8	7	2	0	305	18	18	12	10		305	7860.84	6.82
looks	4831	down	373	21	10	9	18	3	277	6	5	11	13		277	6412.21	6.41
looks	4831	him	482	21	13	22	61	6	12	266	58	7	16		266	5618.76	8.32
looks	4831	back	350	24	17	12	21	5	218	4	11	13	25		218	3768.00	6.01
looks	4831	to	967	116	146	106	12	3	204	153	40	88	99		204	3682.21	16.82

Once the most common collocations/re-collocations were identified automatically for both corpora, they were filtered again using Smadja's method [88]. This meant examining whether the collocation phrase's U-score, k-score and p-strength was greater than the tuple:

$$\{U_0=10, k_1=1, p_0=1\}$$

P_0 was later found to be approximately 1.75 for all the collocate phrases examined and thus could be changed to said value. However it was not necessary as all collocate phrases' p-strength values examined were greater than this value, thus fitting the criteria for acceptance. An example can be seen in Table 12 for the audio description corpus's open class words 'looks' and 'door' collocate phrases. Table 12's phrases 'he looks at', 'she looks at', 'he opens the door' and 'she opens the door' have been rejected due to u-scores being < 10.

Table 12 Shows the flagging of collocation phrases for the nucleate words 'looks' and 'door' in AD corpus with respect to Smadja's tuple $\{U_0=10, k_1=1, p_0=1\}$. The phrase 'he looks at the' is rejected as U<10 (7.04).

Source	Word (w)	Freq(w)	Phrase No.	Freq Phr	U-Score	K-Score	P_1	Phrase (Ph)	Flag
AD	looks	1797	AD1-1	316	8554.49	24.66	3.00	looks up	1
AD	looks	1797	AD1-2	124	1291.04	14.81	2.99	looks up at	1
AD	looks	1797	AD1-3	35	87.65	14.24	2.83	looks up at the	1
AD	looks	1797	AD1-4a	8	28.00	6.47	2.94	and looks up at the	1
AD	looks	1797	AD1-4b	7	49.80	4.74	2.93	he looks up at the	1
AD	looks	1797	AD1-4b	4	35.44	5.78	2.65	she looks up at the	1
AD	looks	1797	AD2-1	349	19426.44	24.66	1.88	looks at	1
AD	looks	1797	AD2-2	88	584.99	18.48	2.85	looks at the	1
AD	looks	1797	AD2-3a	22	40.60	8.33	2.98	and looks at the	1
AD	looks	1797	AD2-3b	7	7.04	4.16	2.79	he looks at the	0
AD	looks	1797	AD2-3c	9	4.29	2.67	2.84	she looks at the	0
AD	door	1356	AD3-1	617	27637.56	39.86	2.85	the door	1
AD	door	1356	AD3-2	68	380.49	7.51	2.93	opens the door	1
AD	door	1356	AD3-3a	10	9.04	3.67	2.86	he opens the door	0
AD	door	1356	AD3-3b	9	11.36	7.95	1.84	and opens the door	1
AD	door	1356	AD3-3c	6	3.20	7.44	2.80	she opens the door	0

Collocation Analysis: Examining Neighbouring Word Positions

Although COLLOCATOR was able to automatically provide the most frequently collocating phrases with respect to the ±1 neighbouring positions of the nucleate open class words and their relocations, the system was not able to find the collocates/re-collocates of the most frequently collocating phrases with respect to the ±2-5 neighbouring positions of the nucleate open class words. Thus, phase 2 involved developing and implementing a method to the words that strongly collocated in the ±2-5 positions (outlined in Figure 16), with respect to the nucleate open class words.

Results from COLLOCATOR, showing the frequency of neighbouring words to frequent nucleate open class words, were examined in terms of frequent words in the ±2-5 positions (see Table 13). High frequency occurrences of words in the ±2-5 positions were noted along with their respective Frequency value, U-scores and k-scores. The *p-strength* (significant relative position) [See 3.3.2.1] of each term was also calculated. These statistics were compared with the tuple $\{U_0=10, k_1=1, p_0=1\}$ outlined by Smadja [88] and if any of these conditions were not satisfied then the collocation phrase was rejected. For example, in Table 13, 'looks *x* and', 'looks *x* to' and 'looks *x* a' were all rejected because their P-strength values were < 1. It must be noted that the grounds for rejecting a collocate phrase were not solely based on the {10, 1, 1} tuple but also on how frequent the phrase was; the AD corpus having a threshold >50 instances and the SC corpus >80.

Table 13 Results of a collocation of the word 'looks' in the Screenplay corpus. Values shaded grey correspond to high instances of a collocate word in a certain position around the nucleate word. The P-strength in this case is calculated for position '2'. ±1 positions are greyed out as they are not considered in this analysis.

Nucleate	Freq (f)	Word W	Freq (W)	-5	-4	-3	-2			2	3	4	5	P_1-Pos2	U-score	K-score
looks	4831	at	2147	27	57	55	5	0	1237	595	47	71	53	1.00	144032.01	37.49
looks	4831	he	1227	56	47	49	30	771	2	52	60	86	74		47171.41	21.38
looks	4831	up	900	30	25	21	41	13	715	17	6	16	16		43487.80	15.65
looks	4831	the	2903	235	299	518	197	9	13	666	481	227	258	1.87	40227.81	50.74
looks	4831	around	407	7	7	12	12	0	354	7	3	4	1		10921.21	7.01
looks	4831	she	533	23	31	17	15	357	0	12	23	34	21		10331.41	9.22
looks	4831	like	396	16	8	7	2	0	305	18	18	12	10		7860.84	6.82
looks	4831	and	1144	92	76	69	41	344	14	174	97	126	111	0.68	7622.24	19.92
looks	4831	down	373	21	10	9	18	3	277	6	5	11	13		6412.21	6.41
looks	4831	him	482	21	13	22	61	6	12	266	58	7	16	2.91	5618.76	8.32
looks	4831	back	350	24	17	12	21	5	218	4	11	13	25		3768.00	6.01
looks	4831	to	967	116	146	106	12	3	204	153	40	88	99	0.93	3682.21	16.82
looks	4831	his	721	53	60	137	35	1	3	182	127	44	79	1.95	3185.89	12.51
looks	4831	a	971	125	140	142	23	2	21	149	107	129	133	0.95	3005.89	16.89
looks	4831	her	487	25	34	65	44	2	3	182	62	34	36	2.74	2365.81	8.41

This analysis was conducted on all the frequent open class words outlined in Table 1 for both corpora and this resulted in a set of collocation phrases with *unknown* words between the interspersing collocate words. For instance the word 'looks' for the AD corpus yielded the phrases 'looks *ww* at', 'looks *xx* the' and 'looks *yy zz* at' (*ww, xx, yy,* zz indicate distinct words).

Having found collocation phrases in the ±2-5 positions for all the frequent open class words the next phase was to search for the interspersing unknown words. Unitex was employed to locate the interspersing words because it was capable of locating *all unknown* words in a phrase (*Ukn*) through the use of *<LEXI>*[18] as a wildcard. Hence, Finite State Automata (FSA) graphs were drawn in Unitex for each open class word's unknown collocation phrase (see Figure 18 for an example) and the results were exported.

Figure 18 A Finite State Automata graph used in Unitex to extract unknown words from the collocation phrase 'looks *xx* at' in the Corpora and some results of the search.

Produces Results

looks ACCUSINGLY at	looks around at	looks around at
looks across at	looks around at	looks askance at
looks again at	looks around at	looks askance at
looks again at	looks around at	looks askance at
looks again at	looks around at	looks back at
looks again at	looks around at	looks back at
looks again at	looks around at	looks back at
looks again at	looks around at	looks back at
looks annoyed at	looks around at	looks back at
looks anxiously at	looks around at	looks back at
looks anxiously at	looks around at	looks back at
looks appraisingly at	looks around at	looks back at
looks archly at	looks around at	looks back at
looks around at	looks around at	looks back at
looks around at	looks around at	looks back at
...	...	...etc

After separating the text by delineating the space characters, a Macro was used to tally up the *Ukn* words from the Unitex results. If the instances (Ins_x) of the *Ukn* words were $Ins_x<5$ for the AD corpus and $Ins_x<10$ for the SC corpus (come from manual inspection of the results), then the collocation phrase was rejected. If AD: $Ins_x>5$ and SC: $Ins_x>10$ then the *Ukn* words' original collocation phrase was extracted. E.g. in the case of Figure 18 the phrase 'looks around at' was extracted but 'looks askance at' was not.

The extracted collocation phrases were than entered into COLLOCATOR to elicit the U- and k-scores and allow us to calculate P-strength. This was done in two stages:

1. Firstly the nucleate open class words were entered into COLLOCATOR with each associated *Ukn* word and the relative statistics were extracted.

---

[18] LEXI is Greek for 'word' and is used here to denote 'any word'

a. IF $\{U_1, K_1, P_1\} \geq \{10, 1, 1\}$ then the phrase was used in stage 2.
b. ELSE the phrase was rejected.

2. Secondly, based on the results of stage 1, the whole collocate phrase including the nucleate word, the *Ukn* word and the collocating word were entered into COLLOCATOR.
   a. IF $\{U_1, K_1, P_1\} \geq \{10, 1, 1\}$ then the phrase was extracted.
   b. ELSE the phrase was rejected.

For example in the case of the collocation phrase 'looks *ww* at' - 'looks' being the nucleate word, '*ww*' being the *Ukn* word and 'at' being the collocating word – firstly all instances of 'looks' and '*ww*' were entered into COLLOCATOR and the results accepted/rejected according to the statistics. Then *accepted* collocate phrases were re-collocated and accepted/rejected accordingly.

Table 14 A table showing the two stages of the flagging of collocation phrases for the collocation phrase 'looks *xx* at' in the AD corpus, with respect to Smadja's tuple $\{U_0=10, k_1=1, p_0=1\}$. The phrase 'looks again' has been rejected as $P_1<1$ (0.51) In the case of 'looks around at', phrase is rejected as U<10 (5.16).

	Source	Nucleate	Freq (f)	Word (W)	Freq Phr.	U-score	K-score	P_1	Phrase (Ph)	Flag
STAGE 1	AD	looks	1797	down	218	2680.56	6.08	3.00	looks down	1
	AD	looks	1797	around	147	1174.81	4.04	2.98	looks around	1
	AD	looks	1797	back	111	380.69	3.01	2.97	looks back	1
	AD	looks	1797	round	72	299.36	1.89	2.99	looks round	1
	AD	looks	1797	over	60	103.20	1.55	2.95	looks over	1
	AD	looks	1797	out	69	80.49	1.81	2.80	looks out	1
	AD	looks	1797	again	51	32.09	1.29	0.51	looks again	0
STAGE 2	AD	looks around	117	at	18	5.16	3.18	2.73	looks around at	0
	AD	looks back	69	at	41	101.49	8.94	2.97	looks back at	1
	AD	looks down	177	at	90	443.00	11.48	2.99	looks down at	1
	AD	looks out	32	at	10	5.60	2.96	2.96	looks out at	0
	AD	looks over	36	at	23	35.61	7.21	2.97	looks over at	1
	AD	looks round	59	at	20	22.20	5.87	2.97	looks round at	1

Once these two stages had identified and filtered a set of collocate phrases for the ±2-5 positions for all the frequent open class words, these collocate phrases were entered into COLLOCATOR as *nucleate* phrases to examine the neighbouring language of the phrases and then re-collocated to elicit the most frequent collocates of the phrases. Once again the {10, 1, 1} tuple was used to filter results. The results of this study can be seen in section [3.3.2.2].

Downward/Upward Collocation: Results

The next stage involved examining the collocates to see if any of the open class words collocated with each other. Sinclair [[87]: p.116] describes two types of collocation– *upward* and *downward* collocation. The upward collocation is the collocation of the node (nucleate) term with a more frequent word, whereas downward collocation is the collocation of the node term with a less frequent collocate term. Upward collocation is not as common statistically as downward

collocation but the words are usually elements of 'grammatical frames' whilst downward collocations provide semantic analysis of the word being analysed.

The collocate diagrams, showing re-collocations (see Appendix A), were manually examined – the collocates were examined first as they had the greatest concentration of *very* frequent words with respect to the nucleate words. The results of this analysis can be seen in Table 7 to Table 10 where upwards collocates are represented with a ▲ and downwards collocates are represented with ▼ . It was found that most of our examined open class words in the corpora collocated upward with each other.

3.3.3 IV. Generalise/Join Collocations

This step generalises the collocation phrases based on the candidate nucleate words for both corpora and produces resultant FSA diagrams. Figure 19 shows the steps of the method for this stage.

V. **JOIN & GENERALISE COLLOCATIONS**
- a. FOR EACH WORD S^{OPEN} IN BOTH CORPORA COPY ALL $\{Y_N\}$
  - **i.** **FIND** Maximal Common Overlap /*Common repeated word throughout Fargues [30]*/
  - **ii.** **JOIN AT** Maximal Common Overlap
  - **iii.** **SIMPLIFY** Graph
    - **1.** **REMOVE** duplicate redundant nodes /*Fargues [30]*/
  - **iv.** **REPEAT** Steps i-iii until no redundancies remain.
  - **v.** **COPY** Resultant Graph.
- b. DRAW RESULTANT GRAPH AS FINITE STATE AUTOMATON

Figure 19 Pseudo code of the method for systematically joining and generalising collocations.

3.3.3.1 Generalise/Join Collocations: Method, Definitions and Tools

Finite State Automata: a model of behaviour composed of states, transitions and actions [140] [142]. In this case they are used to represent restricted phrases of language that contain paths that can be followed to make a coherent phrase.

Two principle operations allow us to generalise the collocations I) and II) and place then in a Finite State Automata form as defined by Sowa in [91] and [92]:

I) ***Join:*** Let ***c*** and ***d*** be any two concepts of **u** whose types and referents are identical. Then **w** is the graph obtained by deleting ***d***, adding ***c*** to all co reference sets in which ***d*** occurred and attaching to ***c*** all arcs of conceptual relations that had been attached to ***d***.

II) ***Simplify:*** If conceptual relations ***r*** and ***s*** in the graph **u** are duplicates, then one of them may be deleted from **u** together with all its arcs.

Maximal Common Overlap: the maximal graph, which is the common restriction of two corresponding sub graphs of the two graphs. It must be a connected graph [30].

In [30] Fargues et al. discuss applying Sowa's work [91] to model natural language semantics. We have drawn on the algorithm that simplifies conceptual graphs in order to automatically produce Finite State Automate (FSA) for collocations (see Figure 19 for algorithm). The method for joining collocations is as follows: [[30] pg. 71-72]

- Copy a graph
- Join two conceptual graphs:
    - Form the maximal common overlap from the two graphs.
    - Attach the pending parts remaining in the two graphs to the maximum common overlap.
- Simplify a graph by suppressing the redundant occurrences of identical edges in a graph.

3.3.3.2 Generalise/Join Collocations: Results

The full results of the join algorithm for all candidate nucleate collocations can be seen in Appendix B and on the accompanying **CD**: Additional Thesis Material, in the form of FSA. Examples of the nucleate 'walks' simplified collocation FSA for the AD and SC corpora can be seen in Figure 20 and Figure 21 respectively.

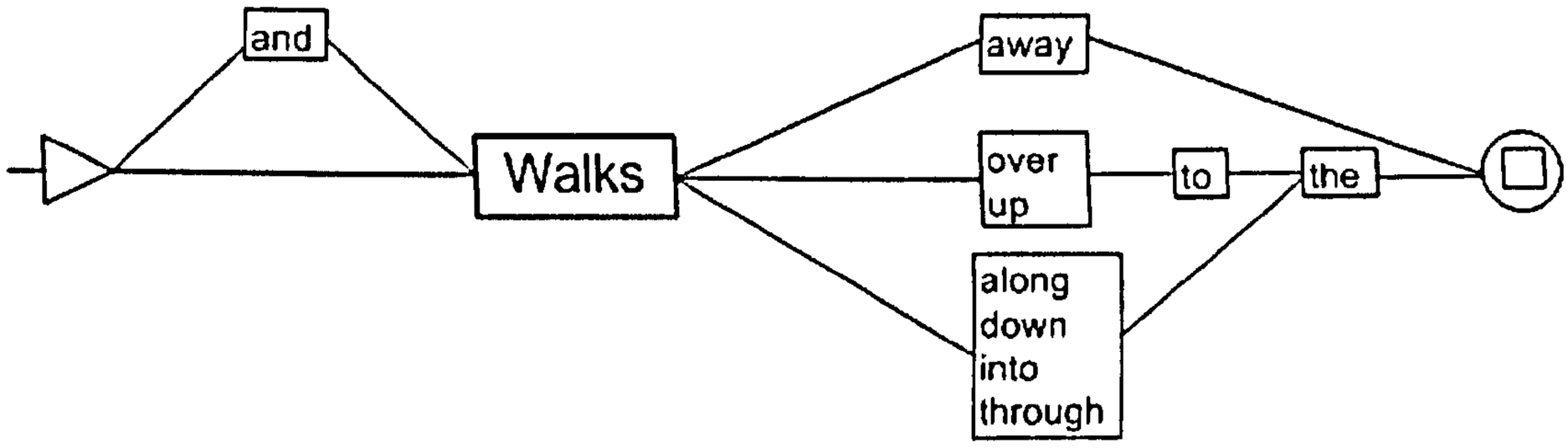


Figure 20 An FSA representing the Local Grammar of the word 'walks' in the AD corpus produced from the collocate phrases of 'walks' using Fargues et al's [30] Join algorithm.

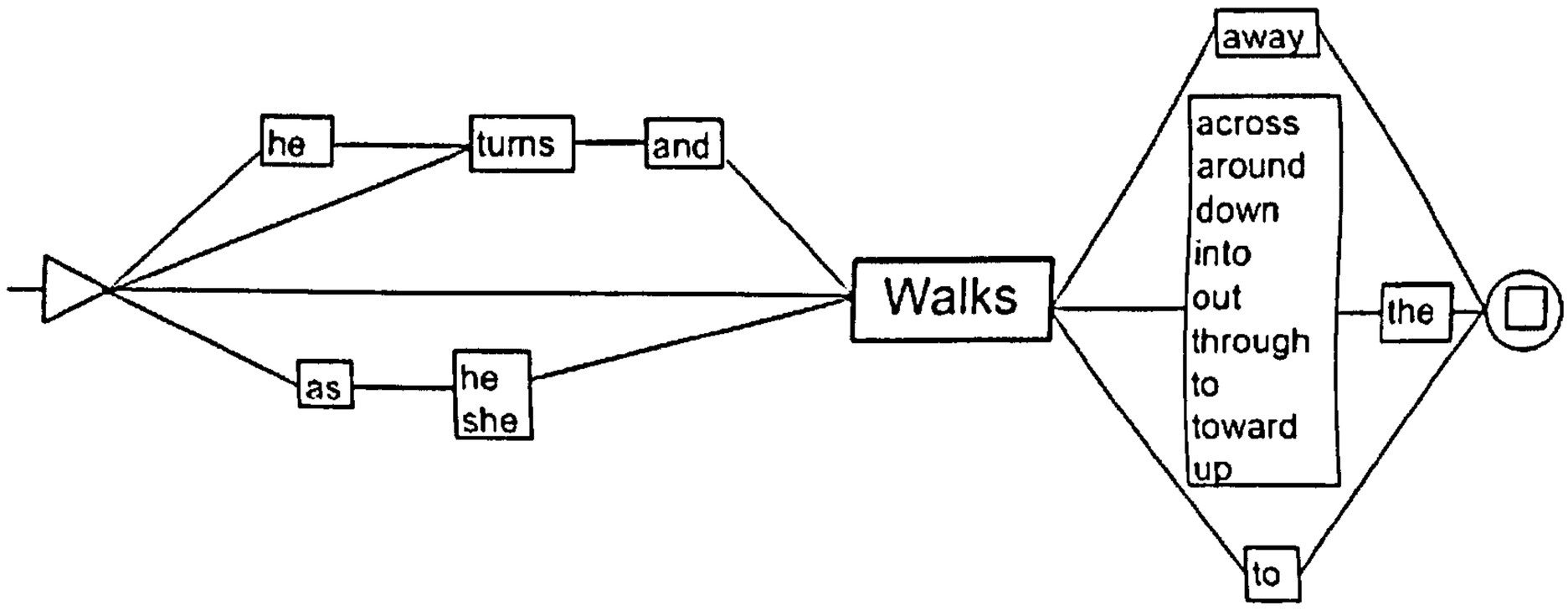


Figure 21 An FSA representing the Local Grammar of the word 'walks' in the SC corpus produced from the collocate phrases of 'walks' using Fargues et al's [30] Join algorithm.

3.3.3.3 Generalise/Join Collocations: Detailed Results and Discussion

This section provides a running example of how the collocation phrases for nucleate 'looks' were joined and generalised into one FSA for 'looks'. Also included is a discussion about the use and formation of the algorithm from Sowa [91] & [92] and Fargues [22] and problems this method may face, such as *over* generalisation.

Unitex allowed us to draw these graphs automatically using the "CONSTRUCT FST Text" function. However, the graphs were too complicated due to the fact Unitex's inbuilt grammatical parser would produce *all* grammatical instances of the words involved in the collocation phrase. Thus, simplified versions of the FSA graphs were drawn. Figure 22 shows an example of the FSA graph of the collocation phrases of open class word "looks" in the AD corpus. The question of how to link the collocation phrase FSA graphs and remove redundancy remained.

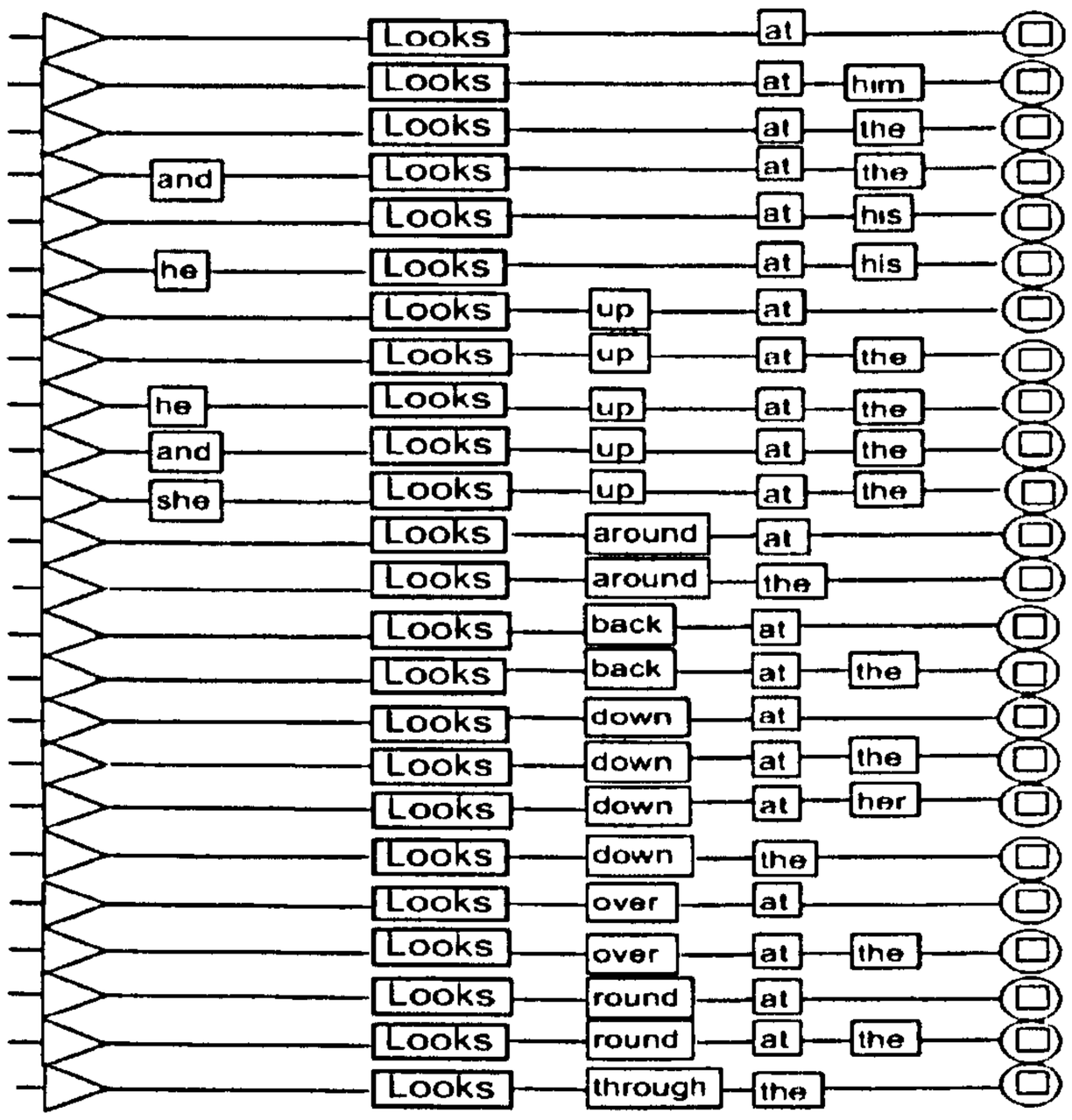
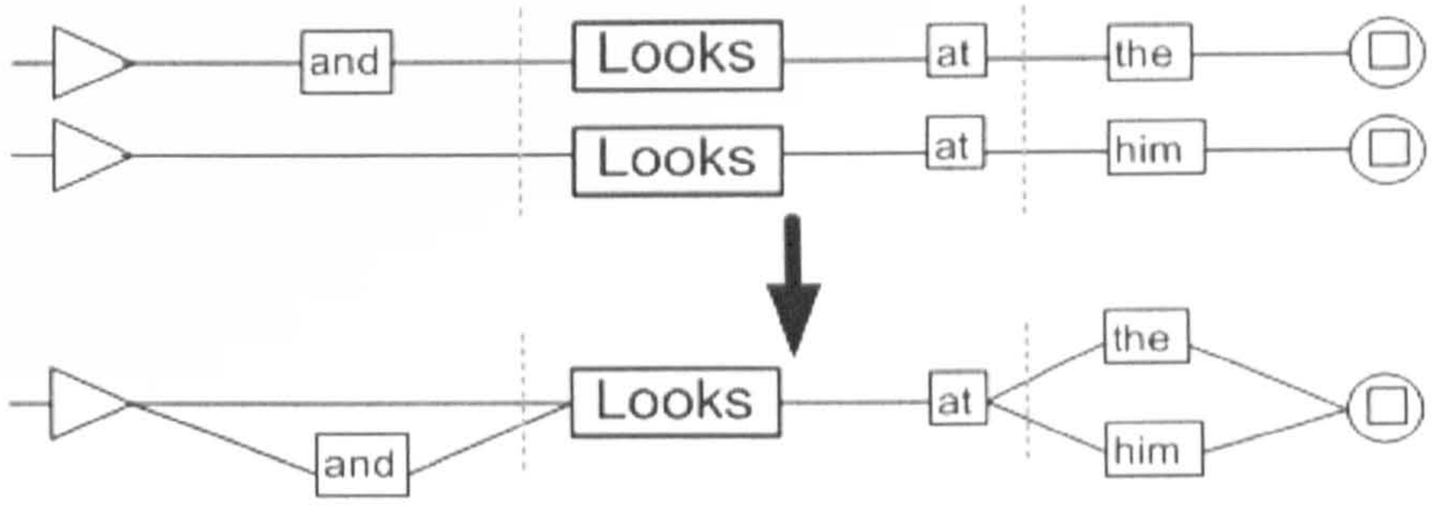


Figure 22 FSA graphs of the collocation phrases for the open class word 'looks' in the AD corpus as seen in tables Table 7 & Table 8.

Removing Redundancies using Fargues et al. Join Algorithm

Manually, the task of how to link the collocation phrase FSA graphs and remove redundancies seemed trivial. For instance, in the case of 'and looks at the' and 'looks at him', the graphs could be joined to represent the same two phrases. The new graph is still logically equivalent to the original.



Although it is an intuitive step in terms of human judgement to join the two FSAs, the question of how to algorithmically, and automatically, join the two graphs presents itself.

In terms of linking collocation phrases at their nucleates, this work considers Sowa's *Join* and *Simplify* canonical formation rules for conceptual graphs. In terms of conceptual graphs the join rule merges identical concepts: "two graphs may be joined by overlaying one graph on top of the other so that the two identical concepts merge into a single concept", and the Simplify rule deletes duplicates: "when two relations of the same type are linked to the same concepts in the same order, they assert the same information; one of them may therefore be erased." [[91]: pg.92]. In [30], Fargues et al. discuss the representational and algorithmic power of conceptual graphs for natural language semantics and knowledge processing. They emphasise the main properties of the conceptual graph model by applying them and comparing them to the properties needed to model natural language semantics. In their work, they discuss the join and generalisation algorithms that have been adapted directly from [91]. They state that it is possible to build new conceptual graphs from existing conceptual graphs by applying the rules seen in 3.3.3.1.

In our work the interest lies with the algorithmic nature of the formation rules. There is evidence here that these formation rules can be directly applied to the problem at hand: how to algorithmically link the collocations to develop generalised FSA graphs for the candidate nucleates. This work is ***not*** represented in the form of conceptual schema/graphs but the FSA graphs here are representing language structures and therefore we can consider Fargues' thoughts on "modelling natural language semantics". Thus, this work has adapted the formation rules and join algorithm to systematically develop FSA graphs from the collocations.

Figure 23b shows an example of the algorithm, based on [30], in action with respect to the collocate phrases of the open class word 'Looks' in the AD corpus. Firstly the maximal common overlap of the phrase is found (in this case, the word 'looks') and joined at that node. Then the graph is simplified by removing the redundant, overlapping edges/nodes of the graph. The maximal common overlap of the graph is again found ('at'), joined, and then simplified removing redundant nodes. This is repeated until a final finished local grammar FSA graph, containing the minimum number of *allowed* paths that represent the initial collocate phrases, is formed.

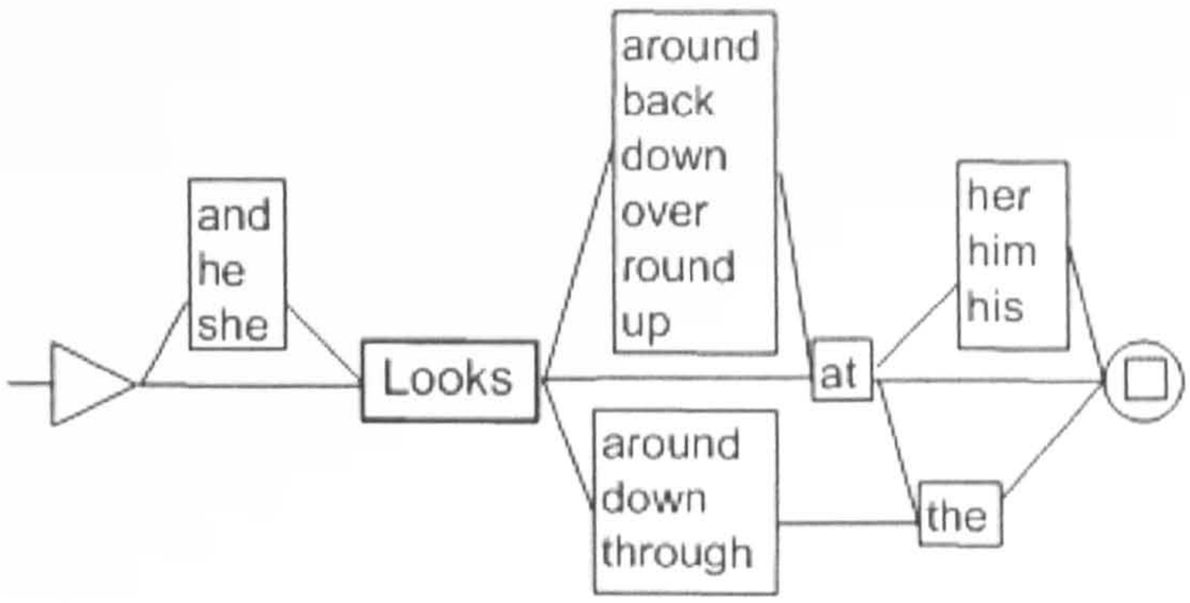


Figure 23 An FSA representing the Local Grammar of the word 'looks' in the AD corpus produced from the collocate phrases of 'looks' using Fargues et al's [30] Join algorithm. See **CD**: Additional Thesis Material for all FSA Local Grammar representations.

Possible Problems using Sowa's Formation Rules and Fargue's Algorithm

There is strong evidence to suggest that Sowa's formation rules for Canonical graphs and Fargue's Join algorithm can be adapted to algorithmically join the most significant collocations of a nucleate and remove redundancies to produce a Finite State Automaton. However, the produced FSA is not an exact representation of the original set of collocation phrases. That is, some of the phrases generated by the FSA do not exist in the original set of collocate phrases. The problem of *over-generation* of phrases exists. For instance from Figure 23 the phrases 'looks back at him' and 'she looks around at her' can be generated. These phrases are not in original set of collocate phrases. In this instance all the generated phrases are meaningful (or canonical), but it may be the case that a generated phrase is *not meaningful*. NB. With respect to the FSAs generated for the candidate words in both corpora, *no* non-meaningful phrases have been produced from the join algorithm in this work.

The discussion however becomes, whether these over-generated phrases are an acceptable side effect of this process, or whether they should somehow be restricted. It depends on what the FSA representations are to be used for. For instance, if the FSA are to represent a strict local grammar for the candidate words (See Section 3.4.2 for discussion) then the FSA should represent only the original phrases with no over-generation. However, if the FSA are to be put to use in text analysis or information extraction, e.g. data mining for information about film events, then the over-generated phrases may be acceptable, even beneficial as long they are meaningful.

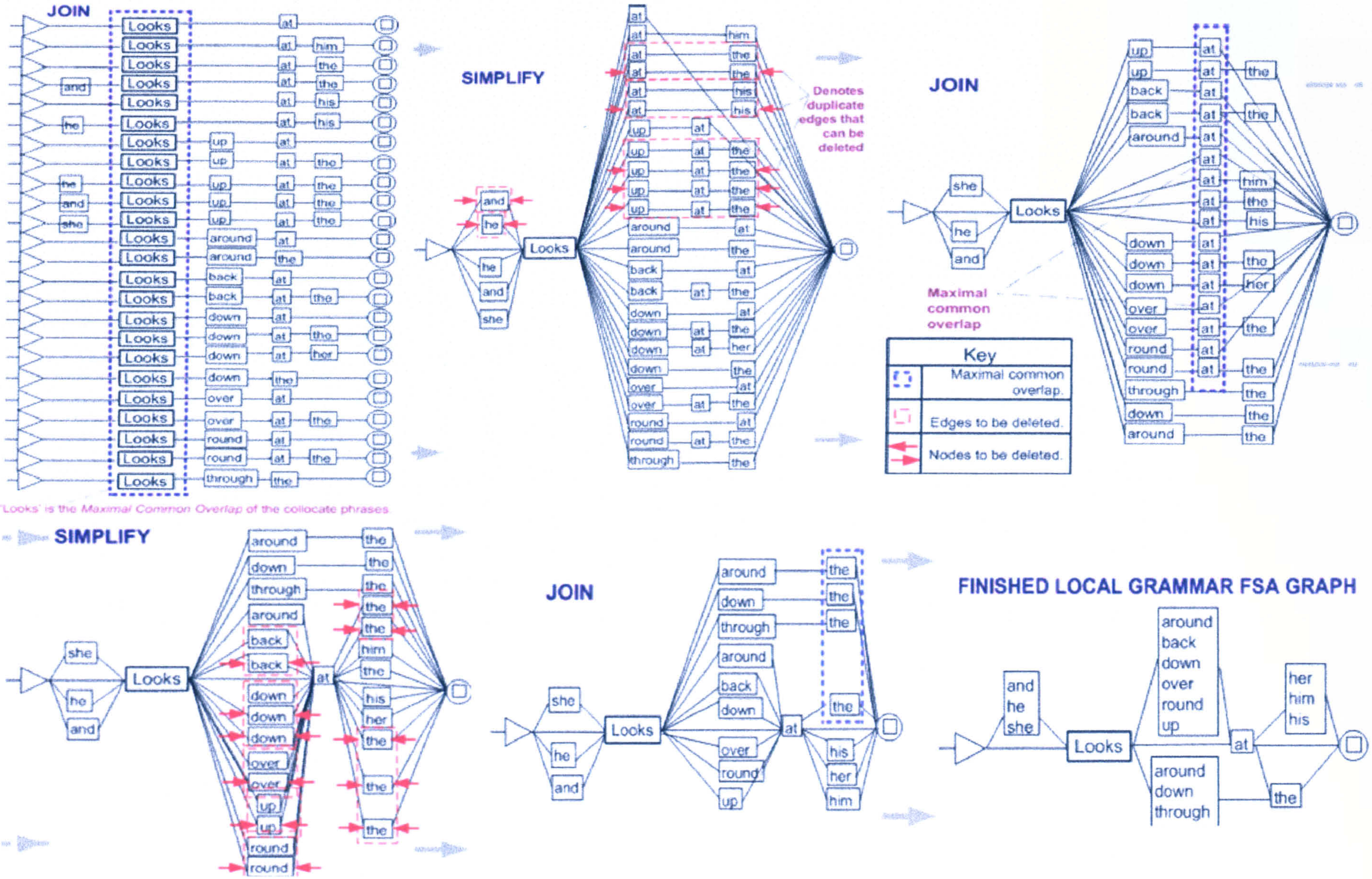


Figure 23b An example of how the collocate phrases for the open class word 'Looks' can be joined, and their redundancies removed, to form an Local Grammar FSA diagram according to Sowa's [84.99] Formation rules.

3.3.4 V. Expanding the Collocations

This step expands the collocations by examining what other frequent nucleate words exhibit the same frequent patterns when exchanging the candidate nucleate word for a 'wildcard'. Figure 24 shows the steps of the expansion method for this stage.

VI. EXPAND COLLOCATIONS

a. **SEARCH** for FSA***n*** in TC_1 using Corpus Analyser /* TC = Test Corpus */
b. **IF** FSA *NW* = VERB /* *NW* = Nucleate Word //
   i. **REPLACE** *NW* with inflectional class /* e.g. for "looks" inflectional class <look> = looks, looks, looked, looking */
   ii. **SEARCH** for FSA /* FSA= The Finite State Automata for all candidate nucleates*/
   iii. **RECORD** the number of instances
c. **ELSE IF** FSA *NW* = NOUN
   i. **REPLACE** *NW* with plural class /* e.g. for "door" plural class <door> = door, doors */
   ii. **SEARCH** for FSA
   iii. **RECORD** the number of instances
d. **ELSE IF** FSA *NW* = 'toward'
   i. **REPLACE** *NW* with <toward> /* = toward, towards */
   ii. **SEARCH** for FSA
   iii. **RECORD** the number of instances
e. **ELSE** proceed to next step
f. **ADD** Synonym set $\mathbf{S}_n$ to *NW*, /* $\mathbf{S}_n$ = synonyms of FSA *NW* found using a Thesaurus*/
   i. **SEARCH** for FSA
   ii. **RECORD** number of instances
g. **REPLACE** FSA *NW* with "<LEXI>" /*<LEXI> represents a wildcard that retrieves any one word in a given slot*/
   i. **SEARCH** for FSA in Test Corpus TC_1
   ii. **RECORD** number of instances
h. **EXAMINE** concordances for instances of recurrent frequent phrases $x > 25$ for > 5000 results or $x > 0.5\%$ of total instances
   i. **RECORD** the instances of *x*.
   ii. **EXTRACT** nucleates from x and store in *NNW* /**NNW*= New Nucleate Words*?
   iii. **ADD** NNW to NW of FSA
i. **REPEAT** steps a-h with each FSA for both training corpora (original corpora)
j. **REPEAT** steps a-i with TC_2

Figure 24 Pseudo Code for expanding the nucleate word set for all nucleate collocation FSA.

3.3.4.1 Expanding the Collocations: Method, Definitions and Tools

The ***generalised collocation Finite State Automata*** (FSA) for each of the nucleates words (NW) was tested on the ***Test Corpus*** that was set aside for each of the AD and SC corpora. Each FSA was placed into a *corpus text analyser* (Unitex in our case) with a test corpus pre-loaded. The nucleate word of the FSA was replaced, in sequence, with

a. { **a_1.** It's Inflectional Class (if a VERB),
   a_2. Its plural class (if a noun),
   a_3. {towards, toward} if NW = Towards }

b. A wild card: <LEXI>

c. The nucleates' synonyms[19].

[19] The synonyms of all the LG Node words were found using the Oxford College Thesaurus 2000 Edition and the online thesaurus http://thesaurus.reference.com/ last accessed 12/10/05

The ***Inflectional Class*** refers to any conjugated variances of a verb, e.g. the inflectional class for 'go' contains <go>= {go, going, goes, went}.

The ***plural class*** refers to the plurals of nouns and the noun itself, e.g. <door>= {door, doors}, <octopus>= {octopus, octopi}. The wildcard **<LEXI>** refers to *any* word being replaceable in that slot e.g. a <LEXI> bus may return 'a *red* bus', 'a *full* bus', 'a *double-decker* bus' etc.

Concordance or ***concordance listing*** a way of displaying the occurrences of a chosen word with its surrounding contexts [[13] page 26]. Concordance: a citation of parallel passages in a book, especially in the Bible [143]. In this case, sections of the corpora containing the same word were compared, usually with the nucleate word in the centre of the sentence, aligned for all instances.

The instances, their *concordances* and number of instances of each FSA were recorded for $\mathbf{a_{1\text{-}3.}}$, **b.** & **c.** If the instances of the results' recurrent frequent phrases $x > 25$ for > 5000 results or $x >$ 0.5% of total instances the total number of instances for ($\mathbf{a_{1\text{-}3.}}$, **b.** & **c)** then we record x. For all recorded instances of phrases x, we copy and save new nucleate words. Then remove redundancies (duplicate words) to produce a list of new nucleate words. A list of new nucleate words and original word is placed into FSA. The *predominance test* examines whether the original nucleate and/or its synonyms is the most predominant (most frequent) nucleate in that FSA. All nucleates are placed into a corpus analyser (Unitex), comparing frequencies of all instances of each nucleate term.

3.3.4.2 Expanding the Collocations: Results

A set of new nucleate words was found for each collocation FSA nucleate. A series of bar charts for all FSAs for both test corpora was created showing the percentage of words that appear in the node slot of the local grammar when the <LEXI> wildcard is used in place of the node word of the local grammar in Unitex. Figure 25 shows the percentages of the most frequent words to fill the 'looks' node slot of the AD and SC 'looks' FSA in both test corpora. In all cases, for both corpora, 'looks' or the inflectional class of 'looks' (<look>) is the *most predominant word* in the central node slot of the local grammars. Also the synonym of 'looks', <stare>, is the second most frequent word in the central node slot. Other synonyms, such as <glance>, <gaze> and <glare>, are also present but less frequent. There are also a few other verbs that appear in the central node slot, e.g. <smile>, <go> and <sit> and also nouns and prepositions, e.g. door, up, out, down, over, but they are less frequent. Figure 26 shows the new FSA for the nucleate word 'looks' and its alternative nucleates. In this case, looks, is the predominant word in that slot.

All the expansion of collocation results and predominance test values can be seen on the accompanying **CD**: Additional Thesis Material.

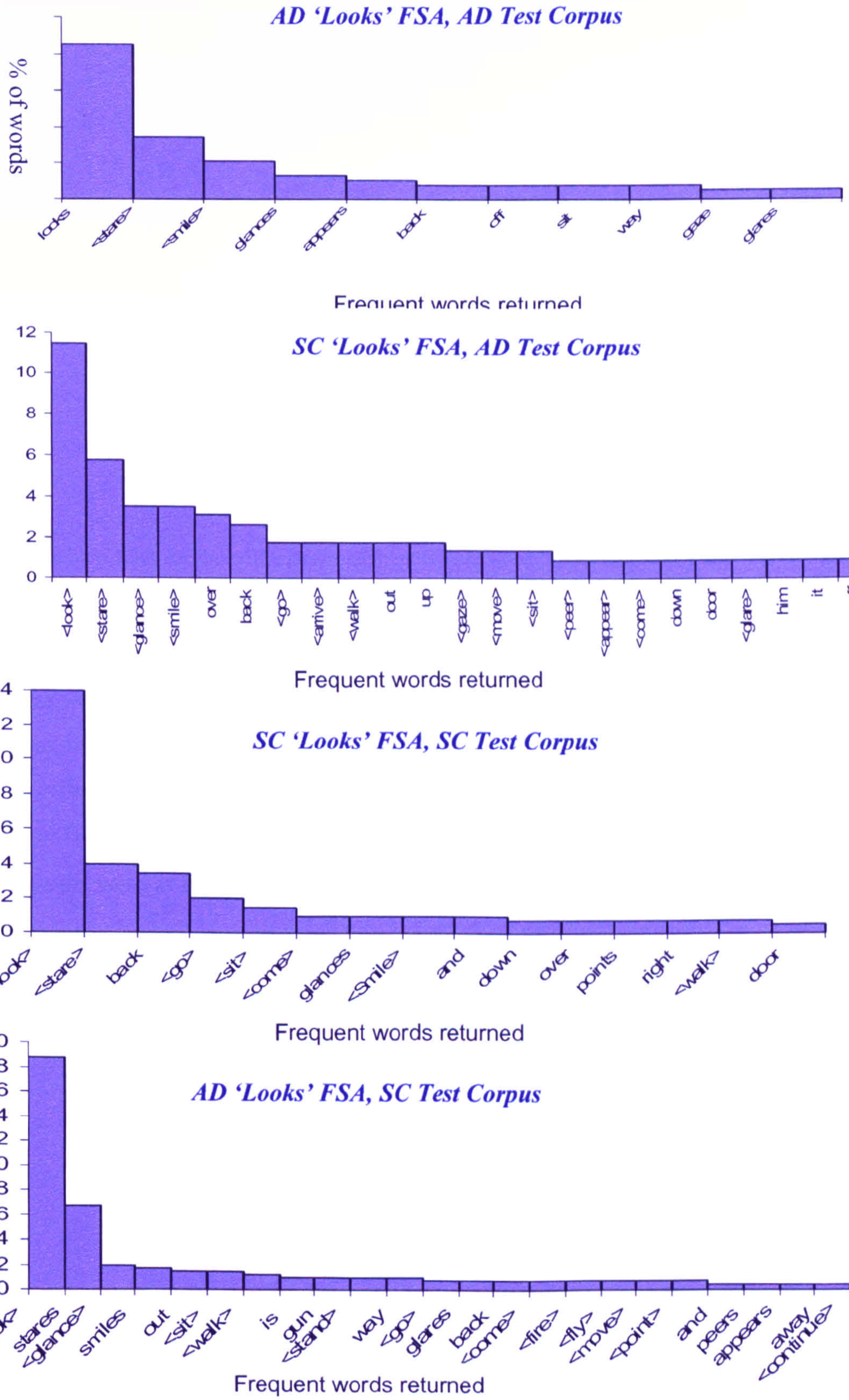


Figure 25 Bar charts of the percentage of the most frequent words that appear in the node slot of the local grammars for 'looks' in both test corpora when the node word 'looks' is replaced by <LEXI> in Unitex.

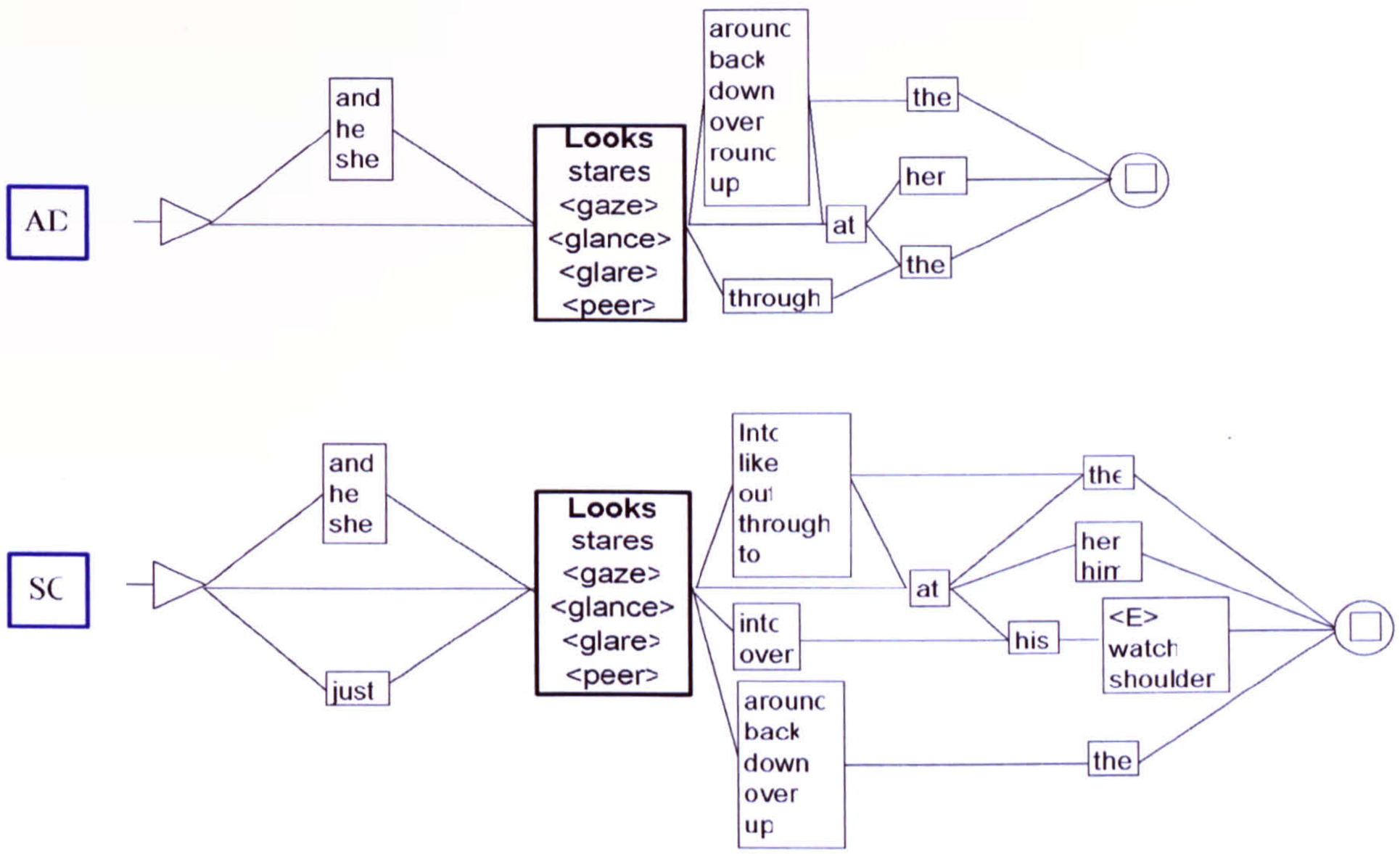


Figure 26 New FSA for the collocations of the nucleate word 'looks' including alternative nucleate words.

Overall, the results did not always return the *original* node word of the local grammar as the most frequent word in the node slot. In some cases the original node word of the local grammar was frequent, but not the most frequent, e.g. results for 'behind', 'room' and 'turns', but in other cases it was infrequent, such as in the cases of the local grammars for 'around' and 'towards' for both corpora. Table 15 shows the overall results of the predominance test. 67 % of the local grammars' node words were at least 'dominant' (√ √ ,√ &√ ⊙) and 54% were clearly dominant (√ &√ √). 8% showed no dominance (x) and those were the words 'around' and 'towards'.

Table 15 A table representing the results of the local grammar central word predominance test for all local grammars on both test corpora.

Key					
Predominant	√√	Dominant but Infrequent	√⊙	Not Dominant	x
Dominant	√	Not Dominant but Frequent	⊙⊙		

Original Central Word	TC AD	LG AD	TC AD	LG SC	TC SC	LG SC	TC SC	LG AD
Around	x		x		x		⊙	
Away	√		√		√		√ ⊙	
Behind	⊙		⊙		⊙		√ ⊙	
Towards	⊙		⊙		x		⊙	
Door	√		√		√		√ √	
Room	√		√		√ ⊙		√ ⊙	
Looks	√		√ √		√ √		√ √	
Takes	√		√		√		√ √	
Turns	√		⊙		⊙		⊙	
Walks	√		⊙		⊙		√ √	
Eyes	√ √		√ √		√ √		√ √	
Head	√ √		√ √		√ √		√ √	

3.3.3.3 Expanding the Collocations: Detailed Results and Discussion

Running Example of the Nucleate Word 'looks': Figure 27 provides a step-by-step illustration of the locating alternative nucleates method and the predominance test for the Nucleate Word (NW) 'looks'. 'Looks' is replaced with the inflectional class: <look>, <look> and the synonyms of 'looks' and the wildcard <LEXI>. The instances of all these searches were recorded. In the case of the wildcard <LEXI>, the concordance of 1890 results were examined and the most frequent recurring NW's instances $\boldsymbol{x}_n$ were recorded where $\boldsymbol{x}$>10 (x=10: ~0.5% of 1890 instances).

The privileged position test provided information about the percentage of the node word instances with respect to the total instances returned by a particular collocation FSA with the <LEXI> wild card replacing the FSA's nucleate word. As an example, for the AD collocation of the word 'looks', when nucleate word 'looks' was replaced by <LEXI>, Unitex returned 1890 instances of matching patterns. Out of those 1890 patterns 222 (11.75%) of those instances had the word 'looks' as their nucleate. When the inflectional class <look> was searched for 360 (19.05%) instances were located and 600 (31.75%) instances were located when the inflectional class and synonyms of 'looks'[20] were searched for. Privileged position test results can be seen in Table 16.

[20] The synonyms for looks were: <gaze>, <gape>, <gawk>, <glimpse>, <glare>, <glance>, <peep>, <peek>, <watch>, <stare>, <examine>, <view>; found at [141] http://thesaurus.reference.com/ accessed 12/10/05

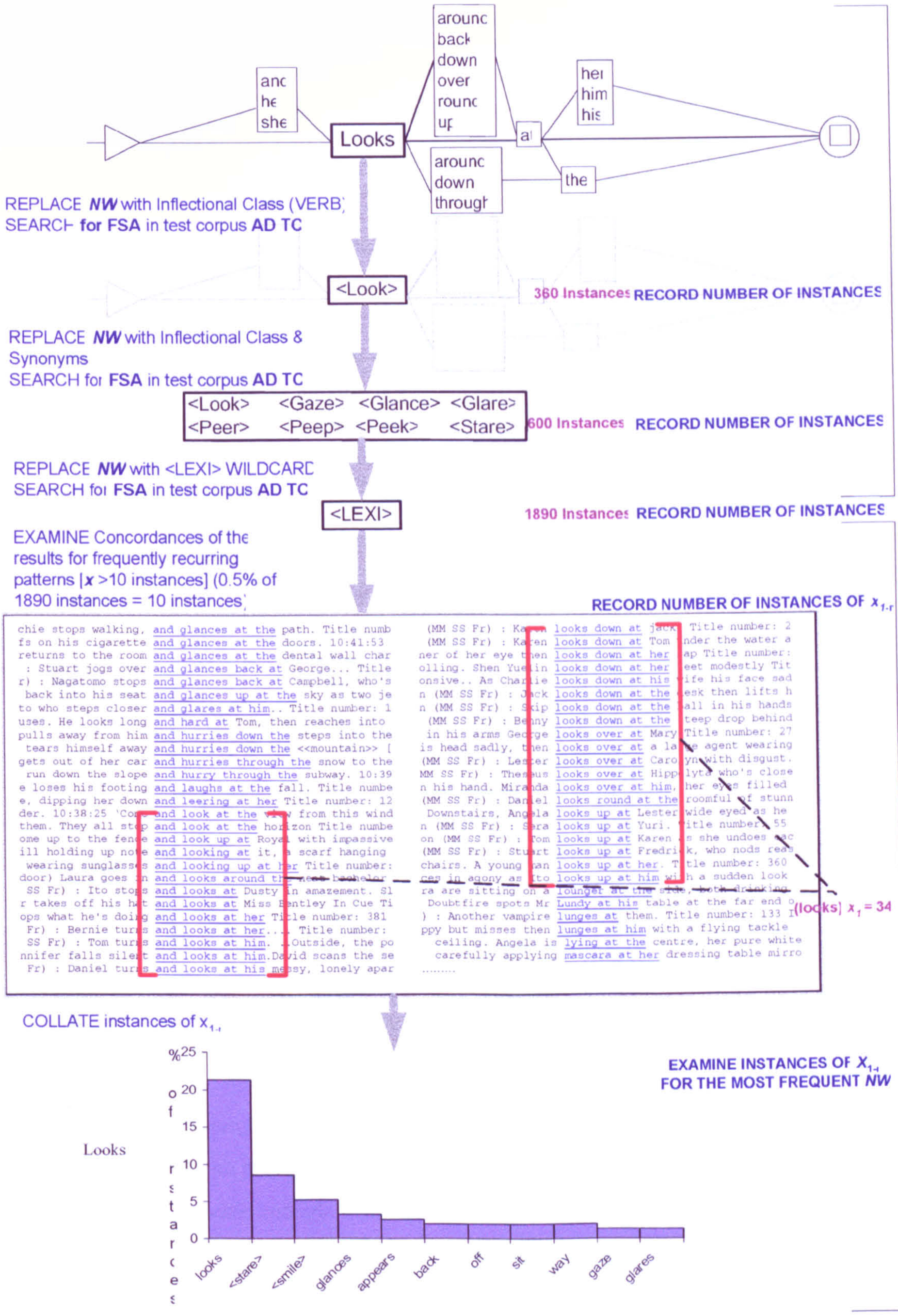


Figure 27 A step-by-step running example for collocation FSA of 'looks' in the Audio Description Test corpus to see whether 'looks' has alternative nucleates and whether it is the predominant node word.

Table 16 A table showing the percentage of instances of the node word, the node word's inflectional class (where available) and the inflectional class and synonyms with respect to the total instances returned. **TC** =Target Corpus, CP= Candidate Nucleate Collocation Phrase.

Original Node Word (OW)	Word Type	TC AD	CP AD	TC AD	CP SC	TC SC	CP SC	TC SC	CP AD
Around	OW %		0.40		0.73		0.66		0.61
	OW & Synonyms %		0.40		0.93		1.42		1.36
Away	OW %		8.04		17.79		17.10		5.35
	OW & Synonyms %		8.04		17.79		17.10		5.58
Behind	OW %		10.14		9.46		7.12		9.46
	OW & Synonyms %		12.16		10.14		7.39		10.14
Towards	OW %		1.41		9.08		1.97		1.03
	OW & Synonyms %		9.18		23.20		31.49		8.16
Door	OW %		2.41		2.34		2.12		2.16
	OW & Infl. Class %		2.51		2.44		2.12		2.33
	OW & Synonyms %		3.02		2.95		2.85		2.90
Room	OW %		1.37		1.34		1.16		1.15
	OW & Infl. Class %		1.39		1.36		1.16		1.16
	OW & Synonyms %		3.11		3.00		1.97		2.53
Looks	OW %		11.75		7.98		8.25		12.27
	OW & Infl. Class %		19.05		12.96		13.99		20.78
	OW & Synonyms %		31.75		21.68		20.36		30.07
Takes	OW %		14.22		11.48		10.42		11.11
	OW & Infl. Class %		14.22		11.48		10.42		11.24
	OW & Synonyms %		17.33		16.39		14.01		14.73
Turns	OW %		3.83		3.79		1.56		1.55
	OW & Infl. Class %		4.64		4.51		1.95		1.91
	OW & Synonyms %		5.21		4.51		1.95		1.92
Walks	OW %		5.94		1.80		0.33		3.49
	OW & Infl. Class %		8.80		2.70		1.48		6.01
	OW & Synonyms %		9.57		2.88		2.18		8.46
Eyes	OW %		*80.00*		*93.75*		*81.82*		*82.46*
	OW & Synonyms %		*100*		*93.75*		*81.82*		*82.46*
Head	OW %		*100*		*100*		*95.85*		*93.26*
	OW & Synonyms %		*100*		*100*		*95.85*		*93.26*

'Eyes' and 'head' seem to be the strongest collocation FSA; in other words, in this analysis, they give positive results on average 87% of the time for 'eyes' and 97% for 'head'. The other FSA do not seem to be *concise* enough when considering their nucleate words' phrases. It seems the phrases in the FSA do not only correspond to their nucleate words and are more generic. On average the 'looks' and 'takes' FSA return phrases, corresponding to themselves and their synonyms, 17.6% and 13.1% respectively. This means that those collocation FSA are only correct ~15% of the time. In the case of the 'around' FSA, the FSA only returned instances that had 'around' as their nucleate word 0.5% of the time, clearly *not* a strong collocation.

There was only slight evidence to suggest that one collocation FSA gave more results than the other. For instance the SC FSA for 'away' and 'towards' gave more positive results than their AD

FSA counterparts. In the case of the 'looks' FSA there is slight evidence to suggest that the AD FSA work better than the SC FSA. There is not a substantial difference in either case however.

Predominance and Privileged Position Test Discussion: Overall it seems that the collocation FSA need 'improvement' with respect to restricting the language that they contain and represent. Adding synonyms and the inflectional class to the nucleate word seems to improve performance of the candidate local grammar FSA but only marginally in most cases. The predominance test has shown that the FSA found by this research may be expanded to include more language from the same grammatical type but even those are restricted. In any case the predominance test has provided words, from the corpora, that can be interchanged with the nucleate word to form new *expanded* collocations. The privileged position test has shown that the node word is interchangeable with words of a similar meaning and that this may improve the accuracy of the collocation FSA. It must be noted that the collocations are generated automatically however and do not have the benefit of human judgment. There may be no substitute for distinguishing the meaning and context of a phrase manually.

3.3.5 VI. Measure Variance of Collocations Against the BNC

This step measures the variance of the frequency of the collocation FSA phrases for the nucleate words against the BNC. Figure 28 shows the steps of the method for this stage.

VI. MEASURE VARIANCE OF COLLOCATIONS AGAINST BNC

a. **COMPUTE** Frequency FSA_{FL} (w) || x_1(w)|| y_1(w) /*Frequency of the FSA_{FL}, phrase x_1 of nucleate of FSA_{FL} Phrase and y_1 of other word of FSA_{FL} phrase in $CORPUS_{FL}$*/

Frequency FSA_{GL} (w) || x_2(w)|| y_2(w) /*Frequency of the FSA_{GL}, phrase x_2 of nucleate of FSA_{GL} phrase and y_2 of other word of FSA_{GL} phrase in $CORPUS_{GL}$*/

Relative frequency $f_{FL}(FSA_{FL}) = FSA_{FL}(w)/N_{FL}$ /* N_{FL} = Total Number of words in $CORPUS_{FL}$*/

Relative frequency $f_{FL}(x_1) = x_1(w)/N_{FL}$

Relative frequency $f_{FL}(y_1) = y_1(w)/N_{FL}$

Relative frequency $f_{GL}(FSA_{GL}) = FSA_{GL}(w)/N_{GL}$ /* N_{GL} = Total Number of words in $CORPUS_{GL}$*/

Relative frequency $f_{GL}(x_2) = x_2(w)/N_{GL}$

Relative frequency $f_{GL}(y2) = y_2(w)/N_{GL}$

T-Score $FSA_{FL} = (f_{FL}(FSA_{FL}) - (f_{FL}(x_1) f_{FL}(y_1)/ N_{FL})) / SQRT(f_{FL}(FSA_{FL}) + (f_{FL}(x_1) f_{FL}(y_1)/ N_{FL}^2))$

T-Score $FSA_{GL} = (f_{GL}(FSA_{GL}) - (f_{GL}(x_2) f_{GL}(y_2)/ N_{GL})) / SQRT(f_{GL}(FSA_{GL}) + (f_{GL}(x_2) f_{GL}(y_2)/ N_{GL}^2))$

Mutual Information $I(FSA_{FL}) = \log2 (f_{FL}(FSA_{FL}) / f_{FL}(x_1) f_{FL}(y_1))$

Mutual Information $I(FSA_{GL}) = \log2 (f_{GL}(FSA_{GL}) / f_{GL}(x_2) f_{GL}(y_2))$

b. **IF** FSA_{FL} T-Score ≥ 1.65 → RECORD FSA_{FL} is 'Likely to co-occur with nucleate'

IF FSA_{FL} T-Score >> 1.65 → RECORD FSA_{FL} is 'Very likely to co-occur with nucleate'

IF FSA_{FL} T-Score < 1.65 → RECORD FSA_{FL} is 'Unlikely to co-occur with nucleate'

c. **IF** FSA_{FL} T-Score ≥ FSA_{GL} T-Score → RECORD FSA_{FL} 'Differs from general language'

ELSE RECORD 'No Significant Difference from general language'

1.1 **IF** FSA_{FL} MI >> 1 AND FSA_{GL} MI <2 → RECORD FSA_{FL} 'Differs from general language'

1.2 **ELSE** RECORD 'No Significant Difference from general language'

1.3 **REPEAT** steps a-d with each FSA for all collocations

Figure 28 Pseudo Code to measure the variance of nucleate collocation FSA phrases against general language (BNC).

3.3.5.1 Measure Variance of Collocations Against the BNC: Method, Definitions and Tools

Variance of the collocation phrases compared to general language, in this case the British National Corpus (BNC) was measured by making the frequencies of the corpora's collocation phrases relative and comparing T-scores and mutual information values.

Firstly, the *absolute frequency* of the instances of all the collocation phrases in each corpus were found. The frequency of the same collocation phrases were found in the BNC using the BNCs free trial query system [144]. Then the phrases' respective *relative frequencies* were calculated. Using the relative frequencies as the *probability* of the phrases within a corpus the *Mutual Information* of each phrase was calculated as well as each phrase's *T-score*.

The ***Mutual information*** (MI) of two random variables is a quantity that measures the independence of the two variables. In our case MI, $I(x;y)$, compares the probability of observing word x and word y *together* (the joint probability) with the probabilities of observing x and y *independently* (chance) [20]&[21].

$$I(x;y) \equiv \log_2 \frac{P(x,y)}{P(x)P(y)} \qquad [\eta]$$

Word probabilities, $P(x)$ and $P(y)$, are estimated by counting the number of observations of x and y in a corpus, $f(x)$ and $f(y)$, and normalizing by N, the size of the corpus. Joint probabilities, $P(x,y)$ are estimated by counting the number of times that x is followed by y, $f(x,y)$, and normalizing by N. [[21] pgs119-120]

For example $I(looks\ at;\ the)$ has a mutual information score of 11.41 $\{\log_2((88\times N_{sc})/(349\times 111184))\}$ where N_{sc} = 1,971,950 words and instances 'looks at' = 34, 'the' = 111184 and 'looks t the' = 88

The way that MI works is that:

- If there is a genuine association between x and y, then $P(x,y)$ will be much larger than chance $P(x)P(y)$ and consequently $I(x;y) >> 0$
- If there is no interesting relationship between x and y, then $P(x,y) \approx P(x)P(y)$ and thus $I(x;y) \approx 0$
- If x and y are in complementary distribution, then $P(x,y)$ will be much less than $P(x)P(y)$, forcing $I(x;y) << 0$

In this way MI allows the examination of the significance associations between the words and phrases in the corpora.

T-score is a measure of the '*t-test*' or the '*student's test*'. ***t-test*** "tells us how probable or improbable it is that a certain constellation will occur". The test looks at the difference between observed and expected means (frequencies) scaled by the variance of a sample of measurements, and tells us how likely one is to get a sample of that mean and variance assuming that the sample is drawn from a distribution with mean μ [[60] pg163].

T-statistic:

$$t = \frac{\bar{x} - \mu}{\sqrt{\frac{\sigma^2}{N}}} \qquad [\theta]$$

Where $\bar{x}$ is the sample mean, σ is the sample variance, N is the sample size and μ is the mean of the distribution

The use of the t-test, to find words whose co-occurrence patterns best distinguish between two words, was suggested by Church and Hanks and t-scores are calculated using an extension of the *t-test* [60]:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \qquad [\iota]$$

In this analysis Church's t-test was calculated not to distinguish between two separate collocate phrases, but instead to distinguish between its collocate phrase and its independent constituent phrases and words. Hence:

$$t = \frac{f(xw) - \frac{f(x)f(w)}{N}}{\sqrt{f(xw) + \frac{f(x)f(w)}{N^2}}} \qquad [\kappa]$$

Where $f(xw)$ are the instances of the word w occurring with x and, N is the sample size [60].

3.3.5.2 Measure Variance of Collocations Against the BNC: Results

The collocation FSA represent the most frequent phrases of the candidate nucleate words in the corpora. These phrases however were, for the most part, found to be more abundant in the film script corpora than in natural language.

Table 17 Shows T-score and Mutual Information values for collocate phrases common to the SC and AD corpora, compared to the BNC corpus, and the number of instances of the phrases in each corpus.

Phrase (Ph)	Inst w AD	Inst w SC	Inst w BNC	AD MI	SC MI	BNC MI	AD T-score	SC T-score	BNC T-score
looks up	316	714	80	6.26	5.01	1.78	17.54	25.89	6.33
looks up at	124	185	26	5.48	5.51	5.96	10.89	13.30	5.02
looks up at the	35	32	10	2.03	1.62	2.67	4.47	3.81	2.66
looks at	349	1235	1091	5.94	5.49	4.21	18.38	34.36	31.25
looks at the	88	252	399	1.86	1.86	2.59	6.81	11.49	16.67
and looks at the	22	32	20	3.24	2.65	0.93	4.19	4.76	2.13
looks around	90	287	37	6.27	5.39	2.92	9.36	16.54	5.28
looks around at	8	36	0	3.34	4.46	N/A	2.55	5.73	N/A
looks back	35	218	108	3.59	3.87	3.30	5.42	13.76	9.34
looks back at	34	85	21	6.79	6.10	5.22	5.78	9.09	4.46
looks back at the	16	27	8	2.77	2.49	2.65	3.41	4.27	2.38
looks down	100	271	55	5.05	4.34	2.41	9.70	15.65	6.02
looks down at	72	125	16	6.36	6.34	5.80	8.38	11.04	3.93
looks down at the	29	47	6	2.54	2.74	2.63	4.46	5.83	2.05
looks over	16	171	32	2.78	4.00	1.12	3.42	12.26	3.05
looks over at	20	43	4	7.15	5.47	4.58	4.44	6.41	1.92
looks over at the	3	11	0	1.12	2.18	N/A	0.93	2.59	N/A
looks through	16	40	17	3.09	2.15	0.90	3.53	4.90	1.91
looks through the	12	23	6	3.44	3.35	2.54	3.14	4.33	2.03
the door	617	1971	12266	3.59	4.64	3.13	22.77	42.61	98.06
opens the door	68	143	66	6.95	7.32	8.30	8.18	11.88	8.10
turns to	299	691	441	4.28	4.23	2.22	16.40	24.88	16.49
the room	254	1097	7705	2.66	2.71	2.14	13.42	28.05	67.85
of the room	43	234	1491	3.34	3.71	2.67	5.91	14.13	32.53
out of the room	21	72	387	6.49	6.01	7.04	4.53	8.35	19.52
living room	11	205	721	8.72	9.53	7.32	3.31	14.30	26.68
the living room	17	121	415	4.48	3.39	3.25	3.94	9.95	18.23

The t-test results gave more varied results. The T-score values were on average much higher than the MI values and the SC and AD phrases did not on average exhibit similar results. However it must be acknowledged that this is a different statistic than Mutual Information. With the t-test the likelihood that one phrase is more common than the last is being examined. Therefore we are examining how likely it is for certain phrases to occur.

With the exception of the phrase 'looks over at the' in the AD corpus (T-score=0.93), all the difference is much greater than 1.65 standard deviations in the AD and SC corpora. With respect to the phrases in the BNC, the T-scores are all greater than 1.65 standard deviations. This

indicates that the phrases are all very likely to occur in all corpora with respect to their constituent words (e.g. 'looks' and 'at' → 'looks at') and their constituent phrases ('looks at the' and 'and' → 'and looks at the').

In some cases the BNC T-score values were much higher than the SC and AD T-scores for a given phrase. For example the BNC T-scores for 'out of the room' and 'the living room' were approximately two times bigger than the SC and AD T-scores. This indicated that the phrases are very frequently used in everyday language, but since the T-scores were high in the AD and SC corpora also, this indicated that the phrases were also commonly used in the film script corpora. Overall the t-test demonstrated that the likeliness of the phrases in the AD and SC corpora differed from the BNC.

To show that the language used in the AD and SC corpora differs from that used in the BNC (natural language) the ideal result would be a MI value of 0-1 in the BNC, indicating no interesting relationship, corresponding to a high MI value (>>1) of the same phrase in the AD and/or SC corpora. We only have that in a few instance of the MI value for the FSAs being much higher than the MI of the BNC, for instance in the case of 'looks up', 'and looks at the', 'looks over' and 'look through' the AD and SC MI values are >2 and the BNC MI values are ~1.

3.3.5.3 Measure Variance of Collocations Against the BNC: Discussion

Mutual Information allows us to see how *similar* the distribution of each phrase is to that of each other corpus' phrase. The higher the MI value the stronger the association between the involved words in a phrase in a given corpus. Thus, comparing MI values of certain phrases across corpora (which can be done since the MI values are normalised) gives an indication of how similar the language use of a given phrase is in one corpus to another. A *high* MI value will help support the argument that, because the words of a phrase are frequently co-occurring, the phrases can be included in developing local grammars. It also gives us information about how dissimilar a phrase's use is in a corpus with respect to that of natural language or the BNC corpus. MI also compares the AD or SC collocate phrase to the number of occurrences of that phrase in the BNC.

The aim is to show that the language used in the AD and SC corpora differs from that used in the BNC (natural language). The ideal situation is a MI value of 0-1 in the BNC, indicating no interesting relationship, corresponding to a high MI value (>>1) of the same phrase in the AD and/or SC corpora. However, a better indication of dissimilarity is the T-score. Church et al state: "Mutual Information is better at highlighting similarity; t-scores are better for establishing differences amongst close synonyms."[[21] pg 143]

In Church et al. [21] the t-score was calculated by:

$$t \equiv \frac{P(w|x) - P(w|y)}{\sqrt{\sigma^2(P(w|x)) + \sigma^2(P(w|y))}} \qquad [\lambda]$$

Where $P(w|x)$ and $P(w|y)$ are the probabilities of a word w occurring given an occurrence of x or y respectively [12].

A simplified version of Church et al. [21] t-score was calculated as follows:

$$t = \frac{f(xw) - f(yw)}{\sqrt{f(xw) + f(yw)}} \qquad [\mu]$$

Where $f(xw)$ and $f(yw)$ are the instances of the word w occurring with x or y.[60]

Church et al. [21] state that, the T-score allows for negative statements to be made about the strength or significance of the words in a phrase, which could otherwise not be made with Mutual Information values. They compare two *similar* phrases 'strong support' and 'powerful support' to each other, within the same corpus. They point out that a difference of at least 1.65 standard deviations is necessary to ensure that the difference is real and not due to chance. *P(powerful support)* was greater than *P(powerful) P(support)* but only by *one* standard deviation ($t \approx 0.99$) leaving about a 30% chance the difference is a fluke. Comparing *P(powerful support)* with *P(strong support)* gave a highly significant result, $t \approx -13$, showing that *P(powerful support)* is thirteen standard deviations less likely than *P(strong support)* [[20]pg 123].

In this analysis Church's t-test was calculated but not to distinguish between two separate collocate phrases, instead to distinguish between its collocate phrase and its independent constituent phrases and words, i.e. P('looks at') ≠P('looks') P('at'). In our case the t-test is:

$$t \approx \frac{\frac{f(xw)}{N} - \frac{f(x)f(w)}{N^2}}{\sqrt{\frac{f(xw)}{N^2} + \frac{f(x)f(w)}{N^4}}} = \frac{f(xw) - \frac{f(x)f(w)}{N}}{\sqrt{f(xw) + \frac{f(x)f(w)}{N^2}}} \qquad [\nu]$$

The t-test served to show that a collocation phrase's constituent elements are *not* independent and do form a collocation. The t-test and T-scores also served the purposes of testing the phrase *significance* for a given corpus (to see whether it is significantly above chance) and allowed the comparison of the *phrase significance* to that of natural language (BNC corpus).

3.4 Collocations FSA as Templates for Information Extraction

It is our belief that the expanded, generalised collocation FSA of the AD and SC corpora can provide *specific* information about film content: they can provide film *events*. In turn, we believe

that the collocation FSA and frequent neighbouring language can be used to develop templates that can be used to extract events and/or other meaningful information about films. Thus, we examine the surrounding language of the FSA for any idiosyncrasies i.e. frequent patterns of repeating phrases, grammar or formatting.

3.4.1 Collocations FSA as Templates for Information Extraction of Film Content

The generalised collocation FSAs were formed systematically, based on the frequency of repeating patterns; no semantic content was considered. Frequently recurring semantic units, such as proper names and nouns, of places and people, would not have been returned by the collocation method [31] (part **III.** of method).

The re-collocation of the phrases found in each FSA gave us an insight into the frequent language used in conjunction with the phrase. For example 'he', 'she', 'him', 'her', 'a/the' {curb, corner, door} and 'his' {shoulder, neck, watch} were all frequent words neighbouring the collocation FSA phrases. This prompted the manual examination of the surrounding language of the FSA in the corpora, using Unitex, to see what language and grammar types were frequent in proximity to the collocation FSA phrases. The FSA for each nucleate was placed into the corpus analyser Unitex (as whole FSA functions) and a concordance was created for each corpus' FSAs.

The concordance analysis revealed frequently collocating *types* of word around the collocation FSA phrases. These word types were grouped according to grammar. For example proper nouns, pronouns and nouns were frequently in proximity of the "Looks" FSA phrases. These proper nouns, pronouns and nouns referred to persons or *characters*[21] in the films, e.g. phrases such as "*Tom* looks at the *Preacher*". Thus the category "Character" was created. In the case of the FSA for "Door" and "Room" there were many verbs preceding the collocation FSA phrases. Thus the category "Actions" was chosen to represent them. The collocation phrases frequently involved the words 'his' and 'her' that were mostly followed by nouns. The nouns predominately referred to objects and consequently the category "Objects" was formed. In the case of the Screenplay FSA for "INT", "EXT", "Day" and "Night", proper nouns referring to places were in close proximity to the phrases >75% of the time. As a result the category "Location" was used to represent them.

Therefore, the categories: "Character", "Object", "Action" and "Location" were formed. It must be noted that the categories were also inspired by the generalisation rules and categories discussed in the work of Sowa [91] [92], i.e. Abstract Transfer (A-Trans) and Physical Transfer (P-Trans)

---

[21] Here the word *character* is not a trivial choice as it implies, not only human characters (persons) but other species e.g. aliens, animals or animated physical objects; generally any sentient life form.

(Schank [84], [85]) were considered as "Action" and Sowa's generalisation of pronouns and names to 'person' was considered for "Character"[22]. Figure 29 depicts the FSA for "Looks"[23] with the appropriate categories found in the concordance analysis added. All the diagrams depicting the FSAs and the categories can be seen in Appendix C.

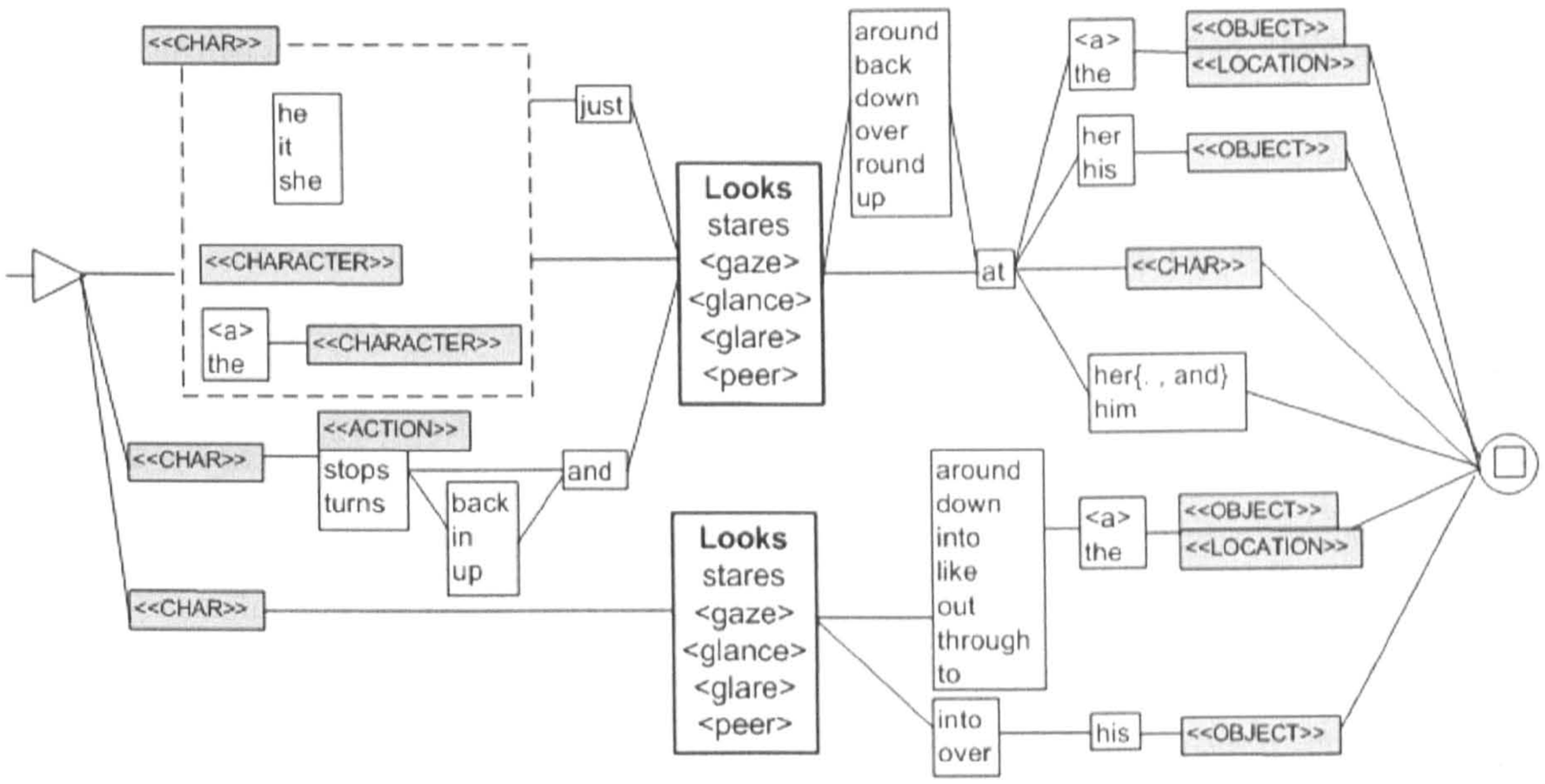


Figure 29 A Finite State Automata diagram for the phrases in the FSA for "Looks" (joined and simplified for both corpora) with the frequent neighbouring language categories CHARACTER, OBJECT, LOCATION and ACTION included. FSA diagrams were produced for all nucleates' FSA phrases and their frequent neighbouring language and can be seen in Appendix C and on **CD**: Additional Thesis Material.

Having located the categories in close proximity around the collocation phrases it became apparent that the collocation phrases *could* be used to extract information from the films scripts such as locations, characters and objects in scenes, and certain actions or happenings performed, as per the categories. More importantly *Events* could possibly be extracted. For instance, the collocation phrases for the open class words "Eyes" and "Head" frequently involved actions such as characters *opening* their eyes and *shaking/nodding* their heads. This was thought to indicate some non-verbal communication between characters. In the case of "Looks" and "Turns", typically, a character 'looks at' or 'turns to' another character or object. In other words a character was focussing their attention on something. Other possible events such as characters changing location (collocations of "door" and "walks") and scene changes ("INT", "EXT", "day" & "night") were also observed.

Having begun to observe evidence of events in the film scripts the question arose: would it be possible to extract these events? The Audio Description scripts were accompanied with time codes and the screenplays time and acts could be matched to the video through closed-caption

---

[22] The word Character was used instead of person, as it was more relevant to movies.

[101] or simple line numbers [83]. Thus, it could be possible to extract the events based on the textual description in the script and the relevant time code from the script. Also of importance, was the fact that the possible events were frequent occurrences of phrases containing other important film information such as what characters are on-screen at which time, the location of the scene, the time of day, what objects are involved in the scene and what actions or happenings are occurring. Hence possible film information surrounding the collocation FSAs could also be extracted. Therefore, from the concordances, the collocation FSA phrases and the categories chosen, *templates* for information extraction were developed. The next section outlines these templates with respect to the events and film elements that could be extracted from the Audio Description scripts and screenplays, based on the expanded collocation FSAs.

3.4.1.2 Focus of Attention Event Template.

As mentioned the collocations of the open class words "Looks" and "Turns" seemed to indicate that a character was focussing their attention on another *character*, an *object* and, less frequently, on a *location* (e.g. the tower, the kitchen). This was observed through the concordances of the "Looks" and "Turns" FSA phrases as well as the privileged position study in section [3.3.3] where, on average, >60% of the concordances of 'looks at' and 'turns to' (the most frequently recurring phrases for both open class words' FSAs), involved a character 'looking at' or 'turning to' another character. If we include objects and locations as elements of a character's 'focus of attention' in the concordances, the instances increase to approximately ~80% focus of attention. This led to the labelling of this particular event the "Focus of Attention" event.

In the same light as the FSA phrases for "Looks" and "Turns" having a *focus of attention* on something, the FSA for "Takes" also had characters focussing their attention, this time predominately on objects. The most recurring phrase shown by the re-collocations of "Takes" is 'takes a deep breath', however, the concordances have many instances of phrases describing characters taking objects e.g. 'Almasy takes the spade' and 'she takes the rope'. This does show a focus of attention of a character towards an object and so is included as a focus of attention event.

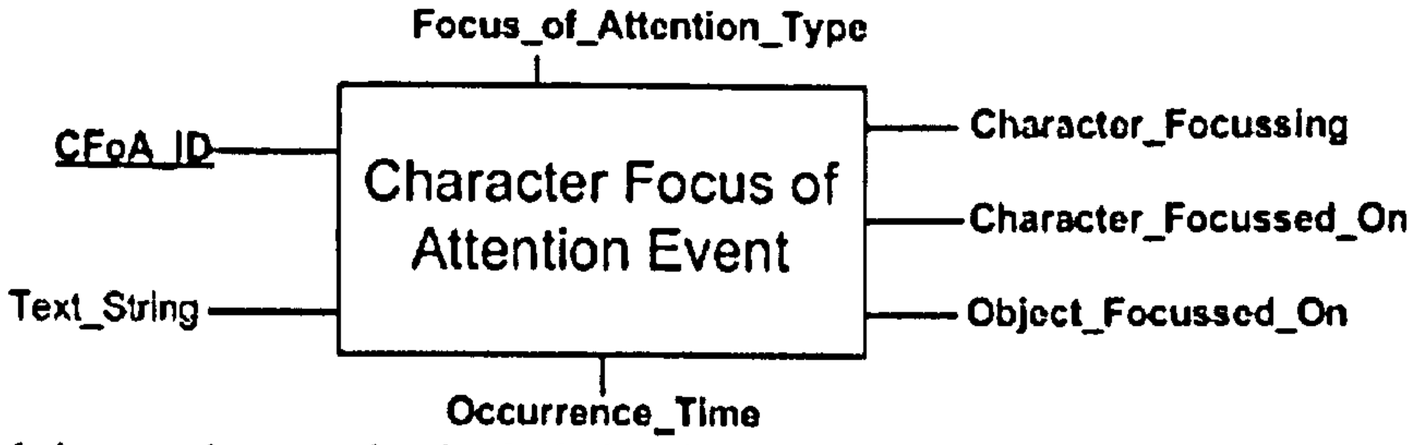
The template design was based on what textual information about film content could be extracted from the collocation phrases. FSA diagrams such as Figure 29 were constructed around the FSA phrases and sample text was gathered from the corpora using Figure 29 as the search criteria to provide examples of information that could be gathered for the focus of attention event.

It was found that information such as what character was *performing the focussing* (Character_Focussing) and which character or object was being *focussed on* (Character_Focussed_on). Also the time at which the event was occurring could be extracted.

---

[23] Figure 29 includes synonyms of looks that frequently returned in Section 3.3.4

This led to the diagrammatical representation of the template as an entity, for the event itself, and a series of attributes that could be captured, see Figure 30. An Identification code was added to differentiate each event (CFoA_ID), the type of focussing, i.e. actively focussing on something: ACTIVE (looks, turns) or passively showing something attention: PASSIVE (takes) to the character or object was recorded and the entire string containing the FSA phrases, the film



elements and time code was also included to be extracted.

Figure 30 Represents the 'Character Focus of Attention' event and its attributes that was used in the Information Extraction template to extract the characters' focus of attention event.

Many iterations of the event were modelled with many more attributes capturing things such as what accompanying actions, if any, were present with the FSA phrase and any secondary or tertiary characters involved. But these attributes were omitted, as they were deemed unnecessary.

3.4.1.3 Change of Location Event Template

On inspection of the collocation phrases for "Walks", "Room" and "Door" and their neighbouring text, both manually (concordances) and systematically (re-collocations) there seemed to be an indication that characters were changing location. There were instances of characters moving from one room to another (walking in and out of rooms and areas), entering and exiting rooms (going through doors) and generally moving: through (orchards, rooms, forests), up (ladders, stairs), down (hills, sand dunes), along (the wall, the desert dunes) etc.

Therefore, a template was developed for *characters changing location,* which was based on what textual information about film content could be extracted from the FSA phrases of the nucleates "Walks", "Room" and "Door". An FSA diagram was constructed around the FSA phrases (see Appendix C and **CD**: Additional Thesis Material) and sample text was gathered from the corpora using the FSA diagram as the search criteria to provide examples of information that could be gathered for 'Change of location' event.

Information such as what character was changing location, the initial and final locations of the character and what state of motion the character was in: ENTERING, LEAVING or WITHIN/ON (State_of_Character) were found to be frequent and could readily be extracted. Also the time at which the event was occurring could be extracted. This led to the diagrammatical representation of the template as an entity, for the event itself, and a series of attributes that could be captured,

see Figure 31. An Identification code was added to differentiate each event (CoL_ID), the type of location change, i.e. an ACTIVE change of location or a description of a location change (DESCRIPTIVE) was recorded and the entire string containing the FSA phrases, the film elements and time code was also included to be extracted. In some cases information such as other characters in the scene (Secondary_Character_Involved) and other actions involved, whilst change of location was occurring, were also available and thus also chosen for extraction.

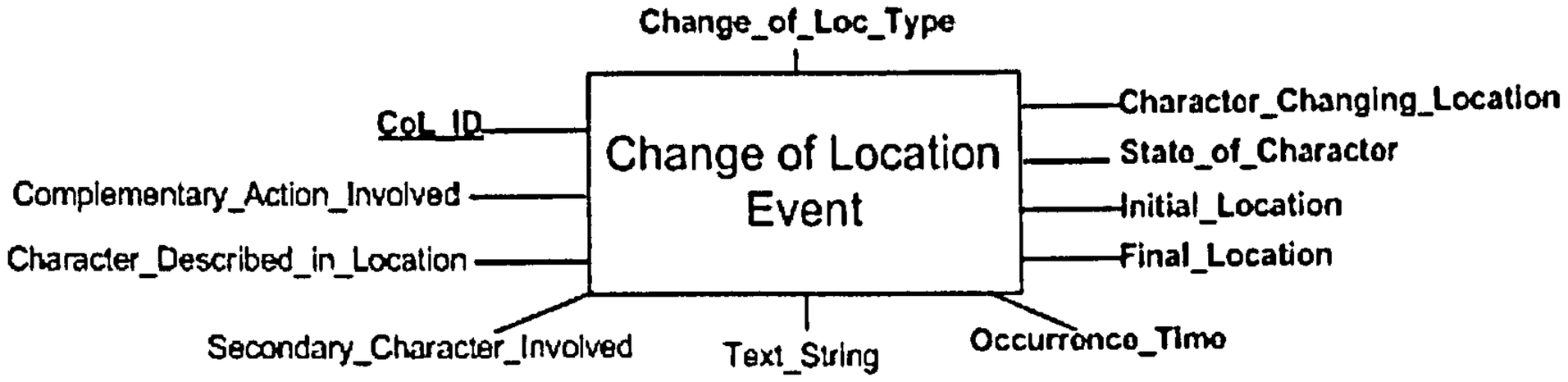


Figure 31 A diagram representing the 'Change of Location' event and its attributes that was used in the Information Extraction template to extract the character changing location event.

3.4.1.4 Non-Verbal Communication Event Template

The collocation phrases for the nucleates "Head" returned >85% of the time the action 'shakes' or 'nods' i.e. 'he shakes his head', 'shakes her head'. For the nucleate "Eyes" >60% of returned instances involved an *action* before the phrase, usually 'opens' or 'closes' and sometimes 'rolling/rolls' (his/her eyes). This was reflected when a FSA of the collocation phrases "Eyes" and "Head" was used as search criteria in Unitex and the category ACTION was searched for in close proximity to the respective collocation phrases in the results. The instances of ACTION words around the FSA phrases were high, specifically the words 'shakes', 'opens' and 'closes'.

Intuitively, the idea that characters may be communicating non-verbally through the *motion* of their eyes or heads led us to believe this could be conceived as a non-verbal communication event. It was speculated that it could be possible to capture information such as who is communicating non-verbally, using which body part (Body_Part_Involved) and what respective action was involved: shaking, nodding, opening, closing etc. The time code of when the event occurs was captured and an ID code for the event was also given (NVC_ID).

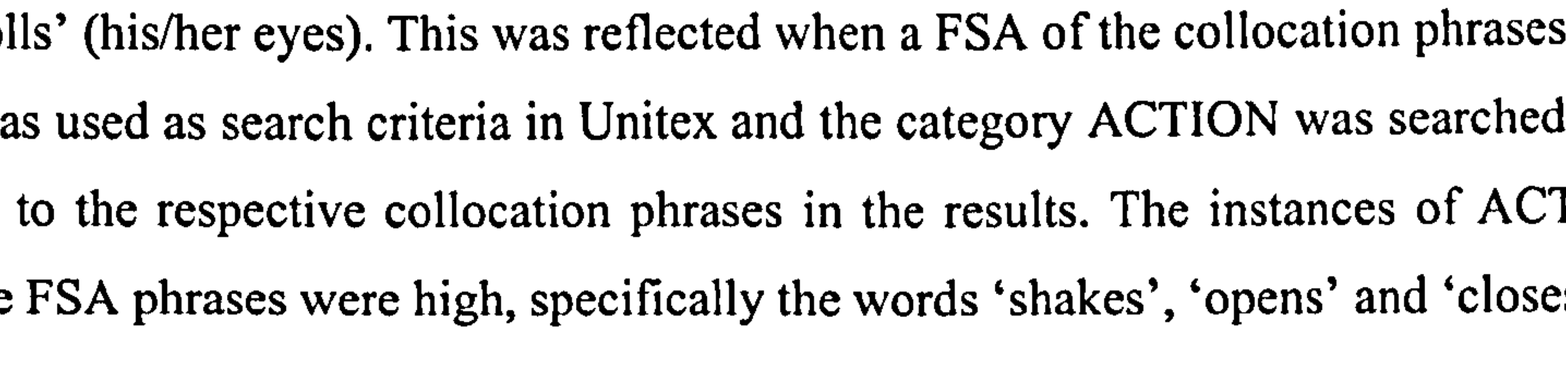


Figure 32 A diagram representing a 'Non-Verbal Communication' event and attributes used in the Information Extraction template to extract information about characters communicating non-verbally.

3.4.1.5 Scene Change Event Template

The nucleate words for the SC corpus "EXT", "INT", "Day" and "Night" largely indicate when a scene is changing and what the new location is (EXT→ Exterior, INT→ Interior) and what time of day it is (daytime, night time, dusk, dawn, afternoon, twilight etc.). Over 90% of instances both in the concordance and collocation analysis of the SC corpus show these open class words being used in that way. Almost all instances of the nucleate words and collocation phrases are accompanied by a new location, e.g. a place (the amphitheatre), country, city, building, room name (kitchen, hotel bar, motel room), etc.

Also some of the nucleate word "Room"'s collocation phrases provide information about scene change. In the audio description there are often descriptions of rooms alone on one line. For example, the phrases "03:01:02.44 The Hotel Bar," or "00:01:02 A Motel room," are not uncommon. These phrases are simply a description of a new room, area or location that has been entered. This provides information of a new location and the film time that it has been moved to.

These accompanying locations and the nucleate words themselves, provide spatial and temporal film element information that can be captured with this event. This allows scene change information to be captured such as, whether the new scene is an interior or exterior location, if the new scene is in the daytime or night-time and what the new location is. A text string providing all the aforementioned information and the time code, which it occurs, is also captured. The event is given a unique ID (ScCh-ID) and its time code is captured (Occurrence_time).

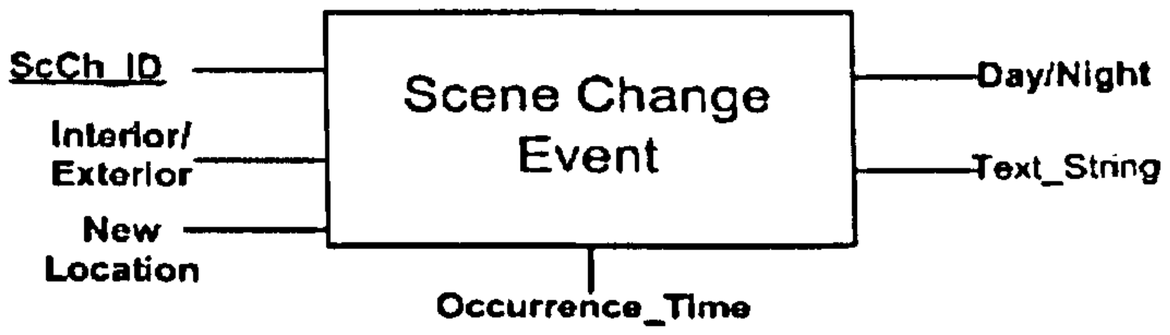


Figure 33 A diagram representing a 'Scene Change' event and its attributes used in the Information Extraction template to extract information about scenes changing and their new locations.

3.4.1.6 Overall Templates and Entity Relationship Diagram

Having described the events that can possibly be extracted using the collocation phrases we needed to put this in perspective with the overall picture of narrative that we already had. The model of narrative based on Chatman's [18] ideas of narrative was considered. Our model (Figure 9) of Chatman's work follows the concept of a story, consisting of existents and events with many manifestations. This is what the domain of film presents: a story (the film's story) consisting of existents and events and being manifested as video data and two types of film script (screenplays and audio description).

Figure 34 shows an entity relationship diagram that attempts to represent film in terms of Chatman's idea of story being composed of events and existents. It also represents manifestations

of films through video data and two types of script: the screenplay and the audio description. Figure 34 speculates that we can capture information about the events "**Character Focus of Attention**", "**Change of Location**", "**Non-verbal Communication**" and "**Scene Changes**" and the existents: characters, locations and objects, which exist in the film scripts.

The question then became, how much information about the events and existents could be extracted and ultimately how much information about the film's *story* could be extracted? Using the aforementioned events as templates for information we set about trying to answer these questions by developing heuristics based on the templates. Chapter 4 provides some answers

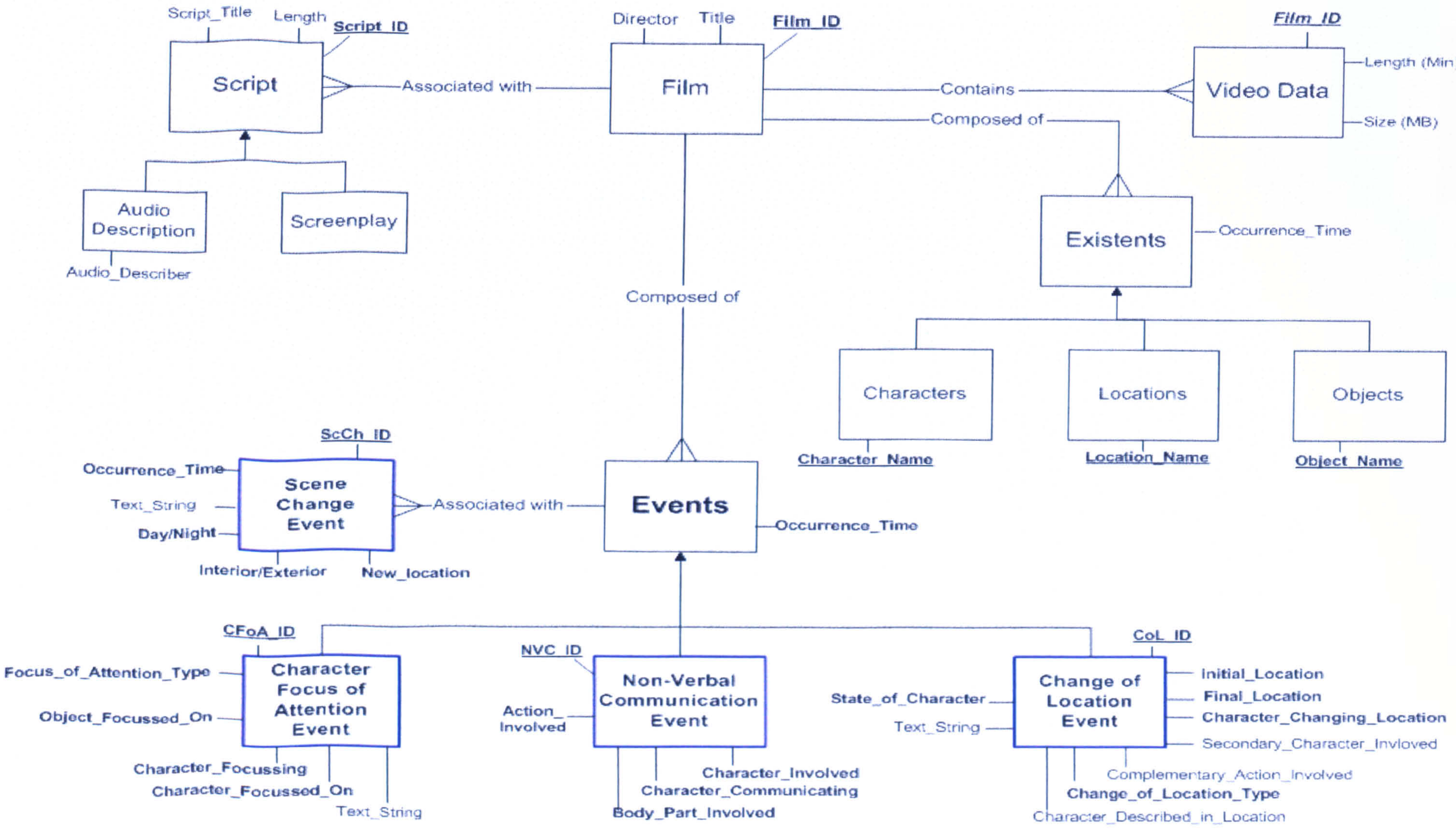


Figure 34 An Entity Relationship diagram depicting entities associated with film and film scripts and their relationships. The diagram incorporates the events: 'Character focus of attention', 'Non-verbal communication', 'Change of location' and 'Scene change' formed from the nucleate word collocations of both corpora.

3.4.2 Discussion: Does a Special Language or LSP Exist in the AD and SC Film Scripts?

Sections 3.3 and 3.4.1 have shown that there is a statistically high regularity of certain collocation phrases in the film script corpora. When compared to general language (BNC), the collocation phrases, for the most part, showed that the nucleate words in the collocation phrases co-occurred, much more frequently in the film script corpora, than in general language, (see Table 17). That is to say that there was a high difference in deviation when comparing the collocations' MI and t-test values. In fact there were no instances of some of the film script collocation phrases in the BNC. Also it was seen that both corpora, independently, had above ten open class words in the top 100 open class words compared to 2 open class words in the BNC's 100 most frequent words.

Further to this, Appendices B and C and section 3.4.1 show that there is a high similarity and regularity of collocation phrase use between the AD and SC corpora. This indicates that there is information that is repeated in film scripts and that is common to all film scripts. This has been represented as four types of event that are common to film (Appendix D and **CD**).

These idiosyncrasies that are present in both corpora, and are used in different measures to general language, are indications of a language for special purpose, an LSP, being used by writers of both audio description scripts and screenplays. The collocation FSA diagrams, the four event templates and the diagrams in Appendix D also suggest that there are local grammars LGs in the film scripts also. The idiosyncrasies in the AD and SC provide evidence to support the notion of a film script sub-language. Evidence of a sub-language will provide further evidence of LGs and an LSP. Thus the next section explores whether there is evidence of a sub-language in film scripts.

3.4.2.1 The Link between Sub-Language and Local Grammar

Harris says that a ***sub-language*** is subsets of the sentences of a language that satisfy the grammar without involving any sentence that is outside the subset [44]. Harris [1968, p152] defines sub-language as: "certain proper subsets of the sentences of a language may be closed under some or all of the operations defined for a language, and thus constitute a sub-language of it". There are certain factors, which help categorise sub-languages [12]:

- Limited subject matter.
- Lexical, semantic and syntactic restrictions.
- 'Deviant' rules of grammar.
- High frequency of certain constructions
- Text structure.
- Use of special symbols.

Sub-languages have certain characteristics that can be considered the 'analyses steps', which need to be followed to ascertain whether a language can be regarded as satisfying the requirements of the sub-language model. These characteristics can be considered the analysis steps to satisfy the requirements of the sub-language model. Harris states: "if we take as our raw data the speech and writing in a disciplined subject-matter, we obtain a distinct grammar for this material" [[12], p.235]. In this case the raw data was film scripts and the subject matter: narrative. This work examined these film scripts in terms of the characteristics that define a sub-language and considers these characteristics as 'analyses'. If a Sub-language for texts that describe films can be shown to exist, then that can be considered an extended local grammar. In this case, the sub-language factor *–limited subject matter–* referred to film scripts, with the 'subject' being a film's story. The subject matter was limited in the sense that we are limited by the constraints of a story. Some events *must* occur for a story to continue and cause-effect chains of events do exist, linking these events. This work's collocation analysis of the corpora yielded some encouraging results and exhibited most of the factors that helped categorise a sub-language.

The frequency analysis of the corpora and the comparison of the corpora to the BNC yielded some *lexical restrictions*, e.g. 'looks', a very frequent word in both corpora, was predominately followed by 'at' and predominately preceded by a proper noun or pronoun. Also in the Screenplay (SC) corpus the terms 'INT/EXT' were only used to describe whether a scene is internal or external and the term 'cut' was very frequent in the SC corpus, predominately used in edit-effects such as the change from one scene to another, e.g. 'cut to: INT'.

Certain *semantic* restrictions, specific to the corpora were also discovered from concordances and collocations, i.e. terms or lexical units that only had meaning in the context of film. For instance the edit-effects are used frequently in films and do not necessarily have meaning elsewhere in natural language. E.g. 'cut to: INT', 'fade in/fade out'. Also, certain directional descriptions, i.e. the description of actions that may be happening in a scene may be considered semantic restrictions, for example, '(over comlink)' is used to describe the fact that a character speaking over a comlink.

Although the collocation analysis of each corpus revealed some *syntactic* restrictions, by simply visually examining the scripts and screenplays certain syntactic restrictions seem evident. For instance, the screenplays are arranged in a certain format, syntactic restrictions –such as a character's dialogue appearing on a new line after their name and filming directions being written in brackets– are but a few examples of formatting in screenplays. In the case of the audio description scripts the semantic restrictions were not that evident and were discovered through collocation analysis. The audio description corpus contained on average more descriptive open class words than the screenplay corpus as well as more action words. This was possibly due to the more *descriptive* nature of audio description. This led to syntactic restrictions in the form of short descriptive sentences.

Collocation analysis of the corpora led to the belief that *'deviant'* rules of grammar were not common in the corpora. This may be because the language used in the corpora does not deviate much from universal grammar, due to the language being mostly dialogue and description of the scenes. On further analysis however it was observed that certain frequently used syntax, that appeared mostly in the screenplay corpus did deviate from universal grammar. For instance, many edit-effects in both corpora had no subject in the sentence or phrase, e.g. 'fade to black' or 'dissolve out'. Some directional information also deviated from the universal grammar: 'Security guard (into phone)' or simply the phrase '(over comlink)' followed by some instruction.

Both corpora revealed a high frequency of certain *constructions or lexical patterns* that were quite similar. This was discovered predominately from the collocation analysis, where the most frequent neighbouring words of each analysed phrase or word were found. Both corpora contained a high frequency of the constructs 'looks at the', 'he turns to', 'out of the room', 'the/a man in the' and 'shakes his head'.

As far as *text structure* is concerned there is a definite, separate text structure to both corpora. The audio description is structured in time intervals; every hour of the film is separated into a different part, using the time code. The text itself is structured with a 'cue' that tells the audio describer when to start speaking followed by the audio described text. In the case of a screenplay the text is structured in acts, scenes and/or shots segments. The text itself is arranged a title and author followed with a location and possibly a temporal description and then either a character name and dialogue or a description of the scene or character or in many cases both.

Evidence was found of the consistent use of *special symbols* in the corpora. In the case of audio description many tags used in the process were removed from the corpus. However these tags can be considered special symbols. For instance before any audio describer speaks the text from a script the tag <NORMAL> is used. The time code is also in a NB:NB:NB format e.g. 01:00:50 which means fifty seconds into the film. The screenplay has its own use of symbols, such as the constant use of brackets for descriptive directions and '-##', '<<' precedes and follows some edit-effects e.g. '-##BLACK AND WHITE SEQUENCE##-', or '<<COLOUR SEQUENCE>>'.

3.4.2.2 Conclusion: Are FSA Collocation Phrases Local Grammars and is there Evidence of an LSP?

The results provided in this chapter do provide evidence to support the hypothesis that there is an LSP in film scripts and the collocation phrases could be considered Local Grammars. The analysis conducted on both the SC and AD corpora show strong evidence that these texts, Screenplays and Audio Description scripts, do conform to the model of a sub-language and thus could be considered languages for special purpose. The link between sub-language and local grammar

made by Huston and Sinclair [48] where they considered small (but significant) sub-languages, and sub-language descriptions, as extended local grammars, strengthens the argument that the collocation phrases are local grammars.

The evidence of and LSP and Local Grammars can be seen in both corpora, results such as the collocation phrases, the collocation finite state automata diagrams and the Mutual Information and T-test analyses. Overall the evidence for the existence of local grammars in these texts is present: the repetition of lexical patterns, strong statistical evidence, and evidence towards the existence of a sub language or language for special purpose and collocations FSA that may be considered 'local grammar' finite state automata.

Thus, there is evidence to support the claim that the collocation FSA found systematically, and for the most part, automatically by this analysis can be considered *local grammars* of film scripts and that Audio Describers and screenwriters use a language for special purpose.

3.5 Conclusions

Initially we asked the questions: what information about film content do these audio description scripts and screenplays texts contain? Are the ways in which this information is expressed sufficiently regular to allow for reliable information extraction, and the automatic generation of Information Extraction (IE) templates and algorithms? Corpus linguistics, specifically collocation analysis, was employed to help answer these questions as it explores patterns of language use, is empirical and has been used in information extraction to populate databases of 'event structures' derived from corpora. Representative corpora of audio description scripts and screenplays of different genre, style and length were gathered from various sources. Corpus linguistics was predominately used to procure significant collocations from the AD and SC scripts. These collocations were expanded and represented as FSA and tested against general language (the BNC). It was found that the expanded collocation FSA fit into four main events.

Thus, in answer to: what information do audio description scripts and screenplays provide about film content? We were able to find patterns of repeating language, idiosyncrasies in language used in AD and SC texts, evidence of local grammars, evidence of a language for special purpose and four event types: 'Focus of Attention event' (FOA), 'Change of Location event' (COL), 'Non-Verbal Communication event' (NVC) and 'Scene Change events' (ScCh), which are common to AD and SC film scripts.

We claim that this chapter shows that the language used in audio description and screenplays contains idiosyncrasies and repeating word patterns, specifically an unusually high occurrence of certain open class words and certain collocations involving these words, compared to general

language. The existence of these idiosyncrasies means that the generation of information extraction templates and algorithms can be mainly automatic. These idiosyncrasies are critical for our work; they provide the basis of analysing film content. If we did not have these idiosyncrasies, we could not show that the information in the language of film scripts was expressed with sufficient regularity to allow for reliable information extraction, and the automatic generation of Information Extraction (IE) templates and algorithms. The regularities *are* present and therefore it may be possible to reliably extract film content information from film scripts.

We also claim that there are four types of event that are commonly described in audio description scripts and screenplays for Hollywood films: FOA, COL, NVC and ScCh. For each of the four events, we provide a template with attributres and a formal description of its associated collocations in the form of a Finite State Automaton. This is important for film video data analysis and may be useful to film scholars, narratologists and the video data analysis community. It is also usfeul in defining a 'set of things' in film and may further the concept of a film ontology.

Following on from this chapter, Chapter 4 addresses the question: are the ways in which this information is expressed sufficiently regular to allow for reliable information extraction, and the automatic generation of Information Extraction (IE) templates and algorithms? To do this, it explores the hypothesis that information extraction of film content can be conducted on the AD and SC corpora based on the four event templates: FOA, COL, NVC and ScCh and that a database of information about film can automatically be populated from these events to produce novel applications with respect to accessing film data.

4 Automatic Extraction of Film Content from Film Scripts

This chapter seeks to explore how the basic knowledge from Chapter 3 can be applied to support novel applications that rely on machine-processable representations of film content. Continuing the question from Chapter 3, Chapter 4 asks: are the ways in which this information is expressed, sufficiently regular, to allow for reliable information extraction, and even the automatic generation of Information Extraction templates and algorithms. Chapter 4 also enquires: what kinds of novel applications may be enabled by machine-processable representations of film content-based on the four events identified in Chapter 3? To answer these questions, an IE system is developed. Algorithms, based solely on the four templates identified in Chapter 3 are designed and implemented using a Natural Language Processing text analysis system. To evaluate the system, a Gold Standard data set is gathered from an hour of film clips from five Hollywood films. The Gold Standard data set is used to evaluate the system by examining its performance, through precision and recall, against the segments of AD and SC scrips for the five film clips. The system is used to populate a database of information for the four event types for 193 AD and SC film scripts. Based on the results of the database, potential novel applications are outlined for the querying, retrieval, browsing/navigation, display and manipulation of film content.

Section 4.1 defines Information Extraction (IE) and how it is used in the work. In 4.2 the design and implementation of our IE system, as well as the tools used to develop it, is detailed. The heuristics used are outlined and the implementation and its issues are specified. Section 4.3 describes the evaluation of the system and includes the gathering of the Gold Standard data set. Precision and recall statistics are defined and used to measure the system's performance. Section 4.4 interprets statistics of the data gathered for the 193 film scripts and explores potential novel applications that could be developed from this information. A discussion follows, § 4.5, on how this work allows for the analysis of aspects of film content that have not been analysed automatically before and the progress we have made in crossing the semantic gap. Whilst the focus here is on automatically extracting film content information from the AD and SC film scripts, and generating a database of film content information, it is important to note that results here also serve to validate our findings of four events and of an LSP, in particular.

4.1 Information Extraction

Information extraction (IE) is a technology dedicated to the extraction of structured information from texts to fill pre-defined templates [105] and is a type of information retrieval whose goal is to automatically extract structured or semi-structured information from unstructured machine-readable documents [145]. For our purposes this definition is most relevant as it refers to 'filling pre-defined templates'. These predefined *IE templates* refer to a set of entities or structures that are considered *slots* for relevant information the user wishes to automatically extract from the relevant texts. Cowie and Wilks [24] state:

> "IE is the name given to any process which selectively structures and combines data which is found, explicitly stated or implied, in one or more texts. The final output of the extraction process … can be transformed so as to populate some kind of database." [ibid pg. 1]

Information Extraction Systems are automated systems that can extract pertinent, structured information from large volumes of general texts. These large volumes of general texts can refer to corpora. Unlike an information retrieval system, rather than indicating which documents need to be read by a user, an IE system extracts pieces of information that are salient to the user's needs. It can be said that a typical application of IE is to scan a set of documents written in a natural language and populate a database with the information extracted. Typical subtasks of IE are: *Named Entity Recognition:* Extracts things such as entity names (for people and organisations), place names, temporal expressions, and certain numerical expressions; and *Coreference:* identification chains of noun phrases that refer to the same object and *Terminology extraction*: finding the relevant terms for a given corpus.

Our work seeks to populate a database of film content information from film script texts, and to structure and combine the film content data. Hence, an information extraction system is employed to do that.

Hobbs described a generic Information Extraction system as a "cascade of transducers or modules that at each step add structure and often lose information, hopefully irrelevant by applying rules that acquired manually and/or automatically" [Hobbs 1993, pg. 87] as quoted in [34]. FASTUS[24] [46] is an IE system for extracting from free text in English. FASTUS's key idea is to separate processing into several stages or "cascade". FASTUS uses five levels of processing: Stage 1 names and other fixed form expressions are recognised. Stage 2, noun and verb groups,

[24] Finite State Automaton Text Understanding System

prepositions and some other tokens are recognised. Stage 3, certain complex noun and verb groups are constructed. Stage 4, patterns for events of interest are identified and corresponding "event structures" are built. In Stage 5, distinct event structures that describe the same event are identified and merged. These merged event structures are used in generating database entries. In this work our method (chapter 3) and system specification is comparable to FASTUS's processing levels and our collocation FSA can be considered "event structures", merged into four film templates.

FASTUS was very effective when tested on the MUC-4 dataset and was only outperformed by one other system. The Message Understanding Conference (MUC) provides a formidable framework for the development of research in the area of IE systems [70]. The MUC is organised by NIST [146] which is a competition-based conference and has provided Gold Standard datasets (in 5 domains) for the evaluation of Information Extraction techniques and systems since 1987.

IE Evaluation metrics have evolved with every MUC [34]. The starting point for the development of these metrics was the Information Retrieval metrics precision and recall. For IE, given a system response and an answer key prepared by a human, the system's *precision* was defined as the number of slots it filled correctly, divided by the number of fills it attempted. *Recall* was defined as the number of slots it filled correctly, divided by the number of possible correct fills, taken from the human-prepared key. All slots were given the same weight [56]. A ***Gold Standard*** set of data, for a particular domain, is an 'accepted' set of data in that domain, against which IE system outputs data can be compared and evaluated. An example is the MUC gold data sets. We develop a Gold Standard data set for the four film events.

Thus, this work seeks to develop an IE system from the templates in Section 3.4.1 that will allow us to populate a database about film content. There are comparisons to be made with the FASTUS system where "Patterns for events of interest are identified and corresponding "event structures" are built... distinct event structures that describe the same event are identified and merged. These merged event structures are used in generating database entries" [[46] pg 1.], which is what our research in Chapter 3 does.

The rest of this chapter examines how an IE system was developed for the 4 film events described in 3.4.1 allowing the automatic population of a film content database.

4.2 Design and Implementation

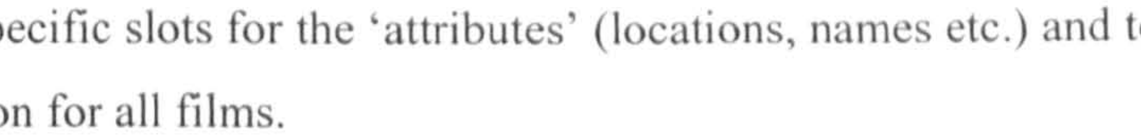
A system was developed to extract information about the four event types, detailed in Figure 34 (Section 3.4.1). The design, implementation, evaluation and testing of the system including the discussion of programming and design issues and software used, are described in this section.

4.2.1 Conceptual Design: Outline of Heuristics

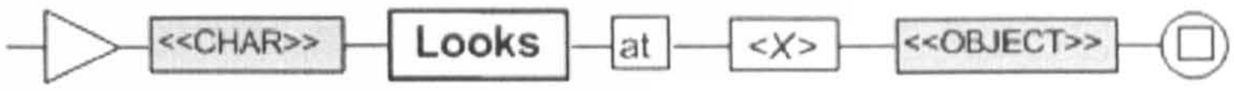
Having interpreted collocation data into event types, a set of heuristics, to extract information about the Focus of Attention (FOA), Change of Location (COL), Non-Verbal Communication (NVC) and Scene Change (ScCh) events in film scripts, was developed. The heuristics were based on the four templates (Section 3.4.1) and the collocation FSA/LGs (Figure 29) in chapter 3.

The heuristics' pseudo-code was developed and outlined to locate strings of text that contain the collocation phrases and then to locate the surrounding words and sort them into the categories. The categories of *words* were: OBJECT, CHARACTER, LOCATION and ACTION, which were found, on manual analysis, to be the most frequently collocating *types* of words. In terms of grammar the categories were: pronouns, proper nouns, punctuation and articles, as deemed necessary. The heuristics also serve to resolve issues such as pronounification or named entity resolution, i.e. assigning correct character names to pronouns that reference them (he, she, him, her, them). The heuristics allow for information about film to be mapped to a specific event (FOA, NVC, COL and ScCh), with specific slots for the 'attributes' (locations, names etc.) and to populate a database of such information for all films.

Using the diagrams from Appendix C as a guide it was possible to systematically 'read off' heuristics from the collocation FSA in order to develop heuristics for the extraction of *elements* or attributes of the events such as character, object and location names, times of day, scene type, actions involved, etc. This was done by intuitively, manually assigning the collocation FSA diagrams in Appendix C to one of the four events (FOA, COL, NVC and ScCh). For example, assigning "looks", "turns" and "takes" FSA diagrams to the FOA event, (see Appendix C for FSA event assignments).

Once the collocation FSAs were assigned to events, heuristic pseudo-code was systematically developed based on which collocation phrases were associated with which categories of information. For example the phrase: 

is providing information about a 'character focussing their attention on another character'. And

similarly 

is providing information about a character focussing their attention on an object. As a result, they were associated to the FOA event. Each FSA in each event was developed as a heuristic separately. A set of rules for each cluster of phrases in the FSA diagrams was developed. Once these were in place they were grouped with respect to what information they could extract. What follows are simplified versions of the heuristics for the FOA, COL, NVC and ScCh events. Full heuristics can be found in Appendix E and on **CD**: Additional Thesis Material.

Focus_of_Attention Event (FOA)

/**Art* = {the, a, an}*/

LOCATE strings [‘**looks***[25] (around, back, down, over, round, up) at’ || ‘**looks*** (into, out, down, through, to) the’

|| ‘**looks*** (into, over) his’

||‘**turns*** back (to, toward, towards)’ || ‘**turns*** (from, off)’

||‘**takes*** (*Art*)’]

IF string = [‘**takes*** (*Art*)’] then `Focus_Of_Attention_Type` = PASSIVE

ELSE `Focus_Of_Attention_Type` = ACTIVE

SEARCH left in text for first reference of a *character* and fill the `Character_Focussing` slot.

SEARCH right in text for first reference to a *character* or an *object* and fill the

`Character_Focussed_On` OR `Object_Focussed_On` Slot.

Change_of_Location Event (COL)

/*<E> = null string or empty space*/

LOCATE strings [‘(turns and, <E>) **walks*** (away, to)’ || ‘(turns and, <E>) **walks*** (across, along, around, down, into, out, through, over) (*Art*)’ || ‘(turns and, <E>) **walks*** (to, towards, toward) ’

|| ‘opens *Art* **door***’ || ‘(goes, moves) to (his, her) **door***’ || ‘*Art* **door*** behind (him, her)’

|| ‘across *Art* (control, living, throne, <E>) **room***’ || ‘out of *Art* (control, living, throne, <E>) **room***’ ||

|| ‘ (and, he , she, <E>) (turns, backs, walks, walk) **away**’|| ‘back **towards** *Art*’]

IF string = [‘(turns and, <E>) **walks*** (across, along, around, away, down, through, over, to)’ || ‘across *Art* (control, living, throne, <E>) **room***’] then `State_of_Character` = WITHIN/ON

IF string = [‘(turns and, <E>) **walks*** (into)’ || ‘opens *Art* **door***’] then `State_of_Character` = ENTERING

ELSE `State_of_Character` = LEAVING

SEARCH left in text for first reference of a *character* and fill the `Character_Changing_Location` slot.

FOR string [‘*Art* **door*** behind (him, her)’] **SEARCH** right in text for first reference to a *character* or an *object* and fill the `Character_Changing_Location` Slot.

---

[25] Where looks* = all the variants and synonyms of ‘looks’ *accepted* in section 3.3.4, e.g. {stare, glance, gaze, peer, glare}. This notation has been applied to all candidate nucleate open class words.

Non-verbal Communication Event (NVC)

LOCATE strings ['(he, she) (shakes, nods) (his, her) **head***' || '(he, she) (opens, closes) (his, her) **eyes***]

IF string = ['(he, she) (shakes, nods) (his, her) **head***'] then `Body_Part_Involved` = 'head' AND `Action_Involved` = 'shakes'

IF string = ['(he, she) opens (his, her) **eyes**'] then `Body_Part_Involved` = 'eyes' AND `Action_Involved` = 'opens'

ELSE IF string = ['(he, she) closes (his, her) **eyes**'] then `Body_Part_Involved` = 'eyes' AND `Action_Involved` = 'closes'

SEARCH left in text for first reference of a *character* and fill the `Character_Communicating` slot.

Scene Change Event ScCh

LOCATE strings ['(-, --, <E>) **day***' || '(-, --, <E>) **night***']

SEARCH left in text for first string ['INT' || 'Interior'] then `INT/EXT` = 'Interior'

SEARCH left in text for first string ['EXT' || 'Exterior'] then `INT/EXT` = 'Exterior'

IF string = ['(-, --, <E>) **day***'] then `Time_of_Day` = 'day'

ELSE `Time_of_Day` = 'night'

SEARCH for string between `INT/EXT` AND `Time_of_Day`.

IF found string = `Location` slot

Figure 35 Simplified heuristics for extraction of information for the four film events: FOA, COL, NVC and ScCh from AD and SC film scripts.

Frequently occurring idiosyncrasies were incorporated into the heuristics, such as *articles* preceding objects and characters: {*the* man, *a* cat, *an* umbrella}, *possessive pronouns* preceding objects and characters {*his* hat, *her* purse}, *pronouns* preceding central phrases {*he* walks into the bar, *she* turns to her, Tom looks at *him*} and *punctuation* ending phrases {Rich grabs her arm,}. The heuristics were tested, semi-automatically, and it was possible to extract information on the four events as well as what time they occurred. The test was conducted on the film the "English Patient" using the search function of Unitex where the relevant concordances found was placed in EXCEL and macros were written with respect to the heuristics to extract the relevant information.

Having put the heuristics into practice, manually for the most part, the heuristics' pseudo-code was ready to be implemented into a system that would allow the audio description and screenplay scripts to be analysed in terms of gathering information concerning the events FOA, COL, NVC and ScCh and at what time these events occurred in the film.

4.2.2 Implementation Design

In the implementation of the heuristics certain formatting issues, specific to AD or SC film script issues arose as well as generic Natural Language Processing NLP issues such as pronoun resolution. This section presents some of these issues and their solutions. See Appendix E and **CD**: Additional Thesis Material, for all four complete heuristics.

4.2.2.1 Issues with Time-codes

Each of the heuristics required the extraction of a time-code or some sort of time reference for estimating, as accurately as possible, the time of an event with respect to the overall time length of the film. The AD scripts all had time code at the beginning of each line or on the line before and this fact was utilised to extract the time at which each event occurred. The time code format differed for certain companies' AD scripts as can be seen in Figure 36. When an event is located the system traverses to the left of the line until a specific format of numbers is located depending on the format of the time-code.

ITFC: <OverallInTime>10:20:04.06</OverallInTime>
<OverallOutTime>10:20:06.18</OverallOutTime><ScriptText>
<normal>Johnny enters the bar first. [Indian Fighter [175]]

BBC: In Cue Time(HH MM SS Fr): 10:02:51:05
Out Cue Time(HH MM SS Fr): 10:02:53:08
Luis looks at her with warm expression. [Daytrippers [165]]

RNIB: 01 03 27 She gently removes a bottle from near the top of the rack, and inspects the label. [The 6th sense [168]]

Figure 36 Excerpts from BBC, ITFC, RNIB AD scripts showing the formats for time-coding video.

In the case of screenplays the scenes and acts of a film were numbered but there was no line to line time reference or reference to when an event or shot occurred. Consequently, for the Screenplays *line numbers* were imposed onto the text files to allow us to situate the time at which an event occurred. The total number of lines was outputted automatically for each screenplay.

1445. INT. POLICE BULL PEN - MINUTES LATER
1446. Tatum has joined Sidney. The sheriff's door opens and Billy is led out
1447. by a coupla UNIFORMS. Burke and Dewey appear in the door watching
1448. Tatum comfort Sidney.

1449. OUT OF EAR SHOT [Scream [159]]

Figure 37 Excerpt from the screenplay for the film Scream [159] showing how line numbers were imposed to estimate time.

Thus a time reference for a screenplay was estimated by:

$$\frac{LineNumber}{TotalNumberofLines} \times FilmLength \quad [\xi]$$

4.2.2.2 Locating Unnamed Characters and Objects

This issue arose when nouns or proper nouns referred to characters or objects and were preceded by articles or pronouns. For instance, when faced with phrases such as '*the* man walked to *his* car', then 'the man' is the character and 'his car' is the object. Code was written to identify these circumstances and place 'the man' and 'his car' into the correct slots for respective events.

SEARCH to the left of the 'nucleate phrase cue'

/*NB nucleate phrase cue refers to the cue of event e.g. 'looks at' for FOA*/

IF string = [(the, an, a, his, her)] then record 'the, an, a, his, her' AND the following character string

SEARCH to the right of the 'nucleate phrase cue'

IF string = [(the, an, a, his, her)] then record 'the, an, a, his, her' AND the following character string

4.2.2.3 Resolving References to Characters

Throughout the collocation FSAs there are references to characters through pronouns, {he, she, his, her, him...}. An issue of resolving which characters the pronouns refer to (pronounification) arose.

In the case of the pronouns 'he', 'him', 'she' and 'her' ('her' only with a full stop, punctuation or 'and' after it, i.e. *not* possessive 'her') an undergraduate student working from our work for his final year project, Darren Johnson, was able to solve the issue of mapping the correct characters to an event by creating an array of all the capitalised words, bar any from a stoplist, whose frequency>5 and then creating an array of all events and *character names*. The 'source' (character names or pronouns doing the focussing) and 'destinations' (character names, nouns or pronouns being focussed on) of the events were then processed by matching capitalised words in the proximity of the event or, if a pronoun, accessing the array to find the last mentioned character; dubbed a 'proximity search'.

SEARCH for all capitalised words in the text, frequency > 5 /*5 found as a threshold by trial and error*/

COMPARE against a stop list of excluded capitalised closed class and 'film' words *EX* e.g. {TV, At, SCRIPT, FORMAT, INT, No, When, Day, Cut etc.}

IF capitalised word = any member of *EX* Ignore

ELSE accept capitalised word as Character name AND record line number or time-code.

The system also resolves instances such as 'looks at his/her reflection/self' for the destination of the FOA event, by accessing the array and finding the last mentioned character. In the case of the screenplays, due to a memory problem in the text analyser used, only the first 1000 instances of the capitalised words with instances>5, not in the *EX* list, are taken as the character name list for a film.

In the case of *possessive* pronoun resolution, e.g. {his, her, their}… followed by an object, the proximity search was used again to find the possible character referred to. However it was usually a character preceding the 'nucleate phrase cue'.

4.2.2.4 Miscellaneous Implementation Issues

Many issues arose which were specific to the text analyser used, VisualText [147]. Some issues did arise with the formatting of screenplays however. The representative corpus called for different types of screenplays, e.g. different drafts and versions from different years (films from the nineteen forties to the present). This meant that some pre-processing needed to be done and that line numbers had to be added to all screenplays to allow the time references to be calculated. Different formats for the ScCh cues had to be accounted for which was done manually by examining the screenplays.

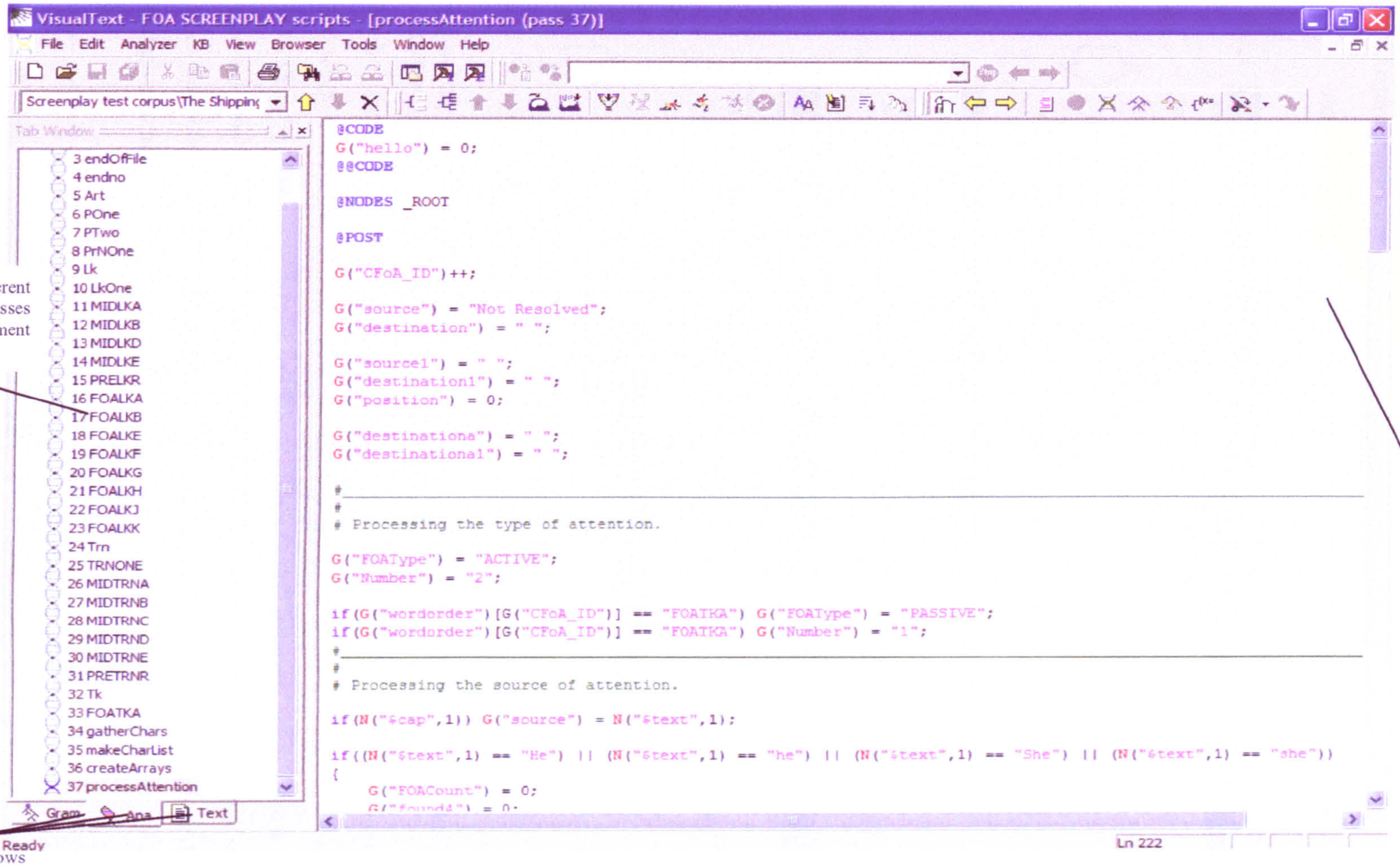
Because we did not want to analyse *dialogue* in screenplays as it did not provide descriptive information, we had to find ways of automatically excluding it. Text formatting was employed to try and exclude the dialogue such as taking advantage of the screenplays' formatting by excluding indented text after a character's name or ignoring words between quotes for the film transcript scripts. However, this only was possible for approximately 40% of the screenplays. The AD scripts contained minimal dialogue. We also had to pre-process the ITFC scripts as they contained many XML tags which VisualText could not process.

4.2.3 Implementation with VisualText

The heuristics were implemented using a text analysis system VisualText [147] developed by Text Analysis International Inc. It is an Integrated Development Environment for building deep text analysis applications. VisualText features NLP++, a C++ -like programming language for elaborating grammars, patterns, heuristics, and knowledge but with specializations for Natural Language Processing. It is possible to use the VisualText® Integrated Development Environment (IDE) to automatically populate databases with the critical content buried in textual documents.

The heuristics were inputted into VisualText's NLP++ as a set of 'passes' through sets of rules in each 'pass'. A 'pass' works by moving iteratively through the text document and firing its rules when identifying the unit it is searching for. For his final year project "Video Browsing and Retrieval Using Audio Description" [52], undergraduate 3rd year CIT student Darren Johnson was asked to implement the 'Focus of Attention' heuristics into VisualText. The aim of his work was to allow the extraction of FOA events, automatically, from audio description Scripts and to analyse the results to explore whether the FOA events exhibited patterns in films or film genres. Johnson was able to follow the FOA heuristics outlined above and in Appendix E and successfully develop a VisualText NLP++ program to extract the FOA events from BBC and ITFC audio description scripts[26] and output the results in a database. He then went on to try and improve the system using personal judgement for how the heuristic rules should be implemented. An example of the visual output of VisualText for the FOA heuristics' implemented rule set can be seen in Figure 38.

[26] As provided by the TIWO project [96]



(a) Shows different text analysis passes through the document to be analysed

(b) Code window. Here we see the code of pass 39 'process Attention'

Tab window: shows the grammatical tab, analyzer tab and the text tab.

Figure 38 Screenshot of the FOA event information extraction passes developed in VisualText's NLP++ IDE. On the left window (a) are the NLP++ passes through the analysed text and tabs for the viewing the grammatical analyzer, the analyzer (current view) and the texts under analysis. The middle window (b) shows the code of the specific pass selected, in this case the 'processAttention' pass.

Each pass outlined in the "Analyzer" tab follows a rule from the FOA heuristics (apart from the 'gatherChars', 'makeCharList' 'createArrays' and 'processAttention' which deal with creating arrays of information and processing that information.) A typical analyzer pass FOALKA is demonstrated in Figure 39.

FOA_LK_A = PRE_LK_R + MID_LK_A **//*CHAR → OBJ, LOC*//**

Figure 39 Demonstrating the analyser FOALKA (pass 18) FOA heuristic of the FOA information extraction system and what other analyser passes it relies on.

Each complete pass works as follows: A script is selected to be analysed, the analyser traverses the script and locates the character strings being searched for (FOALK{A-K}, FOATRN and FOATKS)[27], all character names (all words with a capital letter except those in a stop list), and all relevant time codes (or line numbers). The information is then put into an array, sorted in terms of *when* an event occurs and *who is involved* then information is 'posted' to a text file (excel file) and given a unique ID for every pass.

[27] See Appendix E heuristics and **CD** for details

This method was adopted for all events, (NVC, ScCh and COL), but developed for the specific information to be captured for relevant event. NVC events sought out who was doing the non-verbal communication and which body part was involved with it associated motion. ScCh events searched for where the new scene location was, what time of day and whether it was an internal or external scene. Change of location events sought out whether the characters' were going into, out of a location or whether they were in, within or on a certain location and where possible the initial and final locations of a character. Excerpts from tables with information for the events that was extracted using VisualText can be seen in Table 18.

Table 18 A set of 4 tables depicting typical results of information extracted for the 4 events FOA, ScCh, NVC and COL produced automatically using VisualText's NPL++ IDE from 4 different film scripts "Daytrippers"(AD)[165], "American Beauty" (AD), "The English Patient" (SC) and "Shrek"(AD)[155].

CFOA ID	Focus Type	Character Focussing	Character Focussed On	Object Focussed On	Occurrence Time	Text String
FOAAD5	ACTIVE	Jim		(their) car	00:15:22:24	00:15:22:24 looking at
FOAAD6	ACTIVE	Carl (He)	Jim		00:23:32:07	00:23:32:07 turns to see

ScCh ID	INT\EXT	Location	Time of Day	Line No.	% Film Time	Text String
ScChAD1	Interior	FITTS HOUSE - RICKY'S BEDROOM	Night	2	0.07	INT. FITTS HOUSE - RICKY'S BEDROOM- NIGHT
ScChAD10	Exterior	SALE HOUSE	Day	322	11.25	EXT. SALE HOUSE- DAY

NVC ID	Body Part	Character Communicating	Action Involved	Occurrence Time	Text String
NVCAD9	Head	Madox	shakes	00:25:09:16	00:25:09:16 shakes his head
NVCAD10	Eyes	Caravaggio	closes	00:25:34:20	00:25:34:20 closes his eyes

COL ID	Character State	Character Changing Location	Occurrence Time	Text String
COLAD2	ENTERING	Donkey	00:06:26:17	00:06:26:17 runs into
COLAD3	LEAVING	Donkey	00:10:51:06	00:10:51:06 walks out
COLAD4	WITHIN\ON	Donkey	00:11:55:04	00:11:55:04 walking over to

Having established a method to automatically extract information about events from texts that describe films we then set about creating a database of film information.

4.3 Evaluation and Collecting a Gold Standard Data Set

Having collected a data set of certain film elements and events from film scripts, the question arose of how to evaluate the collected film data set. We wished to evaluate in terms of precision and recall, with respect to the events seen in the films and, by extension, the film scripts. It was decided that a 'Gold Standard' of some of the events would be collected, i.e. humans would record when events were occurring in a set of certain film clips. All evaluation and Gold Standard data can be seen on the **CD**: Additional Thesis Material.

4.3.1 Generating a Gold Standard Data Set

A ***Gold Standard*** set of data, for a particular domain, is an 'accepted' set of data in that domain, against which IE system outputs data can be compared and evaluated and so we set about to get our own Gold Standard data set.

Five, 12 minute film clips were chosen from films that we owned and that we had *at least one of* the audio description or screenplay scripts. This came to about 60 minutes (1 hour and 50 seconds) of video clips for the films "The English Patient" [170], "The World is Not Enough" [156], "Oceans 11" [169], "High Fidelity" [169] and "Out of Sight" [164]. The idea was to allow 12 people with varying backgrounds (including art and film students, book publishers, web designers, computing students, economics students and PhD candidates) to watch the five clips, pausing when they saw an event (a character focussing their attention on someone/something, a character communicating non-verbally, a character changing location or when a scene changes) and recording the time at which this event was occurring, as well as any other relevant information, in pre-prepared tables. Each person was given a set of characters' images and names that would appear in each of the film clips and were asked to look out for only *two* of the four events, because early trials showed four events were too much for people to watch for. The evaluators were asked to provide their names, their occupations, what level of film expertise they possessed and were given the choice of turning the volume of the clips OFF or ON. The full set of instructions can be seen in Appendix D. In total twelve people were given the evaluation and split into two groups A and B consisting of seven and five people respectively. Group A was given the character's focusing attention and non-verbal communication events and Group B was given the characters changing location and scenes changing events. Table 19 shows an excerpt of a table for the event 'character is focussing attention' as filled in by an evaluator for the film clip "Out of Sight" [164]. The tables given to the evaluators were based on the tables seen in Table 18.

Table 19 An example of some 'Character focussing on something or someone' event information gathered from the 13m 04s film clip of the film "Out of Sight" [170] by an evaluator.

Character Focusing	Character Focused on	Object Focused on	Time
Jack	Prison guard		00:00:10
Karen		on phone	00:00:47
Karen		Car	00:00:54
Buddy	Karen		00:01:13
Jack & Prison guard		Outside	00:01:25
Jack		Vase	00:01:32
Karen		Ground	00:01:53

The evaluators' results, for detecting events in the five clips for both groups A and B, were collected and entered into spreadsheets. The spreadsheets were then searched for events that were found by at least three evaluators. These events were judged to be 'the same' if both the description of the event was similar and the time the event occurred was similar. The events agreed upon by the evaluators were then 'distilled' into one event. These 'distilled' events had one description which was a combination of the evaluators' descriptions and one time code, which was calculated by taking the median of the time codes identified by the evaluators. This method was repeated for events at least four evaluators agreed upon (see Table 20).

Table 20 An event from "The World is not Enough" agreed on by six evaluators is distilled into a *single* event trying to capture as much information as possible from the six separate events found by evaluators.

Character Focussing	Character Focussed on	Object Focussed on	Film Time
James	Jones	Chip (locater card)	01:05:58
Dr. Jones		Chip	01:05:55
Dr. Jones		Locator Chip	01:05:55
Dr Jones		Circuit Board	01:05:55
Dr Jones	James Bond	Locater chip	01:05:55
Jones	Bond	Card	01:05:56

Character Focussing	Character Focussed on	Object Focussed on	Film Time
Dr Jones	James Bond	Locator Chip (card)	01:05:55

Once all the events were distilled, each event in each film was numbered. Now, two sets of 'Gold Standard' events for the five film clips existed, the events three evaluators (3+) and events four evaluators (4+) agreed upon. Having gathered the Gold Standard events for the film clips it was now possible to compare them to the events automatically extracted from the segments of the audio description scripts and screenplays for respective film clips. See **CD** for full results.

4.3.1.2 Automatically Extracting Events from the Film Script Segments of the Evaluation Video Clips.

The sections of the audio description and screenplay scripts, that matched the selected five film clips, were located (through watching the film and manually aligning the script with the film clip). Then, the script segments for the five films were placed into the VisualText analysers for the four events. The results were outputted to separate spreadsheets and then collated.

Due to the diverse formatting nature of the film scripts[28] the collated results had to be processed to match the time code of the films. All film clips were 'time'-aligned with the audio description time codes and the line numbers of the screenplays. There were also issues involved where the script did not correspond exactly to the film script. For Example, the Screenplay for "The English Patient" was the 'final draft' version and contained an *extra* scene which was later moved. The video clip did not contain this scene. Any results found in such scenes were ignored. This now allowed a direct comparison to the Gold Standard events.

The automatically extracted events were checked for *false positives*, i.e. when a VisualText analyser result is found but that result is not an event. Quite a few false positives were identified for the screenplays where the phrases searched for by the analyser occurred in dialogue such as in a question "What are you looking at Jim?" and statements such as "Tom has been running around all day". Other false positives included phrases in every day language such as "He takes a deep breath/ a few steps". The false positives were labelled as such and removed. Results found in dialogue were ignored.

After false positives were identified formatting and time code as well as formatting issue resolved, a set of events for each film clip, extracted from the relevant film script segments, was produced, ready to be compared with the Gold Standard results. Results for the film script segments for "The English Patient" video clip can be seen in Table 21. It must be noted that event 'FOAAD9' was 'IGNORED' due to the phrase "He takes a few paces", which was deemed a false positive and thus removed.

---

[28] Audio description scripts come in 3 formats due to the 3 companies associated with the TIWO project that provided scripts: BBC, ITFC and RNIB formats

Table 21 Shows results that the VisualText FOA analyser retrieved for the screenplay, RNIB and ITFC AD script segments for "The English Patient" film clip, see **CD**: Additional Thesis Material for full results.

Event ID	Focus Type	Character Focusing	Character Focused on	Object Focused on	Event Time	Text String
EPRNF1	ACTIVE	Kip		(tangled) wires	01:52:46	02:52:46 stares at
EPRNF2	PASSIVE	Kip		(the) cutters	01:52:55	02:52:55 takes the
EPRNF3	PASSIVE	Caravaggio		(an) open	01:54:53	02:54:53 Holding an
EPRNF4	PASSIVE	Hana		(the) stretcher	01:54:53	02:54:53 carry the
FOAAD1	PASSIVE	Hardy		(a) rope	01:52:42	00:15:05:02 grabs a
FOAAD2	ACTIVE	Kip		(the) bomb	01:52:51	00:15:14:01 stares at the
FOAAD3	PASSIVE	Kip		(the) pliers	01:53:33	00:15:56:17 has the
FOAAD4	ACTIVE	Kip (He)	Hardy		01:53:33	00:15:56:17 looks up at
FOAAD5	ACTIVE	Hana		(her) 0	01:54:54	00:17:17:08 looks at her
FOAAD6	PASSIVE	Hana		(an) umbrella	01:55:00	00:17:23:12 holding an
FOAAD7	PASSIVE	Caravaggio		(a) gun	01:58:08	00:20:31:13 takes a
FOAAD8	ACTIVE	Madox	Almasy		02:01:03	00:23:26:07 looks at
FOAAD9	PASSIVE	Almasy		(a) few	IGNORED	00:24:41:11 takes a
FOAAD10	ACTIVE	Almasy	Almasy		02:02:18	00:24:41:11 turns back to
FOASC1	ACTIVE	Hardy		(his) watch	01:52:48	4012 looks at his
FOASC2	ACTIVE	Kip	Hana		01:54:26	4142 looks at
FOASC3	ACTIVE	Kip	Hana		01:55:39	4154 looks at
FOASC4	PASSIVE	Madox	HANDFUL	sand	02:01:22	4230 takes A

4.3.1.3 Comparing Gold Standard to Automatically Generated Events in terms of Precision and Recall

In this analysis it was deemed essential to employ human judgment to evaluate the accuracy or 'hit rate' of our system due to the nature of discrepancy concerning what an event and/or film element is (see Chatman discussion, Section 1.1 & 1.2). Thus evaluation measures were engaged that would allow people to judge, in five separate instances of film, what events were present. These events were guided by our four event categories selected from collocation phrases (§3.4) and were seen as 'Gold Standard' events. The relevant film script segments of the film clips were passed through the VisualText analysers for all four events and results were checked for false positives and aligned to the correct time codes of the film.

To compare the two sets of results the Gold Standard results were used as a base and the VisualText analyser results were compared to them. The VisualText results' time codes were manually matched to ±5 seconds. Then the description and other fields of the event were compared; if the description matched this was classed as an 'Identical Event'. Within the VisualText or 'System Output' events, a 'False Event' was identified as an event the VisualText analyser extracted that whose attributes were *not* identical to the respective Gold Standard event

with a similar time code. That is, for a Gold Standard event, "Tom looks at the cat" the extracted FOA event is "He looks at the cat", for the same time code, and the VisualText analyser falsely identifies "He" to be a proper noun other than 'Tom' then this is considered a False Event. All 'Identical' events were highlighted and the respective VisualText result 'Event ID' was placed next to the Gold Standard event (see Table 22). The comparison was conducted twice with Gold Standards of people who agreed on 3+ events and people who agreed on 4+ events. This can be seen in detail on accompanying **CD**: Additional Thesis Material.

Table 22 Shows an excerpt of a table that compares Gold Standard events to the VisualText System Output Events for the film clip "Oceans 11"[169]. Matching or 'Identical' events are highlighted.

	Character Focusing	Character Focused on	Object Focused on	Event #	Film Time	Sys Event ID
	Linus & Rusty	Tess Ocean		1	00:42:45	
	Linus & Rusty	Tess		2	00:43:00	
	Daniel	Rusty		3	00:43:32	EPRNF1
	Daniel		Chips in hand	4	00:43:32	
	Tess		Painting	5	00:44:42	FOAAD1
Matching	Benedict		Painting	6	00:45:05	FOAAD2
Events	Tess	Benedict		7	00:45:14	
	Tess		CCTV Camera	8	00:45:23	FOAAD3
	Saul		Sweet/Mint/Gum	9	00:45:44	FOAAD4

4.3.2 Precision and Recall

After comparing the Gold Standard to the VisualText results the accuracy of the system compared to humans' was calculated using statistics for precision and recall.

$$precsion = \frac{EventsIdentifiedByHumans \,\&\, TheSystem}{TotalSystemOutput - (FalseEvents + DoubleEvents)} \times 100 \qquad [\pi]$$

$$recall = \frac{EventsIdentifiedByHumans \,\&\, TheSystem}{GoldStndardEvents} \times 100 \qquad [\varsigma]$$

Table 23 Shows precision and recall results and the number of respective events for each film clip, for both the events, at least three and at least four evaluators agreed upon for the Focus of Attention Events. The precision and recall statistics are presented both as absolute numbers and percentages.

Focus Of Attention 3+	English Patient	World Is Not Enough	Oceans 11	High Fidelity	Out Of Sight
Identical Compared Events	10	3	7	4	14
Gold Standard Events	41	35	21	20	16
System output Events	17	7	10	6	30
False System Results	7	4	3	2	16
PRECISION	**10/17**	**3/7**	**7/10**	**4/6**	**14/30**
RECALL	**10/41**	**3/35**	**7/21**	**4/20**	**14/16**
PRECISION %	**58.82**	**42.86**	**70.00**	**66.67**	**46.67**
RECALL %	**24.39**	**8.57**	**33.33**	**20.00**	**87.50**
Focus Of Attention 4+	**English Patient**	**World Is Not Enough**	**Oceans 11**	**High Fidelity**	**Out Of Sight**
Identical Compared Events	11	3	6	3	10
Gold Standard Events	34	29	17	14	11
System output Events	17	7	10	6	30
False System Results	6	4	0	3	20
PRECISION	**11/17**	**3/7**	**6/10**	**3/6**	**10/30**
RECALL	**11/34**	**3/29**	**6/17**	**3/14**	**10/11**
PRECISION %	**64.71**	**42.86**	**60.00**	**50.00**	**33.33**
RECALL %	**32.35**	**10.34**	**35.29**	**21.43**	**90.91**

The largest discrepancy in Table 23 is seen in the results for "Out of Sight" where for audio description and screenplay segments of script the VisualText analyser extracted 30 events as opposed to the average 10 events for the other four film clips. Also the Gold Standard events were the least ([3+]=16, [4+]=11) for this clip. This led to high recall and low precision values for this clip unlike the *low* recall and higher precision for the other clips. The reason for this was mainly due to the type of scenes involved in the 13 m 03 s of the "Out of Sight" film clip. The scenes were mostly dialogue and long and *descriptive* scenes which made for longer segments of descriptive text in the screenplay segment. These descriptive scenes may be rich in focus of attention phrases which would explain the high number of automatically extracted system output events. In the case of the Gold Standard results being lower than the other films, the scenes in the film clip were long and full of dialogue, in other words not that much happened *action* wise. The clip consisted mostly of dialogue between the two main characters in a car boot. Thus, the evaluators may have not picked up on the subtle focuses of attention on items and people as readily or maybe not much happened to be picked up on.

What these results indicate is that, possibly, the genre of a film will give a different weighting on the different types of event. For instance a romantic film may have more FOA events than a

comedy. This is not clear however as there are only five films in our results. The next sets of results are for the other three events COL, ScCh and NVC.

Table 24 Sets of tables depicting the precision and recall value for the events at least three and at least four people agreed upon for the five film clips and scripts for the events 'Change of Location', 'Scene Change' and 'Non- Verbal Communication'. NB For the NVC event there were only Precision and Recall results available for "The English Patient" [164].

Change of Location 3+	English Patient	World Is Not Enough	Oceans 11	High Fidelity	Out Of Sight
PRECISION	10/22	6/14	6/13	2/7	4/13
RECALL	10/13	6/7	6/12	2/7	4/7
PRECISION %	45.45	42.86	46.15	28.57	30.77
RECALL %	76.92	85.71	50.00	28.57	57.14

Change of Location 4+	English Patient	World Is Not Enough	Oceans 11	High Fidelity	Out Of Sight
PRECISION	5/22	4/14	2/13	0/7	4/13
RECALL	5/6	4/5	2/12	0/5	4/5
PRECISION %	22.73	28.57	15.38	0.00	30.77
RECALL %	83.33	80.00	16.67	0.00	80.00

Scene Change 3+	English Patient	World Is Not Enough	Oceans 11	High Fidelity	Out Of Sight
PRECISION	9/10	5/7	8/11	9/12	5/8
RECALL	9/23	5/9	8/12	9/22	5/12
PRECISION %	90.00	71.43	72.73	75.00	62.50
RECALL %	39.13	55.56	66.67	40.91	41.67

Scene Change 4+	English Patient	World Is Not Enough	Oceans 11	High Fidelity	Out Of Sight
PRECISION	9/10	5/7	8/11	9/12	3/8
RECALL	9/19	5/7	8/12	9/22	3/7
PRECISION %	90.00	71.43	72.73	75.00	37.50
RECALL %	47.37	71.43	66.67	40.91	42.86

Non-verbal Communication 3+	English Patient
PRECISION	3/4
RECALL	3/5
PRECISION %	75.00
RECALL %	60.00

Non-verbal Communication 4+	English Patient
PRECISION	1/4
RECALL	1/3
PRECISION %	25.00
RECALL %	33.33

Overall, as can be seen from Table 23 and Table 24, precision was much higher than recall for the 'Focus of Attention' and 'Scene Change' events, but for the 'Change of Location' events recall was higher than precision. Not much could be ascertained for the 'Non-Verbal Communication' event as there was such a low yield of results due to the fact the phrases "opens/closes his/her

eyes" and "nods/shakes his/her head" were not common in the scripts. Table 25 shows the average precision and recall values for the events.

Table 25 Average precision and recall (%) values for the four events for events agreed upon by at least three and at least four people and the difference in average values for *no pronoun resolution* and *strict pronoun resolution*.

***Strict* Pronoun Resolution**	**Gold Standard Events Agreed upon**	**Focus of Attention**	**Change of Location**	**Scene Change**	**Non-verbal Communication**
PRECISION %	3+	54.29	40.58	75.00	25.00
RECALL %	3+	28.57	60.87	46.15	12.50
PRECISION %	4+	47.14	21.74	70.83	75.00
RECALL %	4+	31.43	45.45	50.75	21.43

***No* Pronoun Resolution**	**Gold Standard Events Agreed upon**	**Focus of Attention**	**Change of Location**	**Scene Change**	**Non-verbal Communication**
PRECISION %	3+	64.29	35.48	72.92	25.00
RECALL %	3+	33.83	47.83	44.87	12.50
PRECISION %	4+	54.29	20.31	68.75	75.00
RECALL %	4+	36.19	39.39	49.25	21.43

4.3.2.1 Increasing Precision and Recall Values

The fact that this is a prototype system allows for a great deal of improvement in terms of precision and recall. Though it can be argued that no retrieval system can have 100% precision and recall when human judgement directs the search queries, (as can be demonstrated in any Google search where a series of *possible*, rated results are returned), the precision and recall values can certainly be improved upon. This section describes possible methods of improving the precision and recall of our system.

Further Natural Language Processing

In our analysis of the VisualText results we were strict with pronoun associations for the results of the evaluation (Gold standard vs. System Output) e.g. if in the system output event time and description were correct and pronoun/proper noun was not correct we did not consider the event. Thus better Pronoun resolution (correct association of pronouns with their respective proper nouns) would increase precision results. Also plural pronouns such as 'them', 'they' were not resolved and resolving them would increase recall.

Issues arose where an adjective was extracted instead of a noun, for example in the phrase "the hidden gun" the adjective 'hidden' would have been returned instead of the noun 'gun'. Also in the case of the verb phrases (FOA and COL events) adverbs in the phrase stopped phrases such as

"Tom walks slowly across the hall" and "Sarah looks longingly at Karl" being extracted by the analyser. This could be resolved by including a grammatical analyser, such as GATE [152], to recognise adjectives and adverbs in certain positions around noun and verb phrases involved. This would increase recall. This was not implemented due to time restrictions.

A grammatical analyser would also be able to distinguish between the noun and verb versions of certain 'body part' nouns. For instance if we consider the nouns: 'face', 'head', 'hands', 'eyes' and 'arms' as nouns related to body parts then they are equally frequent as verbs: "turns to *face* the crowd", "It *heads* out the door", "He *arms* the weapon" and "she *eyes* him up and down". Resolving this could possibly increase system precision giving more 'precise' event extraction.

Refining the Heuristics

The system 'missed' certain events because the patterns involved were not available for the VisualText analyser to detect. The phrases the heuristics are based on only include the most frequent synonyms of the node word in the collocation phrases. Thus, certain words were not included as they were infrequent in the corpora for the four film events. The other synonyms of the node words (such as synonyms for the node words: 'looks', 'takes', 'walks', 'door' and 'room') may increase the 'hits' and increase recall. Alternatively, removing certain frequent synonymous words may increase precision as certain commonly used words, such as: <go> for 'walks' in COL, and <have> for 'takes' in FOA, are used in different contexts than intended in their respective VisualText event analysers. The decision must be balanced based on frequency analysis in each case.

Utilising other Film Texts and Standardised Versions of Film Scripts

Using other sources of collateral film texts such as film subtitles, plot summaries and other versions of the film scripts may increase recall, as there would be more data to mine film element and event information from. However, the other texts may have to undergo the same process in terms of finding frequently collocating phrases relevant to the events and there would be new formatting and script alignment issues to consider.

We would have to standardise the scripts being used for analysis ensuring that the version of the film we are considering (DVD, Cinema reel, uncut version of film, directors cut, international version vs. US domestic version) matches whichever version of the film script we are dealing with. For example the screenplays we are dealing with have multiple versions: (early, first,

second, third and final drafts, shooting scripts, post production scripts etc.), which means that there are different omitted, edited and additional scenes to consider for each version of each script. Audio description scripts are usually developed for their respective DVD versions which may differ from the cinema releases which the screenplays are derived from. These issues may help increase precision and recall overall and will be considered in future work.

Issues exist, other than the system's performance or the choice of collateral media, such as the issue of human judgement. In this case we took the evaluators' judgement as to what constitutes an event to be the Gold Standard. However, people did not tend to agree as to what constituted an event all the time, some people saw more events and some saw none. Even when people saw an event at a given time in the film there was not always correlation concerning *what* they saw. The choice of evaluators was deliberately varied however. All 12 people had different backgrounds and different levels of film 'understanding'[29]. For instance there were film students, art graduates, computing PhD candidates, book publishers, undergraduate computing students and undergraduate economics students. We wanted to try and be subjective and see if such a diverse set of people would see similar events, which they did but perhaps not similar enough.

Another issue is that what a screenwriter considers an event, may not translate directly to what is seen on the screen. There is sometimes a need for a screenwriter to specifically write-in a subtle facial expression or movement or gaze of a character that is important to the scene or overall plot but which may not be so obvious to us. Chatman [18] speaks about an author's communication to an audience and how the audience is required to 'fill in' certain story gaps themselves and not all the elements of the story can be presented within the time allocated for a book or film: the audience must figure certain parts out themselves with the story elements presented to them in the narrative. This leads to the inference that people see different elements of the story than others and each person will have their own unique perspective or image of the film as a whole. This makes getting a Gold Standard difficult and even harder to ascertain a matching set of film elements.

Overall, evaluating the system helped identify certain issues and refinements that could be made to the VisualText analyser passes to increase precision and recall such as pronoun resolution issues and formatting issues. These issues were implemented and analysis of the 2 corpora of film scripts began.

---

[29] Documented through the questionnaire with the question FILM EXPERTISE: High, Medium or Novice

4.3.3 Discussion

The Gold Standard provides us with a set of event data, generated by humans, for four film events in five video clips from popular Hollywood films. This provides a basis to compare our IE system's performance, and our heuristics, to. The Gold Standard data also helps validate the four film events that chapter 3 has systematically found by *proving that they exist*. In other words the fact that 12 humans can search through five, 12 minute film clips and find multiple instances of all four events provides evidence that these events: FOA, COL, ScCh and NVC are common in films.

Overall the system manages to recognise correctly ScCh and FOA events the best, with approximately 70% and 55% precision respectively. Recall is low overall with COL and ScCh being the highest at approximately 48% each. Thus, the system misses a lot of information and there are *near* misses, i.e. the system returns a hit for the wrong reason but it is a correct hit.[30]

Table 26 Overall approximate averages of the four film events' precision and recall statistics when comparing our system performance to the Gold Standard data.

Approx Overall Averages	Focus of Attention	Change of Location	Scene Change	Non-verbal Communication
PRECISION %	55	30	70	50
RECALL %	30	48	48	17

It must be noted however, that we wished to obtain as much information about films (automatically and *objectively*) from the collocation FSA diagrams in Chapter 3 and section 3.4.1 as possible and did not allow for any human judgement or intuition to be used when developing and implementing the heuristics. Thus, any improvements that could have been implemented from the design stage that involved human judgement to improve system performance were not implemented as they were not derived from the collocation analysis when identifying the four events. Darren Johnson's work did involve some human judgement however and made for increased precision (~5%) and recall (~10%).

There is information in the scripts that is 'lacking' however. For instance about 10% of the Screenplay corpus did not have interior or exterior scene change cues or day and night time cues. This was due to the versions of the screenplays we had acquired being diverse, to allow for a representative corpus. For instance, films that were *transcribed* did not have such information in their scene changes and sometimes did not mark scene changes at all. Other films were written

---

[30] Called a 'Contains' extracted instance by [56]

earlier, i.e. 1950s, 1960s where the precedent for adding scene changes in screenplays may not have existed then. Other versions of the scripts both AD and SC did not have that much descriptive information available (e.g. transcripts, early drafts and some films were more descriptive than others in the case of AD scripts) there was not always the information the four events represented, in a textual form, available to be extracted. There is also the information missing due to limited heuristics, information that would have been relevant to the four events. For instance in the case of the COL event words such as 'scrambles' and 'goes' are not included as they were not found to be frequent in the collocation analysis. These words however would have produced more instances (improved recall) for the COL. They may have affected precision negatively however as they may not refer to 'change of location' because 'goes' is a very frequent word used in different contexts.

Overall the system, as it is, performs with low recall and average precision but there is much scope for it to perform better with improvements of the heuristics, the pre-processing and the implementation. Even with no 'tuning' or improvements we still get encouraging precision and recall results from the heuristics individually. The point of this chapter was to examine what happened if we followed the collocation FSA and template data from Chapter 3, directly and objectively so rather than try and improve the IE system performance we are instead more interested in potential applications that can be developed from the IE system results

4.4 Towards Novel Applications for Accessing Film Video Data

The aim of this section is to consider how machine-processable representations of film content, produced by our IE system, can contribute to bridging the semantic gap for film content, and develop novel applications to access film video data. Having completed the heuristics and implementation, a data set of film event information for 193 feature films was extracted. We considered how we could utilise the data set results to develop novel applications for accessing film content. It seemed plausible that the data set could be used for applications, based on the statistics and event information that could be automatically extracted from the database of results. We speculated that the data set could help humans and machines understand film content in terms of the narrative structure of films, and stories and that information about film events could provide information about film structure. We were also interested in how the IE heuristics described in this work could be applied to film retrieval and/or browsing applications and in developing tool-benches to aid audio describers and screenwriters.

Thus, this section explores that statistics that can be automatically extracted from our data set , which allows us to reason about film content. The full data set can be seen on the accompanying **CD**: Additional Thesis Material – "DATABASE of extracted events from AD and SC Film Scripts" file. All graphs in this section can also be found on the **CD**.

4.4.1 Film Event Data Set: Overall Statistics

The implemented heuristics in VisualText for our four events were applied to our entire corpus of AD scripts and SCs. Though not all 193 film scripts could be processed for all four events (due to corrupt script file segments, varied formatting of older scripts and non-presence of certain event cues e.g. int and ext) over 92% of film scripts were processed for the four events. This gave 756 results in total for both corpora (*COL:* 219, *FOA:* 229, *NVC:* 195 & *ScCh:* 113). Each of the four event types had a different weighting of number of events (see Table 27), with the COL event producing the most events on average (~95 per film) followed by ScCh event (~94) and the NVC event the least (~5 per film).

Table 27 Depicts the average number of events for each film event for both AD and SC corpora

	AD	SC	Overall
Av COL	46.90	124.69	95.00
Av FOA	55.43	102.73	84.68
Av NVC	3.91	6.15	5.30
Av ScCh	N/A	94.39	94.39
Average Number of events overall			69.84

Consequently, we could say that film characters change location (or are in/on location) on average 95 times per film, focus their attention on other characters or objects on average 84 times a film, shake their heads and open/close their eyes on average five times per film and that there are on average 94 scene changes per film. These statistics are not conclusive and are based on 193 films but give us an insight into what *kind* of information can be automatically gathered from films. It can be argued that the weighting of these film events, in a film, may be able to provide information to a human or a machine about films in a macroscopic way, e.g. for audio describers, how many events and scenes to *expect* for films or a specific film, for a film researcher into genre (explored later), how many FOA, COL, NVC, ScCh events to a film and more importantly for a machine to be able to categorise films based on the number of film events extracted.

As an example of what kind of overall statistics can be gathered automatically using the templates of the four events, let us consider an example of the film "American Beauty" [163]. Both the

BBC audio description script and screenplay of the film were available for analysis. Table 28 shows the number of each event found for *both* the AD and SC scripts in the film as well as the total number of events. All the instances of the events were *above* average, especially in the case of FOA and NVC. We speculate that this could have been because this was a drama with much character interaction.

Table 28 The number of instances of our 4 events in the film "American Beauty" [163]

Event	Number of Instances
FOA	207
COL	140
NVC	14
ScCh	119
Total	**480**

Plotting the event occurrences against time reveals more information about the film, see Figure 40. Specific segments of the film have a greater density of certain events than others. For instance, at 10-20 minutes and 30-40 minutes there are numerous scene changes, more than any where else in the film and at 55-85 minutes there is a greater density of focus of attention events. These *denser* sequences of events may indicate some sort of intensity in the film.

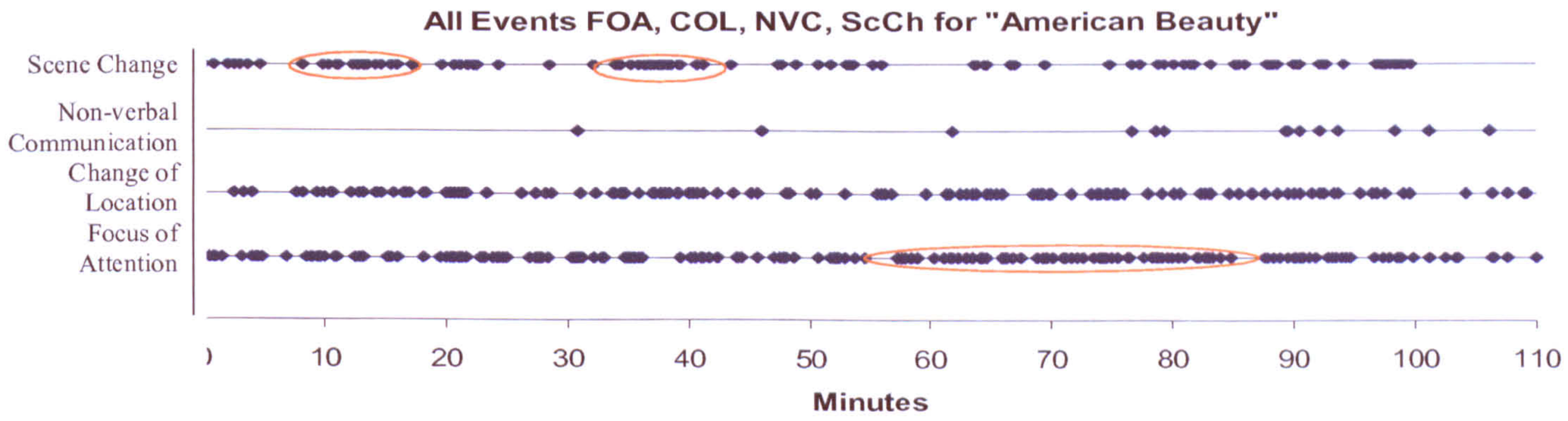


Figure 40 An example of the instances of the 4 events FOA, COL, NVC and ScCh over time for the film "American Beauty" [163].

If we examine a level deeper into the event instances and separate what *types* of each event instance can be automatically extracted, the *density* of event instances becomes more pronounced. Figure 41 shows the subtypes of the four events that can be found. The case of change of location the character may be entering, leaving, within or on a location; Non-verbal communication may occur with eyes or the head and there are different types of scene change, day and night time scenes and unnamed scenes or scenes in space. Figure 41 again reveals that at 10-20 minutes and

30-40 minutes there are numerous scene changes, more than any where else in the film and that there is most focus of attention at 65-85 minutes. It also reveals that there is a lot of movement or location changing between the 35-45 minute segment.

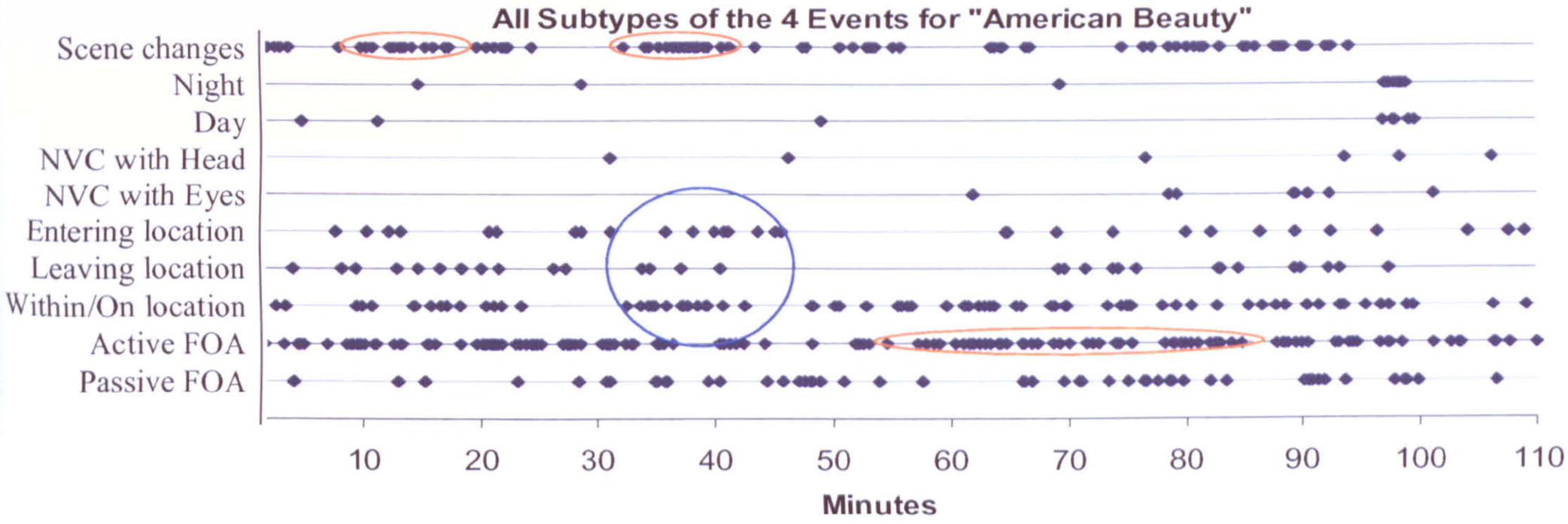


Figure 41 An example of the type of information that can be automatically extracted from the 4 events for the film "American Beauty" [163].

Graphs such as Figure 40 and Figure 41 could be used to read off information about a film and possibly identify dramatic (high density of FOA, NVC) or action scenes (fast scene changes and COL). The graphs could also be used to compare films (different weightings of different events) and perhaps categorise them. These ideas are explored in the sections that follow.

4.4.2 Focus of Attention Event: Information about Characters' On-Screen, Items and Characters' Focus of Attention

The FOA event performed with approximately 55% precision and 30% recall. It was found that there were approximately 85 Focus of Attention events per film and out of those events approximately 65% were *active* and 35% were *passive* FOA events.

The Focus of Attention event provides us with film content information about characters: *what* they are focussing on (objects), *whom* they are focussing their attention on and when a character is on-screen. FOA events have been categorised as either 'Active' or 'Passive'. *Active* refers to whether a character is actively, or consciously, focussing their attention on something or someone and passive refers to whether a character is holding/carrying/taking something, e.g. Tom carries a spade, Hana takes the rope, and knows it is there but is not focussing directly on the object (or

person). Figure 42 shows us an example of the instances of active and passive Focus of Attention over time for three scripts of "The English Patient" [164] (two AD, one SC). The active Focus of Attention is densest at 50-80 minutes. This coincides with the section in the film where the hero's (Almasy's) memories of when he was in love with Katherine are recounted.

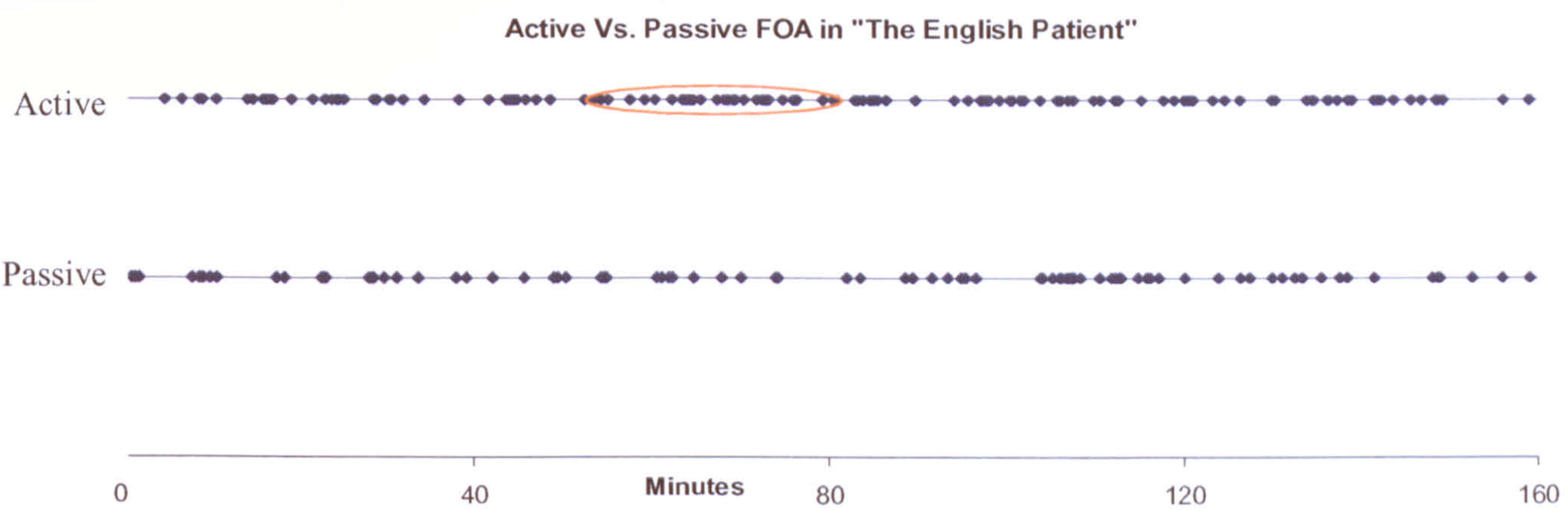


Figure 42 Shows instances of active and passive focus of attention in "The English Patient" [164] Vs. time.

An aspect of film content that can be captured by the FOA event is knowing when a character is on-screen. Further to that we can also explore what that character is focussing their attention on. Using the film "The Mummy" [171] as an example, Figure 44 depicts which character is focussing on which other character at what time and Figure 43 shows which characters focus the most, and possibly reflects which characters are on-screen the most. For instance, Evelyn (the heroine) is the character that has the most FOA instances and focusses mostly on the characters: Connell (hero/love interest), Jonathan (her brother) and Imhotep (villain), shown in Figure 44. Whereas Connell focusses mostly on Evelyn, Beni (comic relief/Imhotep's henchman) and Jonathan.

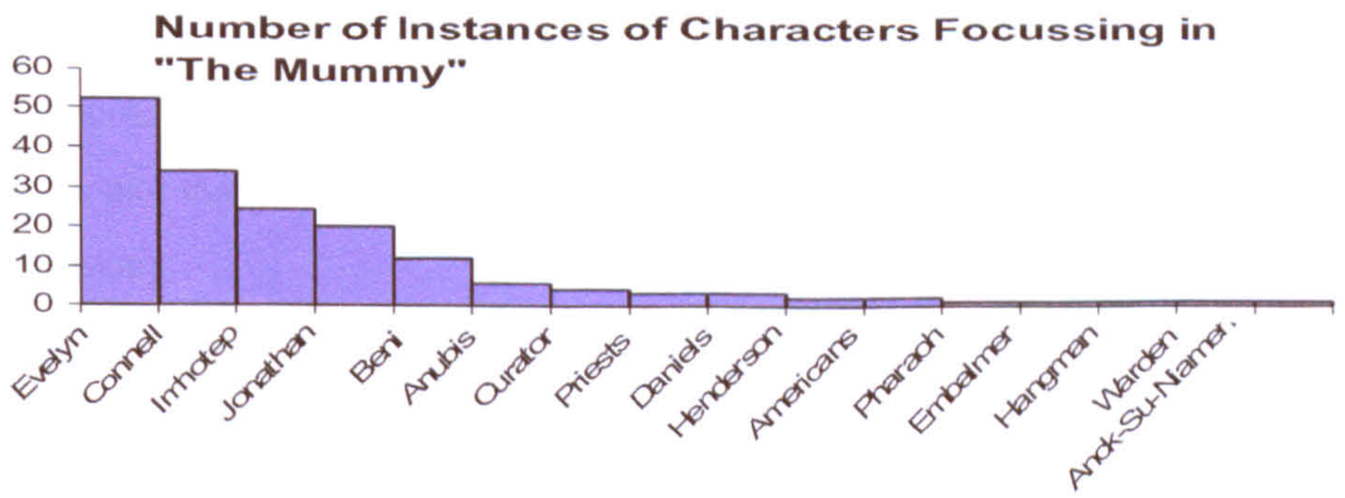


Figure 43 Depicts the number of instances of characters focussing on something or someone in "The Mummy" [171] and (possibly) the characters that appear most on-screen in the film.

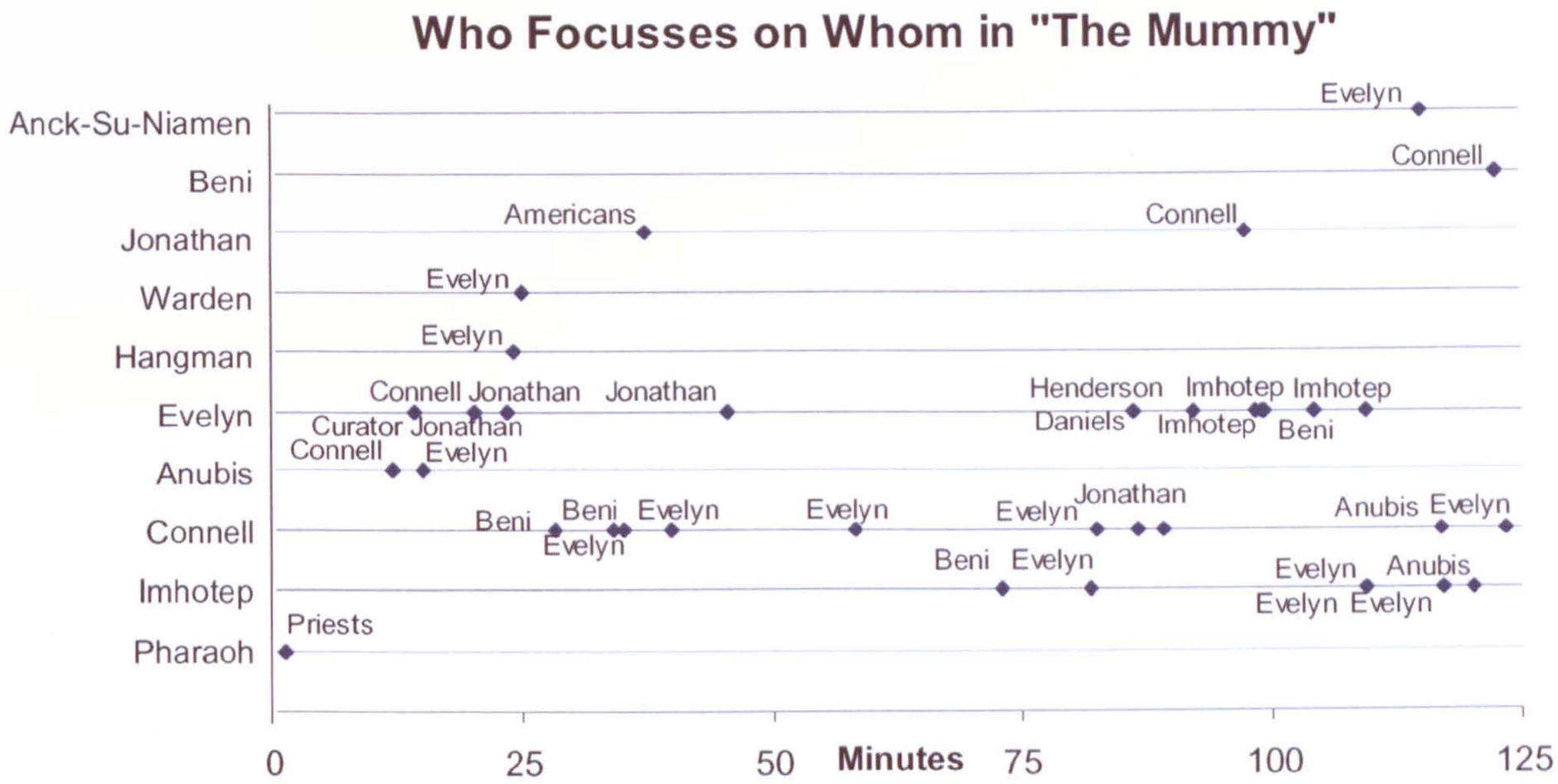


Figure 44 Shows which character is focussing on which other character at what time in "The Mummy".

The Focus of Attention event can also provide film content information about items and objects that characters use or are integral to the plot. This information also shows us what kinds of objects are used in the film. Figure 45 and Figure 46 show instances of objects extracted automatically by the FOA event in "The Mummy"[171] and who is focussing on them at what time. The majority of the items focussed on are building parts (window 5, wall 3, altar 1) and buildings (the temple).

"The Mummy" is set partially in ancient times and mostly in the 1940s in Egypt, thus some of the items extracted reflect that (sarcophagus 4, hieroglyphics 3, flame lit torch 3, scarab beetles 2, mummies 1 etc). The fact that specific items from a given area (Egypt) or period (Ancient Egypt, the 1940s) are extracted supports Chatman's [18] ideas of *cultural codes*, as part of a narrative where the specifics of a culture are interlaced in a narrative. There are also some objects that are integral to the plot that are extracted, such as 'the key box' and 'black book' and others that may help the plot along.

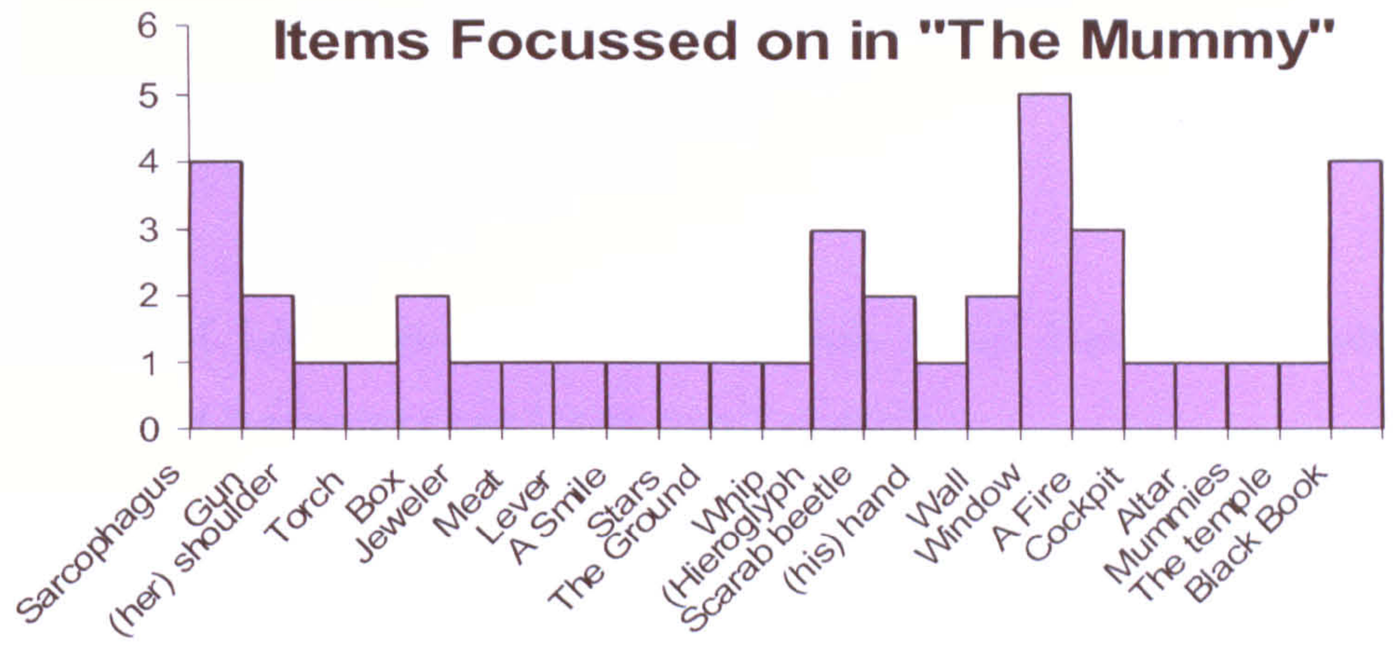


Figure 45 Shows the instances of the items that characters focus their attention on in "The Mummy".

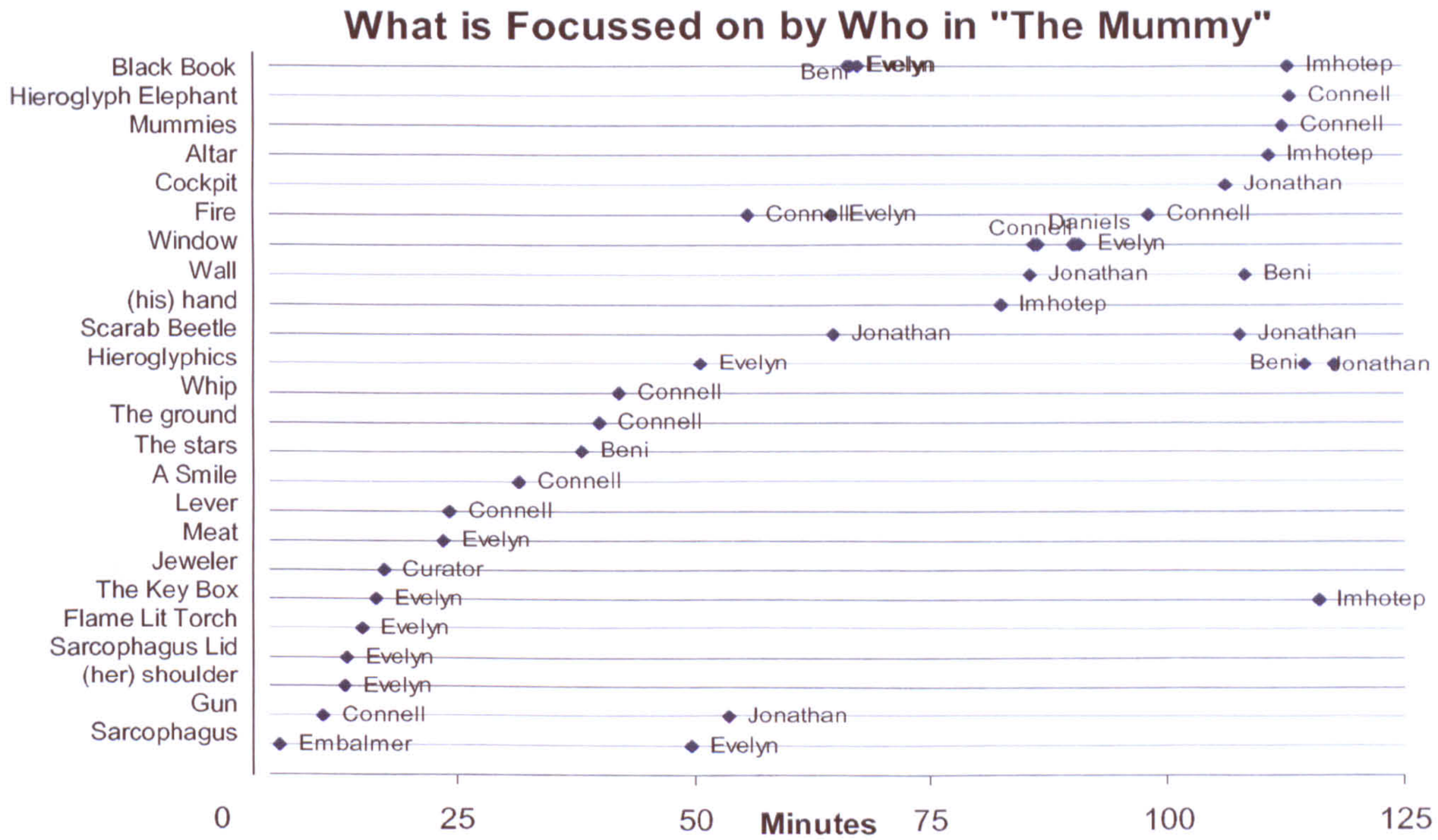


Figure 46 Depicts what objects, body parts and buildings are focussed on by whom in "The Mummy".

In terms of providing information about film content, the FOA event provides information about *existents* in film, where existents refer to anything that *exists* in a story or film, characters, objects, and settings [18]. The FOA event also provides information about cultural codes in stories. The most important film content information extracted automatically by the FOA event is information about characters: information about their presence on-screen, their interactions with

characters and objects and their focus of attention. The FOA statistics may be useful to film students and scholars and to audio describers, who may need to describe a FOA event to the visually impaired, as it marks the FOA instances in a film.

4.4.3 Change of Location Event: Information about Rooms Frequent in a Film and when Characters are Changing Location

The Change of Location event performed with low precision: approximately 30% precision and ~50% recall. It was found that there were approximately 95 changes of location events per film and out of those approximately 65% were characters within or on a location and ~19% were leaving a location and ~17% entering a location. The COL event provides us with information about characters changing rooms, either leaving or entering rooms or when a character is *within* a location or *on* a location at a given time. Figure 47 demonstrates the Change of Location of the characters in the two scripts of the film "High Fidelity" [162] (AD and SC scripts).

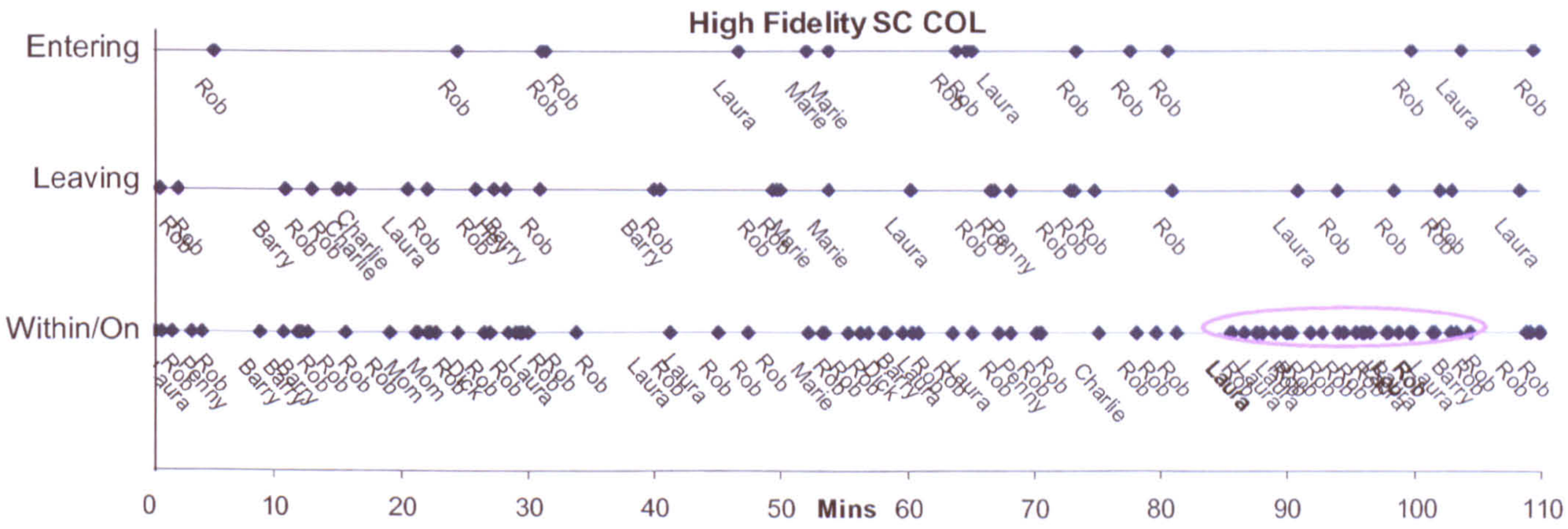


Figure 47 Changes of Location: entering, leaving or within a location, for characters in "High Fidelity" [162]

The COL event also provides us with information about which *rooms* characters were changing to, coming from, or which rooms characters were in. COL events provide us with the most frequent rooms or areas visited in a film and cues to when a character is at a door and about to perform a function (open, close, slam, lock the door). The cues for such information are most common when a character is entering a room (and when a character is leaving a room, but this is rarer). Figure 48 and Figure 50 show the rooms frequently visited by the main characters at a certain time in the films "The English Patient"[164] and "High Fidelity" [162]. The rooms visited

indicate the main locations used throughout the film. For instance, in the "English Patient" the film is mostly set in a 'room/bedroom', in the 'garden' and in general rooms of a house: kitchen, library and corridor. In the case of "High Fidelity" the film seems to be set mostly in a room/bedroom and in an apartment and a bar. In both cases there are many 'door' cues, indicating a lot of activity between rooms.

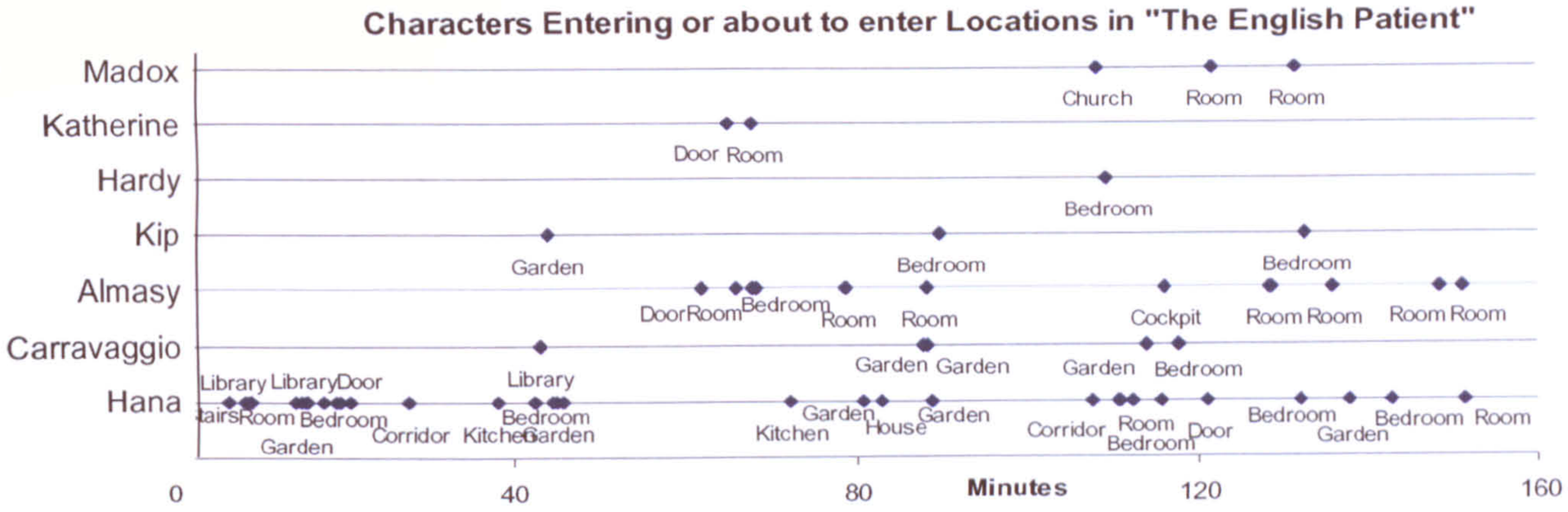


Figure 48 Common locations a character enters or is about to enter in "The English Patient" against time.

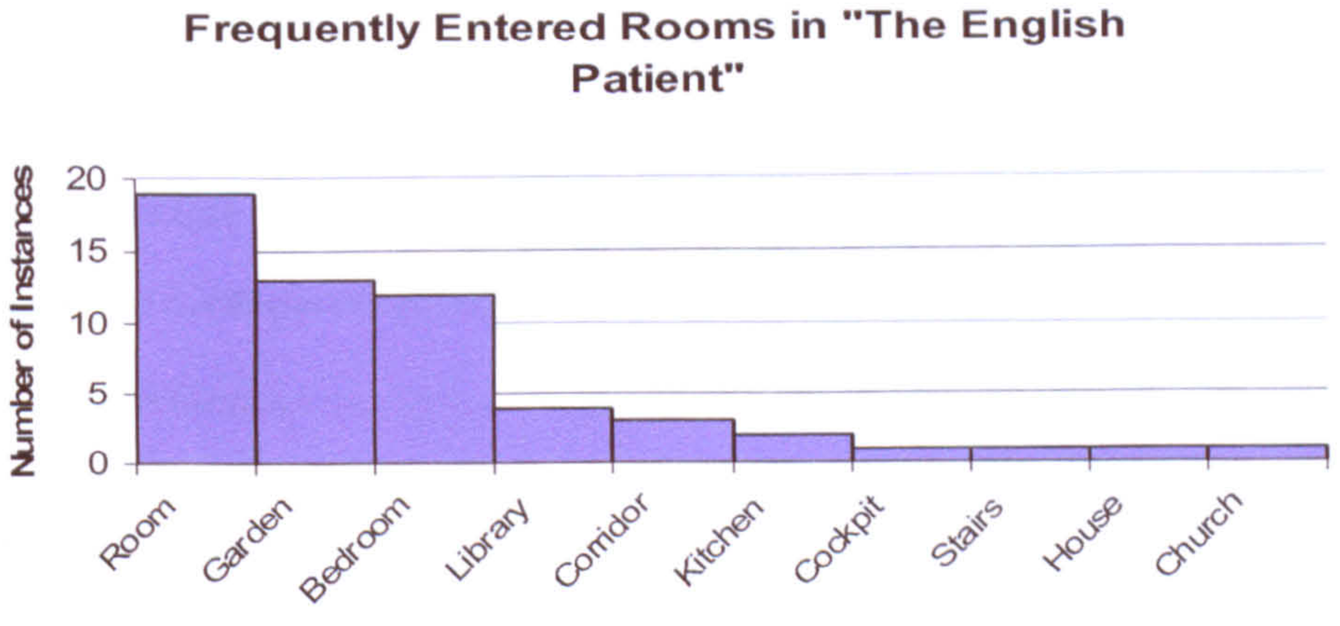


Figure 49 The rooms that are most entered in "The English Patient" [164].

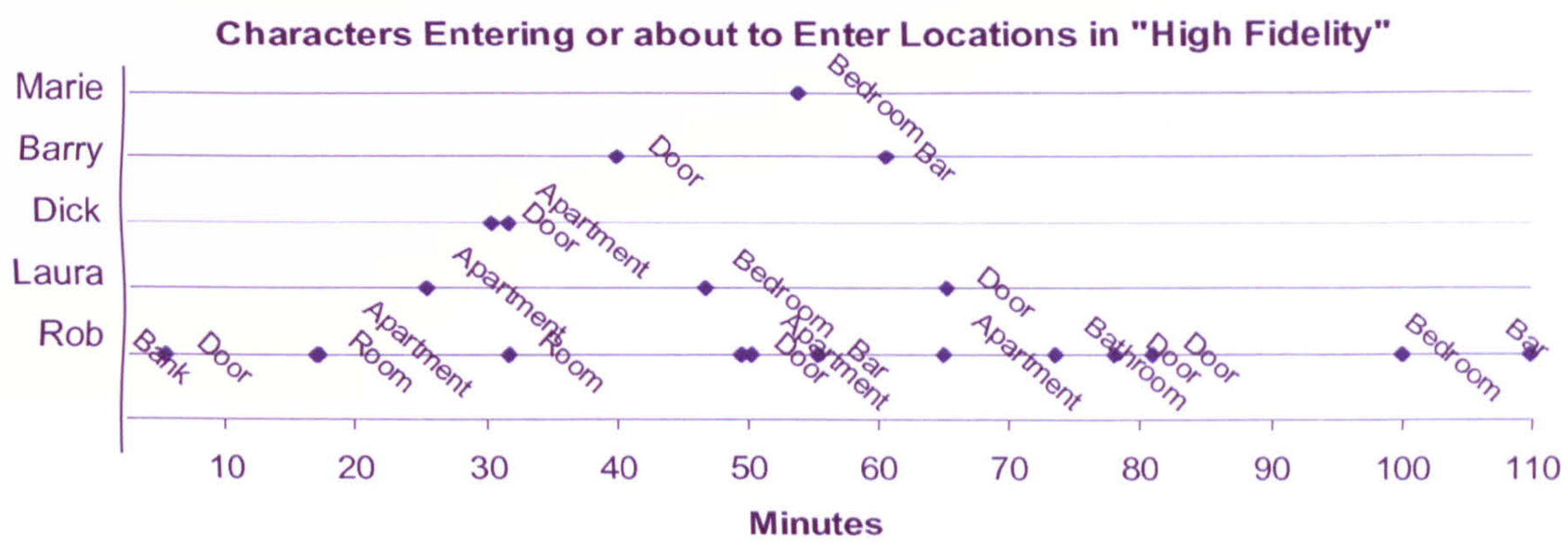


Figure 50 Shows common locations a character enters or is about to enter in "High Fidelity" against time.

The COL event, like the FOA event provides information about *existents* in film, specifically characters and settings [18]. It also provides information about what types of room are frequently visited by characters in films, providing the setting and backdrop for the film. Thus, from the COL event the following film content information can be gathered: when characters are on-screen, when they are changing location (entering, leaving or within a room or area), and the settings or locations in a film.

4.4.5 Non-Verbal Communication Event: Information about when Characters are Communicating Non-Verbally

The NVC event performed with approximately 50% precision and low recall ~17%. It was found that there were approximately five instances of non-verbal communication per film which involved the head (shaking, nodding) or the eyes (opening, closing) with an approximately 50:50 split on average. The NVC event provides us with information about when a character is communicating with a certain body part, non-verbally, at a specific time in a film. Figure 51 depicts at which point a character is communicating non-verbally in the film "Sixth Sense" [168]. The main character Cole seems to be communicating non-verbally most with a series of head shakes and nods.

The film content information being provided here is firstly when a character is on-screen and secondly when a character is communicating non-verbally. These statistics may be useful to film students but may be of great use to audio describers as they specify NVC that will need describing to a visually impaired person.

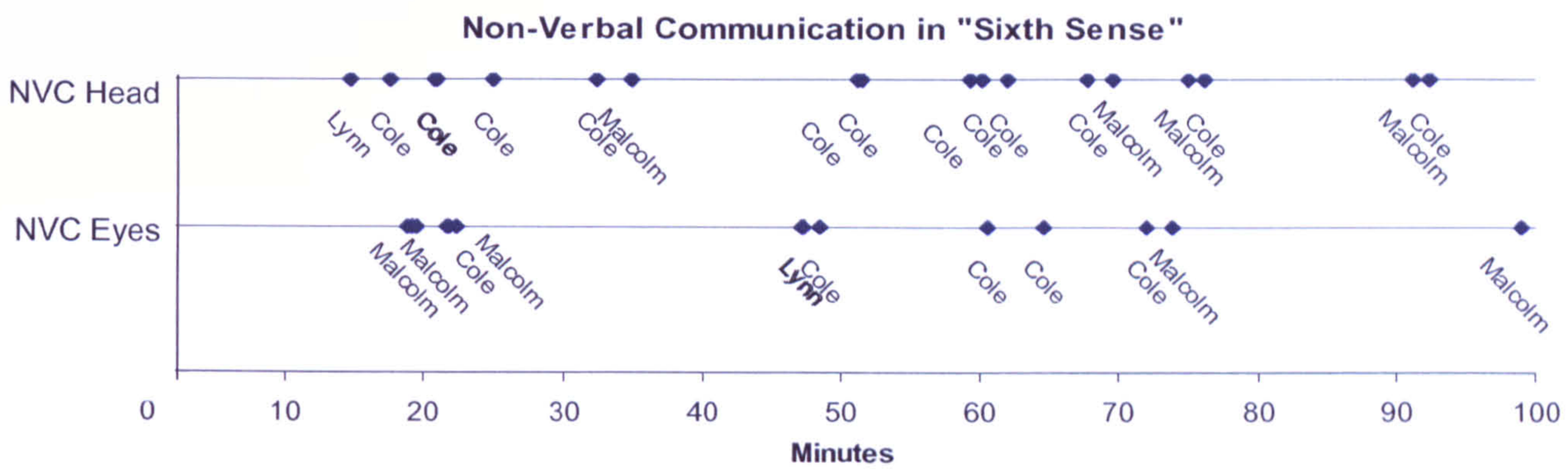


Figure 51 The instances of non-verbal communication for characters against time in "Sixth Sense" [168].

4.4.6 Scene Change Event: Information about Scenes in Films

The scene change (ScCh) event captures information about film. As can be seen below ScCh captures whether a scene is internal/external, its location, occasionally what year or date it is, and the time of the day it is (e.g. dusk, day, night or afternoon). Sometimes it will not give a time of day but a further piece of information, e.g. 'SPACE', indicating that the scene is in outer space, or 'CONTINUOUS' indicating that it is a continuous shot or scene; this is referred to as N/A in Table 29. It must be noted that this information is available for only the SC corpus due to AD scripts not having any scene change information. Below we see a typical scene change 'marker' taken from "The English Patient"[164] indicating that we have cut to an *internal* a scene at the *Ambassador's residence* in *Cairo* in *1939* and it is *night* time.

INT. AMBASSADOR'S RESIDENCE. CAIRO, 1939. - NIGHT

This information is extracted automatically by our system with above 70% precision and >51% recall (see evaluation). Statistics for films' scenes can be extracted at a Macro and 'in film' level. It was found that there were approximately 95-100 change of scene events per film and out of those events approximately 39 scenes were at night and ~61 in the daytime with ~65 interior scenes, ~42 exterior scenes and 7-8 miscellaneous scenes (e.g. set in space). On average ~60.5% of film scenes are day scenes and ~42% are night scenes with ~61% being interior and 39% being exterior scenes (Table 29).

Table 29 The average number of Night, Day, other (N/A), Interior and Exterior information and average percentages of the attributes for all the films in the corpora.

	Night	Day	N/A	Interior	Exterior	%Night	%Day	%Interior	%Exterior
Average Number of Attributes	38.69	60.57	7.80	64.94	42.15	39.18	60.42	61.22	38.78

These statistics may be useful in themselves to students analysing films and narrative and provide information about the 'discourse' of films or how they are presented to us: mostly in the day and inside. A computer system may associate the pattern [INT <LOCATION> DAY/NIGHT] exclusively to film allowing scene information to be extracted automatically.

From the scene change event, location, time of day, whether the scene is internal or external and combinations of this information can be extracted. This can all be considered film content information. Figure 52 to Figure 55 depict combinations of this information. Figure 52 depicts the instances of the day and night cues from the screenplay of the film "8 Legged Freaks" [160]. Other 'times of the day' cues are grouped with the day and night cues e.g. afternoon, noon, dawn with the *day* cue and, evening, dusk, sunset with the *night* cue. In the case of Figure 52 there is a clear progression from night (spiders arrive) to day, night and finally morning at the film's end.

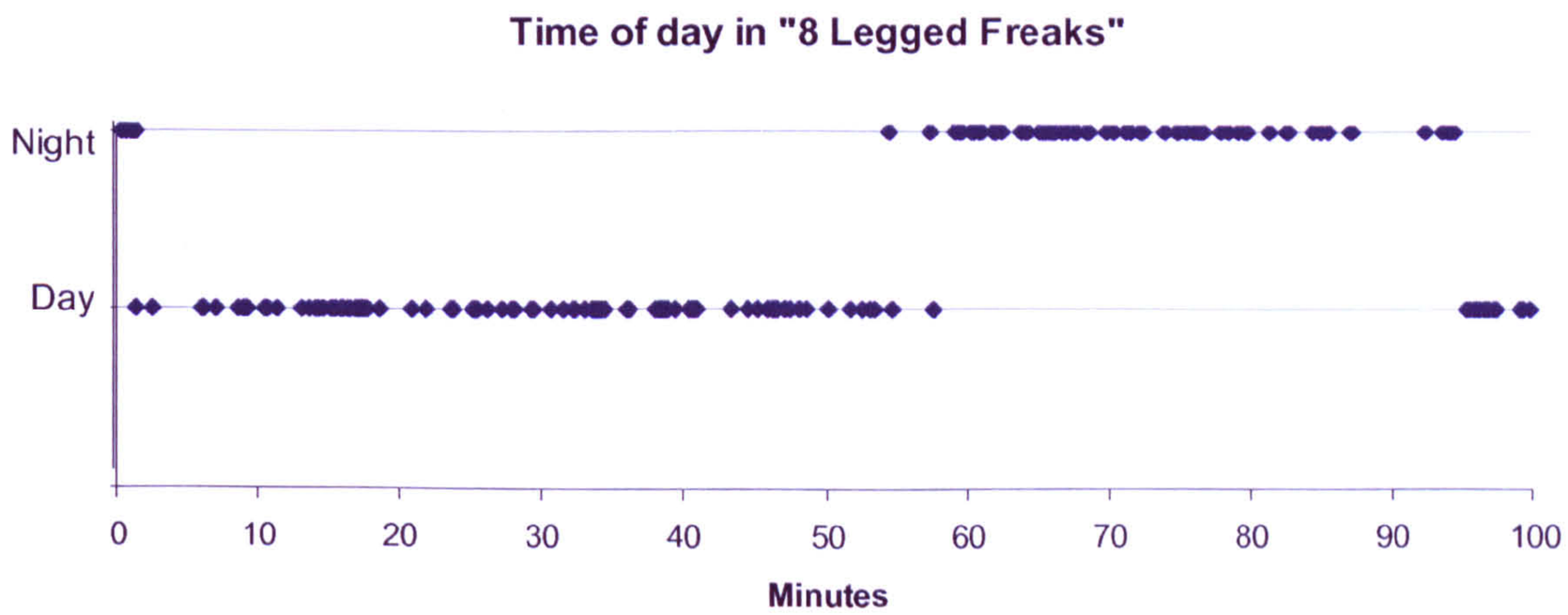


Figure 52 The time of day in "8 Legged Freaks" [160] at given scene changes in the film.

Figure 53 and Figure 54 depict instances of the most frequently visited locations in the film "The Mummy" that have been extracted from the ScCh events. The most frequent location in the film is the 'Underground Chamber' at the 'City of the Dead: Humanaptra'. Figure 54 depicts at what time of the day the locations are visited and at what time in the film and Figure 55 shows which location instances are internal or external scenes.

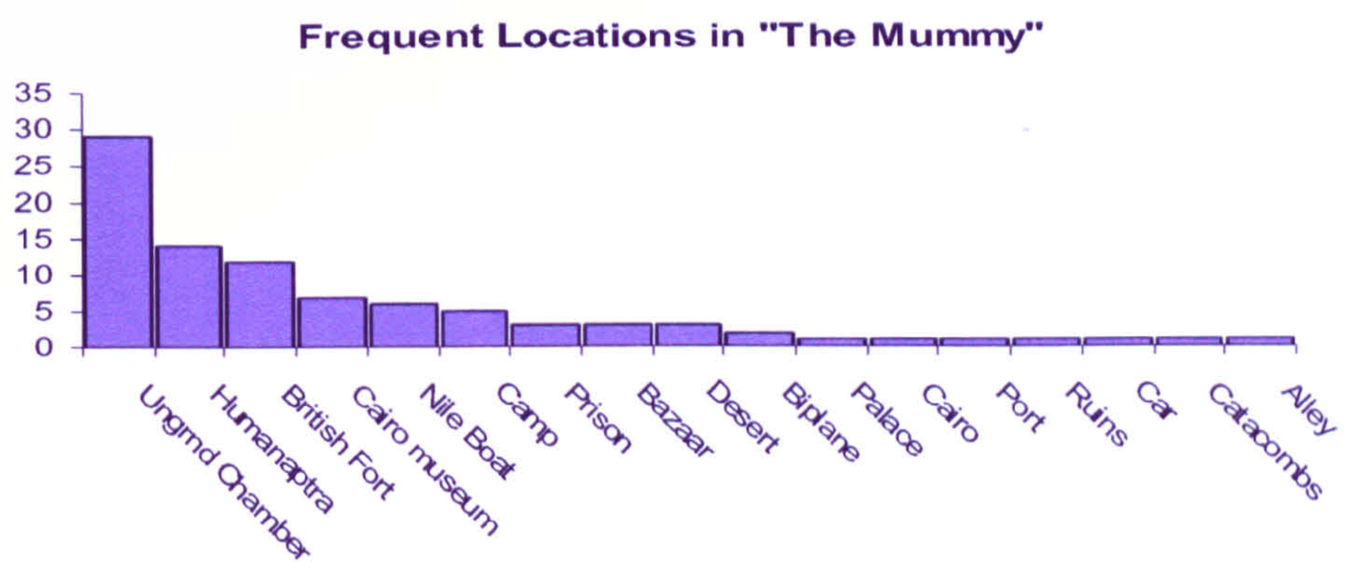


Figure 53 The instances of the most frequent locations in "The Mummy".

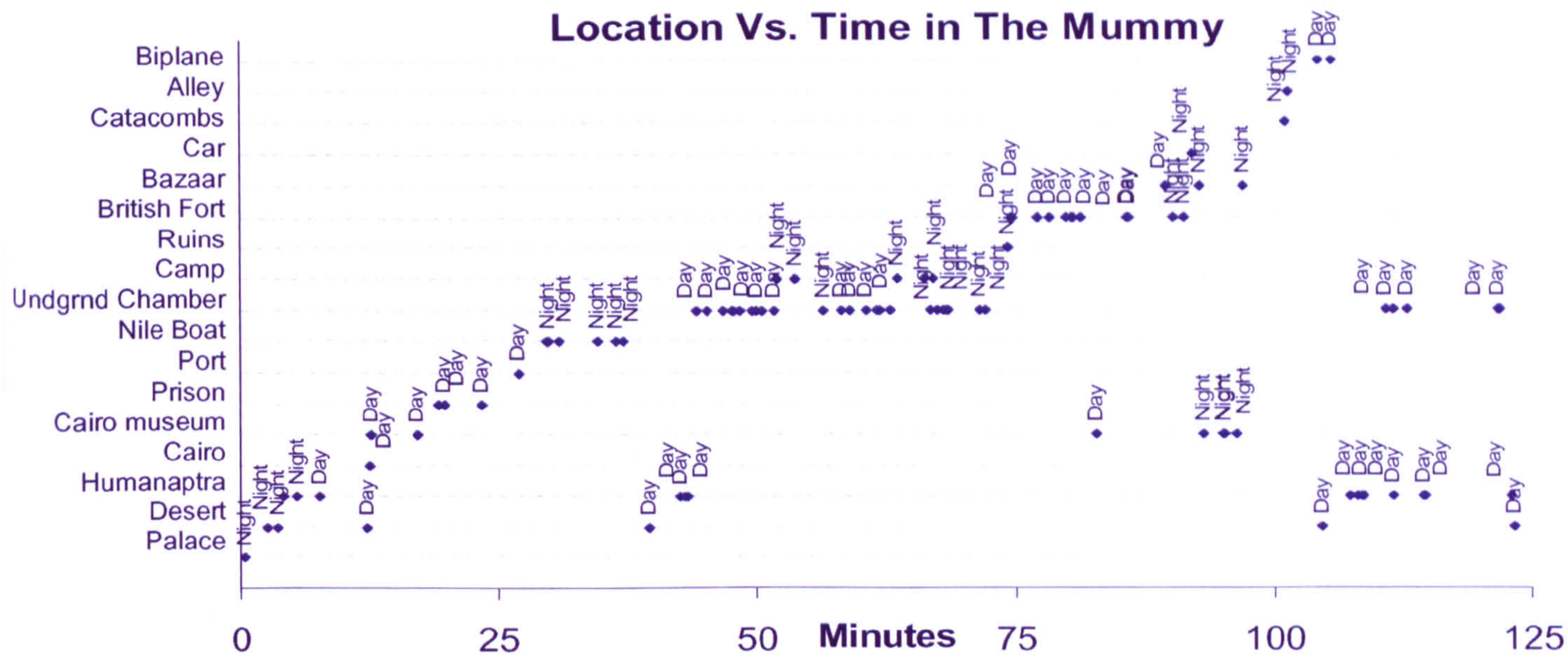


Figure 54 Frequent locations in "The Mummy" [171], what time they appear in the film and the time of day it is.

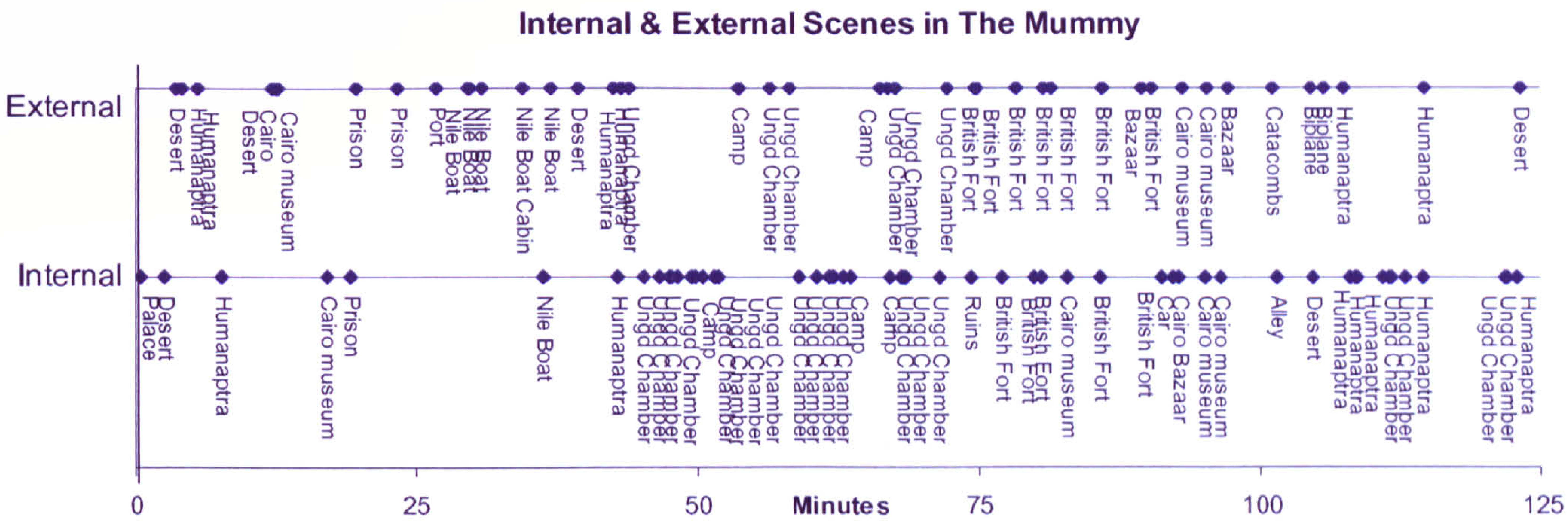


Figure 55 The instances of internal and external scenes and their locations in "The Mummy" [171]

The Scene Change information captures film content in the form of locations, time of day cues, spatial cues for location (internal/external). Again information concerning Chatman's [18] concept of existents can be captured in the form of *settings* (see Figure 9). This information may be useful to film scholars examining specific traits of films e.g. scene changes. It may also be useful to audio describers as a cue for recording a description of a scene change. The ScCh template may also be useful to script writers as part of a tool-bench to automatically layout a pre-emptive formatting template for a scene change.

4.4.7 Event Statistics: Further Inferences

This section describes possible applications and ideas that the statistics of the four events can be used for in order to represent, extract, compare, navigate/browse and generally reason about film content.

4.4.7.1 Genre Classification

The word genre here refers to a category that a film is classed as. Table 30 shows the categories of film used in this study, ranging from action to western. Using the Internet Movie Database [136], we classified the 193 films in our corpora by genre (seen in Table 30).

Most of the films were 'action' films (31%), 'comedy' (18%) and 'crime' films (18%). Due to the large number of 'action' films and the large difference in number to other genre categories, e.g. 1

'western' film and 1 'fantasy' film, we decided to group certain genres and the results of this can be seen in Figure 56.

Table 30 Common genres of film and the number of films in our corpora that are classed as a specific genre, as collected from **www.imdb.com** [136].

Genre	Abbreviation	Genre Representation Number	Number of films of that genre
Action	A	1	61
Adventure	Ad	2	9
Animation	An	3	9
Biography	B	4	6
Comedy	C	5	35
Crime	Cr	6	17
Drama	D	7	35
Fantasy	F	8	1
Horror	H	9	10
Mystery	M	10	3
Musical	Mu	11	0
Romance	R	12	5
Science Fiction	S	13	0
Thriller	T	14	3
War	W	15	0
Western	We	16	1
Total Number of Films			**193**

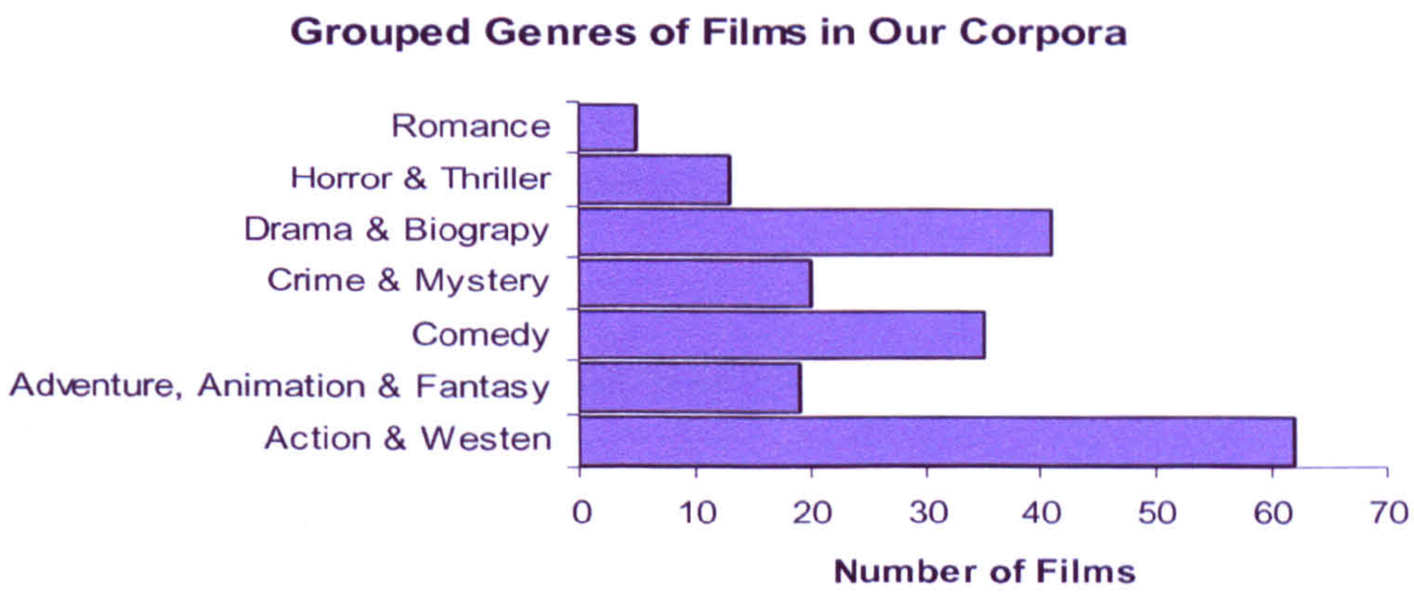


Figure 56 Depicts the grouped genres for the films in the AD & SC corpora

Having grouped the films into genres it was now possible to gather and compare information from the four film events across genres. So, using MS EXCEL macros we were able to automatically gather information per genre from the film corpora with regards to the film events. Table 31 displays the average number of events per grouped genre. On average comedies have

the most number of Changes of Location[31] and action movies, westerns and adventures seem to have the lowest change of location. This is reflected very closely in the Focus of Attention events also. Non-verbal communication seems to be highest in comedies and lowest in action movies as well. Adventures and dramas have the highest number of scene changes, with thrillers and horror having the least.

Table 31 Grouped genre average number of events per film event for both corpora

Genre	Number of Films	Av COL	Av FOA	Av NVC	Av ScCh
Action & Western	64	52.00	48.00	1.00	97.81
Adventure & Animation	19	56.61	57.33	2.89	202.86
Biographical & Drama	47	84.57	76.82	5.05	151.05
Comedy	35	110.83	104.50	8.06	91.18
Crime & Mystery	21	84.75	88.42	5.86	116.59
Romance	8	90.63	74.00	5.50	141.67
Thriller & Horror	12	96.44	91.78	4.56	47.94

We were also able to compare the Scene Change event information across genres but for only the Screenplay corpus (Table 32) as the AD corpus had no INT, EXT, DAY or NIGHT cues. From Table 32 we see that overall we still have a ~60:40 split in favour of DAY scenes and Interior scenes. 'Crime & Mystery' and 'Romance' films have more day scenes (73 and ~86 respectively) on average than the average (60.5) and 'Romance' (~62) and 'Adventure & Animation' (~46) night scenes are higher than average (38 events). 'Drama' (~79.5), 'Romance' (84) and 'Crime & Mystery' (76) films are higher than average in Interior scenes (65 events).

Table 32 Average figures for Scene Change information in both corpora for grouped genres.

Genre	Night	Day	N/A	Interior	Exterior	%Night	%Day	%Interior	%Exterior
Action & Western	39.63	55.90	4.49	58.75	41.26	35.54	53.12	54.42	45.66
Adventure & Animation	46.20	52.80	11.80	63.00	47.80	35.73	43.71	55.64	44.36
Biographical & Drama	37.30	70.80	8.86	79.47	37.52	33.00	54.10	68.82	31.18
Comedy	22.10	58.60	9.79	53.57	37.00	21.83	67.11	61.47	38.53
Crime & Mystery	38.00	85.60	3.33	75.89	51.00	30.49	66.03	60.42	39.58
Romance	61.70	73.00	6.67	84.33	57.00	45.64	48.86	62.22	37.78
Thriller & Horror	25.90	27.30	9.67	39.56	23.44	37.01	35.31	65.62	34.38

Analysing and comparing genres may be of interest to film students and any film enthusiast. It may be possible to provide humans and computer systems with information about the mood or

[31] Perhaps it is the slapstick element of the film.

specific idiosyncrasies of a genre of movies. An example of this can be seen in Table 33 where we compare 'horror' films, where horror is the primary or secondary genre of a film according to IMDB [136].

Table 33 Depicts Z-Score and % of events information about the Scene Change events for 'Horror' films in our corpora

	Event Z-Scores					% Of Event Type per Film				
Film Title	**Night**	**Day**	**N/A**	**Interior**	**Exterior**	**%Night**	**%Day**	**%Interior**	**%Exterior**	**%N/A**
Blade	1.73	0.26	0.00	1.63	0.59	75.00	25.00	75.93	24.07	0.00
Blade Trinity	1.12	0.07	0.00	0.76	0.94	72.99	27.01	65.00	33.33	1.67
Blade 2	0.55	-0.24	-0.09	0.56	-0.41	73.08	26.92	75.71	20.71	3.57
Jaws	-0.98	-0.85	0.00	-1.28	-1.10	40.00	60.00	35.00	65.00	0.00
Terminator	1.70	-0.58	0.00	0.61	1.50	87.91	12.09	59.89	40.11	0.00
The Devil's Advocate	0.62	2.28	0.00	1.80	0.81	43.04	56.96	75.22	24.78	0.00
Final destination 2	-0.95	-0.27	2.74	-0.30	-0.41	22.73	77.27	44.78	21.64	33.58
From Dusk till dawn	-0.86	-0.77	0.00	-1.13	-1.02	50.00	50.00	50.00	50.00	0.00
Halloween 6	-0.23	-0.66	0.00	-0.67	-0.36	72.86	27.14	57.14	42.86	0.00
Halloween Resurrection 8	-0.97	-1.14	-0.38	-1.28	-1.49	90.00	10.00	58.33	33.33	8.33
I still know what you did last summer	0.30	0.10	-0.09	0.19	0.42	62.79	37.21	61.87	34.53	3.60
Scream	-1.09	-1.06	2.10	-0.93	-0.97	33.33	66.67	33.33	20.51	46.15
What lies Beneath	0.02	0.41	0.00	0.09	0.24	52.00	48.00	64.80	35.20	0.00

In 7/13 films of the 'horror' genre have more night than day scenes with ~55% on average night scenes vs. ~44% day scenes. This does not reflect the overall averages for all films in the corpora which are reversed (60:40 in favour of DAY). This may be an idiosyncrasy of horror films: that they are set mostly at night. In terms of film events in general however the number of average COL, FOA, NVC and ScCh events do not differ that much from the overall average, which will not allow us to reason about 'horror' through these statistics.

We can compare the different genre event information to examine for any idiosyncrasies, differences or similarities. For instance, from Table 34 we observe that the 'Romance' genre deviates most from the mean, and is positive (0.737) indicating that there are more scene changes in 'Romance' films and least in the 'Thriller and Horror' genre due to a negative z-score (-0.477). it can also be argued that 'Comedies' have the less COL events than other genres (-0.524), 'Crime and Mystery and 'Biographical and Dramas' films have more FOA events than other films and 'Adventure and Animation' films have the least number of NVC events. It can be said that each genre has a different amount of the four event types and thus we may be able to train a system to distinguish a genre based on statistical event information.

Table 34 Z- Scores of events for each film event for each genre.

Genre	ScCh	COL	FOA	NVC
Action & Western	0.109	0.230	0.992	-0.202
Adventure & Animation	0.038	-0.451	0.663	-0.420
Biographical & Drama	0.212	0.261	1.274	0.381
Comedy	-0.350	-0.524	0.494	-0.245
Crime & Mystery	-0.030	0.180	1.330	0.269
Thriller & Horror	*-0.477*	*0.104*	*0.851*	*-0.104*
Romance	0.737	0.394	0.984	0.186

Having examined films as a whole, with macro statistics, it was also possible to reason about and represent film content for single feature length films. The next section describes representations of film content in the corpora of films.

4.4.7.2 Character & Important Object Presence On-screen

Using the FOA, COL and NVC events it is possible to know when a character is on-screen. Using the FOA Active events it is possible to know when *two* characters are on-screen at the same time. The FOA event will also inform us when an important object is on-screen. Thus, the events can act as locators of characters and objects. It is possible to track the movements of the characters across time using the COL event. In this case we are considering: whenever a character is mentioned in an event, then they are considered to 'appear'. It is possible to combine the FOA, COL and NVC event results for a film to give a better picture of when characters are on-screen.

Figure 57 Instances of when the characters in the film "American Beauty" [163] are on-screen based on 3 events COL, FOA and NVC.

4.4.7.3 Dramatic/Action Scene Classification

The basic principle of this classification is that when the FOA and ScCh events get *denser*, i.e. more of an event per given time than usual, this indicates the possibility of an intense dramatic scene, (FOA: many character interactions and characters looking at or away from each other). Or this indicates an action or chase scene that is fast paced, and may involve many scene changes and objects being focussed on. Figure 58 shows a possible example in terms of the screenplay for the Action/Horror "Blade 2" [174]. There are areas where the scene changes are denser than others and they indicate action scenes, (at 40-55 minutes: Blade fights the villain Nomak, 70-85 minutes: group fight vampires and 99-110 final battle). 95-100 is a drama scene between daughter and father. It can be speculated that the rate of change of scenes is proportional to how fast paced the film is in certain parts.

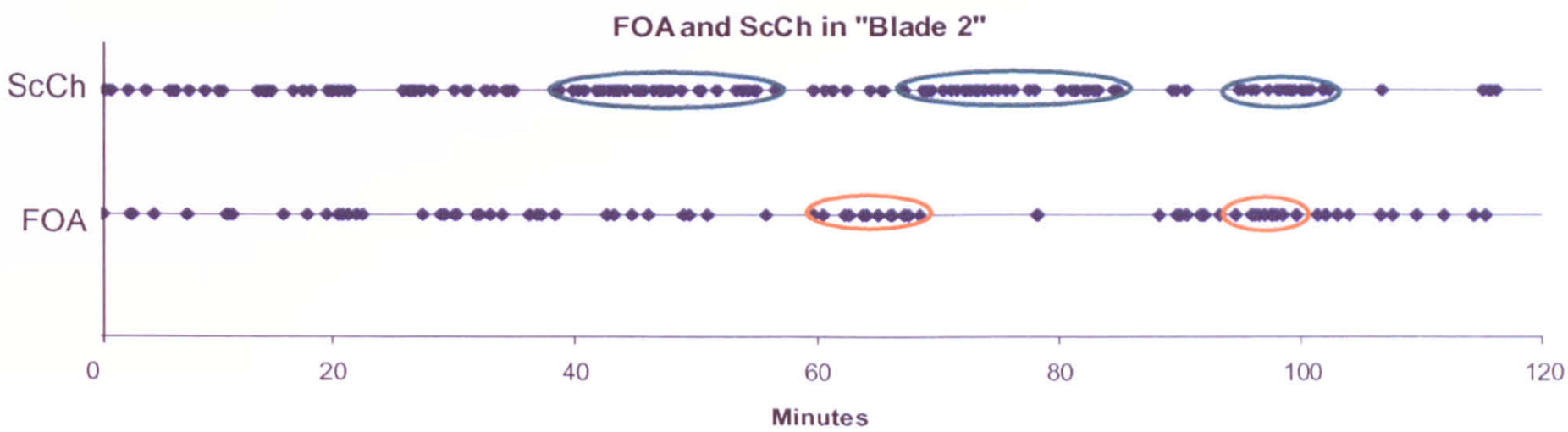


Figure 58 The instances of FOA and ScCh events in "Blade 2" [174] and circled indications of possible action events in the ScCh series and drama events in the FOA series.

4.4.7.6 Preliminary Investigation of Micro-Analysis for Gold Standard Scenes

The four events make it possible to analyse a film scene-by-scene. This may be useful as a way of detecting cause and effect relationships of events in a film. It also provides the opportunity to examine which characters are present in each scene, their relationships between scenes, any items that could potentially be of importance to the story in general or to a specific event, locations and when locations are revisited and by whom. As an example Figure 59 shows the events for an evaluated Gold Standard clip of the film "Oceans 11" [169], which was 11 m 32 s long, where the events were chosen by six evaluators. The events were only chosen if at least three people agreed on an event. 13 scenes are extracted, which have been labelled 's1-s13', and Table 35 shows the events that are contained in each of the 13 scenes. Table 35 represents the COL, FOA and NVC events that occur in the "Oceans 11" clip and makes it possible to see the movements of the characters involved easier as well as the interactions between the characters.

Figure 59 and Table 35 show us that using these statistics it is possible to examine character movements and behaviours visually and perhaps more importantly in a machine-processable manner. We are told what 'room' a character is in, in which scene of the film, at what time, with which other characters and, to some extent, what all those characters are doing and how they interact.

With these micro-level statistics, it is possible to envision new querying applications for each scene of a film that allow more in-depth navigation of characters' movements, locating important objects, knowing what character interactions are occurring at what time and much more.

Table 35 Gives a breakdown of what events occur in each of the 13 scenes in the "Oceans 11" [169] film clip.

		FOA	COL	NVC
S1	Hotel Lobby	Linus & Rusty look at Tess	Tess- Hotel> Lobby	N/A
S2	Fake Casino Vault	Daniel looks at Casino Chips Daniel looks at Rusty	Rusty- Outside>Fake Vault Rusty& Danny > Outside	N/A
S3	Bellagio Gallery	Tess looks at Painting Benedict looks at Painting Tess looks at Benedict Tess looks at CCTV	Benedict Hall - > Gallery Benedict - > Gallery	Benedict Nods
S4	Casino	Saul looks at Mint Saul looks at Bet Table Saul looks at Benedict	N/A	N/A
S5	Restaurant	Tess looks at Daniel Daniel looks at Waiter	Danny- Outside>Restaurant	N/A
S6	Casino	Saul looks at Benedict	N/A	Saul Shakes Head
S7	Restaurant	Daniel looks at Tess Daniel & Tess look at Benedict	Benedict - Casino>Restaurant	Daniel & Benedict Nod
S8	Lobby	Linus looks at Daniel	Daniel - Restaurant> Lobby	N/A
S9	Basher's Room	Basher looks at Stones	N/A	N/A
S10	Demolition	Linus looks at Daniel	Benedict - Crowd > Stage	N/A
S11	Basher's Room	Basher looks at TV	Basher - Room > Corridor	N/A
S12	Meeting Room	N/A	N/A	N/A
S13	Casino Vault	Group looks at Yen	Yen - Box > Fake Vault	N/A

The next section explores potential applications that become possible from what was suggested in Section 4.4.7.

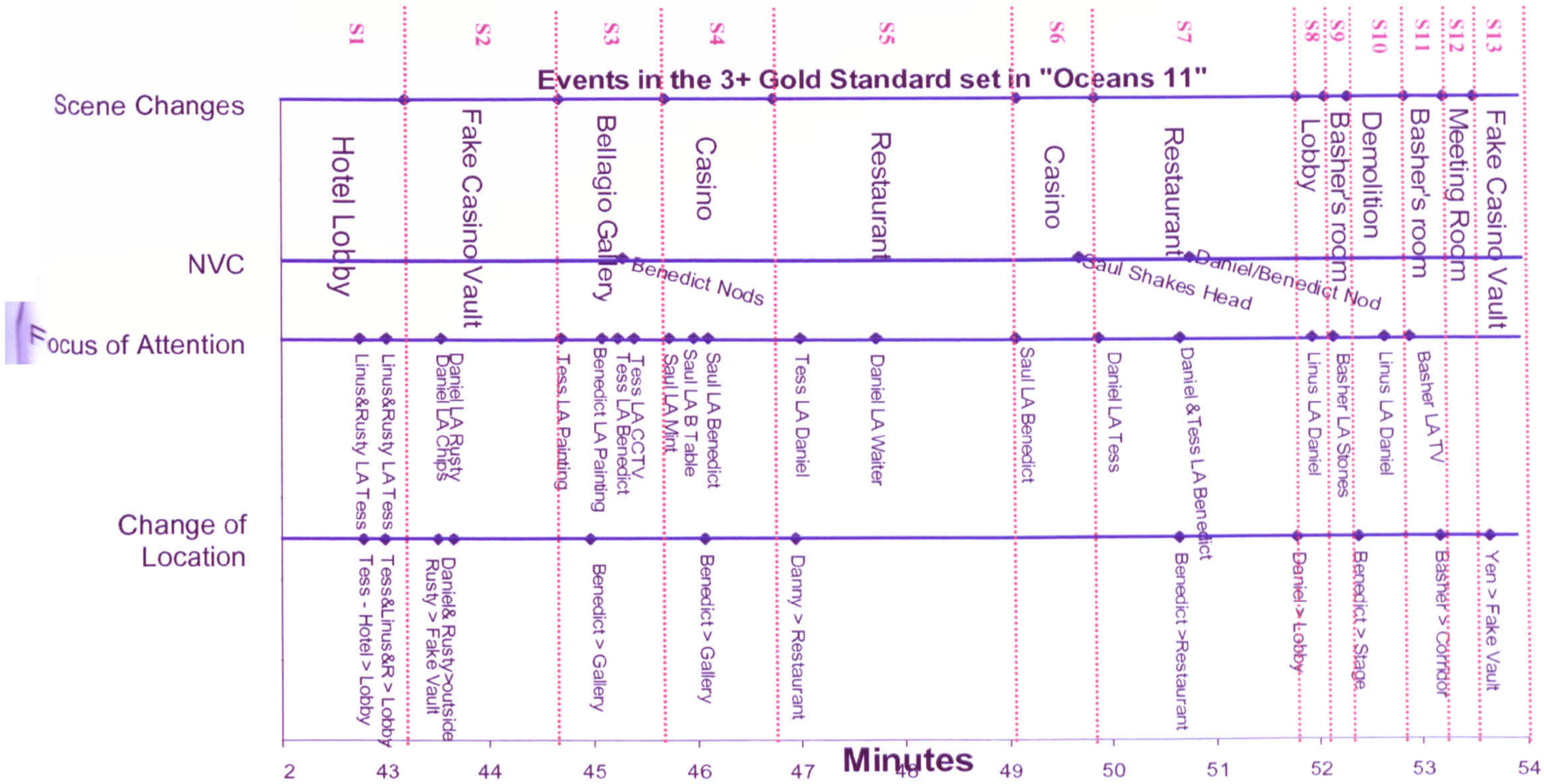


Figure 59 Depicts the Gold Standard instances of COL, FOA, NVC and ScCh events as extracted by six people (see evaluation) in the 11' 32'' clip of the film "Oceans 11" (from 42'30'' to 54 mins.) The clip contains 13 scenes which are marked as s1-s13 between orange lines. Can be found on accompanying **CD.**

4.4.8 Potential Applications from Film Event Statistics

The days of Star Trek's [167] automated, voice activated computer querying system may still be a while off but with the digitisation of more film and video data, the advent of new compression methods for video, faster Internet transmission of data and storage capacities for home PC hard disks growing each year, digital video data is becoming more easily accessible and manageable. The concept of a global media archive of digital video data is nearly upon us. However, what does remain to be solved is the semantic gap for video data that will allow a machine to 'understand' video data the same way a human does.

Bridging the semantic gap for film content may bring forth new applications for home users and film fans and will allow querying of film content at a whole new level. It will allow us to extend real-world applications: from home entertainment systems to querying video content on the Internet. It will allow users to form specific inter-scene and cross film queries for film and TV data such as: show me all the scenes that happen at sunset on a beach or all the scenes where Homer Simpson eats a doughnut. Therefore, with respect to this work, being able to analyse

aspects of film content automatically means, for the first time, we can envision new applications and new tools for researchers and other communities.

4.4.8.1 Accessing Film Data– Retrieval, Browsing, Navigation, Visualisation

Section 4.4.7 has shown possible statistical information about films that can automatically be extracted. Statistics such as when events are occurring and character information statistics: what characters focus on, characters' Non-verbal Communication, locations they are in, items they are holding or focussing attention on. Also, statistics showing information about important objects and items in films, frequently used items; information about cultural codes in films: period of film, styles of objects and period props; frequently visited locations in films and when by whom at what time of day. These statistics, and more, provide possible applications to allow us to reason about film content.

For instance, through examining the clustering of COL, FOA and ScCh events per *x* amount of time in a film we can recognise drama scenes and action scenes. Generally, when the clusters of FOA events are dense for a given time period of the film this possibly indicates a drama scenes and when there are many scene changes in close proximity this usually indicates an action scene. Another possible application of examining groups of events over time in a film is the possible comparison of films, based on different weightings of different events at different points of the film. This lends well to the idea of distinguishing genre. If we hypothesise that different genres of films have different 'weighting' of certain film events then it may be possible to distinguish genre based on these weightings for a given film. It is also possible through this work to recognise when characters are on-screen and with which other characters, and objects or items. This is important as we can 'trace' a character's movements throughout a film allowing us to distinguish certain actions he/she/it/they perform that may be integral to a film's plot progression. We can in effect map the cause and effect chains of events and character motives/goals/behaviours. It is also plausible to be able to automatically segment films in terms of scenes and events.

Through these novel applications that use the statistics it is possible to envision a set of real-world commercial applications. For instance, a page of statistical information could be added to any film database. A likely database that this could be applied to is the populate film database site www.imdb.com [136]. It would not be difficult to add a separate 'In-Film Statistics and Film Information' web page with statistics on the various film events and film statistics (i.e. how many scene changes, day scenes, night scenes, internal scenes, external scenes, film locations there are,

at which times, and where). Also the concept of frequency of change of events per given film time (similar to the 'pace' of the film), through rate of changes of scenes (quickly/slowly), could also be provided to indicate possible categories of scenes. We can also aid the online querying of Internet video search engines (Google [119], Yahoo [120] etc.). The analysis of the collateral texts to the videos on the web (in our case of course Hollywood films) using existing video transcripts, film scripts, closed-caption etc. can lead to much more in-depth queries being processed about the content of the videos and semantic inferences, such as character motives, emotions and similarity of moods of films. Previous work on emotion information [80], [102] in films may help in this.

It may also be possible for TV stations to decide which adverts are best for a film at what time in the film based on the event, character, emotional and possibly genre information that can be extracted. For example, a certain age group may be watching a certain film, e.g. teenagers watching an action movie, couples watching a romance, children watching an animation. Having prior knowledge of what type of content the film is showing at what time and what kind of audience is viewing the film, may allow personalised advertising tailored for a specific film. Applying the event information to advertising in this way may be of interest to TV rating research companies such as http://www.nielsenmedia.com [153] which deal with gathering statistics about viewers' viewing habits.

Knowing the content of a film may allow better parental controls to stop children watching adult or sensitive film events, e.g. instance in thrillers or horror movies. This sort of alert or scrambling could be incorporated into a DVD or digital TV package based on the events and emotions [80], [102] that can be automatically extracted from the relevant film scenes.

We believe that the method described in Chapters 3 and 4 can be applied to other domains of video data with equal success such as police video transcripts, court room transcripts and some sports. Most directly we can apply the method to Television series transcripts. It would be possible to create statistic pages about TV series from TV transcripts written in the same way as screenplays and audio descriptions, for TV websites such as www.TV.com [148]. It would be possible to gain information about TV series in general over a *whole series*: plot and story arc information, character screen presence, cause and effect chains etc.

An example is shown in Figure 60 for the scene changes over the entire first series of the popular TV series LOST [154]. This sort of application is aimed at TV series fanatics (of which I am one) which have a huge fan base and film students and general public.

4.4.8.2 Film Analysis Tool-benches for Film Scholars, Audio Describers and Screenwriters

From the collocation analysis in Chapter 3, and the fact that idiosyncrasies and regularities appear both in the language of the film scripts and in the formatting, it is possible to envision a series of tool-benches to aid film script writers. For audio describers it is possible to consider an automatic *Thesaurus* of words commonly used in AD scripts that can be available to audio describers when they are writing. Also, an *indicator* that would produce a visual representation of statistics of frequently used words in frequently used scene descriptions may be useful for training audio describers. Another training tool for new audio describers could be automatically finding scenes with non-verbal communication that have to be described and suggest words and formatting for given time slot of spoken text (e.g. getting the words down for a description to fit a three second audio pause in the film track). Statistics of 'word use' may also be useful in this context for training audio describers so as not to repeat the same phrases or words many times.

For Screenwriters a Thesaurus would be useful of common ways to describe what is going on, frequently used dialogue, and words that may 'fit' a specific time period or area of the world. Statistics of 'word use' may also be useful in this context also. Pre-emptive formatting tools for formatting screenplays with respect to dialogue, adding special effects, edit-effects and sound effects and describing what is going on in a scene (the preamble) is possible through this work.

The IE results given by the IE system are structured and this means that they are in a machine-processable form. This allows other tool-benches, such as automatic film annotation tools to be developed. However, such tool-benches have to be synchronised to the films' in question and to the annotation language e.g. MPEG-7, which may not be a trivial task.

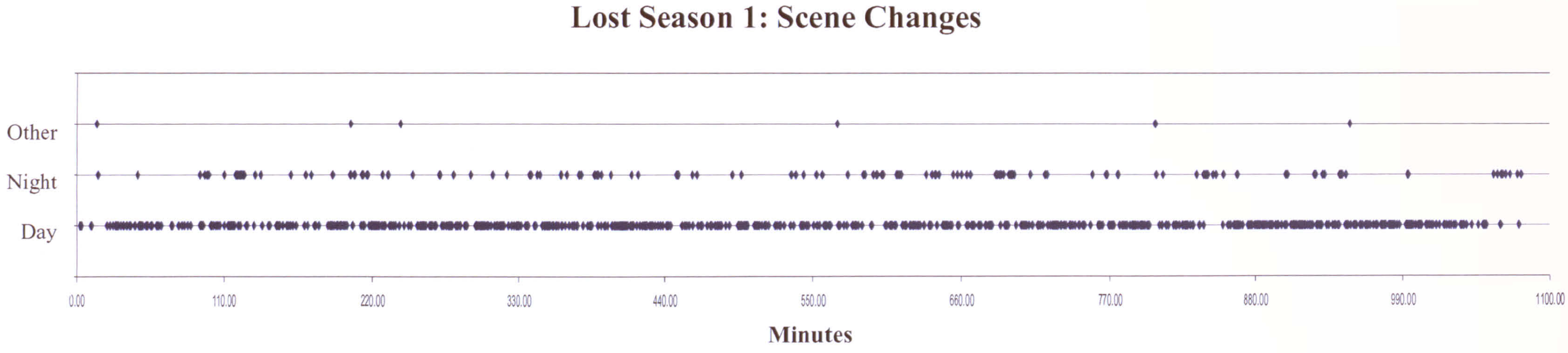


Figure 60 Instances of Scene Change events and what time of day it is (day or night) in the first series of the television mystery series LOST[154] (25 episodes). Can be seen on **CD**: Additional Thesis Material

4.5 Discussion

The aim of this chapter was to explore how the basic knowledge from Chapter 3 can be applied to support novel applications that rely on machine-processable representations of film content. Chapter 4 asked were the ways in which the information from Chapter 3 expressed with sufficiently regularity to allow for reliable information extraction, and even the automatic generation of IE templates and algorithms. Chapter 4 also enquired: what kinds of novel applications may be enabled by machine-processable representations of film content-based on the four events identified in Chapter 3. These questions were answered by expanding the templates in Section 3.4.1 into heuristics to extract four events: FOA, COL, NVC, ScCh and then implementing those heuristics into a text analysis system and applying them to our corpora. Our system was then evaluated using a set of Gold Standard event data gathered by 12 people watching five film clips (60 minutes of video in total) who identified the four events as they watched the film clips. Low recall values were seen for most of our events but there were instances of relatively high precision, > 70% for the ScCh event. Many suggestions on how to improve the precision and recall were made but they were not implemented due to time restraints on completing the PhD and not having the time to program or edit source code for certain natural language and grammar analysis programs (e.g. GATE [152]) which would need adapting for our purposes.

Based on evidence seen in this chapter, we claim that there are four types of event that are commonly described in audio description scripts and screenplays for Hollywood films: Focus of Attention (FOA), Change of Location (COL), Non-verbal Communication (NVC) and Scene Change events (ScCh) and that information about these events will support novel applications for automatic film content analysis. Work from Chapter 3 and this chapter contribute a resulting IE system which, with no further modifications, performed at around 50% precision and 35% recall. We also contribute a database containing information about four types of events in 193 films, extracted by the system automatically, and a Gold Standard data set of the events, for an hour of film scenes. As a consequence of the IE system data set, and as proof of concept, ideas for novel applications are outlined in Sections 4.4.7 and 4.4.8. These novel applications aim to cross the semantic gap for film content in new ways– attempting to get further across the semantic gap than before. This involves making inferences about the event information and its statistics, in terms of film content and narrative/film theory, to elicit as set of mid-level features that, when used alone or in combination, allow the accessing of film video data in new ways in terms of querying, browsing and visualising film content. These contributions we believe will be useful to the video

content analysis community and the novel applications will appeal and be relevant to film scholars, narratologists, audio describers, screenwriters and the video content community.

Overall, from the film scripts we have been able to automatically extract: common events and information about characters, objects, locations, character body parts and their movements, characters' location changes, character actions, characters' focus of attention, scene changes, the time of day it is for a scene and whether a scene is internal or external. This has led to ideas on applications and inferences about extracting film content considered for specific communities: audio describers, script writers, film scholars and film/television information websites, and general community in terms of video browsing/querying of video data. Even though we had low precision and recall and an estimation of screenplay time (i.e. no accurate time code or reference to work from) there was still enough precision to allow these ideas to be considered and explored.

The creation of Gold Standard data validates the proposal of the four event types in chapter 3. That is, the fact that 12 independent viewers with varied backgrounds, for the same five video clips were able to locate and extract instances of the film events, and that these would match to a set of automatically extracted data from *film script text* of the film clips, demonstrates that the four events are frequently found in films. This suggests that these events may be intuitive to film viewers. Not only are the four events frequently found in films but they may be essential film content. It must also be noted that sets of film content data were gathered automatically from 193 films. If these events were not present then such a database would not be possible. The evaluation with the Gold Standard data set also helps strengthen the argument the collocation phrases found in Chapter 3 are common to film scripts and film content in general and that an LSP and local grammars are present in film scripts. The FOA, COL and NVC events seem to describe character interactions, and the phrases used in them, seem to be repeated throughout all screenplays and audio description scripts. The precision statistics, although not high for all events, are further proof that the collocation phrases are frequently recurring in films and may be considered narrative 'building blocks' when considering writing and/or describing film. The events themselves can be considered mid-level analysis features towards bridging the semantic gap for film content. If we recall our model of Chatman's [18] overall picture of narrative (Chapter 1, Figure 9) the concept of *events* and *existents* making up a story in narrative theory is expressed. The four event templates allow us to extract both events (action and happenings) and existents (characters, objects and settings). We believe it may be possible to elaborate this work into a way to model narrative, specifically, the story structure of film content.

5 Contributions, Conclusions and Future Opportunities

The aim of this work was to bridge the semantic gap with respect to the analysis of film content using the novel approach of systematically exploiting collateral texts for films, such as audio description scripts and screenplays. We asked the questions: First, what information do these texts provide about film content and how do they express it? Second, how can machine-processable representations of film content be extracted? Third, how can these representations enable novel applications for analysing and accessing digital film data? To answer these questions three main stages of research were conducted: the analysis of collocations in corpora of audio description scripts and screenplays; the development and evaluation of an information extraction system based on corpus analysis results and the outlining of novel applications based on information extracted from audio description and screenplay scripts.

This chapter discusses the success and usefulness of the research by presenting claims and contributions our work makes available to the video data analysis, film scholar, narratology, screenwriter and audio describer communities (Section 5.1). Section 5.2 discusses how successful we have been in bridging the semantic gap for film content. In summary we believe this work presents the opportunity to bridge the semantic gap for film content in a new way: by exploiting collateral texts that describe film. Section 5.3 presents opportunities for future work. In the short term we wish to make implementation changes to the IE system to improve precision and recall and novel application development. Also, we envision new research that can stem from the results and methods of Chapter 3 and 4. We also visualise longer term research to develop blue sky applications utilising the research and results of this work.

5.1 Claims and Contributions

What follows are a set of claims and contributions that this work makes that may be of interest to the video data community, narratologists, screenwriters, audio describers and indicate real-world applications.

Claims:

1. *The language used in audio description and screenplays contains idiosyncratic, repeating word patterns, specifically an unusually high occurrence of certain open class words and certain collocations involving these words, compared to general language.* Our evidence for this claim comes from the collocation analysis in Chapter 3 where there was strong statistical evidence of repeating lexical patterns and the existence of highly frequent words not frequently used in general language in representative corpora of film scripts. The existence of these idiosyncrasies means that the generation of information extraction templates and algorithms can be mainly automatic. For our work the idiosyncrasies provide the basis of analysing film content. If we did not have these idiosyncrasies, we could not show that the information in the language of film scripts was expressed with sufficient regularity to allow for reliable information extraction, and the automatic generation of IE templates and algorithms. The regularities *are* present and hence it is possible to reliably extract film content information from film scripts.

Language in AD and SC corpora contains idiosyncrasies and regularities

2. *There are four types of event that are commonly described in audio description scripts and screenplays for Hollywood films: Change of Location (COL), Focus of Attention (FOA), Non-Verbal communication (NVC) and Scene Change (ScCh) events.* Evidence for this claim originates in Chapter 3 where templates for four event types were intuitively developed from frequent collocation phrases and possible local grammars. These event templates were validated in Chapter 4 by developing a film content IE system based on the templates and evaluating them with a set of Gold Standard data. It is interesting to note the Gold Standard data set was gathered using the four events as a guide. This claim is important for film video data analysis and may be useful to film scholars, narratologists and the video data analysis community. It is also useful in defining a 'set of things' in film and may further the concept of a film ontology.

Four types of Event commonly described in Hollywood Films

3. *Information about these events will support novel applications for automatic film content analysis.* The ideas for novel film content search applications and script writing tool-benches outlined in Sections 4.4.7 and 4.4.8 at the proof of concept level, are based on the results from the database of the four event types in 193 films. The statistics that can be automatically provided by the IE system, we believe make it possible for these novel applications to be developed and put to use in real world applications for new ways to access and process film content.

Information about Events support novel applications

Contributions:

1. *A set of templates with attributes and a formal description of its associated collocations in the form of a Finite State Automaton (FSA) for each of four events: (COL), (FOA), (NVC) and (ScCh).* These events are important for film video content analysis as they occur frequently in most Hollywood films and allow information about film content to be represented and extracted. The event templates present the opportunity to develop machine-processable representations of film and present extractable mid-level features (existents and events), which can be used to instantiate a narrative model such as Figure 9 or as seen in Section 1.2. This contribution coincides with the need for a film 'ontology' to define a set of things that exist in film and video data and may allow such an ontology to be developed as has been seen in work such as [35], [100]. This contribution may be of interest to film scholars as a way of *quantitative* film analysis and the video data analysis community as mid-level semantic units that are machine-processable in films.

Four Film Events

2. *The extension and testing of an existing, mainly-automated method to generate templates and algorithms for information extraction. The resulting system, with no further modifications, performed at around 50% precision and 35% recall.* This method is important for video content analysis because it enables the extraction of aspects of semantic content that cannot be extracted with current audio-visual feature analysis, see Chapter 2 discussions. The method may be applicable to other film text types, e.g. subtitles and film plot summaries. This may interest corpus linguists as it is an application of the techniques they use on new data: AD and SC corpora. The method may be of interest to many domains which use collateral texts (sports commentating, police video reports, court room proceedings) as the collocation analysis method can be applied to them with minor changes.

Method for Collocation Analysis of Corpora and IE

Database and Gold Standard

3. *A database containing information about four types of events in 193 films, which was extracted by the system automatically and a Gold Standard data set of the events, for an hour of film scenes.* The results of the database, makes feasible a set of applications for film video data accessing and processing not possible until now: applications to access aspects of film video data and for the production of AD scripts and screenplays. The database itself can be considered a shareable resource for other researchers to use and a resource for multimodal fusion of features to compliment audio-visual analysis of film. Seen on accompanying **CD**.

Framework

4. *A framework for analysing film content which synthesises elements of corpus linguistics, information extraction, narratology and film theory.* The framework enriches understanding for automatic film content analysis and provides insights that may be of interest to other domains: evidence of a language for special purpose, local grammars and an empirically-grounded analysis of film content for film scholars.

5.2 Summary and Critique

This research was successful in demonstrating a new approach to crossing the semantic gap for film content. Here, we consider the extent to which we have crossed the gap. In Section 1.2 we noted a number of frameworks to model film content and structure, and specify story structure. To instantiate these models requires extracting information about: characters' goals, intentions, beliefs and emotions; character and event cause-effect relationships; knowledge about objects, locations, character directions and moods; dialogue and action scenes; and film edit-effects. It is these kinds of frameworks, and the representation of film and story content, that we are working towards, as these representations of film/narrative are the kinds of things that can help the interpretation of film, to bridge the semantic gap. This work provides steps towards most of these model instantiation requirements. For instance the NVC, COL and FOA events provide information about characters' goals, intentions and what characters are doing and provide a way to *track* a character's direction, actions, appearance and disappearance off and on-screen. The COL and ScCh events provide information about what location a character is in and the FOA and ScCh events provide information about a character's objects and surroundings. Action and dramatic scenes can also be deduced in a film by examining the speed of change of FOA, COL and ScCh events. With some inferencing, our events are ways of instantiating Section 1.2's models of story structure and film and hence provide strong steps towards bridging the semantic gap. Our work has presented a novel way of bridging the semantic gap through the analysis of film scripts that allows machine-processable events in film to be mainly automatically extracted.

We consider the events and their respective attributes as extractable mid-level features of films and partial building blocks for films. Information about the four events extracted could be used to instantiate a model for story structure in film as can be seen in Figure 9.

There are still research steps that remain for us to cross the semantic gap for film content. We need to conduct a full investigation into inferencing mid-level, high-level and conceptual-level semantics features from the film events and the idiosyncrasies in the film script corpora. Such inferences into semantic features of film content are to allow a greater *coincidence* between machine and human viewer. There is also some value to be gained by exploring other film text types, such as subtitles, closed-caption text and plot summaries [97], as they can provide an even better description of what is happening in a film and can complement the AD scripts and screenplays. Integrating this work with audio-visual analysis of film content will provide a multimodal method for analysing film that will take advantage of audio, visual, textual elements of a film as well as collateral texts. We believe such a multimodal approach will go much further to bridging the semantic gap for film content.

Much research has tried to deal with bridging the semantic gap by processing audio-visual features, such as the work seen in Section 2.2. Few have used text analysis and multimodal fusion of audio/visual/textual features in film to try and bridge the semantic gap. We believe that using multimodal features is more effective in bridging the semantic gap, as video content is multimodal by nature. However, we believe even more strongly that although analysis of text is a mature research field, analysing texts that describe film, is very useful for understanding film content.[32]

In the different stages of our research there are a number of method and implementation steps that could be improved, and made automatic, and a number of manual choices (such as the choice of which collocation phrases to be included in an event type) that could have been made differently.

We believe that our extended, mainly-automatic collocation analysis method performs well and has given us some interesting results that have been utilised to perform somewhat reliable Information Extraction of film content. The method needs more automation however. Work such as [6] and [8] at the University of Surrey is providing steps to automate the method as much as possible. However, we believe that some human judgement is still necessary in such a method and that a complete automation of it may not be possible. "Since movies are the result of an artistic process sometimes causing confusion on purpose, the necessity for human judgement will always remain to some extent." [103]. The way in which the results are represented formally, through Finite State Automata, has proven manageable with respect to representing textual phrases and simplifying them, (see *Join,* Section 3.3.3). However, the FSA do present the problem of over-

[32] As in trying to find what objects and events are involved in video e.g. an ontology of video

generalisation of textual collocation phrases that may cause phrase pathways to be created that do not exist. There may be a better, more robust, way to formally represent the collocation data than FSAs such as a specific phrase formalism resembling the one developed in [102] for emotions. In this work however, FSA served our purposes.

The corpus linguistic method outlined in this work, and its relevant collocations statistics and term frequencies, helped provide evidence of a language for special purpose (LSP) in film scripts used by screenwriters and audio describers. It also provides evidence of Local Grammars (LGs) in films and, in some cases, these LGs can be extended or joined as can be seen in Appendixes B and C. Evidence of a sub-language as defined by Harris as seen in [12] and [44] is also present, strengthening the argument that an LSP is present and LGs exist in the film scripts.

The choice of the four events and their attributes, although manual, were grounded in some intuitive categorisations. Elements seemed to group intuitively, such as in the case of the FOA event there were three local grammars (or FSA) that dealt with characters focussing their attention on items and other characters: 'looks at', 'turns to' and 'takes'. A more systematic way of developing the event categories would have been less subjective but in the absence of such a method we believe the events to be objective due to their statistical derivation. The implementation of heuristics and algorithms into an IE system produced reasonable results in terms of precision and recall and evaluated the events along with their high frequency in films. Also the fact that 12 independent people where able to pick such events out of an hour of feature film clips, with sufficient regularity, is further evidence of the existence of the events in films.

The IE system itself did not perform well, in terms of precision and recall overall, but individually the precision and recall statistics were better. For instance the ScCh event IE system implementation performed with ~70% precision and ~55% recall. The precision and recall statistics can be increased by including certain generic NLP changes, some inferencing and more heuristic rules. However, our goal was to see how objective we could be in the implementation of just the four event templates and attributes. The system still performed adequately with no changes to the templates and heuristics. The Gold Standard data may also be at fault as, even though it was collected by 12 people from different backgrounds, results did vary quite a bit. It may also have been useful examining five *whole* Hollywood films instead of five 12 minutes clips.

We believe that the statistics gathered from the IE system provided a large database of mostly reliable data and provided more evidence for the existence and regularity of the four events in Hollywood film scripts, even though the IE system had low precision and recall. The statistics gathered, once examined, provided ideas for novel applications. The combination of certain statistics and some inferencing on our behalf allowed us to speculate about real-world applications

of the IE system database results. Real-world applications such as statistics information pages on film and TV information websites (www.imdb.com [136] and www.tv.com [148]) and advances video search engine queries (Google [119], Yahoo [120], www.youtube.com [149] etc.). We also envision tool-benches for screenwriters, audio describers and film students from this work. We also believe that this method can be applied to other domains in the same way, domains such as court room video archiving, archives of CCTV footage that are transcribed, sports commentaries and documentaries.

5.3 Opportunities for Future Work

The work has served to cross the semantic gap for film content. However, there is still work to be done to allow a computer system to attain the level of interpretation a human has for film content. Assuming the semantic gap is bridged; this presents opportunities for blue sky systems that have not been available thus far. This section examines future work in terms of short term work to increase precision and recall, long term research that can follow on from this work and blue sky applications that can be developed based on our work.

5.3.1 Near-Future Work: Next Steps

In the short term we will endeavour to improve precision and recall by implementing new heuristics; improving existing ones by adding new nucleate word and synonyms, deal with formatting issues of scripts better, make pass rules in the NLP more strict etc. and investigating the generic NLP problems, such as pronoun resolution, through other research. We will try and find ways of manually improving the four event templates to allow us to extract more pertinent information about films than what is just available in the collocation phrases. We will also try to strictly define the local grammars that exist in the two corpora and attempt to join them where appropriate to allow for a 'local grammar for film scripts' to be defined. This may help the heuristics become 'stricter' as the LGs will be strictly defined to remove over-generalisation of phrases. There is also some value to be gained by exploring other film text types, such as subtitles, closed-caption text and plot summaries [97], as they can provide an even better description of what is happening in a film and can complement the AD scripts and screenplays.

There are still research steps that remain to be done. As mentioned in Section 5.2, we wish to conduct a full investigation into inferencing mid-level, high-level and conceptual-level semantics features from the film events and the idiosyncrasies in the film script corpora. This is to allow a greater *coincidence* between machine and human viewer. Integrating our work with audio-visual analysis of film content can provide a multimodal method for analysing film that will take

advantage of audio, visual, textual elements of a film as well as collateral texts. We believe such a multimodal approach will go much further to bridging the semantic gap for film content.

Currently, two Masters level students at the department of computing, Videsh Lingabavan and Matthew Knight, are working from the results of the database outlined in Sections 4.3 and 4.4. We envision research at PhD level for this research, however, specifically, in the improvement of the method in Chapter 3 and the generation of more heuristics in Chapter 4. Also PhD research may wish to continue the work started here in terms of examining novel applications that can be developed from a database of film content information. More specifically, we would like to see research into analysing other collateral texts that describe film, e.g. subtitles, closed-caption texts that can complement this research in terms of bridging the semantic gap for film content. We can also envision research linking our work and audio visual analysis work to allow for multimodal fusion of film content features.

We can see merit in using this work to add new functions to existing screen writing tool-benches such as "Final Draft" [150]. Final Draft is a piece of word processing software specifically designed for writing movie scripts. Our work could be integrated into such a package to allow statistics of words and events in a script to be available to a writer, i.e. a writer could find out how many FOA, NVC, COL and ScCh they have written and keep track of which locations, characters and objects are written about and where. Also, formatting elements that have been extracted from the IE system may be used in a pre-emptive formatting text capacity such as on mobile phone text messaging.

5.3.2 A Vision of Future Systems that Access Film Content

We believe that the work started here can be considered a 'stepping stone' to a greater level of research that will close the semantic gap as much as possible. Thinking along a much longer timescale we asked ourselves where we would like this work to lead in 20-30 years. We envision new ways of expressing the events we have extracted, once extracted. We envision automatic summaries of films, TV series, court room footage and video footage of sports commentaries, much like work in [49], stemming from this work. We also envision new ways to access video data at a user specific level.

The semantic gap does not allow for computer systems to *understand* film content the way a human does. Closing that gap, as we have attempted to do, opens up the possibility to allow new ways of navigation of films, specific to a user's request. I.e. being able to go to any part of a film specified and to find anything or any character specified. It allows us to examine behaviours, moods and goals of characters in a film. In short we can search for *anything*. Such a system that can allow us to reach that specific a search will be invaluable to any researchers in film, news,

court room footage or sports video, saving hours of manual labour searching for a specific shot or scene in a segment of video.

Since we believe that narrative or story structure in film can be represented as machine-processable 'units' we think that our method can be applied to any text with a story to elicit regular, frequent events, existents, plot information, e.g. newspaper articles, novels, comic books. Once these units have been extracted they can be used a narrative building blocks and possibly capture the essence of stories. With the advent of holographic 3-dimentional technology on the way it is not inconceivable to consider this a new medium for the transmission of TV and a new format for cinema. Our method of extracting events from existing media means that we could automatically produce representations of story structure that can be translated directly into a 3-d representation of the events in question. The Star Trek: The Next Generation [167] 'Holodeck' idea may be attainable, with a little help from us.

Although novel applications have been described in this work, they are the tip of the iceberg and science fiction in terms of querying video data and navigating films, may, with a lot of work, become science fact.

References

[1] Adams, B.; Dorai, C.; Venkatesh, S.; Automated film rhythm extraction for scene analysis. *IEEE International Conference on Multimedia and Expo*, 22-25 Aug. 2001, pgs 849 –852.

[2] Adams B., Dorai C., and Venkatesh S., Towards Automatic Extraction of Expressive Elements for Motion Pictures: Tempo, *IEEE Trans. Multimedia* Vol.4, no. 4, 2002, pgs 472-481,

[3] Adams B, Dorai C, Venkatesh S, Bui HH. Indexing Narrative Structure and Semantics in Motion Pictures with a Probabilistic Framework, *In Proceedings of IEEE International conference on Multimedia and Expo, 2003. ICME'03,* 2003, pgs 453-456.

[4] Ahmad K., Gillam L., Tostevin L., University of Surrey Participation in TREC 8: Weirdness indexing for Logical Document Extrapolation and Retrieval (WILDER), *NIST Special Publications*, 2000

[5] Ahmad, K. & M. Rogers, Corpus Linguistics and Terminology Extraction. *In S.-E. Wright and G. Budin (eds.), Handbook of Terminology Management (Vol. 2).* Amsterdam & Philadelphia: John Benjamins, 2001

[6] Ahmad, K., Gillam, L., and Cheng, D. Textual and Quantitative Analysis: Towards a new, e-mediated Social Science, *Proc. of the 1st International Conference on e-Social Science*, Manchester, June 2005

[7] Allen, R.B. & Acheson, J., Browsing the structure of multimedia stories, *ACM Digital Libraries*, June 2000, pgs 11-18.

[8] Almas Y. and Ahmad K., LoLo: A System based on Terminology for Multilingual Extraction, *Proc. of COLING/ACL'06 Workshop on Information Extraction, Beyond a Document*, Sydney, Australia. 2006

[9] Aner-Wolf, A., Determining a scene's atmosphere by film grammar rules. In *Proceedings IEEE International Conference on Multimedia & Expo*, 2003 , pgs.365-368

[10] Antani R., Kasturi R. & Jain R., A survey on the use of pattern recognition methods for abstraction, indexing and retrieval of images and video, *Pattern Recognition*, Vol. 35, No. 4, 2002, pgs 945-965

[11] Bao J., Cao Y., Tavanapong W., and Honavar V. Integration of Domain- Specific and Domain-Independent Ontologies for Colonoscopy Video Database Annotation. In IKE, 2004, pgs 82–90

[12] Barnbrook G. *Defining Language: A local grammar of definition sentences.* Amsterdam: John Benjamins Publishers, 2002

[13] Biber D., Conrad S. and Reppen R., *Corpus Linguistics, Investigating Language Structure and Use*, 1998, Cambridge University Press

[14] Bordwell D., *Narration in the Fiction Film*, University of Wisconsin Press, 1985.

[15] Bordwell, D. & Thomson, K. *Film Art: An Introduction*, McGraw-Hill, 5th edition, New York, 1997.

[16] Chan CH, Jones GJF, Affect-Based Indexing and Retrieval of Films, *Proceedings of the 13th annual ACM Multimedia international conference 2005* pgs 427-430 Nov 6-11 2005.

[17] Chang, S.-F., The Holy Grail of Content-Based Media Analysis, *IEEE Multimedia*, vol. 9, no. 2, Apr-Jun, 2002, pgs 6-10.

[18] Chatman, S., *Story and discourse: narrative structure in fiction and film*, Ithaca: Cornell University Press, 1978.

[19] Chen Hsuan-Wei, Kuo Jin-Hau, Chu Wei-Ta and Wu Ja-Ling. Action movies segmentation and summarization based on tempo analysis. In *Proceedings of the 6th ACM SIGMM international workshop on Multimedia information retrieval*, 2004, pgs 251–258

[20] Church, K., Gale, W., Hanks, P., and Hindle, D., Using Statistics in Lexical Analysis *In Zernik U., (ed), Lexical Acquisition: Exploiting On-Line Resources to Build a Lexicon*, Hillsdale, NJ: Lawrence Erlbaum, 1991, pgs. 115-164.

[21] Church, K., Gale, W., Hanks, P., and Hindle, D., Using Statistics in Lexical Analysis, 1991, Bell Laboratories and Oxford university Press. Found at http://www.patrickhanks.com/papers/usingStats.pdf last accessed 06/09/05.

[22] Corridoni J., M., Bimbo, A., D., Lucarella D. & Wenxue H., Multi-Perspective Navigation of Movies, *Journal of Visual languages and Computing,* Volume 7, 1996, pgs 445-466.

[23] Coste D., *Narrative as communication*, University of Minnesota press, 1989.

[24] Cowie J. Wilks. Y., *Information Extraction*, In R. Dale, H. Moisl and H. Somers (eds.) *Handbook of Natural Language Processing*. New York: Marcel Dekker, 2000

[25] Christel Michael G., Conescu Ronald M, Addressing the challenge of visual information access from digital image and video libraries. *JCDL.* 2005, pgs 69-78

[26] Crystal D., *A Dictionary of Linguistics and Phonetics: 4th Edition*, Blackwell Publishers, 1997

[27] Dasiopoulou S., V. K. Papastathis, V. Mezaris, I. Kompatsiaris, M. G. Strintzis, An Ontology Framework For Knowledge-Assisted Semantic Video Analysis and Annotation, *in Proceedings of the ISWC 2004 Workshop "SemAnnot"*, Hiroshima, 2004

[28] Dimitrova N., Zhang H-J., Shahraray B., Sezan I., Huang T., Zakhor A., Applications of Video-Content Analysis and Retrieval, *IEEE Multimedia*, vol.9, no..3, July 2002, pgs 42-55

[29] Emmott, C., *Narrative Comprehension: A Discourse Perspective,* Clarendon Press: Oxford, 1997

[30] Fargues, J., Landau M.., Dugourd, A., Catach, L., "Conceptual Graphs for Semantics and Knowledge Processing", *IBM Journal of Research and Development,* Vol. 30, No.1, Jan 1986, pgs 70-79

[31] FINGRID: Financial Information Grid, Research Report, Reference: RES-149-25-0028, University of Surrey, pg.23-24, 2004. Ahmad K, University of Surrey and Nankervis J., University of Essex http://www.esrc.ac.uk/ESRCInfoCentre/

[32] Fischer S., Lienhart, R., and Effelsberg, W., Automatic recognition of film genres, *International Multimedia Conference, Proceedings of the third ACM international conference on Multimedia,* San Francisco, California, United States, 1995, pgs 295–304

[33] Foresti, G.L., Marcenaro, L. & Regazzoni, C.S., Automatic detection and indexing of video-event shots for surveillance applications, *IEEE Transactions on Multimedia*, Vol. 4, No. 4, Dec 2002, pgs 459-471

[34] Gaizauskas R. and Wilks Y., Information Extraction: Beyond Document Retrieval, *Journal of Documentation*, 54(1):70{105, 1998.

[35] Geleijnse G., A Case Study on Information Extraction from the Internet: Populating a Movie Ontology, 2004 http://semanticsarchive.net/Archive/Tc0ZTg4N/Geleijnse-OntologyPopulation.pdf , accessed 11/07/06

[36] Genette, G, *Narrative Discourse: An essay on Method (J.E. Lewin Trans)* Ithaca, NY: Cornell University Press, 1980.

[37] Gross, M., *Local grammar and their representation by finite automata.* In M. Hocy, (editor), *Data, Description Discourse*, pgs 26-28. HarperCollins London, 1993.

[38] Gross, M. *The Construction of Local Grammars*. In Finite-State Language Processing, E. Roche & Y. Schabès (eds.), Language, Speech, and Communication, Cambridge, M MIT Press, 1997, pgs 329-354.

[39] Gross, M., A bootstrap method for constructing local grammars. In *Contemporary Mathematics. Proceedings of the Symposium, 18-20 December 1998, Belgrade, Serbia*, 1999, N. Bokan (ed.), University of Belgrade, pgs 229-250.

[40] Hanjalic A., Lagendijk L. & Biemond J., Automated High-Level Movie Segmentation for Advanced Video-Retrieval Systems, *IEEE Transactions on Circuits and systems for Video Technology*, Vol. 9, no 4, Jun 1999, pgs 580-588.

[41] Hanjalic A., Adaptive extraction of highlights from a sport video based on excitement modelling, *IEEE Transactions on Multimedia*, Dec 2005, Vol. 7, Issue 6, pgs 1114-1122.

[42] Hanjalic A. and Xu L-Q, Affective video content representation and modelling" *IEEE Trans. On Multimedia,* Vol. 7, no. 1, Feb 2005, 143–154,

[43] Hanjalic A., Extracting Moods from Pictures and Sounds: Towards Truly Personalised TV. *IEEE Signal Processing Magazine*, Mar 2006, pgs 90-100

[44] Harris, Z., *A Theory of Language and Information: A Mathematical Approach*, Oxford University Press, 1991.

[45] Herman D., *Story Logic: Problems and Possibilities of Narrative, (Frontiers of Narrative series)*, University of Nebraska press, 2002.

[46] Hobbs J. R., Appelt D.s, Israel D, Bear D., Kameyama M, Stickel M., and Tyson M. *Fastus: A cascaded Finite- state transducer for extracting information from natural-language text. In E. Roche and Y. Schabes, editors, Finite State Devices for Natural Language Processing.* MIT Press, 1996.

[47] Hopcroft, J. E., Motwani R. and Ullman J, D., 2000, *Introduction to Automata Theory, Languages, and Computation (2nd Edition)* Pearson-Addisson-Wesley. Original 1979.

[48] Hunston, S. and Sinclair, J. *A Local Grammar of Evaluation.* In Evaluation in Text: Authorial Stance and the Construction of Discourse, Hunston, S. & Thompson, G. (eds.), Oxford, Oxford University Press: 2000, pp. 75-100.

[49] Hwan Oh, J., Wen Q., Hwang S. and Lee J., *Video Abstraction*, Chapter XIV, in: *Video Data Management and Information Retrieval,* Deb S., Idea Group Inc. 2005 Pg 321-346

[50] Jaimes A, Tseng B., and Smith J., Modal keywords, Ontologies, and Reasoning for Video Understanding. *In Conference on Image and Video Retrieval*, Urbana, IL, July 2003

[51] Jain R., Antani S. and Kasturi R., A survey on the use of pattern recognition methods for abstraction, indexing and retrieval of images and video, *Pattern Recognition*, vol. 35, no. 4, 2002, pgs 945-965.

[52] Johnson D., Video Browsing and Retrieval Using Audio Description, Final Year Project CIT UNIS, 2006

[53] Jung B., Kwak T., Song J. and Lee Y., Narrative Abstraction Model for Story-oriented Video, *In Proceedings of ACM Multimedia 2004,*Oct 10-16 2004, pgs 828-835.

[54] Kang, H, Affective content detection using HMMs, *In procs. of the 11th ACM International Conference on Multimedia*, ACM press, Berkeley, November 4-6, 2003, pgs 259-262.

[55] Lacey, N., *Narrative and Genre: Key Concepts in Media Studies*, Macmillan 2000.

[56] Lavelli A., Califf M. E., Ciravegna F., Freitagz D., Giuliano C., Kushmericky N., Romano L. A Critical Survey of the Methodology for IE Evaluation, *4th International Conference on Language Resources and Evaluation,* 2004

[57] Leech G., Rayson, P. and Wilson, A., *Word Frequencies in Written and Spoken English: based on the British National Corpus*, Longman, 2001

[58] Lehnert, W, G, Plot Units and Narrative Summarisation, *Cognitive Science*, Vol. 4, 1981, pgs 293-331.

[59] Manjunath B. S., Salembier P & Sikora T. (Eds.), *Introduction to MPEG-7: Multimedia Content Description Language*, 2002, John Wiley & Sons

[60] Manning C, D, & Schütze H, *Foundations of Statistical Natural language Processing*, 1999, MIT Press

[61] Mason, O., Automatic Processing of Local Grammar Patterns, *In Proceedings of the 7th Annual CLUK (the UK specialist group for computational linguistics) Research Colloquium*, 2004.

[62] Medioni G., Cohen I., Bremond F., Hongeng S., Nevatia R., Event detection and analysis from video streams, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, , Vol: 23, Issue: 8, Aug 2001, pgs 873-889

[63] Metz, C., *Film Language: A semiotics of the Cinema.* New York: Oxford University Press. 1974a.

[64] Metz, C., *Language and Cinema.* The Hague: Mouton, 1974b.

[65] Mohri M., Local Grammar Algorithms, *IN: Inquiries into words constraints and Contexts*. Antti Arppe et al. (eds.), 2005 pg 84-93

[66] Nack F and Parkes A., AUTEUR: The Creation of Humorous Scenes Using Automated Video Editing, *AAi/AI-ED Technical Report No.113, in IJCAI-95, Workshop on AI and Entertainment and AI/Alife*, Montreal, Canada, Aug 21 1995

[67] Nack F. and Parkes A., The Application of Video Semantics and Theme Representation in Automated Video Editing, *Multimedia Tools and Applications*, Vol. 4, No. 1, January 1997, pgs 57-83.

[68] Nevatia R., Hobbs J., and Bolles B., An ontology for video event representation, *In Proc. of IEEE International Conference on Computer Vision and Pattern Recognition*, volume 7, 2004, pgs 119–128

[69] Ortony A, Clore G L and Collins A., *The Cognitive structure of emotions*, Cambridge Univ. Press, 1988

[70] Poibeau T. A corpus-based approach to Information Extraction, *In Journal of Applied System Studies*, vol. 2, no. 2, 2000.

[71] Poibeau, T., Ballvet Corpus-based lexical acquisition for Information Extraction, *Actes du workshop Adaptive Text Extraction and Mining (ATEM), 17th International Joint Conference on Artificial Intelligence (IJCAI'2001)*, Seattle, 2001

[72] Rasheed Z. and Shah M., Scene Detection in Hollywood Movies and TV Shows. *In Proc. of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. (CVPR'03)*, 2003, pgs 1-6,

[73] Rasheed Z., Sheikh Y., and Shah M., On the use of computable features for film classification, *IEEE Trans. Circuits and Systems for Video Technology*, vol. 15, No.1, Jan 2005, pgs 52- 64

[74] Rodrıguez A., Guil N., Shotton D, M. and Trelles O., Analysis and Description of the Semantic Content of Cell Biological Videos, *Multimedia Tools and Applications*, Vol. 25, No. 1, January 2005, pgs 37-58

[75] Ronfard R., Tran-Thuong T., A framework for aligning and indexing movies with their script, *IEEE International Conference on Multimedia & Expo 2003*, Baltimore, Maryland, July 2003, pgs 6-9

[76] Roth V., Content-based retrieval from digital video, Image and Vision Computing, *special issue on Content-Based Image Indexing and Retrieval*, Vol. 17, no.7, May 1999, pgs 531-540.

[77] Salway and Graham, Extracting Information about Emotions in Films. *Procs. 11th ACM Conference on Multimedia 2003*, 4th-6th Nov. 2003, pgs 299-302. ISBN 1-58113-722-2

[78] Salway A. and Frehen C., Words for Pictures: analysing a corpus of art texts, *In Proc. Of International Conference for Terminology and Knowledge Management TKE*, 2002

[79] Salway A. and Tomadaki E Temporal information in collateral texts for indexing moving images, *in Proceedings of LREC 2002 Workshop on Annotation Standards for Temporal Information in Natural Language*, 2002.

[80] Salway A, Graham M., Tomadaki E and Xu Y., Linking Video and Text via Representations of Narrative, *AAAI Spring Symposium on Intelligent Multimedia Knowledge Management*, Palo Alto, 24-26 March 2003.

[81] Salway A and Tomadaki E, Temporal Information in Collateral Texts for Indexing Moving Images, *Proceedings of LREC 2002 Workshop on Annotation Standards for Temporal Information in Natural Language*, Eds. A. Setzer and R. Gaizauskas, 2002, pp. 36-43.

[82] Salway A. and Ahmad K, Multimedia systems and semiotics: collateral texts for video annotation, *Procs. IEE Colloquium Digest, Multimedia Databases and MPEG-7*, 1998, pp.7/1-7/7.

[83] Salway A., Vassiliou A, and Khurshid A., What Happens in Films?, *In proceedings of IEEE International Conference on Multimedia and Expo.*, 06-06 July 2005, pp. 49-52

[84] Schank, R., C., *Tell me a story: A new look at real and artificial Memory,* Scribner, New York, 1990

[85] Schank, R., C. & Abelson, R. P. *Scripts, Plans, Goals and Understanding: An Inquiry Into Human Knowledge Structures*, Lawrence Erlbaum Associates. 1977.

[86] Shirahama K., Iwamoto K., and Uehara K., Video Data Mining: Rhythms in a Movie, *Procs. IEEE Int. Conf. Multimedia and Expo*, ICME 2004, pgs 1463- 1466.

[87] Sinclair, J., *Corpus, Concordance, Collocation*, London: Oxford University Press, 1991.

[88] Smadja, F., *Retrieving Collocations from Text: Xtract,* In Armstrong, S. (Editor), *Using Large Corpora,* London: MIT Press, 1994.

[89] Smeulders, A.W.M., Worring M., Santini, S. Gupta, A. Jain, R. Content-based image retrieval at the end of the early years, *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 22, Issue 12., 2000, pgs 1349-1380

[90] Snoek C. G. M, *Worring* M, Multimodal Video Indexing: A Review of the State-of-the-art, *Multimedia Tools and Applications*, Vol. 25, Issue 1, Jan 2005,pPgs 5-35

[91] Sowa, J. F., *Conceptual Structures: Information Processing in Mind and Machine*, Addison-Wesley, Reading, MA 1984

[92] Sowa, J. F., Relating Templates to Language and Logic, in *Information Extraction: Towards Scalable, Adaptable Systems*, ed. by M. T. Pazienza, LNAI #1714, Springer-Verlag, 1999, pgs 76-94.

[93] Srihari, R. K., and Zhang, Z., Exploiting Multimedia Context in Image Retrieval, *Library Trends*, Dec 1998, 48(2), pgs 496-520.

[94] Sundaram H., Chang S.F., *Computable Scenes and structures in Films,* IEEE Trans. on Multimedia, Vol. 4, No. 2, June 2002. Pgs 482- 491.

[95] Tavanapong W. and. Zhou J.Y, Shot Clustering Techniques for Story Browsing, *IEEE Transactions on Multimedia*, Vol. 6., No. 4, 2004, pgs 517-527

[96] TIWO, Television In Words project, Grant Number GR/R67194/01 http://www.computing.surrey.ac.uk/ckm/tiwo_project, last accessed 23/02/04

[97] Tomadaki E. PhD Thesis, *Cross-Document Coreference between Different types of Collateral Texts for films*, University of Surrey, 2006.

[98] Traboulsi, H., Cheng, D. & Ahmad, K., Text Corpora, Local Grammars and Prediction, *In Proceedings of LREC*, Lisbon, 26-28 May, 2004, pgs 749-752.

[99] Truong B.T. and Venkatesh S. Determining dramatic intensification via flashing lights in movies. *In procs. IEEE International Conference on Multimedia & Expo*, Tokyo, Japan, 2001, pgs 61–64.

[100] Tuffield M., Millard D. and Shalbolt N, Ontological Approaches to Modelling narrative, *2nd AKT Doctoral Symposium, Aberdeen University*, UK, 25 January 2006

[101] Turetsky R. and Dimitrova N., Screenplay Alignment for Closed-System Speaker Identification and Analysis of Feature Films, *In Proceedings IEEE ICME 2004*, pgs 1659- 1662.

[102] Vassiliou A., Salway A. and Pitt D., Formalising Stories: sequences of events and state changes, *In Proceedings IEEE Conference on Multimedia and Expo, ICME 2004*, pgs 587- 590.

[103] Vendrig J. & Worring M., Systematic Evaluation of Logical Story Unit Segmentation, *IEEE Transactions on Multimedia*, Vol. 4, No. 4, 2002, pgs 492- 499.

[104] Wei C-Y., Dimitrova N., and Chang S.-F., Color-Mood Analysis of Films Based on Syntactic and Psychological Models, *Procs. IEEE Int. Conf. Multimedia and Expo,* ICME 2004, Vol 2, pgs 831- 834

[105] Wilks Y., *Information Extraction as a core language technology*, In Pazienza M-T. (ed.), Information Extraction, Springer, Berlin. 1997

[106] Xu M., Chia L-T. & Jin J., Affective Content Analysis in Comedy and Horror Videos by Audio Emotional Event Detection, *Procs. IEEE Int. Conf. Multimedia and Expo,* ICME 2005, pgs 622- 625

[107] Xu Y., PhD Thesis, *Representation of Story Structures for Browsing Digital Video*, University of Surrey, 2006

[108] Zhai, Rasheed Z, Shah M, Conversation Detection in Feature Films Using Finite State Machines, *IEEE Proceedings of the Pattern Recognition, 17th International Conference on Pattern Recognition (ICPR'04)*, 2004, pgs 458-461

[109] Zhai Y, Rasheed Z, Shah M., Semantic classification of movie scenes using finite state machines, *IEE Proc.-Vis. Image Signal Process, 2005*, pgs 896-901

[110] Zhang, D. and Nunamaker, J. A Natural Language Approach to Content-Based Video Indexing and Retrieval for Interactive E-Learning. *IEEE Transactions on Multimedia,* Vol. 6, No. 3, June 2004, pgs 450-458

[111] Zhu X., Wu A. E. X, Feng A., and Wu L., Video Data Mining: Semantic Indexing and Event Detection from the Association Perspective. *IEEE Transactions on Knowledge and Data Engineering*, Vol. 17, No. 5, May 2005, pgs 665-677.

Websites

[112] Dictionary.com, Online Dictionary, dictionary.reference.com/search?q=movie, accessed 16/05/06

[113] Wordnet, Online thesaurus project, http://wordnet.princeton.edu/perl/webwn, accessed 16/05/06

[114] Wikipedia, Online Encyclopaedia, en.wikipedia.org/wiki/Audio_description, accessed 16/05/06

[115] Oxford English Dictionary, http://www.oed.com/, accessed 10/10/04

[116] Wikipedia, Online Encyclopaedia, en.wikipedia.org/wiki/Screenplay, accessed 16/05/06.

[117] Trecvid; TREC Video Retrieval Evaluation, www-nlpir.nist.gov/projects/trecvid/, accessed 08/05/06

[118] MPEG-7 overview, www.chiariglione.org/MPEG/standards/mpeg-7/mpeg-7.htm, accessed 08/05/06

[119] Google Video search Engine, http://video.google.com/, accessed 09/05/06

[120] Yahoo Video search Engine, http://video.search.yahoo.com/, accessed 09/05/06

[121] Blinkx Video Search Engine, http://www.blinkx.com/ & http://www.blinkx.tv/, accessed 09/05/06

[122] Virage Video Search System, www.virage.com/content/pathways/video_search_iptv/, accessed 09/05/06

[123] VideoQ, Object Oriented Video Search, http://persia.ee.columbia.edu:8080/, accessed 09/05/06, Columbia University

[124] Informedia, Video Understanding Project, http://www.informedia.cs.cmu.edu/, accessed 09/05/06

[125] Fischlar Project, http://www.cdvp.dcu.ie/, accessed 09/05/06, Dublin City University Ireland

[126] Dictionary of Computing http://foldoc.org/foldoc.cgi?query=ontology Accessed 11/07/06

[127] Wikipedia, Online Encyclopaedia, en.wikipedia.org/wiki/Collocation, accessed 15/06/06

[128] Wikipedia, Online Encyclopaedia, en.wikipedia.org/wiki/Corpus_linguistics, accessed 15/06/06

[129] British Broadcasting Corporation, BBC: www.bbc.co.uk/, accessed 17/06/06

[130] ITFC, www.itfc.com/, accessed 17/06/06

[131] Royal National Institute for the Blind: RNIB, http://www.rnib.co.uk/, accessed 17/06/06

[132] Script O Rama; film Scripts, www.script-o-rama.com/, accessed 17/06/06

[133] Simplyscripts; film Scripts, www.simplyscripts.com/, accessed 17/06/06

[134] Internet Movie Script Database; film Scripts, www.imsdb.com/, accessed 17/06/06

[135] Daily Script; film Scripts, www.dailyscript.com/movie.html , accessed 17/06/06

[136] Internet Movie Database: www.imdb.com/, accessed 17/06/06

[137] SystemQuirk, www.computing.surrey.ac.uk/SystemQ/, accessed 18.07.06, University of Surrey

[138] Unitex, www-igm.univ-mlv.fr/~unitex/, accessed 19.06.06, Laboratoire d'Automatique Documentaire et Linguistique

[139] Wikipedia, Online Encyclopaedia, http://en.wikipedia.org/wiki/Standard_score, accessed 06.09.05

[140] Wikipedia, Online Encyclopaedia, en.wikipedia.org/wiki/Finite_state_automata, accessed 22/05/06

[141] Online Thesaurus, http://thesaurus.reference.com/ ,accessed 12/10/05

[142] Dictionary of Computing, foldoc.org/foldoc.cgi?query=finite+state+automata, accessed 13/07/06

[143] Definition from the Online Oxford English Dictionary http://www.oed.com/, accessed 12/09/05

[144] BNC search query system, http://sara.natcorp.ox.ac.uk/lookup.html, accessed 01/09/05.

[145] Wikipedia, Online Encyclopaedia,en.wikipedia.org/wiki/Information_extraction,accessed 14/07/06

[146] NIST, Machine Understanding Conference, www.itl.nist.gov/iaui/894.02/related_projects/muc/index.html, accessed 23/06/06

[147] Text Analysis International Inc., VisualText: http://www.textanalysis.com , accessed 24/03/06

[148] Information about US Television Scheduling and shows www.tv.com, accessed 22/07/06

[149] Online video repository, www.youtube.com, accessed 20/07/06

[150] Script writing software, Final Draft, http://www.finaldraft.com/, accessed 23/07/06

[151] Muvee digital video editing software http://www.muvee.com/, accessed 05/10/06

[152] General Architecture for Text Engineering: GATE, http://gate.ac.uk/, accessed 05/10/06

[153] Nielson Media Research http://www.nielsenmedia.com/nc/portal/site/Public/, accessed 09/10/06

Films and TV Series

[154] J.J. Abrahams, *LOST*, 2004-??, Bad Robot Productions and Touchstone Television

[155] Andrew Adamson and Vicky Jenson, *Shrek*, 2001, Dreamworks SKG

[156] Michael Apted, *The world is not enough*, 1999, Metro-Goldwyn-Mayer

[157] James Cameron, *Titanic*, 1997, 20th Century Fox

[158] James Cameron, *True Lies*, 1994, 20th Century Fox

[159] Wes Craven, *Scream*, 1996, Dimension Films

[160] Ellory Elkayem, *Eight Legged Freaks*, 2002, Warner Brothers

[161] Victor Flemming, The Wizard of Oz, 1939, Metro-Goldwyn-Mayer

[162] Steven Frears, *High Fidelity*, 2000, Buena Vista

[163] Sam Mendes, *American Beauty*, 1999, Dreamworks SKG

[164] Anthony Minghella, *The English Patient*, 1996, Buena Vista/Miramax

[165] Greg Motolla, *The Daytrippers*, 1996, Alliance Communications corporation

[166] Christopher Nolan, *Memento*, 2000, I Remember Productions LLC

[167] Gene Rodenberry (Creator), *Star Trek*, 1966-2005 (and beyond), Paramount Pictures

[168] M. Night Shyamalan, *The Sixth Sense*, 1999, Hollywood Pictures

[169] Steven Sodenbergh, *Oceans 11*, 2001, Warner Brothers;

[170] Steven Sodenbergh, *Out of Sight*, 1998, United International Pictures

[171] Stephen Sommers, *The Mummy*, 1999, United International Pictures

[172] Steven Spielberg, *Saving Private Ryan*, 2000, Amblin Entertainment

[173] Quentin Tarantino, *Pulp Fiction*, 1997, Miramax

[174] Guilliermo Del Toro, *Blade 2*, 2002, New Line Cinema

[175] Andre De Toth, *The Indian Fighter*, 1955, Metro-Goldwyn-Mayer

Film Scripts

ABR.	Genre Category	Genre	BBC: Film Title	Length
An	*Animation*	H/C/A	Buffy The Vampire Slayer	86
A	*Action*	Cr/D/M	Daytrippers	87
R	*Romance*	D	American Beauty	122
D	*Drama*	A/Cr/T	Hard Rain	87
H	*Horror*	**R/C/D**	**High Fidelity**	**113**
T	*Thriller*	**D/F/R**	**It's a Wonderful life**	**130**
C	*Comedy*	D/R	Leaving Las Vegas	111
W	*War*	D	Losing Isaiah	111
F	*Fantasy*	D/R	Love Is A Many Splendored Thing	102
Cr	*Crime*	C/F/R	Midsummer's Night Dream	116
M	*Mystery*	C/D/R	Month by the lake	92
Ad	*Adventure*	C/D	Mrs. Doubtfire	125
S	*Science Fiction*	C	Nine Months	103
B	*Biography*	**A/Cr/R**	**Out of Sight**	**123**
We	*Western*	F/C/D	Pleasantville	124
Mu	*Musical*	Cr/D/T	Red corner	122
		C/D	Royal Tannenbaumbs	109
		C/D/R	Shakespeare in Love	123
		An/Ad/C	Shrek	90
		D	Stealing Beauty	113
		Ad/C/F	Stuart little	84
		B/D/Mu	The Buddy Holly Story	113
		B/We	The True Story Of Jesse James	92
		D/W/A	To End all wars	125
		C/R	Truth About Cats And Dogs	97
		C	**Wag the dog**	**97**
		C/D	Waiting to Exhale	127
		C/R/D	Working Girl	115

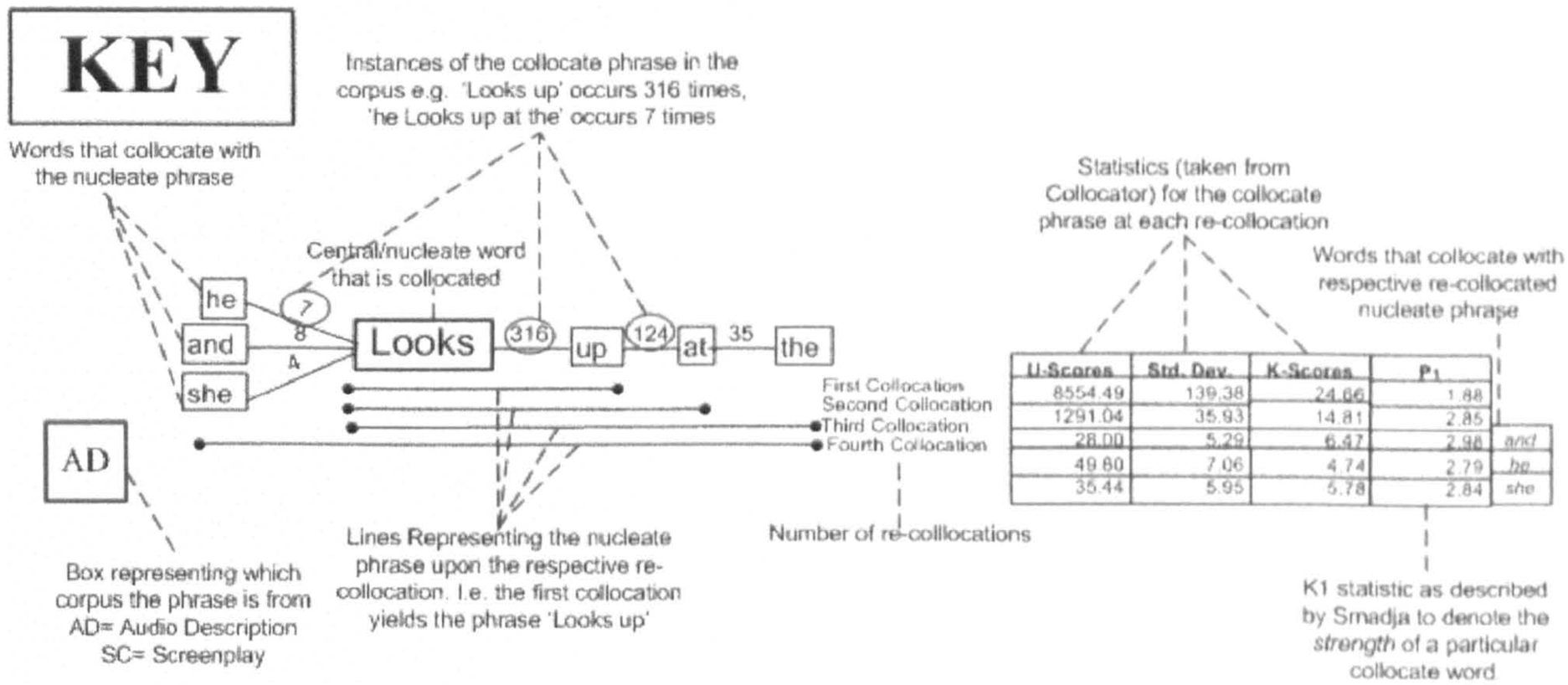
Genre	RNIB	Leng	Genre	ITFC	Len	Genre	Screenplay	Leng
D/M/T	**6th sense**	107	B/D	Amazing Howard Hughes	215	A/C/H	8 Legged Freaks	99
B/D	Iris	91	A/D/W	Apocalypse Now	153	C	Airplane	88
Ad/W/A	The Great Escape	172	**R/D/W**	**The English Patient**	**160**	S/A/H	Alien Resurrection	109
An/C/F	Monsters INC	92	M	Green for danger	91	A/D/S	Armageddon	150
C/Cr/R	Some like it Hot	120	We	Indian fighter	88	C/D/R	As Good as it gets	139
D/R	The horse Whisperer	170	A/F	Jason and the Argonauts	104	A/Ad/C	Austin Powers	94
D/F/T	**Unbreakable**	**106**	**C/D/R**	**Jerry Mcguire**	**139**	A/Ad/S	Back to the future	111
Ad/F/Mu	The Wizard of Oz	101	An/C/S	Lilo and Stitch	85	A/C/T	Bad Boys	118
D/R/W	Captain Corelli's Mandolin	131	Cr/D/M	Midnight garden of good & evil	155	T/D/M	Basic	98
An/S/F	Atlantis	95	A/T/C	Murder of crows	102	A/H/T	Blade	120
C/D/R	Chocolat	121	**A/C/Cr**	**Oceans 11**	**116**	A/H/T	Blade 2	117
An/Ad	Dinosaur	82	C/D	One hot summer night	100	A/Ad/T	Bond 18 (Tomorrow Never dies)	119
R/T/D	Enigma	119	D	One true thing	127	A/T/Ad	Bond 19 (The world is not enough)	128
A/Cr/T	Gone in 60 seconds	117	T/D/M	The pelican brief	141	D/M	Citizen Kane	119
Ad/F	Harry Potter and the Philosopher's stone	152	D/Ad/S	The postman	177	A/Cr/D	Cradle to the Grave	101
Cr/D/T	Insomnia	118	Cr/D	Road to perdition	117	Cr/D	Gangs of New York	167
An/D/R	Lady and the Tramp	76		Robin hood		D/M/S	Gattaca	101
D/R	**The shipping News**	**111**	Ad/C/F	Scooby Doo	88	C/F/A	Ghostbusters	105
A/Ad	Spy Kids	88	C	See no evil hear no evil	103	A/Ad/D	Gladiator	155
An/Ad	The Emperor's New Groove	78	Cr/T	Silence of the lambs	118	D	Good Will Hunting	126
D/H/T	The Others	101	A/F/S	Spiderman	121	Ad/F	Harry potter and chamber of secrets	160
R/D/W	**The English Patient**	**160**	A/D	Submarine	93	A/F	Highlander	116
			D/C/W	Tea with Mussolini	117	**D/F/R**	**It's a wonderful life**	**130**
			R/D	Brief Encounter	86	A/H/T	Jaws	124
			M/T/D	39 Steps	89	**C/D/R**	**Jerry Mcguire**	**139**
			Ad/T/D	Man who Knew too much	120	A/Ad/H	Jurassic Park	127
						A/C/Cr	Lethal Weapon 4	127
						C/F	Little Nicky	90
						Cr/D/M	Memento	113
						A/C/S	Men in Black	98

C	Monty Python Life of Brian	94	A/Ad/T	The Rock	136	C/Cr/D	Get Shorty	105
C/R/D	Notting Hill	124	D	The Shawshank Redemption	142	A/Ad/T	Goldfinger	112
C/Ad/Cr	O-Brother Where art thou	106	**D/M/T**	**The Sixth Sense**	**107**	H/T/F	Halloween 6	88
D/S/T	Outbreak	127	D/S	The Truman Show	103	Ad/F	Harry Potter and Prisoner of Azkaban	141
A/Cr/R	**Out of Sight**	**123**	C/R	There's Something about Mary	119	A/Cr/D	Hostage	113
Cr/D	Pulp Fiction	154	A/Ad/C	Three Kings	114	H/M/T	I still know what you did last summer	100
A/D/S	Planet of the Apes	112	A/D/R	Titanic	194	A/Ad/T	Indianna Jones and temple of Doom	118
A/D/T	Raiders of the Lost Ark	115	An/Ad/C	Toy Story	81	Cr/D/T	Jackie Brown	115
D	Rebel Without a Cause	111	An/A/S	The Transformers	84	H/T	Jeepers Creepers 2	104
Cr/D	Scarface	170	**D/F/T**	**Unbreakable**	**106**	Cr/D/M	JFK	189
B/D/W	Schinler's List	195	**C**	**Wag the dog**	**97**	A/Ad/S	Lost in Space	130
H/M/T	Scream	111	H/M/T	What lies Beneath	130	D/M	Manchurian Candidate 2004	129
D/C/R	Sense and Sensibility	136	Cr/D/T	Wild Things	108	A/Ad/T	Mission Impossible 2	123
D/M/S	Signs	106	A/S/T	X men	104	A/Ad/T	Mission Impossible	110
R/C/D	Sleepless in Seattle	105	C/R	10 things I hate about you	97	F/H/T	Nightmare on Elm Street	91
A/S/Ad	Star Wars Attack of the Clones	142	D/S/T	12 Monkeys	129	Cr/D/T	Phone Booth	81
A/S/Ad	Star Wars The Phantom Menace	133	A/T/D	Air Force One	124	T/S	PI	84
A/S/Ad	Star Wars: A new Hope	121	C/Cr	Analyse this	103	C/D/R	Punch Drunk Love	95
A/S/Ad	Star Wars Return of the Jedi	134	A/D/T	Backdraft	132	A/Ad/C	Rush Hour 2	90
A/T/Cr	Swordfish	99	D	Barton Fink	116	C/D/R	Sideways	123
A/S/T	Terminator	108	A/H/T	Blade Trinity	113	C/Cr/T	Snatch	104
A/Ad/S	Terminator2	137	Cr/D/T	Confidence	97	A/S/Ad	Star Trek X Nemesis	116
A/S/Ad	The Empire Strikes Back	124	A/Cr/T	Cellular	94	A/S/Ad	Star Wars Revenge of the Sith	140
R/D/W	**The English Patient**	**160**	C/D/T	Dr Strangelove	93	Ad/M/T	The Bourne Supremacy	108
A/C/D	The Fugitive	130	C/Ad	Dumb and Dumber	107	D/F/S	The Butterfly Effect	113
M/T/Ad	The Game	128	A/Ad/S	Escape from New York	99	D/R/C	The Cooler	101
A/Ad/D	The Last Samurai	154	D/R/S	Eternal Sunshine of spotless mind	108	D/T/H	The Devil's Advocate	144
Cr/D	The man who wasn't there	116	A/Cr/D	Face off	138	Cr/D	The Godfather part 2	200
A/T/S	The Matrix	136	H/T/F	Final destination 2	90	B/D/H	The Insider	157
A/Ad/C	The Mummy	124	H/A/C	From Dusk till dawn	108	C	The Producers	88
C	The Mystery Men	121	Ad/A/T	From Russia with love	115	**D/R**	**The shipping News**	**111**

APPENDIX A Collocations

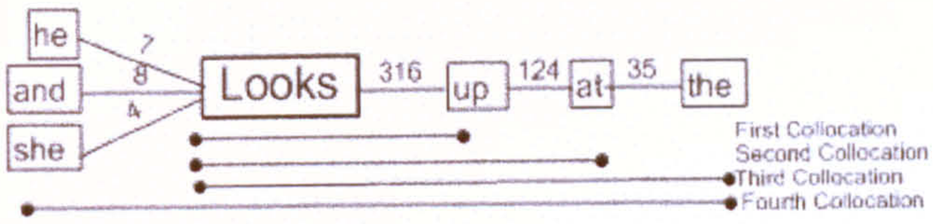
Collocation results of the most frequent Open Class words, based on z-scores, in Both Corpora (SC and AD).

The following are a series of diagrams representing the most frequent collocations (and re-collocations) of the top 10 most frequent open class words within the Audio Description and Screenplay corpora. This analysis was conducted using System Quirk's Collocator.

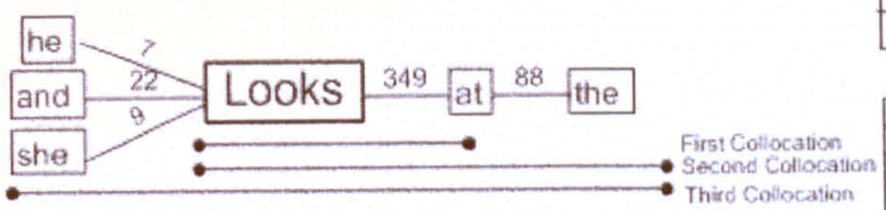


Found on accompanying CD.

AD

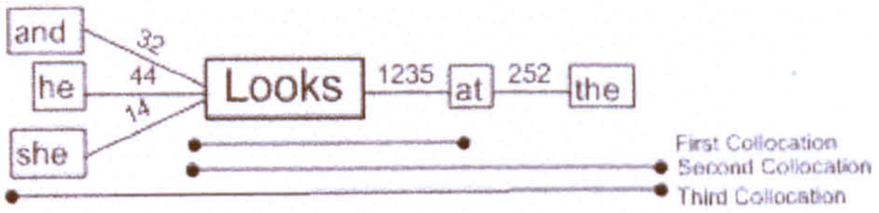


U-Scores	Std. Dev.	K-Scores	P1	
8554.49	139.38	24.66	3.00	
1291.04	35.93	14.61	2.99	
87.65	9.36	14.24	2.83	
28.00	5.29	6.47	2.94	*and*
49.80	7.06	4.74	2.93	*he*
35.44	5.95	5.78	2.65	*she*



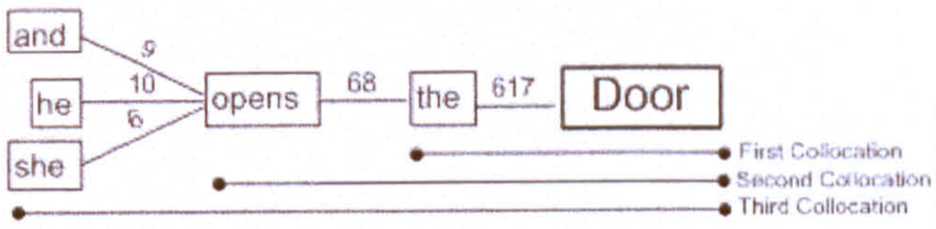
U-Scores	Std. Dev.	K-Scores	P1	
19426.44	139.38	24.66	1.68	
584.99	24.17	18.48	2.85	
40.6	6.37	8.33	2.98	*and*
7.04	2.65	4.16	2.79	*he*
4.29	2.07	2.67	2.84	*she*

SC

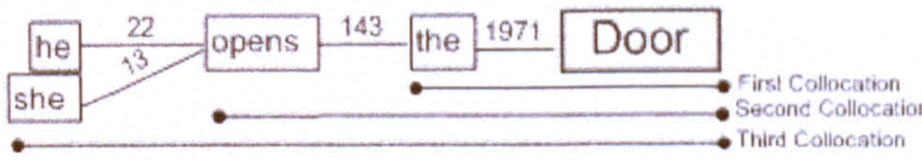


U-Scores	Std. Dev.	K-Scores	P1	
144032.01	379.52	37.50	2.69	
4758.20	68.98	26.72	2.62	
162.69	12.76	8.38	2.91	*he*
79.29	8.90	10.91	2.59	*and*
16.21	4.03	2.57	2.91	*she*

AD

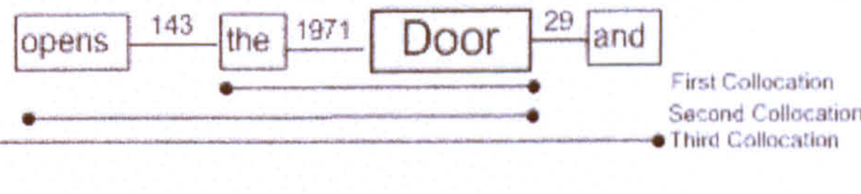


U-Scores	Std. Dev.	K-Scores	P1	
27637.56	164.43	39.86	2.85	
380.49	19.51	7.51	2.93	
9.04	3.01	3.67	2.86	*he*
11.36	3.37	7.95	1.84	*and*
3.20	1.79	7.44	2.80	*she*



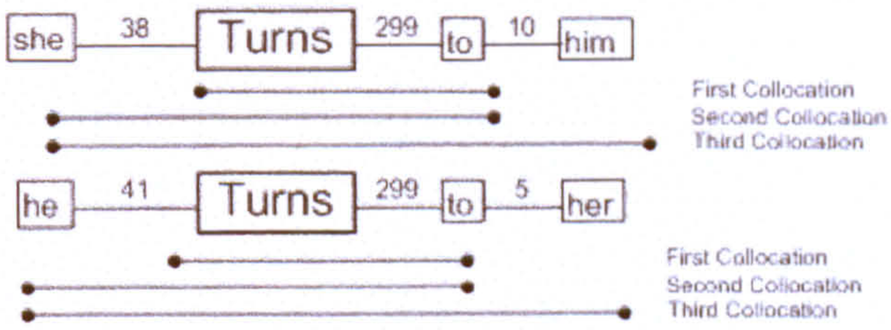
U-Scores	Std. Dev.	K-Scores	P1	
284879.41	533.74	61.96	2.84	
1908.04	43.68	10.29	2.55	
421.44	6.52	5.36	2.70	*he*
13.60	3.69	3.52	2.71	*she*

SC



U-Scores	Std. Dev.	K-Scores	P1
284879.41	533.74	61.96	2.84
1908.04	43.68	10.29	2.55
71.98	8.48	9.73	2.50

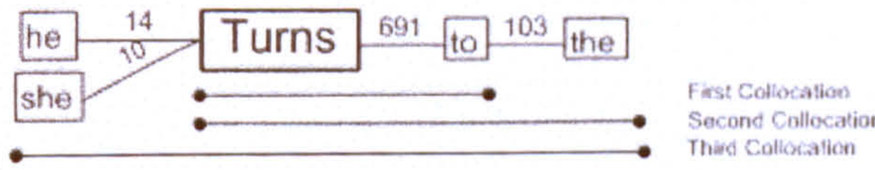
AD



U-Scores	Std. Dev.	K-Scores	P1
7552.20	86.90	19.28	2.88
133.49	11.35	6.60	2.97
3.09	1.76	5.65	2.47

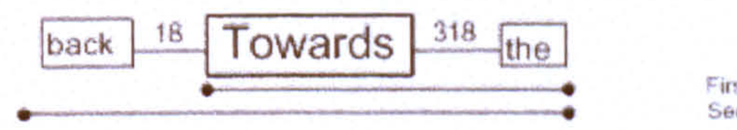
U-Scores	Std. Dev.	K-Scores	P1
7552.20	86.90	19.28	2.88
118.64	10.89	5.29	2.95
9.44	3.07	7.64	2.33

SC



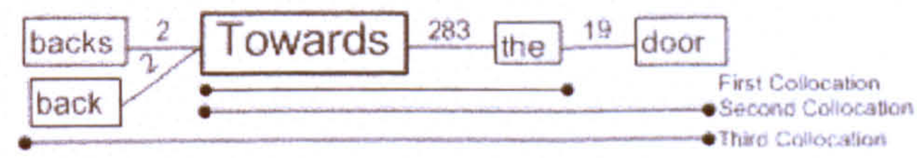
U-Scores	Std. Dev.	K-Scores	P1	
37328.69	193.21	29.34	2.94	
805.01	28.37	27.18	2.07	
15.45	3.93	4.28	2.93	*he*
8.81	2.97	2.04	2.93	*she*

AD



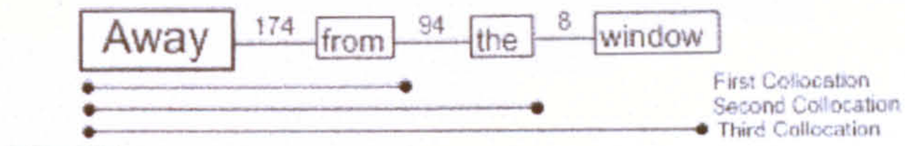
U-Scores	Std. Dev.	K-Scores	P1
11719.60	108.26	39.03	2.90
27.16	5.21	3.10	2.84

SC

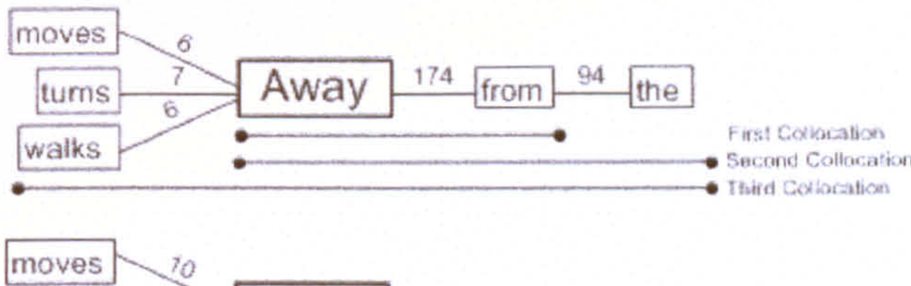


U-Scores	Std. Dev.	K-Scores	P1	
6050.36	77.78	35.97	2.94	
193.61	13.91	23.48	2.62	
0.41	0.69	0.70	2.66	*backs*
0.36	0.60	0.16	3.00	*back*

AD

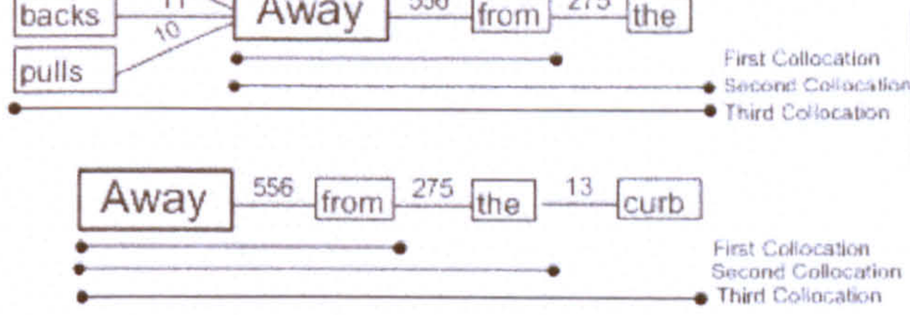


U-Scores	Std. Dev.	K-Scores	P1
2709.29	52.05	9.11	3.00
665.36	28.80	22.61	2.92
5.65	2.39	1.72	2.98



U-Scores	Std. Dev.	K-Scores	P1	
2709.29	52.02	9.11	3.00	
665.36	25.80	2.65	2.92	
4.29	2.07	1.73	2.95	*turns*
3.24	1.79	0.99	3.00	*moves*
3.21	1.79	1.23	2.96	*walks*

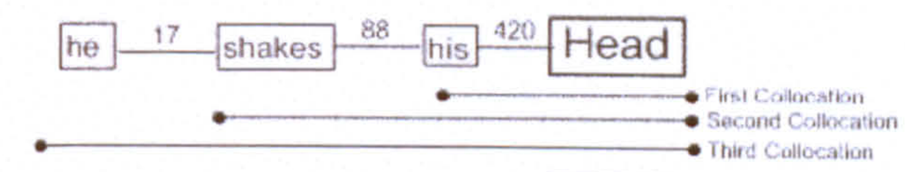
SC



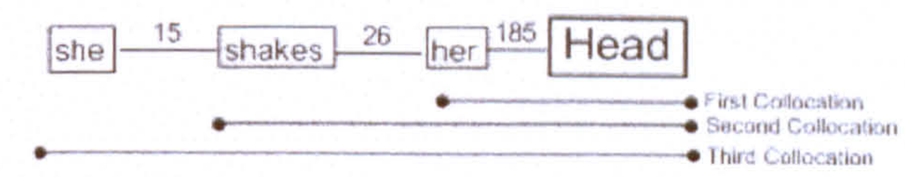
U-Scores	Std. Dev.	K-Scores	P1	
27011.69	164.35	19.19	3.00	
5549.44	74.50	37.71	2.92	
10.64	3.26	1.70	2.94	*backs*
8.64	2.94	1.99	2.86	*moves*
7.81	2.80	3.01	2.76	*pulls*

U-Scores	Std. Dev.	K-Scores	P1
27011.69	164.35	19.19	3.00
5549.44	74.50	37.71	2.92
12.81	3.58	1.55	2.99

AD

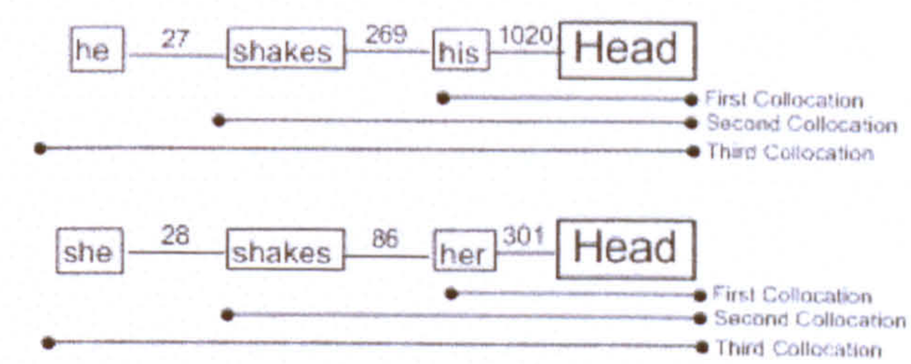


U-Scores	Std. Dev.	K-Scores	P1
14743.44	121.42	24.79	2.99
683.84	26.15	8.12	3.00
22.16	4.71	5.61	2.93



U-Scores	Std. Dev.	K-Scores	P1
2899.85	53.85	12.82	2.98
119.21	10.92	10.84	2.99
18.76	4.33	6.01	2.96

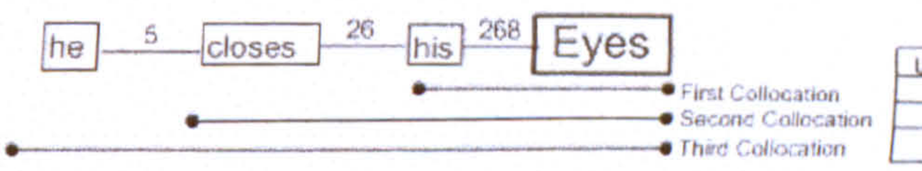
SC



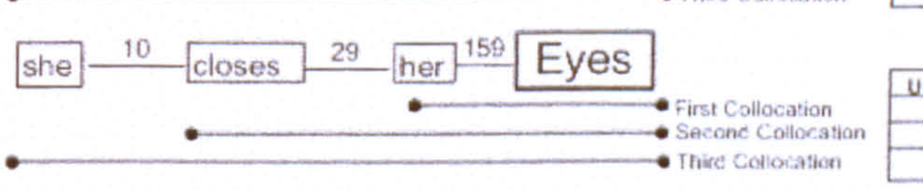
U-Scores	Std. Dev.	K-Scores	P1
87507.61	295.82	36.01	3.00
6079.41	177.97	18.51	2.98
54.44	7.38	6.15	2.76

U-Scores	Std. Dev.	K-Scores	P1
7492.61	285.56	11.31	2.99
632.29	25.15	11.01	2.99
61.89	7.87	9.26	2.91

AD

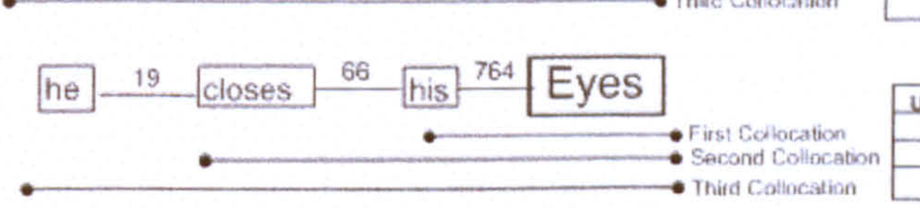


U-Scores	Std. Dev.	K-Scores	P1
5760.49	75.90	21.59	2.99
59.49	7.74	2.92	3.00
2.16	1.47	3.30	3.00

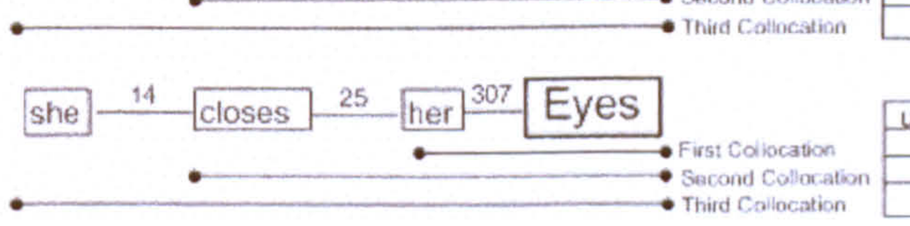


U-Scores	Std. Dev.	K-Scores	P1
1953.36	44.20	14.61	2.99
73.04	8.55	4.99	2.86
8.61	2.93	5.32	2.96

SC

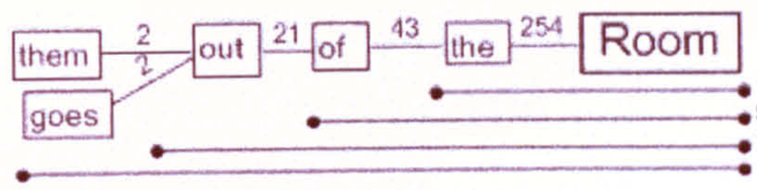


U-Scores	Std. Dev.	K-Scores	P1
47999.84	219.09	34.28	3.00
366.41	19.14	5.43	2.96
22.62	5.71	6.76	2.74



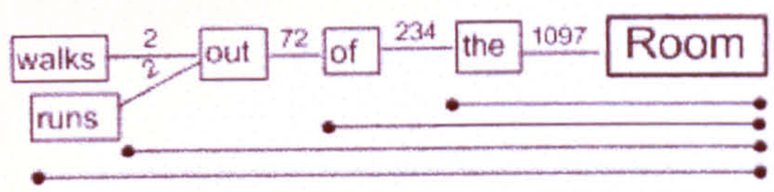
U-Scores	Std. Dev.	K-Scores	P1
7359.61	85.79	16.26	2.99
55.36	4.44	11.41	2.98
16.04	1.01	6.10	2.85

AD



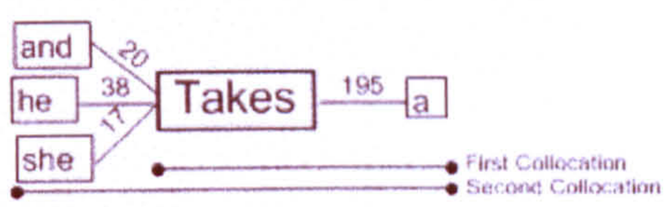
U-Scores	Std. Dev.	K-Scores	P1	
5047.01	71.04	32.88	2.51	
152.89	13.37	8.36	2.98	
39.36	6.27	5.77	3.00	
0.44	0.66	0.41	2.41	*them*
0.36	0.60	0.27	3.00	*goes*

SC



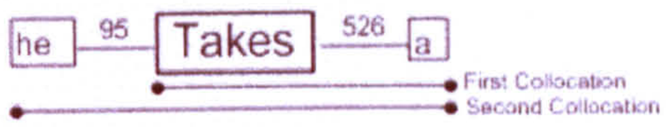
U-Scores	Std. Dev.	K-Scores	P1	
83197.20	288.44	60.80	2.61	
9393.84	66.27	18.89	2.98	
300.36	17.33	6.65	2.99	
4.44	2.11	3.27	2.66	*walks*
3.16	1.78	1.61	2.93	*runs*

AD



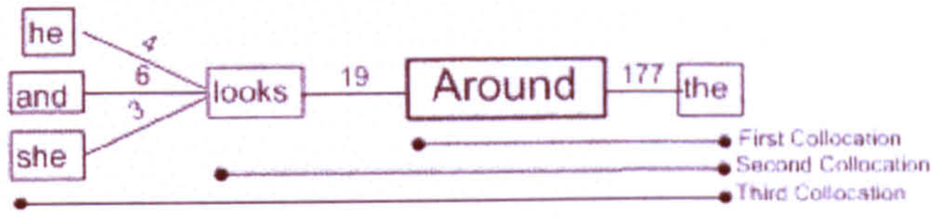
U-Scores	Std. Dev.	K-Scores	P1	
2993.05	54.71	18.08	2.95	
120.76	10.99	7.55	2.98	*he*
32.64	5.71	8.16	2.99	*and*
24.89	6.99	2.80	2.52	*she*

SC



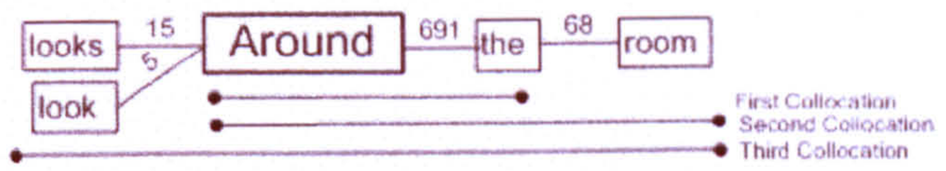
U-Scores	Std. Dev.	K-Scores	P1
21962.55	148.20	28.21	2.97
709.25	26.63	11.86	2.95

AD

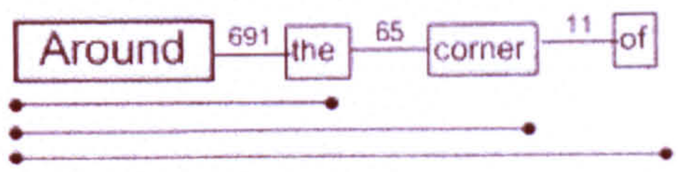


U-Scores	Std. Dev.	K-Scores	P1	
1930.96	43.94	33.69	2.70	
31.40	5.60	5.52	2.85	
3.21	1.79	4.25	2.96	*and*
1.45	1.20	2.71	2.91	*he*
0.11	0.90	1.17	3.00	*she*

SC

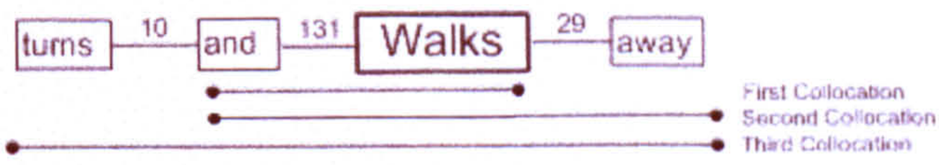


U-Scores	Std. Dev.	K-Scores	P1	
29401.25	171.47	60.19	2.75	
391.09	19.78	6.03	2.98	
20.04	4.48	4.68	2.99	*looks*
21.60	1.47	2.02	2.85	*look*



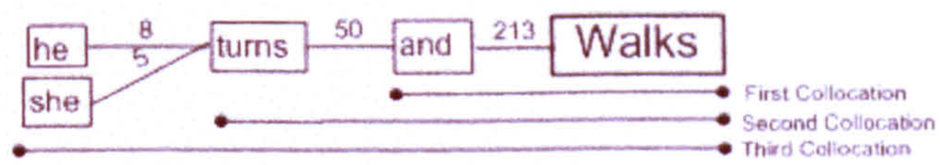
U-Scores	Std. Dev.	K-Scores	P1
29401.25	171.47	60.19	2.75
370.61	19.25	5.27	3.00
6.49	2.91	8.89	2.71

AD

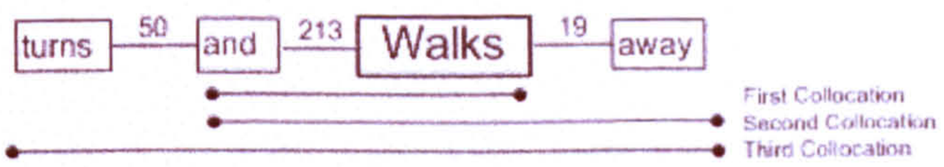


U-Scores	Std. Dev.	K-Scores	P1
1304.64	36.12	14.39	2.92
71.56	8.46	5.48	2.98
8.76	2.96	5.69	2.97

SC

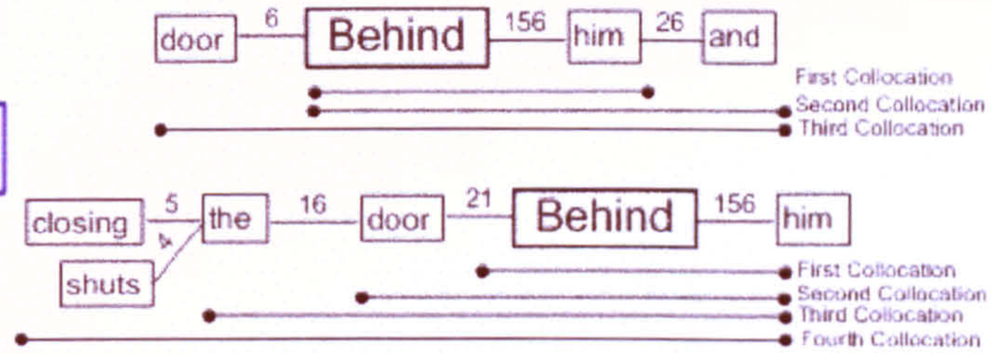


U-Scores	Std. Dev.	K-Scores	P1	
3366.56	58.02	18.75	2.85	
203.8	14.28	7.35	2.94	
28.16	5.31	7.73	2.79	*he*
2.40	1.55	1.44	2.58	*she*



U-Scores	Std. Dev.	K-Scores	P1
3366.56	58.02	18.75	2.85
203.8	14.28	7.35	2.94
34.70	5.90	4.95	2.75

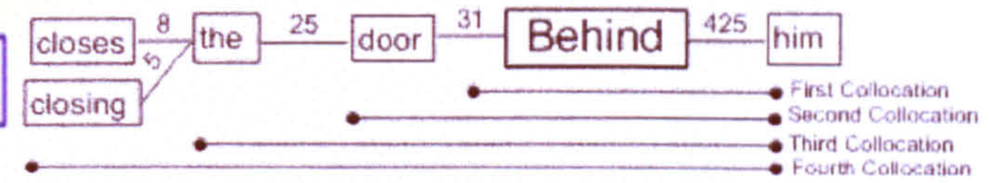
AD



U-Scores	Std. Dev.	K-Scores	P_1	
2119.89	46.04	11.45	3.00	
55.29	7.43	5.06	2.98	
2.24	1.50	1.90	2.97	

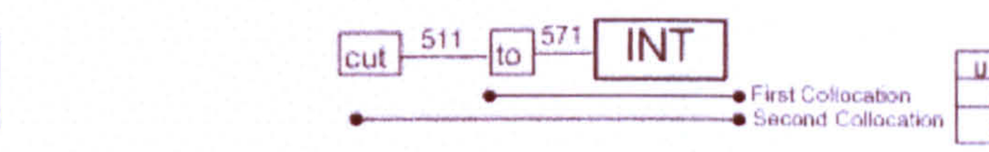
U-Scores	Std. Dev.	K-Scores	P_1	
2119.69	46.04	11.45	3.00	
35.81	5.98	5.08	2.89	
19.21	4.38	5.40	2.90	
2.24	1.50	2.58	2.93	*closing*
1.44	1.20	1.34	3.00	*shuts*

SC



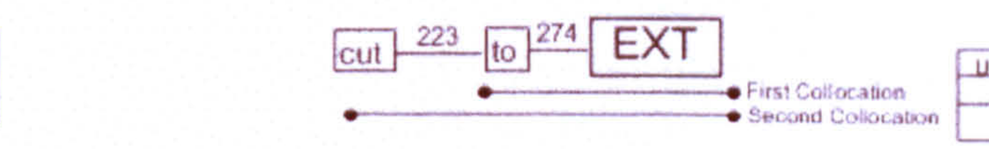
U-Scores	Std. Dev.	K-Scores	P_1	
15527.31	124.61	16.66	2.99	
78.64	8.67	5.45	2.75	
48.49	6.96	11.69	2.89	
5.60	2.37	4.44	2.96	*closes*
2.25	1.50	1.75	3.00	*closing*

SC



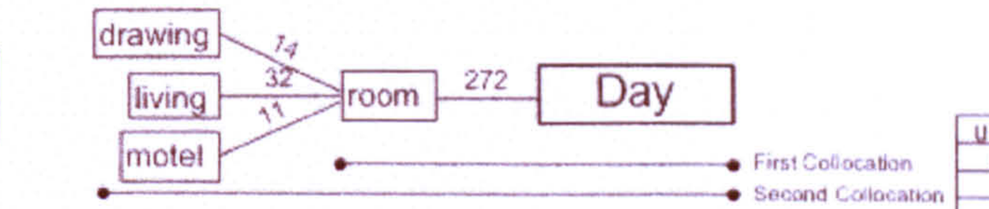
U-Scores	Std. Dev.	K-Scores	P_1
29415.64	171.51	28.48	
21902.69	147.99	29.50	

SC



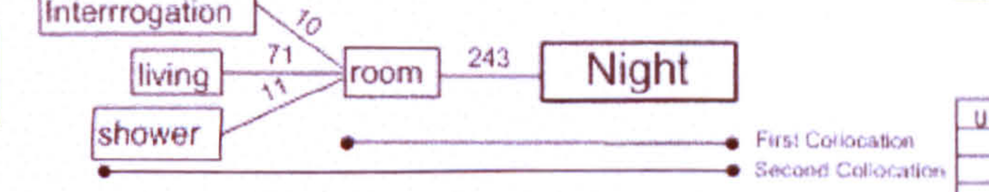
U-Scores	Std. Dev.	K-Scores	P_1
29433.12	171.24	33.04	-
4211.45	64.90	24.22	-

SC



U-Scores	Std. Dev.	K-Scores	P_1	
65753.41	250.51	42.00		
88.49	9.41	4.09		*living*
22.04	4.70	2.40		*drawing*
10.76	3.28	1.05		*motel*

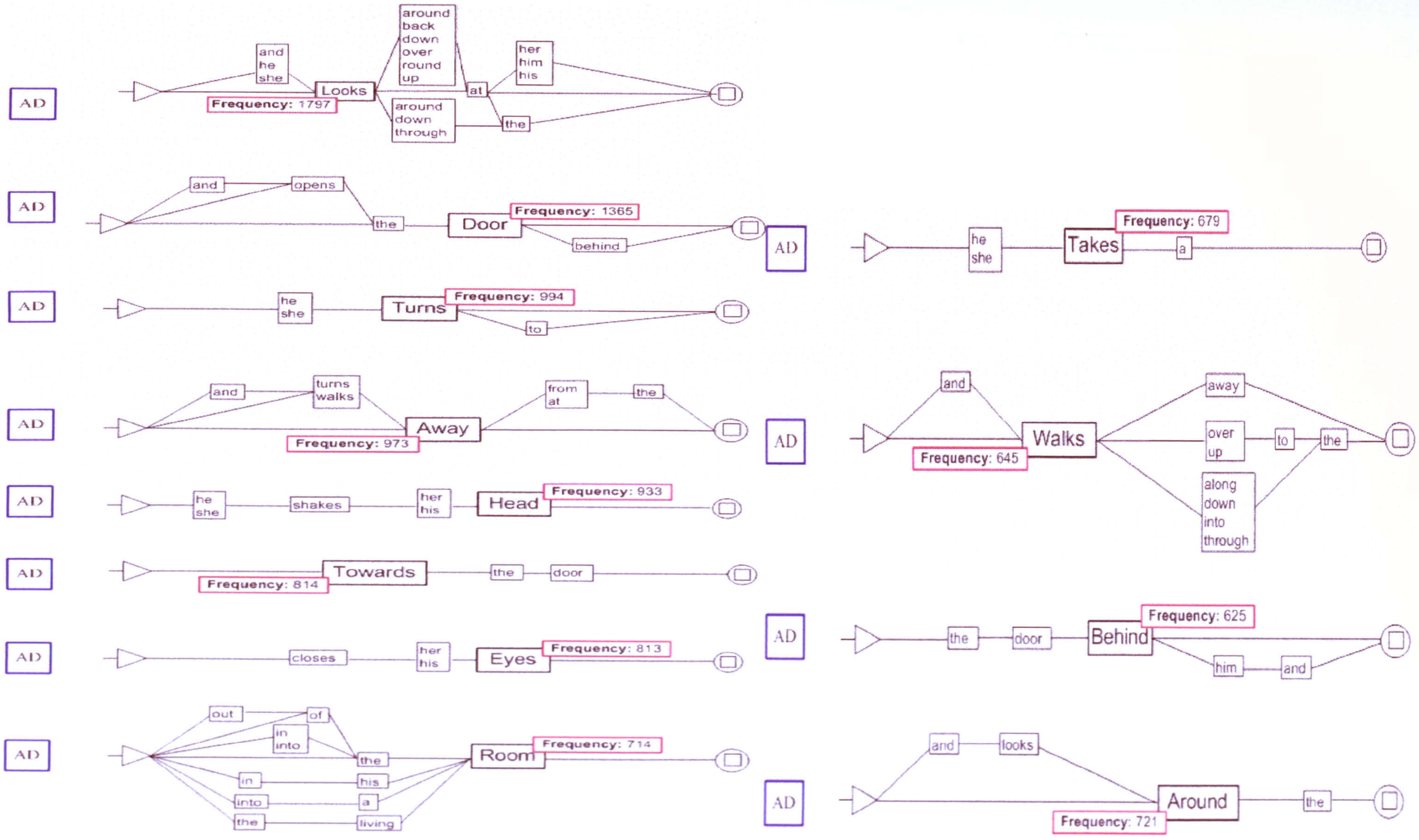
SC

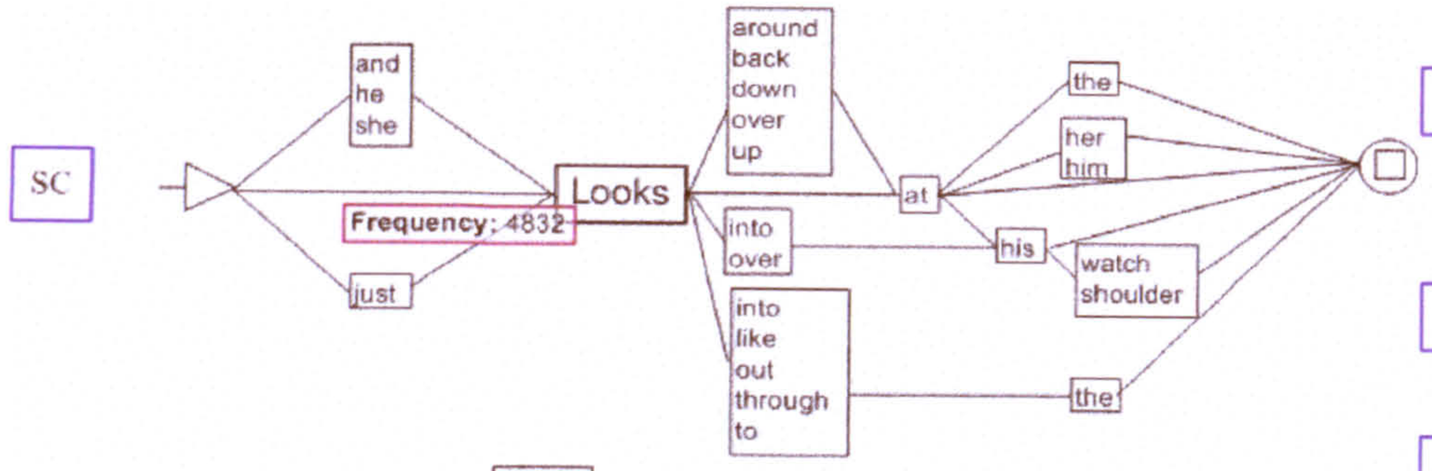


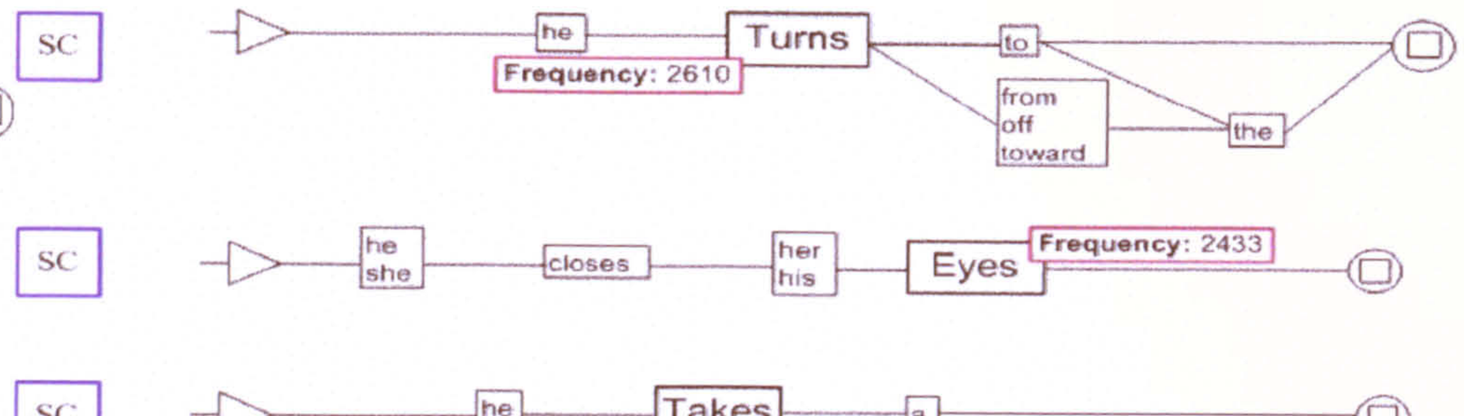
U-Scores	Std. Dev.	K-Scores	P_1	
7223.61	84.99	10.12		
452.96	21.28	8.70		*living*
10.41	3.23	1.34		*shower*
4.41	2.10	0.79		*interrogation*

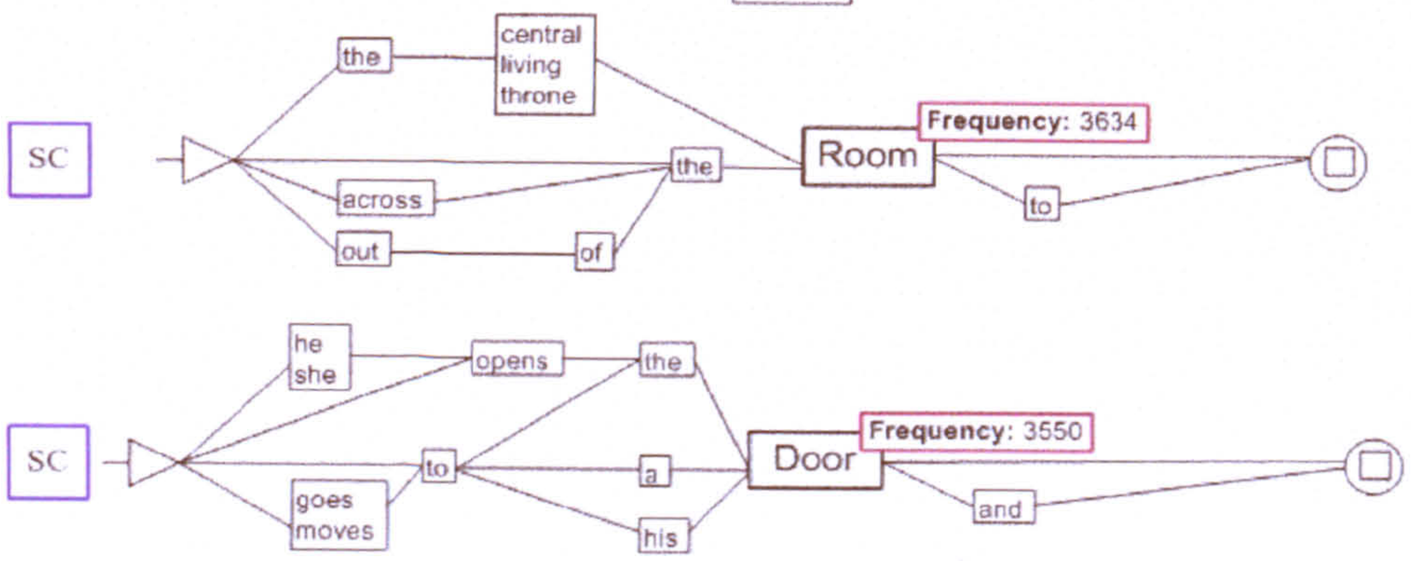
APPENDIX B Local Grammar Finite State Automata Graphs FSA

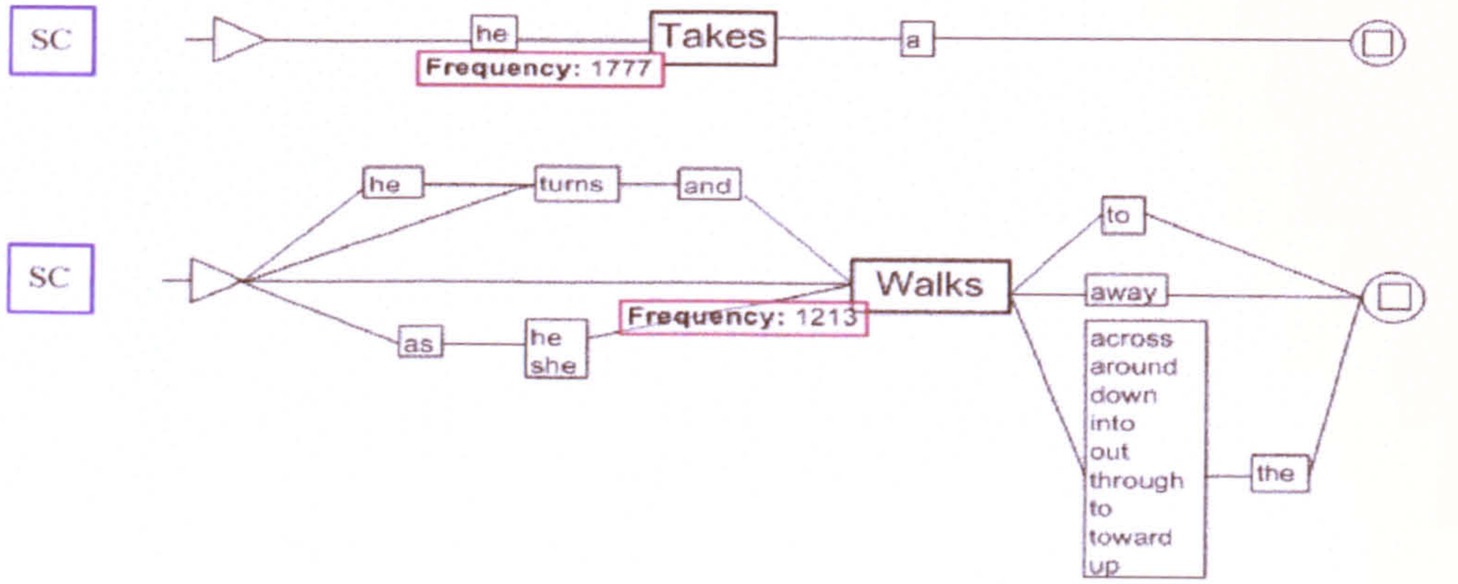
Found on accompanying CD.


AD
and
he
she
Looks
Frequency: 1797
around
back
down
over
round
up
around
down
through
at
her
him
his
the
AD
and
opens
the
Door
Frequency: 1365
behind
AD
he
she
Turns
Frequency: 994
to
AD
he
she
Takes
Frequency: 679
a
AD
and
turns
walks
Away
Frequency: 973
from
at
the
AD
and
Walks
Frequency: 645
away
over
up
to
the
along
down
into
through
AD
he
she
shakes
her
his
Head
Frequency: 933
AD
Towards
Frequency: 814
the
door
AD
the
door
Behind
Frequency: 625
him
and
AD
closes
her
his
Eyes
Frequency: 813
AD
out
of
in
into
the
in
his
into
a
the
living
Room
Frequency: 714
AD
and
looks
Around
Frequency: 721
the


SC
and
he
she
Looks
Frequency: 4832
just
around
back
down
over
up
into
over
into
like
out
through
to
at
the
her
him
his
watch
shoulder
the


SC
he
Turns
Frequency: 2610
to
from
off
toward
the

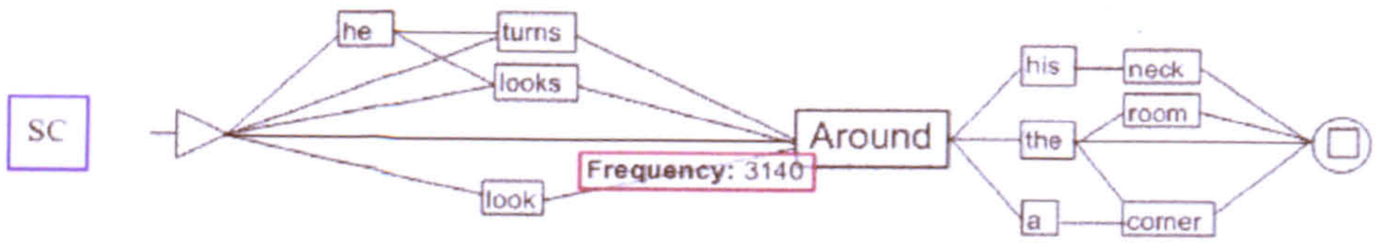

SC
the
central
living
throne
Room
Frequency: 3634
across
out
of
the
to

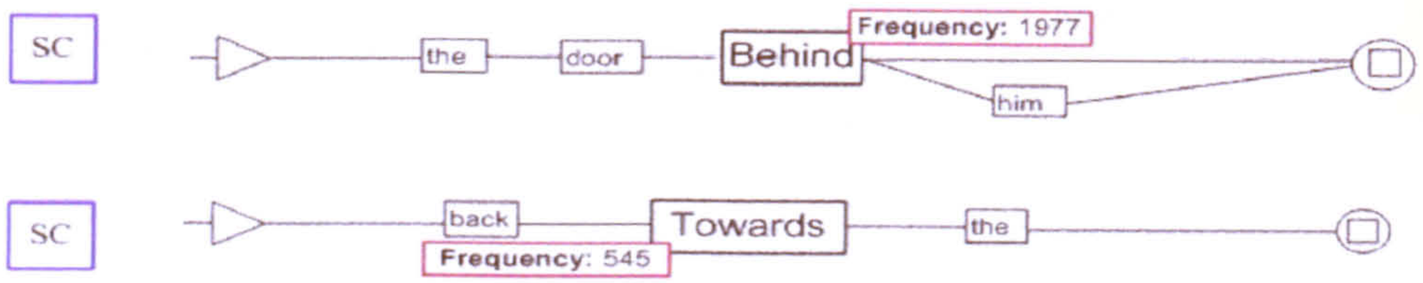

SC
he
she
closes
her
his
Eyes
Frequency: 2433

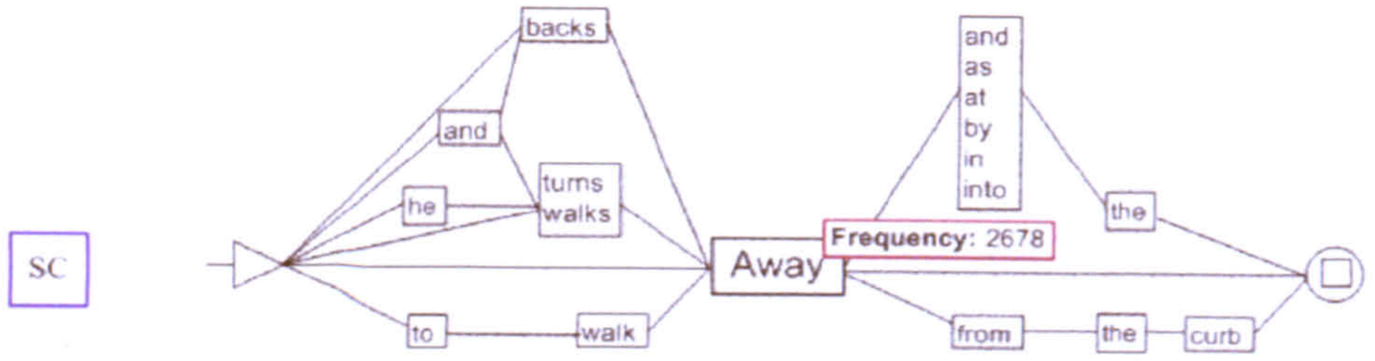
SC
he
Takes
Frequency: 1777
a

SC
he
she
opens
the
Door
Frequency: 3550
goes
moves
to
a
his
and

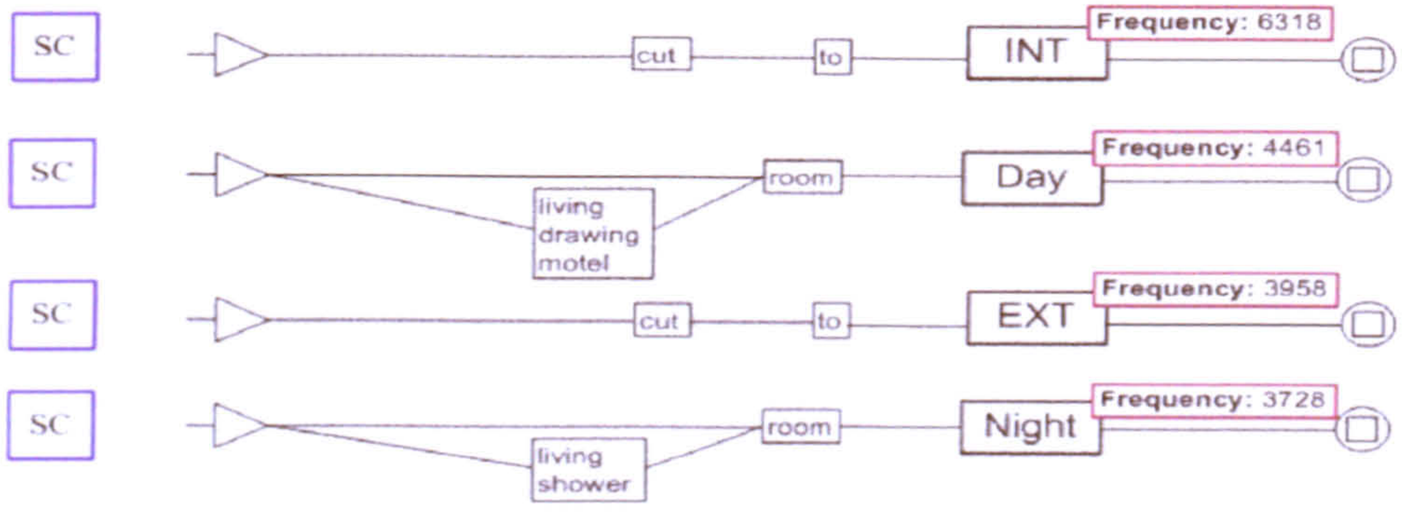
SC
he
turns
and
Walks
Frequency: 1213
as
he
she
to
away
across
around
down
into
out
through
to
toward
up
the

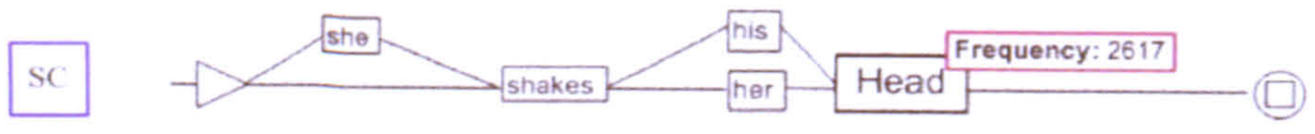

SC
he
turns
looks
Around
Frequency: 3140
look
his
neck
the
room
a
corner


SC
the
door
Behind
Frequency: 1977
him


SC
backs
and
he
turns
walks
Away
Frequency: 2678
to
walk
and
as
at
by
in
into
the
from
the
curb

SC
back
Towards
Frequency: 545
the

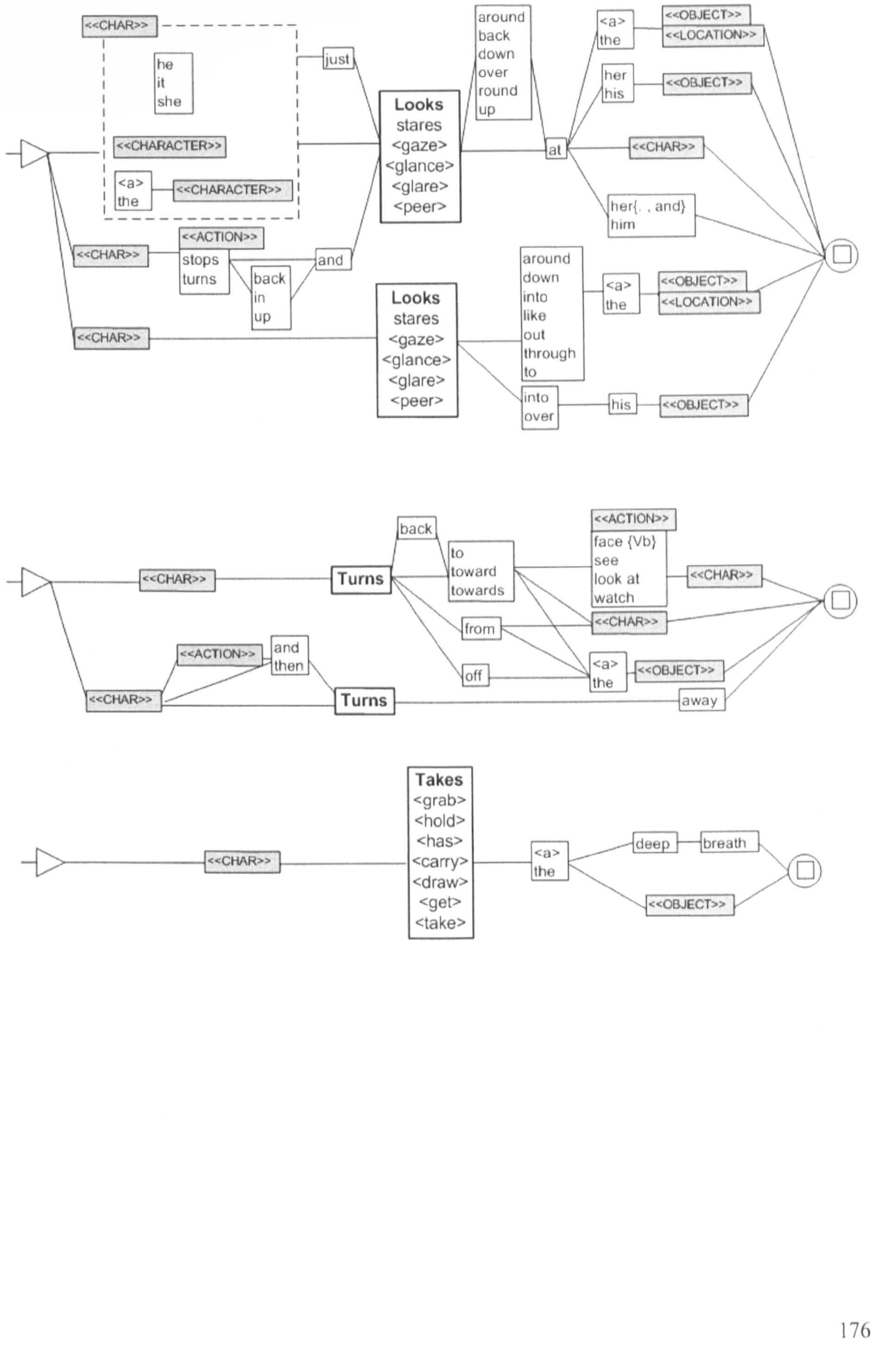

SC
cut
to
INT
Frequency: 6318
SC
living
drawing
motel
room
Day
Frequency: 4461
SC
cut
to
EXT
Frequency: 3958
SC
living
shower
room
Night
Frequency: 3728


SC
she
shakes
his
her
Head
Frequency: 2617

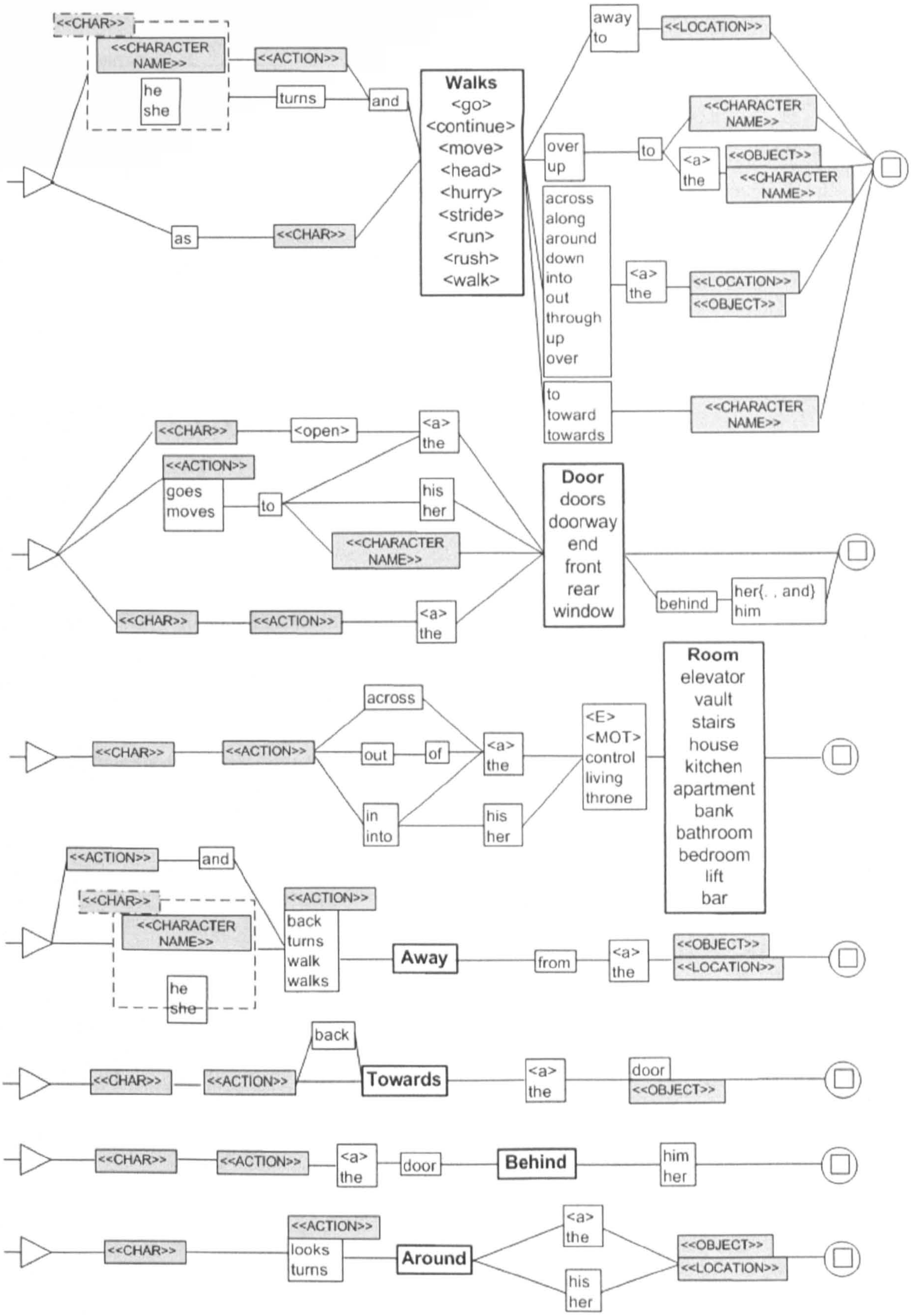
APPENDIX C Local Grammar FSA used in Information Extraction Heuristics

Found on accompanying CD.

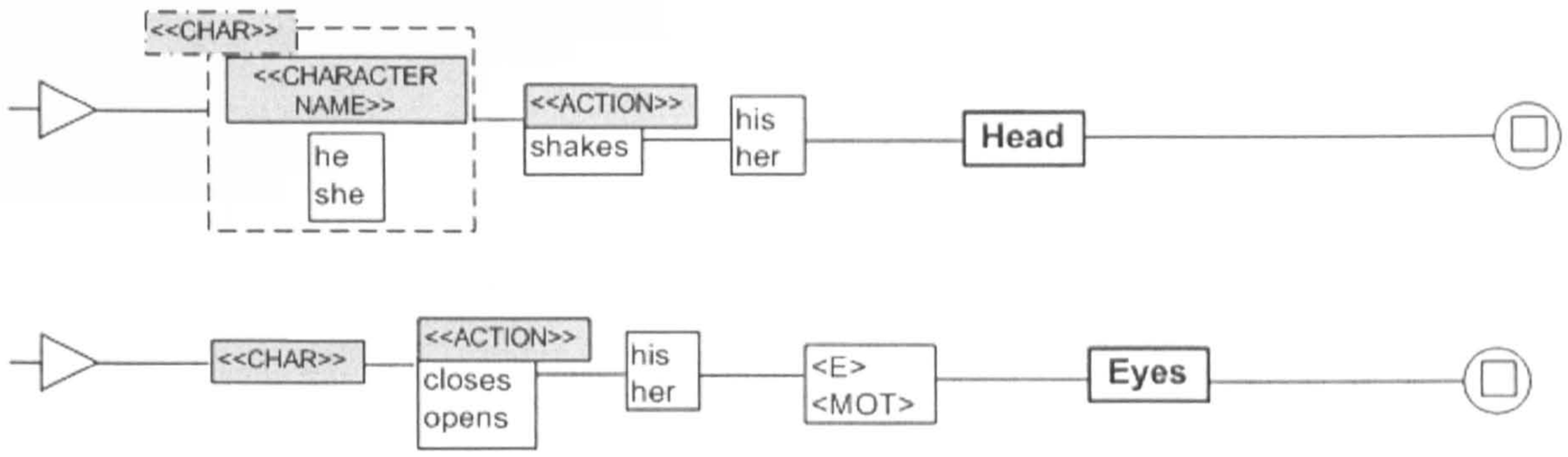
FOCUS OF ATTENTION



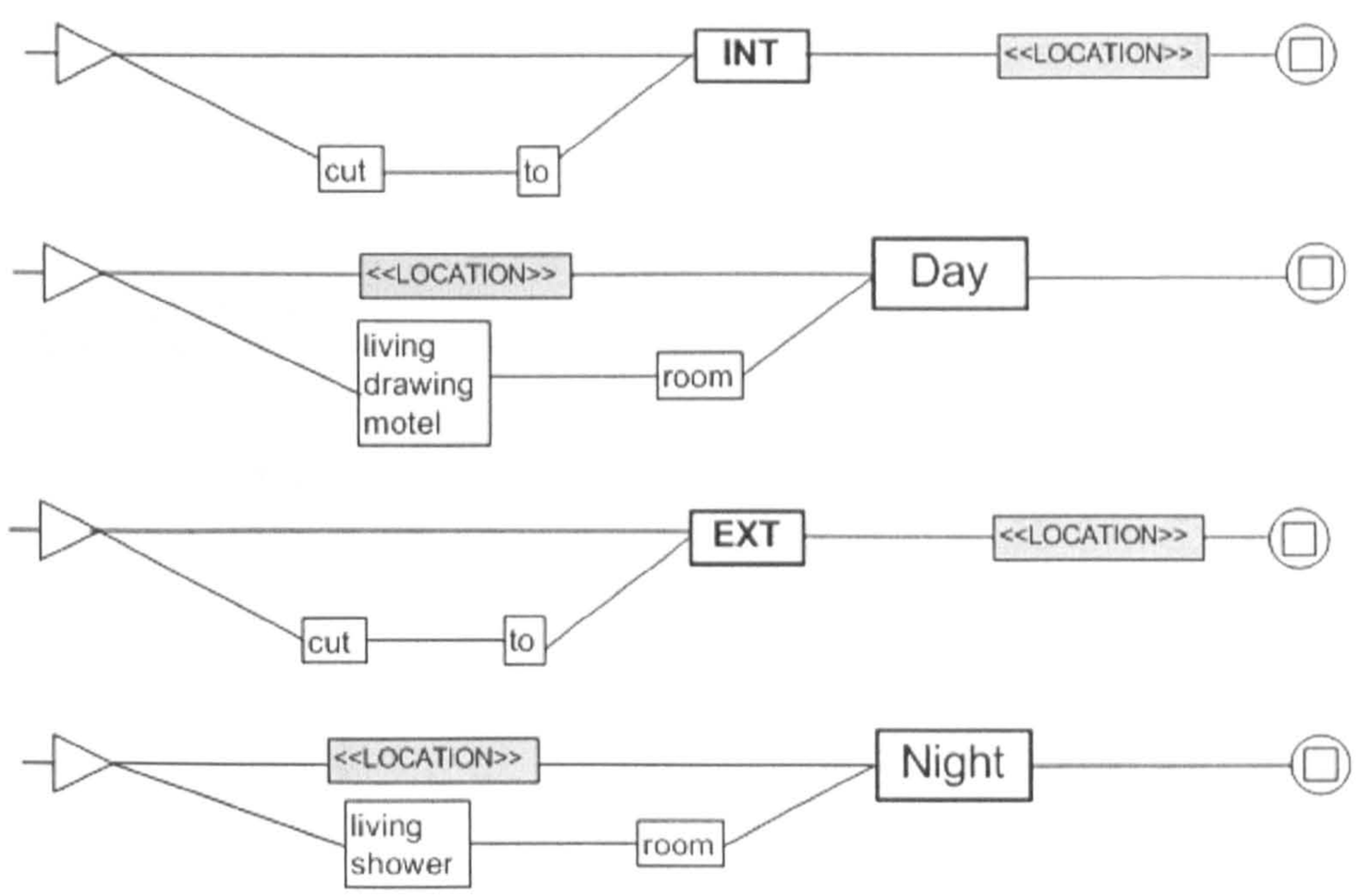
CHANGE OF LOCATION



NON VERBAL COMMUNICATION



SCENE CHANGE



APPENDIX D Instructions for Gathering Gold Standard Event Data.

Found on accompanying CD.

A Group: Focus and Communication

Name:

Occupation:

Film Expertise: HIGH MEDIUM NOVICE (LOW)

Sound: ON OFF

Thank you for taking the time to do this evaluation and analysis

The object of this exercise is to provide a data set to evaluate an Information Extraction system.

Please watch the following ~60 minutes of film scenes (5*12 minute film clips) from five movies (see overleaf). Every time you observe a "**Focus of attention event**" or a "**Non-verbal Communication event**" please pause the film (CTRL+P or  button) and complete the rows of the table provided overleaf. [Some dummy rows have been filled in for you.] You may select whether to have the sound on or off to help you concentrate better on the events occurring.

A focus of attention event is defined as anytime a character is focussing on (e.g. looking at, turning towards) a person (another character in the film) or an object. Generally, whenever a character is focussing their attention on something. The attributes of the ***Focus of attention event*** table are defined as:

Character_Focussing = The character who is doing the focussing on something

Character_Focussed_On = The character being focussed on (**optional**)

Object_Focussed_On = The object being focussed on (**optional**)

Occurrence_Time = The time the focus of attention event occurs at.

NB please fill in at least one of the Character_Focussed_On or Object_Focussed_On attributes or both if you think that is the case.

A non-verbal communication event occurs when a character is communicating without speech to another character using parts of their body e.g. nodding, opening closing their eyes, winking etc. The attributes of the ***Non-verbal Communication event*** table are defined as:

Character_Communicating = The character doing the non-verbal communication.

Character_Communicated_to = Character being communicated to non-verbally.

Body_Part_Involved = The body part being communicated with e.g. eyes, head, hands.

Occurrence_Time = The time the event occurred.

B Group: Changes

Name:

Occupation:

Film Expertise: HIGH MEDIUM NOVICE (LOW)

Sound: ON OFF

Thank you for taking the time to do this evaluation and analysis

The object of this exercise is to provide a data set to evaluate an Information Extraction system.

Please watch the following ~60 minutes of film scenes (5*12 minute film clips) from five movies (see overleaf). Every time you observe a "**Character changing location event**" or a "**Scene Change event**" please pause the film (CTRL+P or  button) and complete the rows of the table provided overleaf. [Some dummy rows have been filled in for you.] You may select whether to have the sound on or off to help you concentrate better on the events occurring.

A change of location event is defined as anytime a character changes location from one area to another, i.e. goes through a door, goes to another room, walks across an open space or a room etc.

The attributes of the ***Change of Location table*** are defined as:

State_of_Character= Whether a character is ENTERING, LEAVING or WITHIN a room when changing location.

Initial_Location = The character's initial location before changing location. (**optional**)

Final_Location = The character's Final location after changing location. (**optional**)

Character_Changing_Location = The character that is changing location

Secondary_Character = If there is a secondary character involved in the location change (**optional**)

Occurrence_Time = The time the focus of attention event occurs at.

NB please fill in at least one of the **Initial_Location** or **Final_Location** attributes or both if you think that is the case.

A Scene Change event occurs when the scene you are viewing changes suddenly to another scene. Usually the location will change suddenly. The attributes of the *Scene Change table* **are defined as:**

Day/Night = Whether the new scene is in the daytime or night-time.

Interior/Exterior = Whether the new scene is exterior or interior.

New_Location= The new location of the scene change

Occurrence_Time = The time the event occurred.

Please watch the following five video clips found on the accompanying CD and fill in the event tables accordingly, pausing when you see a relevant event to record the time. The clips are about 12 minutes each and labelled 01 to 05.

Clip 1:The English Patient	*'01 English Patient.avi'*	12.33
Clip 2:The World is Not Enough	*'02 World is not enough.avi'*	11.46
Clip 3:Oceans 11	*'03 Ocean 11.avi'*	11.36
Clip 4:High Fidelity	*'04 High Fidelity.avi'*	11.50
Clip 5:Out of Sight	*'05 Out of Sight.avi'*	13.04

Please fill in which video clip you are watching in the "FilmClip" section of the tables.

You are requested to use Windows Media Player if you have it.

Once again thank you for taking the time to participate in this evaluation.

APPENDIX E Heuristics

Found on accompanying CD.

Global Word Sets

Art = {the, a, an}
P= {',', '.', and}
PrN_1 = {his, her}
PrN_2 = {him, her***P***}
//* <MOT> is the wild card for any word in that word space*//
//* <E>= null string or empty space*//
//* ACTION is any verb in that word space *//

Focus of Attention

Lk= {<look>, <stare>, <glance>, <gaze>, <peer>, <glare>}
Lk_t = {just, <E>}
Lk_u = {back, in, up, <E>}
Lk_v = {into, over}
Lk_w= {stops, turns, ACTION}
Lk_x = {around, back, down, over, round, up, <E>}
Lk_y = {around, down, into, like, out, through, to}

LK_1 = [***Lk*** + ***Lk_x*** + "at"]
MID_LK_A = [***LK_1*** + ***Art***]
MID_LK_B = [***LK_1*** + ***PrN_1***]
MID_LK_C = [***Lk***+ ***Lk_y*** + ***Art***]
MID_LK_D = [***Lk*** + ***Lk_v*** + ***PrN_1***]
PRE_LK_R = [***Lk_w*** + ***Lk_u*** + "and"]

FOA_LK_A = ***PRE_LK_R*** + ***MID_LK_A*** **//*CHAR → OBJ, LOC*//**
FOA_LK_B = ***Lk_t*** +***MID_LK_A*** **//*CHAR → OBJ, LOC*//**
FOA_LK_C = ***MID_LK_A*** **//*CHAR → OBJ, LOC*//**
FOA_LK_D = ***Lk_t*** + ***MID_LK_C*** **//*CHAR → OBJ, LOC*//**

FOA_LK_E = ***LK_1*** **//*CHAR → CHAR *//**
FOA_LK_F = ***PRE_LK_R*** + ***LK_1*** **//*CHAR → CHAR *//**

FOA_LK_G = ***PRE_LK_R*** + ***MID_LK_B*** **//*CHAR → OBJ *//**
FOA_LK_H = ***MID_LK_B*** **//*CHAR → OBJ *//**
FOA_LK_I = ***Lk_t*** + ***MID_LK_B*** **//*CHAR → OBJ *//**
FOA_LK_J = ***Lk_t*** + ***MID_LK_D*** **//*CHAR → OBJ *//**
//**********//

Trn_w = {back, <E>}
Trn_x = {and, then, ACTION, <E>}
Trn_y = {to, toward, towards}
Trn_z = {face, see, look at, watch, ACTION}

TRN_1 = ["turns" + ***Trn_w***]
MID_TRN_A = [***TRN_1*** + ***Trn_y***]
MID_TRN_B = [***TRN_1*** + ***Trn_y*** + ***Trn_z***]
MID_TRN_C = [***TRN_1*** + ***Trn_y*** + ***Art***]
MID_TRN_D = ["turns from"]
MID_TRN_E = ["turns from" + ***Art***]
MID_TRN_F = ["turns off" + ***Art***]
PRE_TRN_R = [***Trn_x*** + "Turns away"]

FOA_TRN_A = ***MID_TRN_A*** **//*CHAR → CHAR*//**
FOA_TRN_B = ***MID_TRN_B*** **//*CHAR → CHAR*//**
FOA_TRN_C = ***MID_TRN_D*** **//*CHAR → CHAR*//**

FOA_TRN_D = MID_TRN_C **//*CHAR → OBJ*//**
FOA_TRN_E = MID_TRN_E **//*CHAR → OBJ*//**
FOA_TRN_F = MID_TRN_F **//*CHAR → OBJ*//**
FOA_TRN_G = PRE_TRN_R **//*CHAR*//**
FOA_TRN_H = ACTION + ***PRE_TRN_R*** **//*CHAR*//**
//**********//

Tke = {<take>, <grab>, <hold>, <has>, <carry>, <draw>, <get>}

MID_TKE_A = [***Art*** + "deep breath"]
MID_TKE_B = [***Art***]

FOA_TKE_A = Tke + MID_TKE_A **//*CHAR → OBJ*//**
FOA_TKE_B = Tke + MID_TKE_B **//*CHAR → OBJ*//**
//**********//

1. Locate and Count strings ***FOA_LK_A-J, FOA_TRN_A-H*** and ***FOA_TKE_{A-B}***
2. IF ***FOA_LK_{A-J}*** or ***FOA_TRN_{A-F}*** is found, place the string 'ACTIVE' in the "Focus_of_Attention_Type" field ELSE IF ***FOA_TRN_{G-H}*** is found, place 'DISCONTINUED' string in the "Focus_of_Attention_Type" field ELSE IF ***FOA_TKE_{A-B}*** is found, place 'PASSIVE' string in the "Focus_of_Attention_Type" field.
3. IF ***FOA_LK_{E-F}*** or ***FOA_TRN_{A-C}*** Search to the left of string for pronoun || proper-noun ||noun
    a. IF noun examine the article to the left and place the article and noun string into the "Character Focussing" field.
    b. IF pronoun {he, she} traverse line for noun/proper noun associated with pronoun. When found place noun & associated article or proper noun string into "Character Focussing" field.
    c. IF proper noun place the string into the "Character Focussing" field.
4. Search to the right of the string for pronoun, proper noun, noun or article.
    a. IF proper noun found add to "Character_Focussed_On" field.
    b. IF pronoun {PrN_2} traverse lines in the immediate proximity of the string for proper nouns/nouns and add to "Character_Focussed_On" field.
5. IF ***FOA_LK_{G-J}*** or ***FOA_TRN_{D-L}*** Search to the left of string for pronoun || proper noun || noun
    a. IF noun examine the article to the left and place the article and noun string into the "Character Focussing" field.
    b. IF pronoun {he, she} traverse line for noun/proper noun associated with pronoun. When found place noun & associated article or proper noun string into "Character Focussing" field.
    c. IF proper noun place the string into the "Character Focussing" field.
6. Search to the right of the string for a noun.
    a. IF noun found add noun to the "Object_Focussed_On" field.
7. IF ***FOA_LK_{A-D}*** Search to the left of the string for a pronoun, proper noun or noun.
    a. IF noun examine the article to the left and place the article and noun string into "Character Focussing" field.
    b. IF pronoun {he, she, it, they} traverse line for noun/proper noun associated with pronoun. When found place the noun and associated article or proper noun string into "Character Focussing" field.
    c. IF proper noun place the string into the "Character Focussing" field.
8. Search to the right of the string for a pronoun or article.
    a. IF noun found add noun to the "Object_Focussed_On" field.
    b. IF proper noun found add proper noun string to the "Object_Focussed_On" field.
9. IF ***FOA_TRN_{G-H}*** Search to the left of the string for a pronoun, proper noun or noun.
    a. IF noun examine the article to the left and place the article and noun string into "Character Focussing" field.
    b. IF pronoun {he, she} traverse line for noun/proper noun associated with pronoun. When found place the noun or proper noun string into the "Character Focussing" field.

c. IF proper noun place the string into the "Character Focussing" field.

10. IF ***FOA_TKE_{A-B}*** Search to the left of the string for a pronoun, proper noun or noun.
    a. IF noun examine the article to the left and place the article and noun string into the "Character Focussing" field.
    b. IF pronoun {he, she} traverse line for noun/proper noun associated with pronoun. When found place noun & associated article or proper noun string into "Character Focussing" field.
    c. IF proper noun place the string into the "Character Focussing" field.
11. Search to the right of the string for a noun.
    a. IF noun found add noun to the "Object_Focussed_On" field.
12. Traverse back to the beginning of the line for line number or time code. Place line number or time code in the "Occurrence_Time" field.

Non-Verbal Communication

Eys_x = {opens, closes}
Hd_x = {shakes, nods}
NVC_EYS_A = [***Eys_x*** + ***PrN_1*** + "eyes"]
NVC_HD_A = [***Hd_x*** + ***PrN_1*** + "head"]

Locate the strings ***NVC_EYS_A*** or ***NVC_HD_A***

1. IF ***NVC_EYS_A*** is found add the string 'Eyes' to the "Body_Parts_Involved" field and the ***Eys_x*** string to the "Action_Involved" field. ELSE IF ***NVC_HD_A*** is found add the string 'Head' to the "Body_Parts_Involved" field and 'shakes' to the "Action_Involved" field
2. Search for pronoun, noun or proper noun left of the strings.
    a. IF noun search for article to the left of string and add article and noun to "Character_Communicating" field.
    b. IF proper noun add string to the "Character_Communicating" field.
    c. IF pronoun search for proper nouns and nouns (preceded by article) in the immediate proximity before the string. IF found add pronoun string in parenthesis and add the proper noun or article and noun to "Character_Communicating" field.
3. Traverse back to the beginning of the line for line number or time code. Place line number or time code in the "Occurrence_Time" field.

Scene change: Temporal-Spatial Cues

IF **AUDIO DESCRIPTION SCRIPT** and IF ***COL_RM_{A-D}*** is found then:

1. Search 1 space to the left,
    a. IF number NB:NB:NB is found THEN place numeric string in "Occurrence_Time" field and the string ***COL_RM_{A-D*** in "New_Location" field. **END**
    b. ELSE **END**

IF **SCREENPLAY**

1. Locate the strings: '***night***', '***day***'.
    a. IF '***night***' or '***day***' are found place string located in the "Day/Night" field.
2. Search to the left of the string for 'INT.' or 'EXT.'
    a. IF 'INT.' is found place string 'interior' in the "INT/EXT" field. ELSE IF 'EXT.' is found place string 'exterior' in the "INT/EXT" field.
3. Locate string (***x***) from the end of the string 'INT.' or 'EXT.' to the beginning of the word '***day***' or '***night***' and place the string ***x*** in the "New_Location" field.
4. Traverse back to beginning of line for line number. Place line number in "New_Time" field.
5. Place the string, from 'INT.' or 'EXT.', inclusive, to the line number and '.', in "Text_String" field.

Change of Location

Wlk = {<walk>, <go>, <continue>, <start>, <begin>, <move>, <head>, <hurry>, <stride>, <run>, <rush>}
Wlk_v = {away}
Wlk_w = {turns, ACTION}
Wlk_x1 = {into, up}
Wlk_x2 = {across, along, around, down}
Wlk_x3 = {out, through, over}
Wlk_y = {to, toward, towards}
Wlk_z = {over, up}

PRE_WLK_A = [***Wlk_w*** + "and"]

MID_WLK_P = [***Wlk*** + ***Wlk_v***]
MID_WLK_Q = [***Wlk*** + ***wlk_z*** + "to"]
MID_WLK_R1 = [***Wlk*** + ***Wlk_x1***]
MID_WLK_R2 = [***Wlk*** + ***Wlk_x2***]
MID_WLK_R3 = [***Wlk*** + ***Wlk_x3***]
MID_WLK_S = [***Wlk*** + ***Wlk_y***]

COL_WLK_A = ***PRE_WLK_A*** + ***MID_WLK_Q*** //*CHAR → CHAR*//
COL_WLK_B = ***MID_WLK_Q*** //*CHAR → CHAR*//
COL_WLK_C = ***PRE_WLK_A*** + ***MID_WLK_S*** //*CHAR → CHAR*//
COL_WLK_D = ***MID_WLK_S*** //*CHAR → CHAR*//

COL_WLK_E{1-3} = ***PRE_WLK_A*** + ***MID_WLK_R{1-3}*** //*CHAR → CHAR, OBJ*//
COL_WLK_F{1-3} = ***MID_WLK_R{1-3}*** //*CHAR → CHAR, OBJ*//

COL_WLK_G = ***PRE_WLK_A*** + ***MID_WLK_P*** //*CHAR → LOC*//
COL_WLK_H = ***MID_WLK_P*** //*CHAR → LOC*//
//**********//

Dr = {door, doors, doorway, back, end, rear, window}
Dr_x = {goes, moves, ACTION}

PRE_DR_A = ["<open>" + ***Art***]
PRE_DR_B = [***Dr_x*** + "to" + ***Art***]
PRE_DR_C = [***Dr_x*** + "to" + ***PrN_1***]
PRE_DR_D = [***Dr_x*** + "to" + <MOT>'s] //*<MOT>'s = any word ending in apostrophe s*//
PRE_DR_E = [ACTION + ***Art***]

MID_DR_P = [***Dr*** + "behind" + ***PrN_2***]

COL_DR_A = ***PRE_DR_A*** + ***Dr*** //*CHAR*//
COL_DR_B = ***PRE_DR_B*** + ***Dr*** //*CHAR*//
COL_DR_C = ***PRE_DR_C*** + ***Dr*** //*CHAR*//
COL_DR_D = ***PRE_DR_D*** + ***Dr*** //*CHAR*//
COL_DR_E = ***PRE_DR_E*** + ***MID_DR_P*** //*CHAR*//
//**********//

Rm = {room, elevator, vault, stairs, house, kitchen, apartment, bank, bathroom, bedroom, lift, bar, casino, street, road, church, garden, roof, corridor, cage, cockpit}
Rm_x = {<E>, <MOT>, control, living, throne}
Rm_y = {in, into}

PRE_RM_A = ["across" + ***Art***]
PRE_RM_B = ["out of" + ***Art***]
PRE_RM_C = [***Rm_y*** + ***Art***]
PRE_RM_D = [***Rm_y*** + ***PrN_1***]

MID_RM_P = [***Rm_x*** + ***Rm***]

COL_RM_A = ***PRE_RM_A*** + ***MID_RM_P*** //*CHAR*//
COL_RM_B = ***PRE_RM_B*** + ***MID_RM_P*** //*CHAR*//
COL_RM_C = ***PRE_RM_C*** + ***MID_RM_P*** //*CHAR*//
COL_RM_D = ***PRE_RM_D*** + ***MID_RM_P*** //*CHAR*//
//**********//

Aw = {away}
Aw_x = {back, turns, walk, walks}

PRE_AW_A = [***Aw_x***]

MID_AW_P = [“away from” + ***Art***]

COL_AW_A = ***PRE_AW_A*** + ***MID_AW_P*** //*CHAR→OBJ, LOC*//
COL_AW_B = ***Aw*** + ***MID_AW_P*** //*CHAR→OBJ, LOC*//
//**********//

Blnd = {behind}

PRE_BD_A = [ACTION + ***Art*** + “door”]
MID_BD_P = [***Blnd*** + ***PrN_2***]
COL_BD_A = ***PRE_BD_A*** + ***MID_BD_P*** //*CHAR→CHAR*//
//**********//

Twrd = {towards, toward}

PRE_TW_A = [ACTION + “back”]
PRE_TW_B = [ACTION]

MID_TW_P = [***Twrd*** + ***Art***]

COL_TW_A = ***PRE_TW_A*** + ***MID_TW_P*** //*CHAR→OBJ*//
COL_TW_B = ***PRE_TW_B*** + ***MID_TW_P*** //*CHAR→OBJ*//
//**********//

CHAR_CHANGE_ACTIVE = [

1. IF ***COL_DR_{A-E}*** isolate the article or pronoun {his, her}.
    a. Place string “ACTIVE” in the “Change_of_Location_Type” field.
    b. Place string ‘ENTERING’ in the “State_of_Character” field.
    c. Search to the left one more string, IF verb place in the “Complimentary_Action” string.
    d. Place article or {his, her} and word to the right in the “Final_Location” field.
    e. **END**
2. Place the string “ACTIVE” in the “Change_of_Location_Type” field.
3. Search to the left of the string for a verb.
    a. IF verb found, add this string to the “Complimentary_Action” field and search further to the left for a proper noun, noun or pronoun.
    b. IF noun then search for article and add the string to “Character_Changing_Location” field.
    c. IF pronoun search to the left for a proper noun. Add the proper noun to the “Character_Changing_Location” field
4. IF proper noun is found then add string to the “Character_Changing_Location” field
5. IF pronoun is found search to the left for a proper noun and add string to “Character_Changing_Location” field.]

CHAR_CHANGE_DESCRIPTIVE = [

1. Place string ‘WITHIN/ON’ in the “State_of_Character” field.
2. Place the string “DESCRIPTIVE” in the “Change_of_Location_Type” field.

3. Search to the left of the string for a verb.
4. IF verb found, add this string to the "Complimentary_Action" field and search further to the left for a proper noun, noun or pronoun.
    a. IF noun then search for article and add the string to the "Character_Descr_in_Loc" field.
    b. IF pronoun search to the left for a proper noun. Add the proper noun to the "Character_Descr_in_Loc" field
    c. IF proper noun is found then add string to the "Character_Descr_in_Loc" field
5. IF pronoun is found search left for proper noun and add string to "Character_Descr_in_Loc" field.
6. IF noun is found then search to the left for an article or pronoun {his, her} and place article or pronoun and noun strings into the "Character_Descr_in_Loc" field.]

//**********//

1. Locate and count strings: ***COL_WLK_{A-H}, COL_DR_{A-E}, COL_RM_{A-D}, COL_AW_{A-C}, COL_AR_{A-D}, COL_BD_A*** and ***COL_TW_{A-B}***
2. IF ***COL_WLK_{A-B}*** or ***COL_WLK_{C-D}*** then:
    a. **CHAR_CHANGE_ACTIVE**
    b. IF ***COL_WLK_{A-B}*** then place string 'WITHIN/ON' in the "State_of_Character" field. ELSE IF ***COL_WLK_{C-D}*** then place string 'ENTERING' in "State_of_Character" field.
    c. Look to the right of the string for article, pronoun {his, her}, proper noun or noun
        i. IF noun then add the noun to the "Final_Location" field.
        ii. IF proper noun look to the right for a noun or ***P*** = {',','.', and}. IF noun found then add the proper noun and noun to the "Final_Location" field. IF ***P*** found add proper noun to the "Secondary_Character_Involved" field.
        iii. IF {his, her} search right for noun. Add pronoun & noun to "Final_Location" field.
        iv. IF article search to the right for noun or adjective.
            1. IF noun add article and noun to the "Final_Location" field.
            2. IF adjective search next word for noun. IF noun is found add the article, adjective and noun to the "Final_Location" field
3. IF ***COL_WLK_{E1}, COL_WLK_{F1}, COL_AW_{A-C}, COL_BD_A*** or ***COL_TWD_{A-B}*** then:
    a. **CHAR_CHANGE_ACTIVE**
    b. IF ***COL_WLK_{E1}, COL_WLK_{F1},*** or ***COL_TWD_{A-B}*** then place string 'ENTERING' in the "State_of_Character" field. ELSE IF ***COL_AW_{A-C}*** or ***COL_BD_A*** then place string 'WITHIN/ON' in "State_of_Character" field.
    c. Look to the right of the string for article, pronoun {his, her}, proper noun or noun
        i. IF proper noun look to the right for a noun. IF noun found then add the proper noun and noun to the "Final_Location" field.
        ii. IF {his, her} search right for noun. Add pronoun & noun to "Final_Location" field
        iii. IF article search to the right for noun or adjective.
            1. IF noun add article and noun to the "Final_Location" field.
            2. IF adjective search next word for noun. IF noun add article, adjective and noun to the "Final_Location" field.
4. IF ***COL_WLK_{E2}, COL_WLK_{F2}, COL_WLK_{G-H}*** or ***COL_AR_{A-D}*** then:
    a. **CHAR_CHANGE_ACTIVE**
    b. IF ***COL_WLK_{E2}, COL_WLK_{F2},*** or ***COL_AR_{A-D}*** then place string 'WITHIN/ON' in the "State_of_Character" field. ELSE IF ***COL_WLK_{G-H}*** then place string 'LEAVING' in "State_of_Character" field.
    c. Look to the right of the string for article, pronoun {his, her} or proper noun.
        i. IF proper noun look to the right for a noun. IF noun found then add the proper noun and noun to the "Initial_Location" and "Final_Location" fields.
        ii. IF {his, her} search to right for noun. Add the pronoun and noun to the "Initial_Location" and "Final_Location" fields.
        iii. IF article search to the right for noun or adjective.

1. IF noun add article & noun the "Initial_Location"& "Final_Location" fields
2. IF adjective search next word for noun. IF noun add article, adjective and noun to the "Initial_Location" and "Final_Location" fields.

5. IF ***COL_WLK_{E3}*** or ***COL_WLK_{F3}*** then:
   a. **CHAR_CHANGE_ACTIVE**
   b. IF ***COL_WLK_{E3}*** or ***COL_WLK_{F3}*** contains the string 'out' then place the string 'LEAVING' in "State_of_Character" field. ELSE IF ***COL_WLK_{E3}*** or ***COL_WLK_{F3}*** does NOT contain string 'out' place the string 'WITHIN/ON' in "State_of_Character" field.
   c. Look to the right of the string for article, pronoun {his, her} or proper noun.
      i. IF proper noun look to the right for a noun. IF noun found then add the proper noun and noun to the "Initial_Location" field.
      ii. IF {his, her} search right for noun. Add pronoun & noun to "Initial_Location" field.
      iii. IF article search to the right for noun or adjective.
         1. IF noun add article and noun to the "Initial_Location" field.
         2. IF adjective search next word for noun. IF noun add article, adjective and noun to the "Initial_Location" field.
6. IF ***COL_DR_{A-E}*** is found then:
   a. Place the ***COL_DR_{A-E}*** string found in the "Final_Location" field (from article).
   b. **CHAR_CHANGE_ACTIVE**
   c. Place string 'ENTER' in the "State_of_Character" field.
   d. Look to the right of the string for pronoun {him, her***P***}, proper noun or article.
   e. IF {him, her***P***} search to left for Proper noun. Add the proper noun to the "Secondary_Character_Involved" field.
   f. IF Proper noun. Add the proper noun to the "Secondary_Character_Involved" field.
   g. IF article search right for noun. Add article & noun to "Secondary_Character_Involved" field
7. IF ***COL_RM_{A-D}*** is found then:
   a. Place the ***COL_RM_{A-D}*** string found in the "Final_Location" field (from article).
   b. **CHAR_CHANGE_DESCRIPTIVE**
   c. Look to the right of the string for proper noun, article
      i. IF Proper noun add the proper noun to the "Character_Descr_in_Loc" field.
      ii. IF article search to the right for noun or adjective.
         1. IF noun add article and noun to the "Character_Descr_in_Loc" field.
         2. IF adjective search next word for noun. IF noun add article, adjective and noun to the "Character_Descr_in_Loc" field.
8. Traverse back to the beginning of the line for line number or time code. Place line number or time code in the "Occurrence_Time" field.