A Corpus-based Analysis of the Language of Audio Description

Motivations

Multimedia Content Analysis...

- In order to extract information about 'semantic video content', first we need to know what information is generally present in audio description and how it is generally expressed

Narrative...

- We are interested in what essential information must be conveyed in order for a story to be understood

Multimodality / Audiovisual translation

- We would like to characterise the 'end product' of the audiovisual translation. We are interested to see the effects of describers' training and guidelines.

Research Questions

- Is the language used for audio description a special language / sublanguage?
- What kinds of information about a film's story are provided by audio description, and in what regular ways is this information expressed?
- What kinds of things / events are commonly described?

Some factors impacting on the language used for audio description

- It refers to a restricted domain of discourse, which is what can be seen on-screen in films
- It fulfils a communicative function which is to provide enough information, objectively, for the audience to follow the story told by a film, without patronizing the audience by spelling out obvious information that could be inferred
- Specifically, we would expect it to include information about events in cause-effect relationships occurring in space and time, and about the characters involved in the events and their emotional states
- It must combine with existing dialogue which means, among other things, it must be concise
- These requirements are embodied in the training and guidelines followed by the professionals responsible for producing audio description.

Approach

- Functional explanations of language registers, such as special languages, seek to map between the communicative needs of language users and prevalent idiosyncratic linguistic features in the language. The use of language, by trained professionals, to communicate about a restricted field of discourse and for a specific purpose, normally results in a special language characterised by a preponderance of linguistic features that are idiosyncratic in comparison with everyday general language (Hoffman 1984).

Approach

- Following the kind of corpus linguistics approach described by Biber, Conrad and Reppen (1998), the analysis presented here begins to identify and describe a special language – 'the language of audio description' - in terms of statistically significant differences between linguistic features in a corpus of audio description scripts and a general language sample.

Approach

- A corpus-based analysis concentrating on unusually frequent words, word sequences and paradigms in audio description scripts

- The corpus comprises British English audio description of 91 mainly 'mainstream / Hollywood' films.
- No Part-of-Speech tagging, etc. [**More on this later**]

Overview of Method

1) Gather a representative corpus of audio description
2) Identify 'unusually' frequent words, with reference to a general language corpus
3) OPTION A: Examine concordances of these words

   OPTION B: Automatic statistical analysis of frequent word sequences and paradigms [**More on this later**]
4) Infer common kinds of information, or more particularly, common event types

What are the frequent words in the corpus?

Tentative Grouping of Open-Class Words in 300 Most Frequent Words

- Characters and their body parts: *man, head, face, eyes, hand, hands, men, woman, hair, arms, arm, feet, girl, mouth, boy, crowd, shoulder, officer, people, lady, body, police, soldiers, father*

- Actions: *looks, turns, takes, walks, goes, stands, steps, smiles, stares, puts, watches, opens, looking, runs, sitting, comes, picks, sees, holds, wearing, smile, nods, standing, leans, glances, gives, holding, watch, beat, grabs, leaves, falls, reaches, watching, drops, closes, lifts, throws, shakes, passes, run, follows, climbs, kiss, pushes, kisses, walk, lies, staring, carrying*

- Objects and scenes: *door, room, car, window, table, water, bed, house, floor, gun, boat, street, road, ground, horse, phone, desk, hat, office, book, bag, stairs, chair, seat, sky, fire, jacket, bedroom, corridor*

Identifying 'Unusually Frequent' Words

SL/GL Ratio (Ahmad and Rogers 2001)

- Calculated by dividing the *relative frequency* of a word in the special language (SL) corpus by its relative frequency in a general language corpus (GL).
- Relative frequency = the frequency of a word in a corpus by the total number of words in the corpus.
- If SL/GL ratio = 1 then the word is being used ‘normally’
- If SL/GL ratio = 50 then the word is being used relatively 50 times more often in the special language

Log Likelihood

- Another measurement of how much more relatively often a word occurs in one corpus compared with another
- See **http://ucrel.lancs.ac.uk/llwizard.html**

What are the *unusually* frequent words in the corpus?

Words with high SL/GL ratio (and frequency > 30)

SL/GL >100: *saunters, hurries, stares, shoves, clambers, straightens, gazes, kneels, scrambles, leans, glares, nods, periscope, strolls, crouches, tosses, blinks, trots, frowns, hurls, clunk, grabs, pulls, llama, watches, smashes*

SL/GL = 50-100: *unlocks, hauls, staggers, heaves, minion, stumbles, shakes, wipes, hesitates, pats, haired, lowers, pushes, wanders, crawls, grins, glances, flings, picks, flicks, slaps, hugs, smiles, sniffs, glides, scarecrow, sits, slams, rubs, pours, squeezes, diner, postman, spins, shuts, salutes, drags*

SL/GL = 25-50: *rips, walks, climbs, closes, sips, strides, slumps, gallops, flashback, leaps, knocks, throws, fades, stirs, rushes, kisses, tugs, creeps, jumps, dives, shrugs, crashes, lifts, turns, licks, opens, silhouetted, elevator, pauses, swings, sighs, bounces, stops, dials, swims, bangs, presses, slips, removes*

NB. Character names have been removed

Words with high SL/GL ratio (and frequency > 30)

Many of these words appear to be verbs, and in particular troponyms, i.e. verbs that express a particular manner of doing something:

saunters, hurries, shoves, clambers, straightens, gazes, kneels, scrambles, glares, strolls, crouches, tosses, blinks, trots, frowns, hurls, smashes, unlocks, hauls, staggers, heaves, stumbles, wipes, hesitates, pats, lowers, wanders, crawls, grins, flings, flicks, slaps, hugs, sniffs, glides, slams, rubs, pours, squeezes, spins, shuts, salutes, drags, rips, sips, strides, slumps, gallops, leaps, knocks, stirs, rushes, tugs, creeps, jumps, dives, shrugs, crashes, licks, pauses, swings, sighs, bounces, swims, bangs, presses, slips, removes

How are the frequent words used?

Kinds of Information in Audio Description: SUMMARY

- Manual inspection of concordances of unusually frequent words suggests that audio description concentrates on providing information about...
  - characters' appearances
  - characters' focus of attention
  - characters' interpersonal interactions
  - changes of location of characters and objects
  - characters' emotional states

Information About Characters' Appearances

- Often when a character first comes on screen in a film they are introduced in the audio description with a relatively simple description of their appearance, for example their age, clothing or some distinctive feature
- Some common phrases are:
  - *(woman* | *man*) + *in* + an item of clothing, or an age;
  - *(woman* | *man*) + *wearing* + an item of clothing
  - *(woman* | *man*) + *with* + a distinctive physical feature.
  - (*young* | *old* | *elderly* | *short* | *tall*) + (*man* | *woman*)

Information About Characters' Appearances

EXAMPLES

a man in a white T-shirt leans towards Jim

he sees a man and a woman in a red suit walk by

the door of a low-rise brick apartment building opens and a woman in her thirties steps out

a dark-haired man with a moustache stands at the door

an old woman with a pointed nose and wild, white hair stands in a gloomy room

Information About Characters' Focus of Attention

- The words *looks* and *looking* are often used in phrases like *looking at*, *looks at*, *looks up at*, *looks down at*, and *looks around*. These tend to give information about a character's current focus of attention; also *watches as*, *stares at*, *glances at, gazes at*.

EXAMPLES

Corelli looks at his men

Samuel looks at the blue and black picture

Iris looks at John curiously as he puts down his cup

Ricky stares at Frank intently

Aladar gazes at the green valley in wonder.

the girl watches closely as he puts the needle on the record

Information About Characters' Focus of Attention

- Phrases formed with the words *eyes* and *head* also indicate focus of attention.

EXAMPLES

young Parker keeps walking, his eyes fixed on Mary Jane

keeping her eyes fixed on Ben, she walks over

Willard opens his eyes and absently regards the fan

slowly she turns her head to face the door

Information About Characters' Interactions

- Phrases like *turns to*, *shakes hands*, *sits next to*, *their eyes meet*, *puts (his | her) hand on*, *gaze into each others eyes* and *turns away* are perhaps indicative of key moments in an interaction.

EXAMPLES

the captain turns to Gatlin

she turns to Drosoula, who glares at her

Pelagia moves into the room, then turns to look back at him

they shake hands, Diane nods once quickly, smiling

Stitch sits next to an elderly lady and takes her hand

their eyes meet for a moment, then she turns to close the curtain

Jules puts her hand on Ellen's shoulder

Information About Characters' Interactions

- Other key information about how an interaction between characters is proceeding is commonly given with phrases like *smiles at*, *shakes (his | her) head* and *nods* + adverb.

EXAMPLES

Luc smiles at her conspiratorially

as Ellen smiles at her father, her eyes moisten

Tess shakes her head and swallows nervously

Ellen looks steadily at him and shakes her head

Michael nods, tentatively

Prince John nods, approvingly

Annie nods and Tom grins to himself

Information About Locations of Characters and Objects

- Characters' changes of location, typically within a scene, are expressed with phrases including *goes (to* | *into* | *off* | *out)*, *walks (away* | *off* | *out* | *over) to* and *steps (towards* | *into* | *onto)*. These actions may be preceded by *stands up* or followed by *sits down*.

EXAMPLES

quietly she gets out of bed and goes to the window

Luca hangs his head and steps towards a bench with Mary

Corelli looks at him coldly, then turns and walks away

he stands up and goes back into the main room

Information About Locations of Characters and Objects

- The opening and closing of doors are frequently described – *(opens* | *closes) the door*, *door (opens* | *closes)*, and are often connected with characters entering and leaving scenes; similarly when a vehicle *pulls up* or *drives off*.

EXAMPLES

She opens the door to Richard

Chas closes the door, plunging Royal into darkness

the next day, a white car pulls up at the house

Information About Locations of Characters and Objects

- Objects change location, and sometimes ownership, when one character *hands* something to another, when a character *picks up* / *pulls out* an object, and when a character *puts* an object somewhere.

EXAMPLES

she picks up a jar from the kitchen table

she hands him a rucksack

Information About Characters' Emotional States

- Our corpus analysis reveals some commonly recurring phrases that seem to be used to convey characters' emotional states. The most straightforward way this is described is by saying a character *looks* or *is looking* followed by an adjective (e.g. *confused, shocked, surprised, thoughtful, troubled, uneasy, annoyed, puzzled, concerned, dejected*)

EXAMPLES

Mrs Mills looks confused, then recovers herself

Rebecca looks blissfully happy, Samuel doesn't

Noelle's looking faintly embarrassed

Information About Characters' Emotional States

- The words *smiles*, *stares*, *looks* and *walks* all occur very frequently in the corpus and can be modified to indicate an emotional reaction to the events affecting a character:

  smiles + (contentedly | fondly | happily | sadly | shyly | wryly)

  stares + (blankly | coldly | curiously | proudly | uncertainly | in confusion | in disbelief)

  looks + (anxiously | nervously | desperately)

  walks + (briskly | calmly | slowly | stiffly)

EXAMPLES

Vianne turns to Anouk who smiles contentedly

he stares in confusion at the mass of demonstrators

Showing no emotion, he ... walks calmly into the corridor

Information About Characters' Emotional States

- Actions involving characters' *heads*, *faces* and *eyes* also give information about their emotional states:

EXAMPLES

Billy's young face breaks into a wide smile

she leans back and her face crumples in despair

Thurman strides across to the dock, his head held high

Ellen's eyes fill with tears and she smiles sadly at Kate

Conclusions

- We were successful in identifying some idiosyncratic features (unusually frequent words and phrases) of what appears to be a special language of audio description. These features can be explained by considering the audio description's restricted domain of discourse and its need for concision and objectivity.
- The degree of systematicity observed in the audio description scripts should be encouraging to those who produce guidelines for audio description and those who train audio describers. It is also encouraging for those seeking to develop language technologies for ‘assisted audio description’ and for repurposing audio description to index digital video archives.
- Our investigation began to create an empirically-grounded overview and classification of the main kinds of information provided by audio description.

Need for further work

- Corpus-based approaches must be complemented by other ways to understand the processes of audio description. Future investigations would do well to follow the advice of Piety (2004) and analyse audio description with respect to the visual and audio content of the films it describes.
- Would we get similar results with audio description from different countries?
- Our analysis to date has concentrated on characters' states and actions; we need also to analyse spatial and temporal information by looking at the usages of closed-class words; we began to do this for temporal information in (Salway and Tomadaki 2002)

The need and opportunity for further automation

TO BE CONTINUED AFTER A BREAK

+ QUESTIONS / DISCUSSION TIME

Another look at '*looks*'

- From an initial analysis of its concordance, it seems that '*looks*' is used in two distinct ways:

1) To describe a character looking at someone or something. In this case the sequence is:

(CharacterName | *he* |*she*) + *looks* + *(at* | *across* | *around)*

Other words are used this way, e.g. *glances* and *stares*

2) To describe a characters' mental / emotional state. In this case the sequence is:

(CharacterName | *he* |*she*) + *looks* + a word ending in *-ed*

'*man*' and '*woman*'

- From their concordances we see some common patterns around '*man*' and '*woman*'...
  - Both are often preceded by *(the* | *a)*, sometimes with *(young* | *middle-aged* | *old)*, or with *(white* |*grey* | *dark)-haired*
- But also some differences...
  - In between *(the* | *a)* and *man* we see *bearded*, *handsome*, *bald*
  - In between *(the* | *a)* and *woman* we see *brunette*, *beautiful*

What's happening in this analysis of concordances?

1) Identification of recurring word sequences and word paradigms (group of words that appear in the same position in sequences)
2) An interpretation of the kinds of information that a particular word sequence conveys

The need for further automation

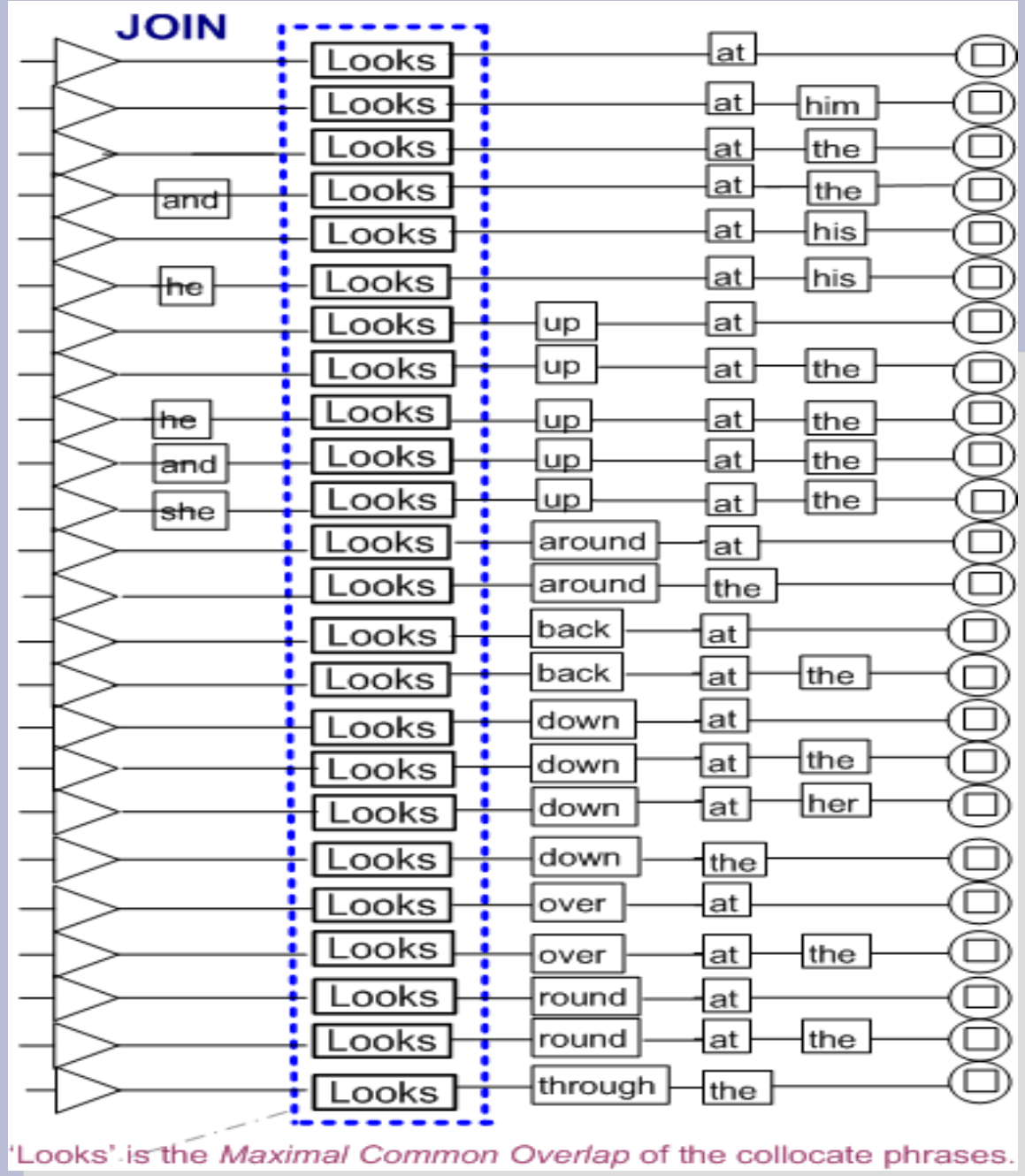
- An automated analysis should do better at identifying recurrent word sequences and paradigms, i.e. 'grammar induction'.
- Automation should mean improvements in:
  - Coverage
  - Detail and Precision
  - Consistency and Objectivity

The opportunity for further automation

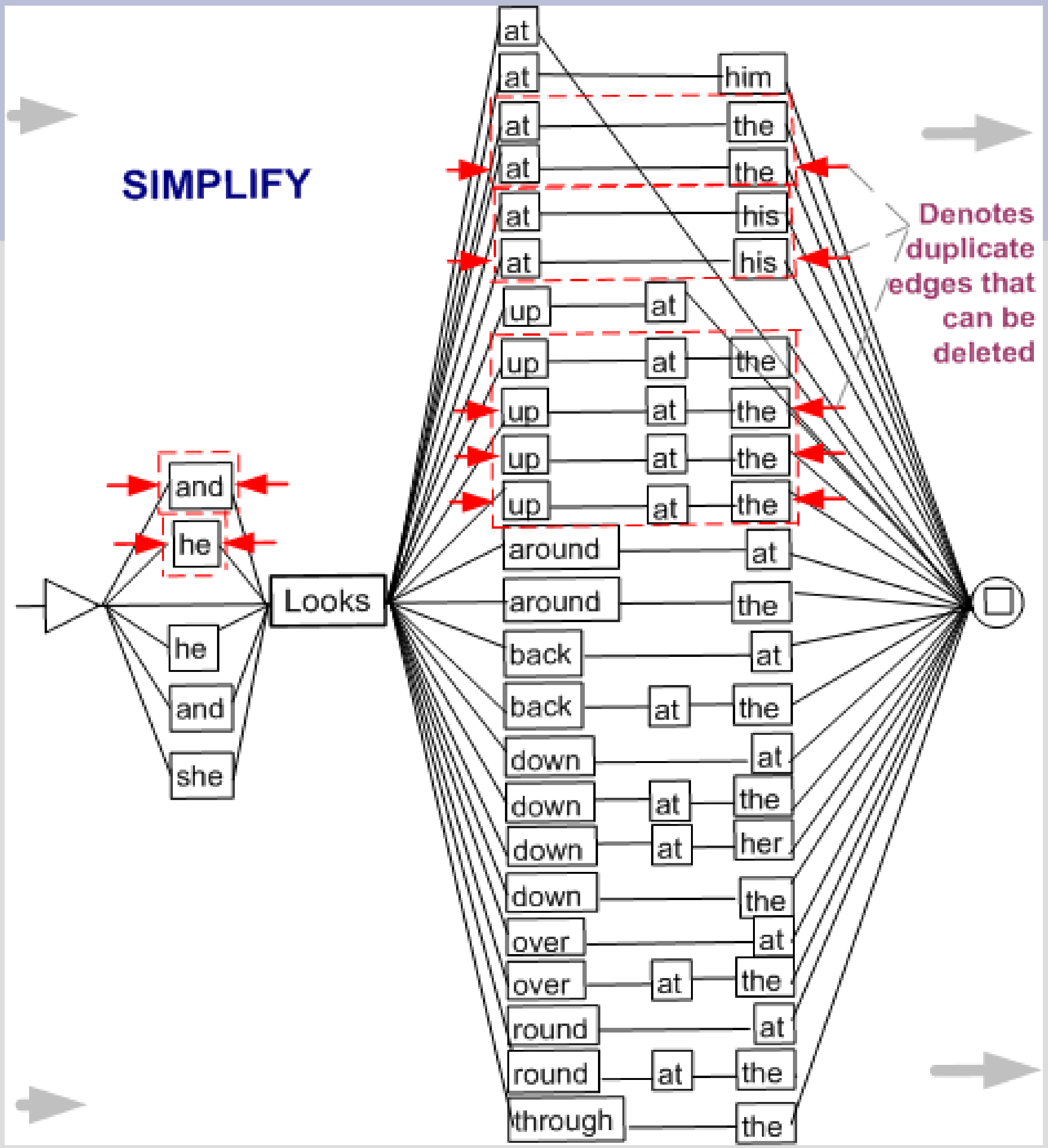
- Two recently proposed techniques that identify and describe frequent word sequences and paradigms:
  - **Collocation/Re-collocation analysis**: applied to the analysis of audio description by (Vassiliou 2006)
  - **ADIOS - Automatic Distillation of Structure:** an algorithm to induce grammar from unstructured / unannotated text data (Solan et al. 2005)

Collocation / Re-collocation Analysis

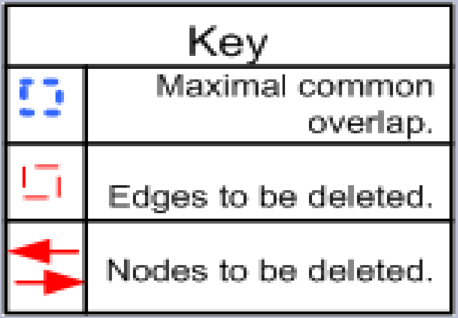
- Vassiliou (2006) adapted and extended an algorithm presented by (Ahmad, Gillam and Cheng 2005), and applied it to analyse a corpus of audio description; see also (Salway, Vassiliou and Ahmad 2005)
- The idea here is to identify collocations – statistically significant co-occurrences – of unusually frequent words, then collocations of the collocations, and so on.
- The algorithm then joins and simplifies collocations with common word sequences, and expands them by searching the corpus for other words that would fit in place of the original unusually-frequent word
- Output was used to design an information extraction system and create a database of film events



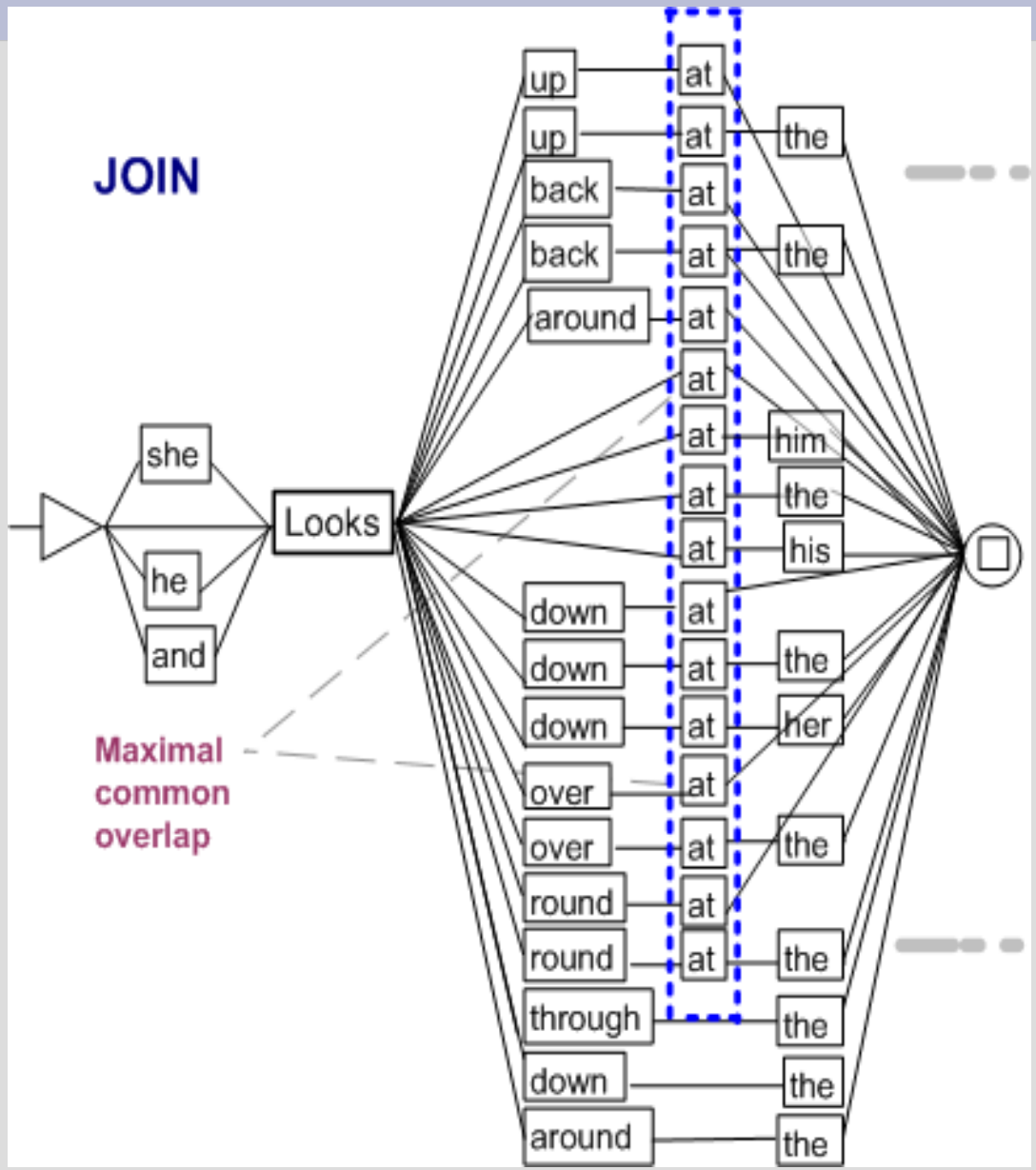
Key	
	Maximal common overlap.
	Edges to be deleted.
	Nodes to be deleted.



SIMPLIFY
and
he
he
and
she
Looks
at
at
at
at
at
at
up
up
up
up
up
around
around
back
back
down
down
down
down
over
over
round
round
through
him
the
the
his
his
at
at
at
at
at
at
the
the
the
the
at
the
at
at
the
at
the
her
the
at
the
at
the
the
Denotes duplicate edges that can be deleted

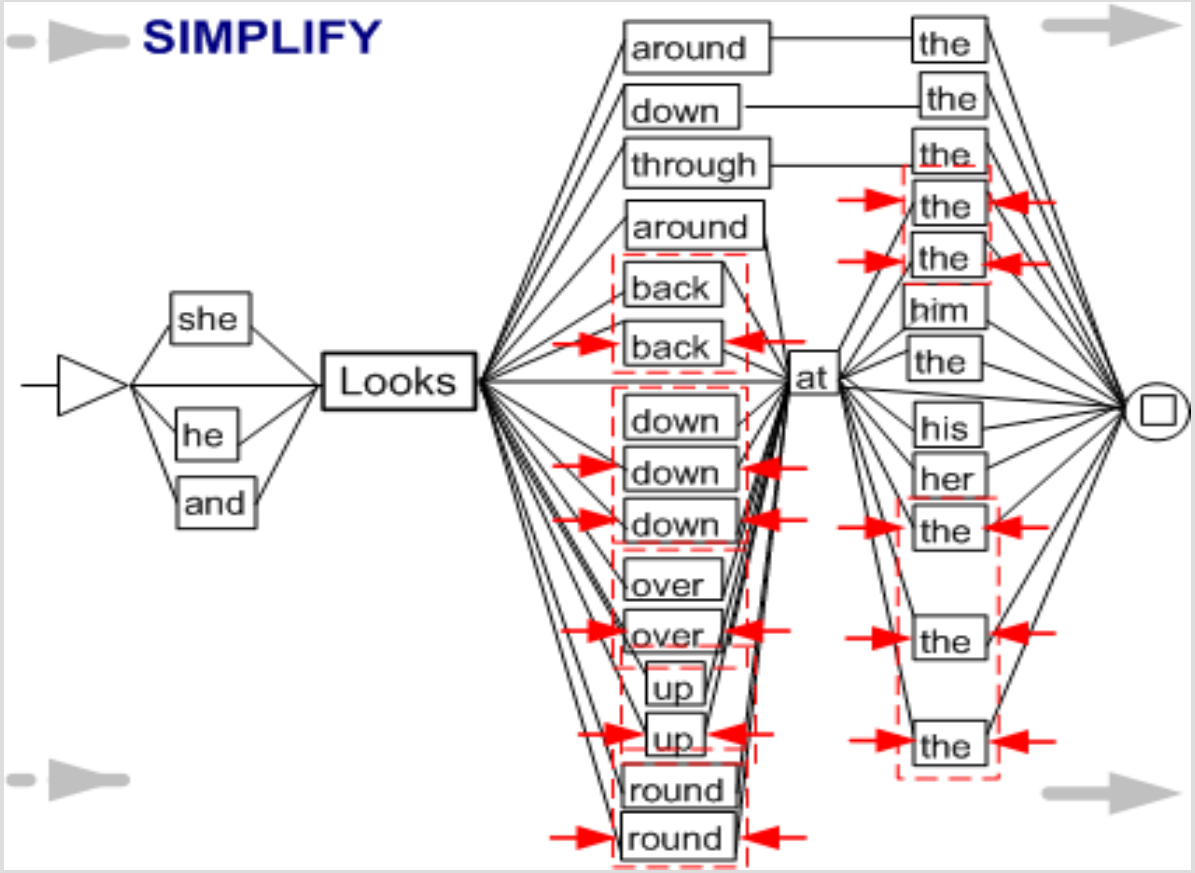


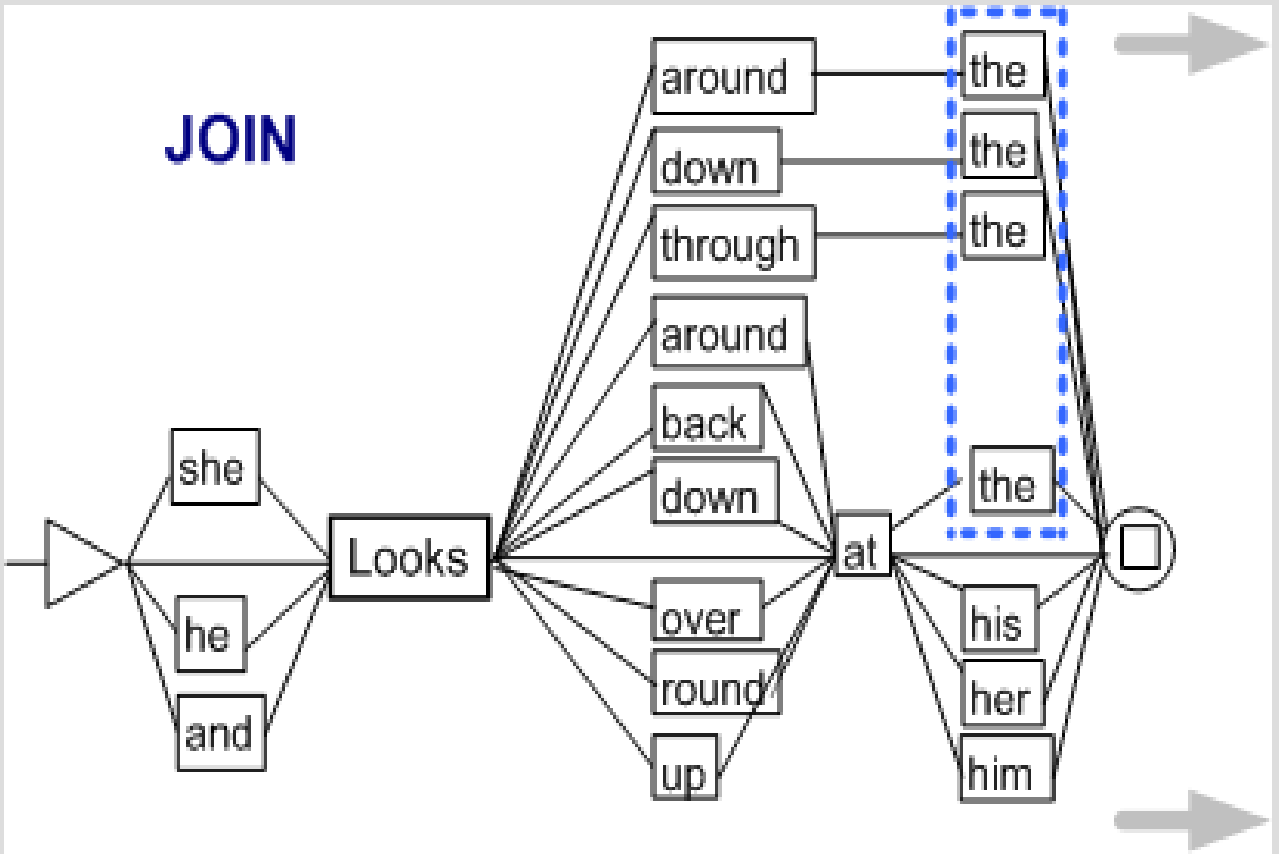
Key
Maximal common overlap.
Edges to be deleted.
Nodes to be deleted.

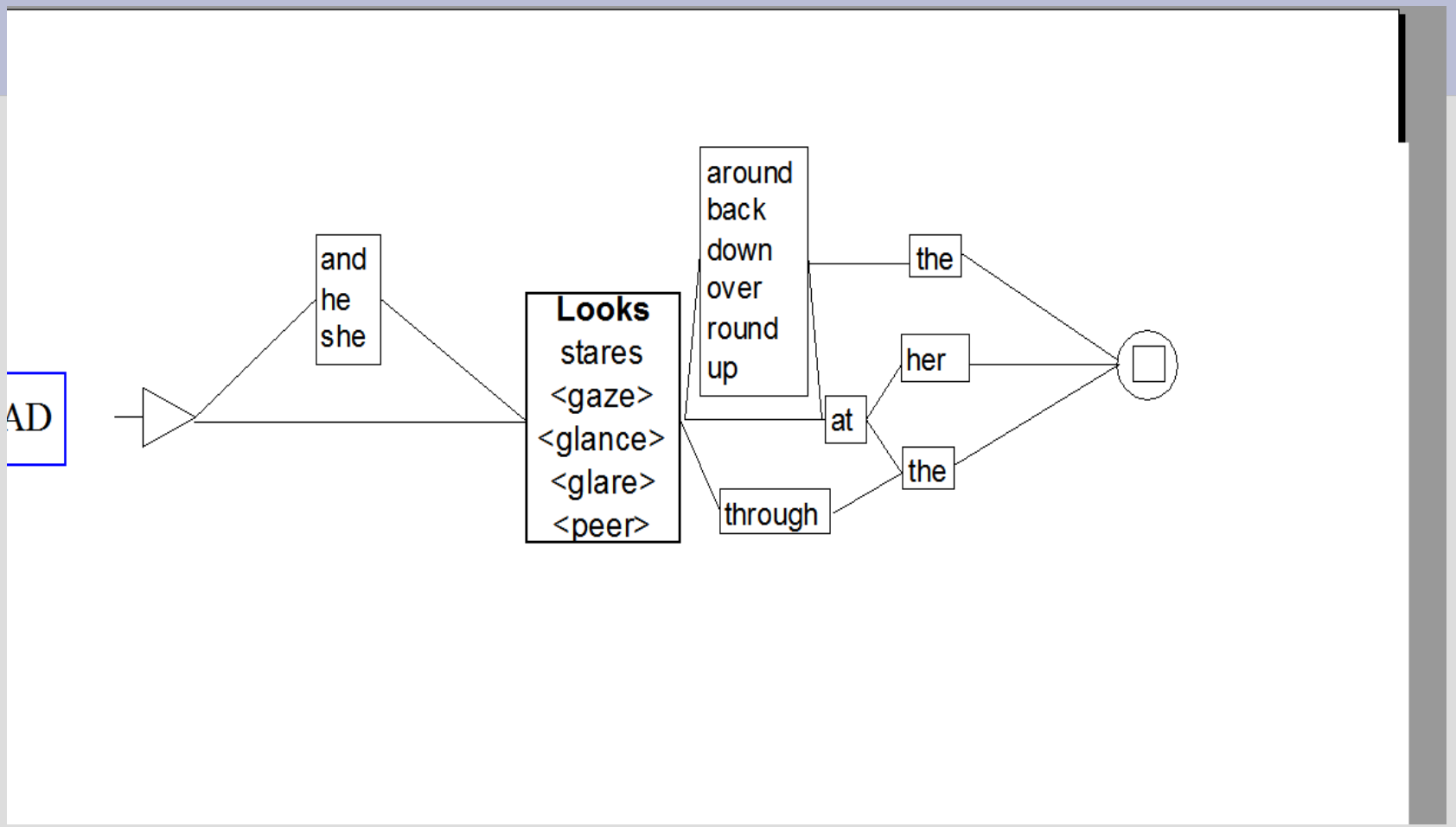


Key	
	Maximal common overlap.
	Edges to be deleted.
	Nodes to be deleted.

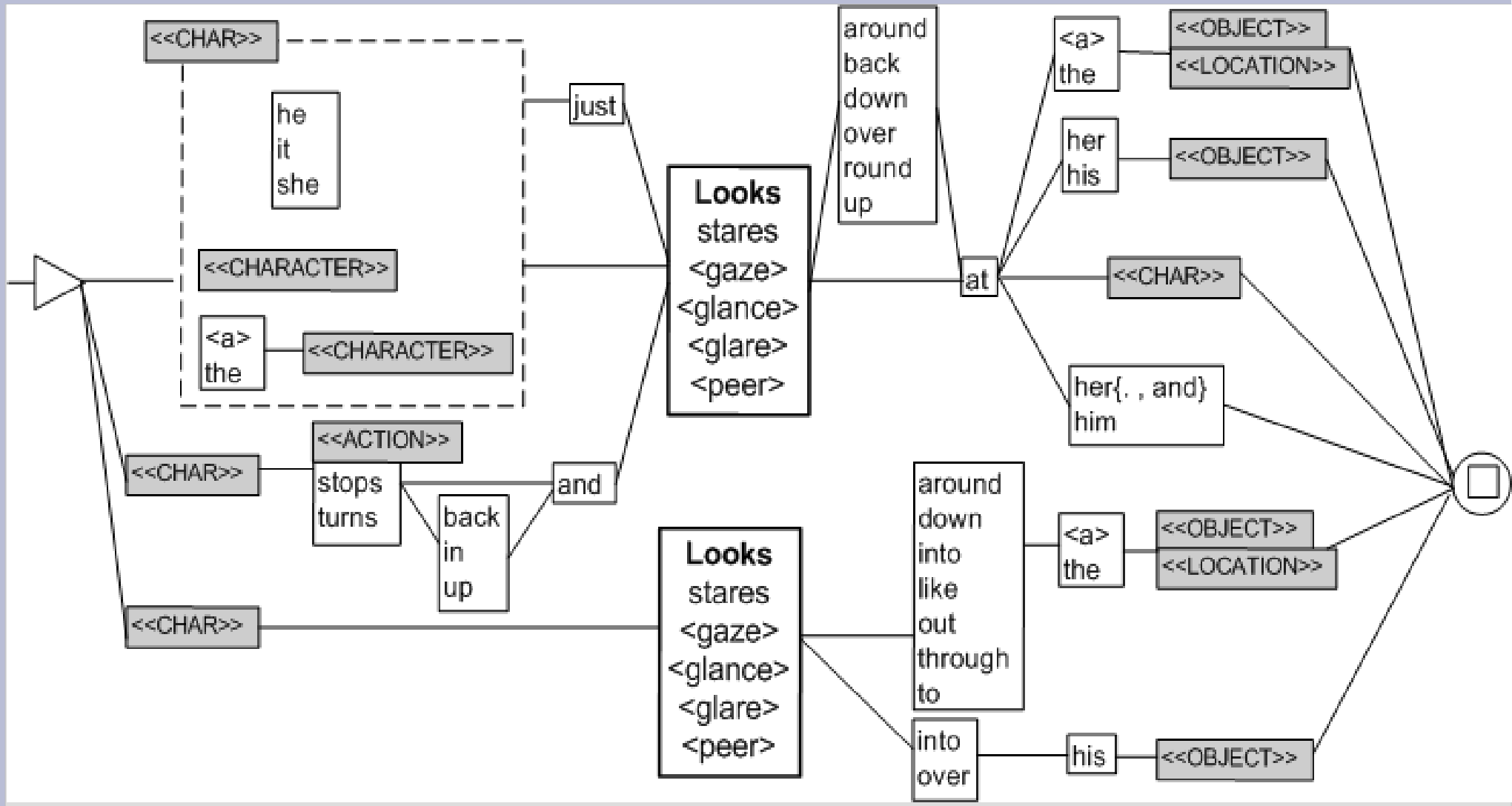
Key	
	Maximal common overlap.
	Edges to be deleted.
	Nodes to be deleted.



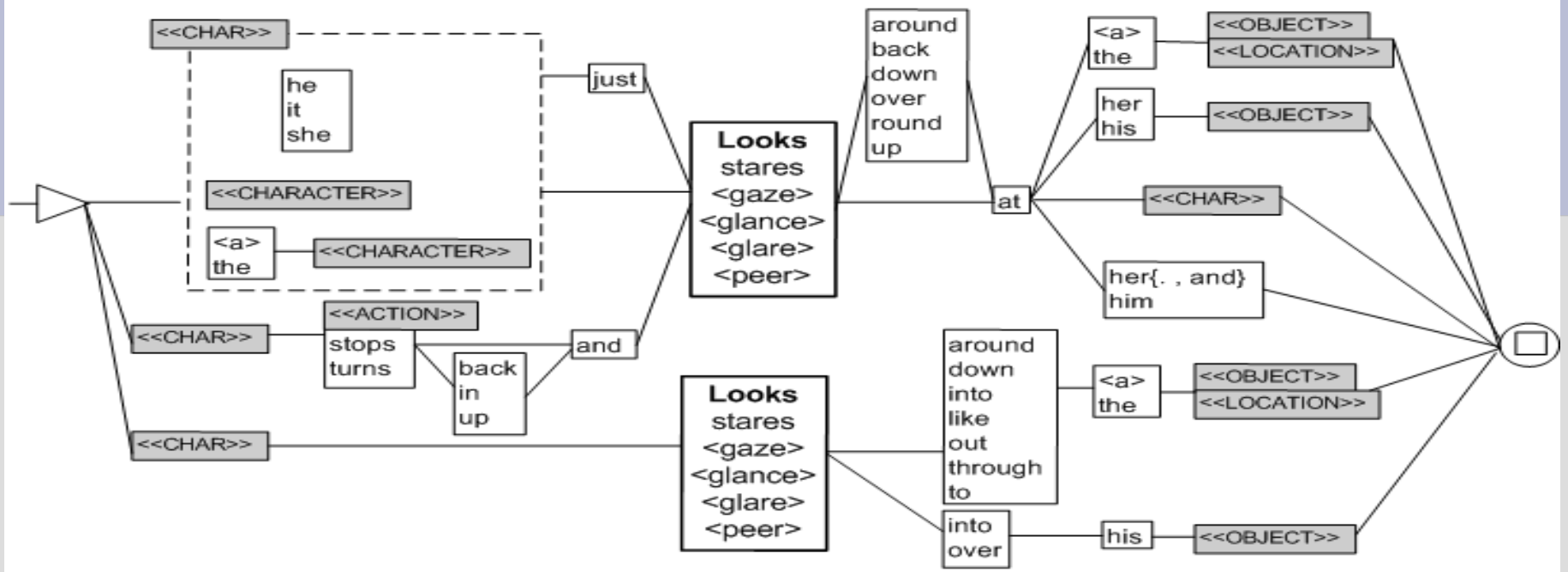

Key
Maximal common overlap.
Edges to be deleted.
Nodes to be deleted.
JOIN
she
he
and
Looks
around
down
through
around
back
down
over
round
up
at
the
the
the
the
his
her
him

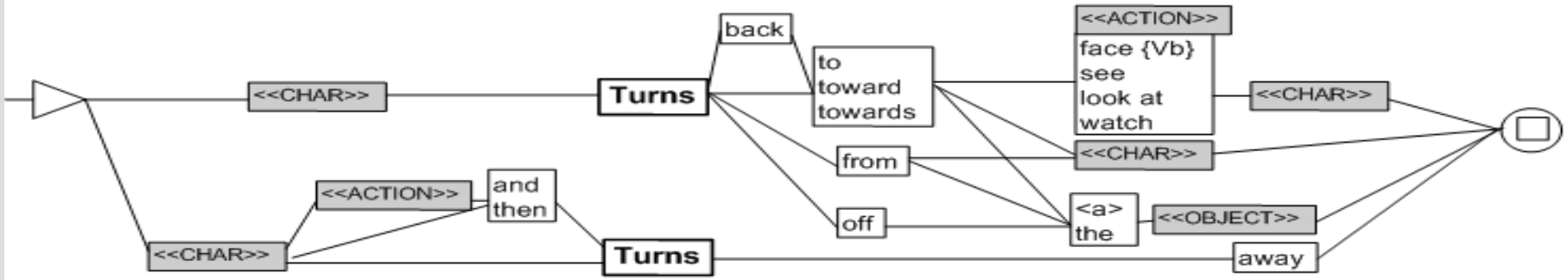

AD
and
he
she
Looks
stares
<gaze>
<glance>
<glare>
<peer>
around
back
down
over
round
up
at
through
the
her
the

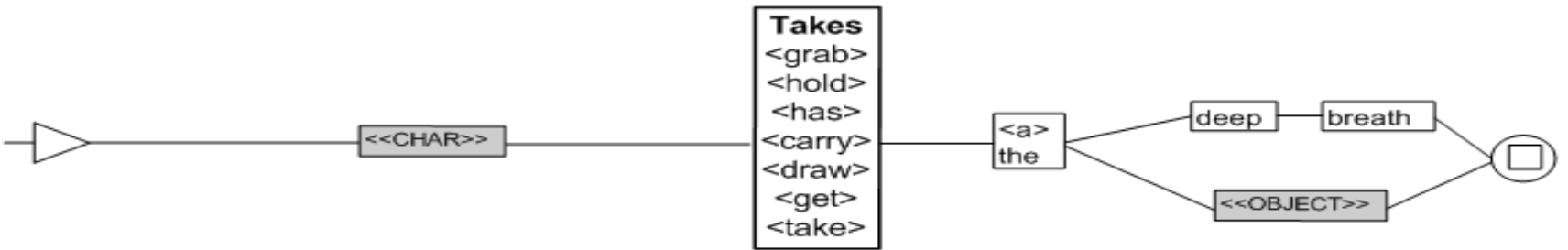
Additional Manual Analysis...



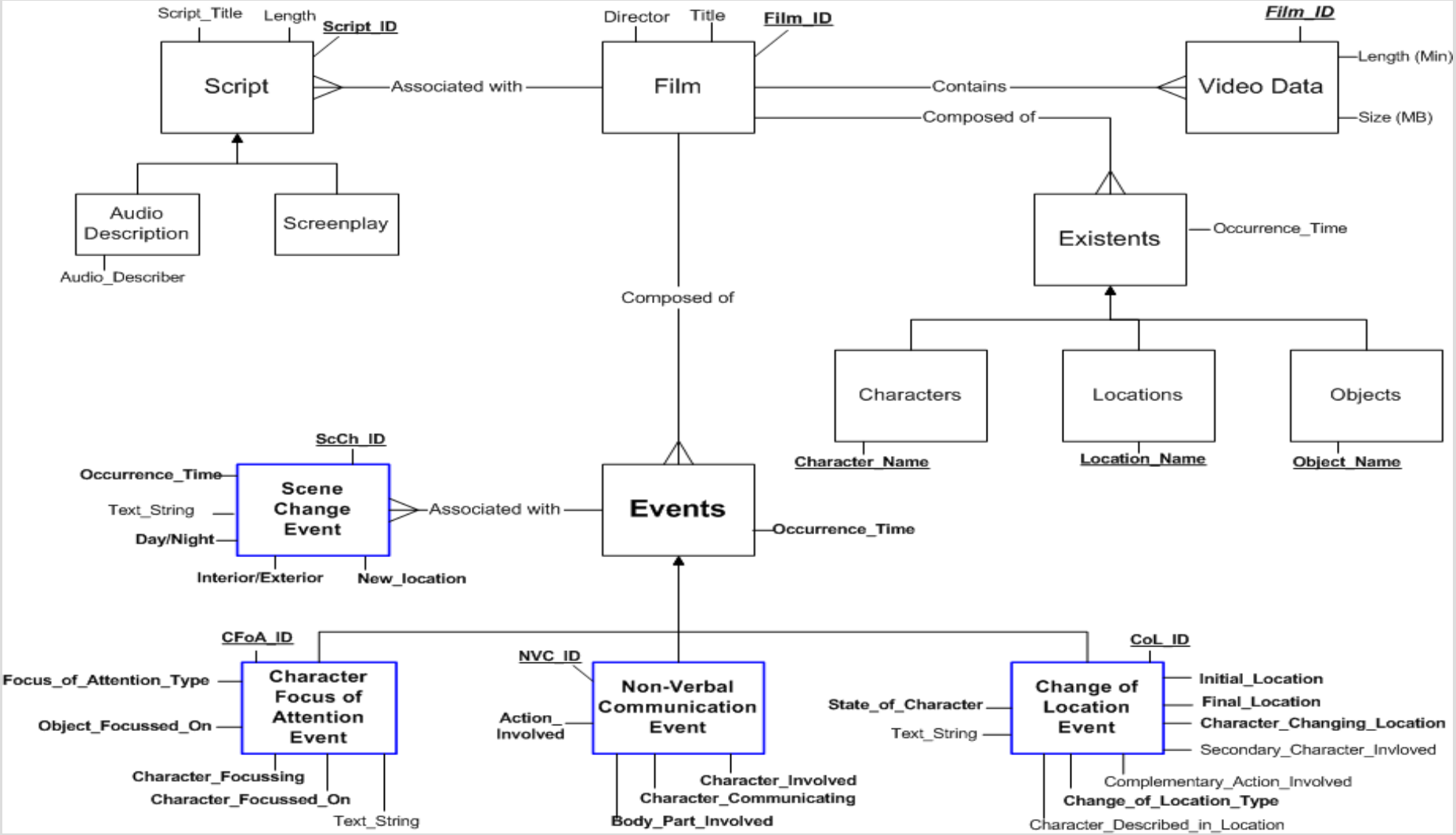
FOCUS OF ATTENTION







Four common types of event



An Automatically-extracted Database of Film Events

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
FOAAD7	PASSIVE	Jim		the gun	00:25:11:36	00:25:11:36 takes

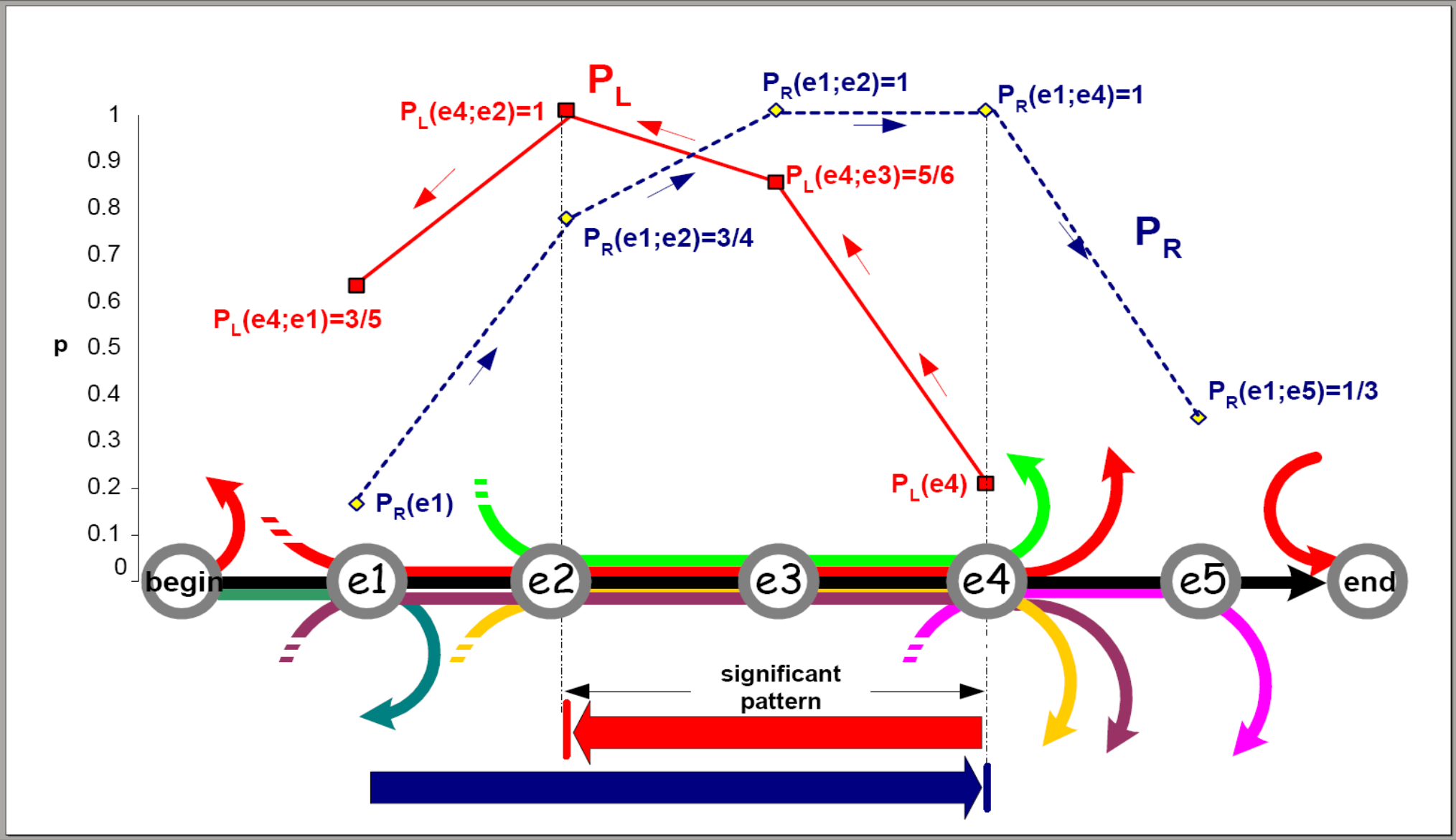
ScCh ID	INT\EXT	Location	Time of Day	Line No.	% Film Time	Text String
ScChAD1	Interior	FITTS HOUSE - RICKY'S BEDROOM	Night	2	0.07	INT. FITTS HOUSE - RICKY'S BEDROOM- NIGHT
ScChAD10	Exterior	SALE HOUSE	Day	322	11.25	EXT. SALE HOUSE- DAY

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

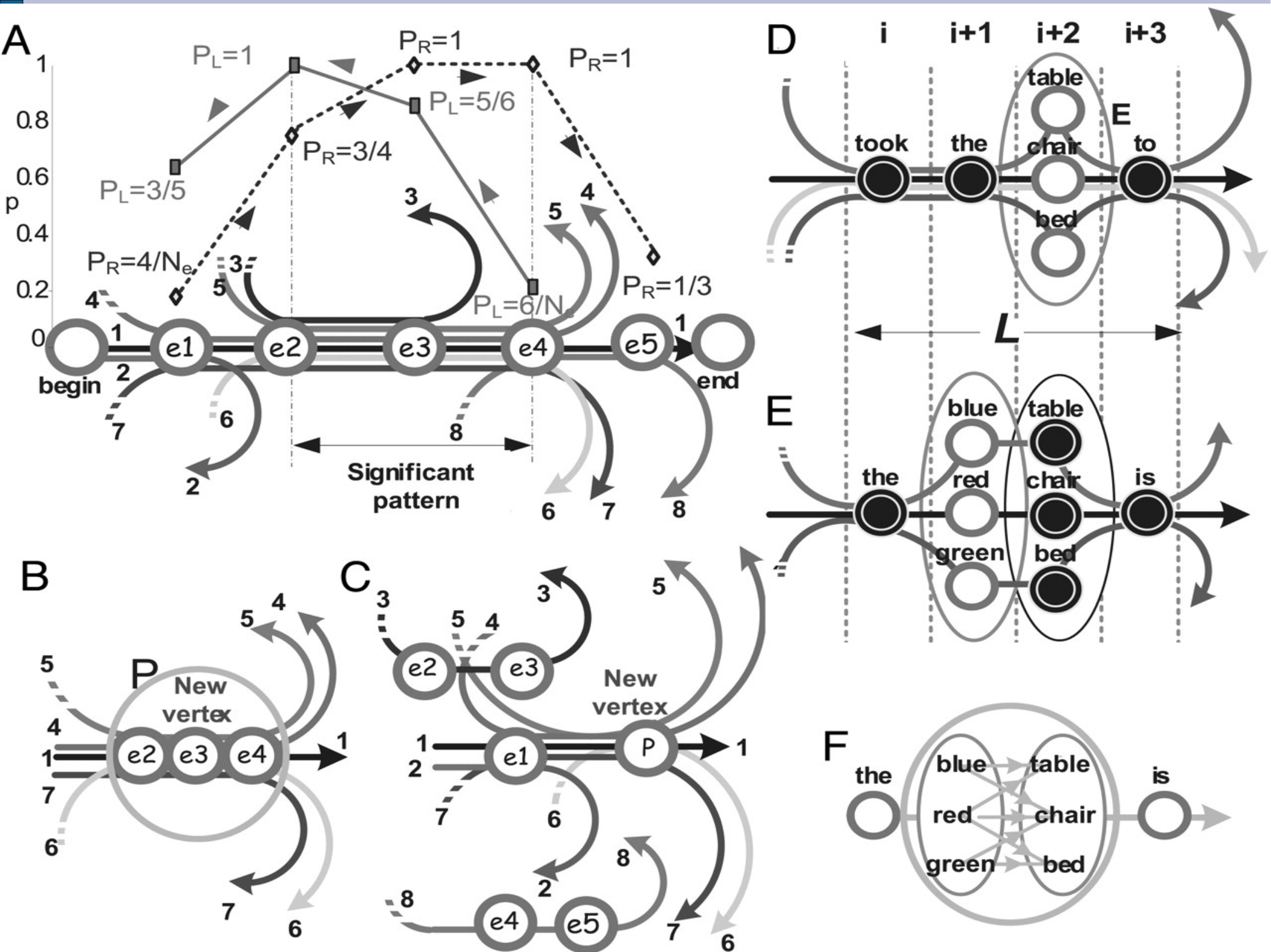
COL ID	Character State	Character Changing Location	Occurrence Time	Text String
COLAD2	ENTERING	Donkey	00:06:26:17	00:06:26:17 runs into
COLAD3	LEAVING	Donkey	00:10:51:06	00:10:51:06 walks out
COLAD4	WITHIN\ON	Donkey	00:11:55:04	00:11:55:04 walking over to

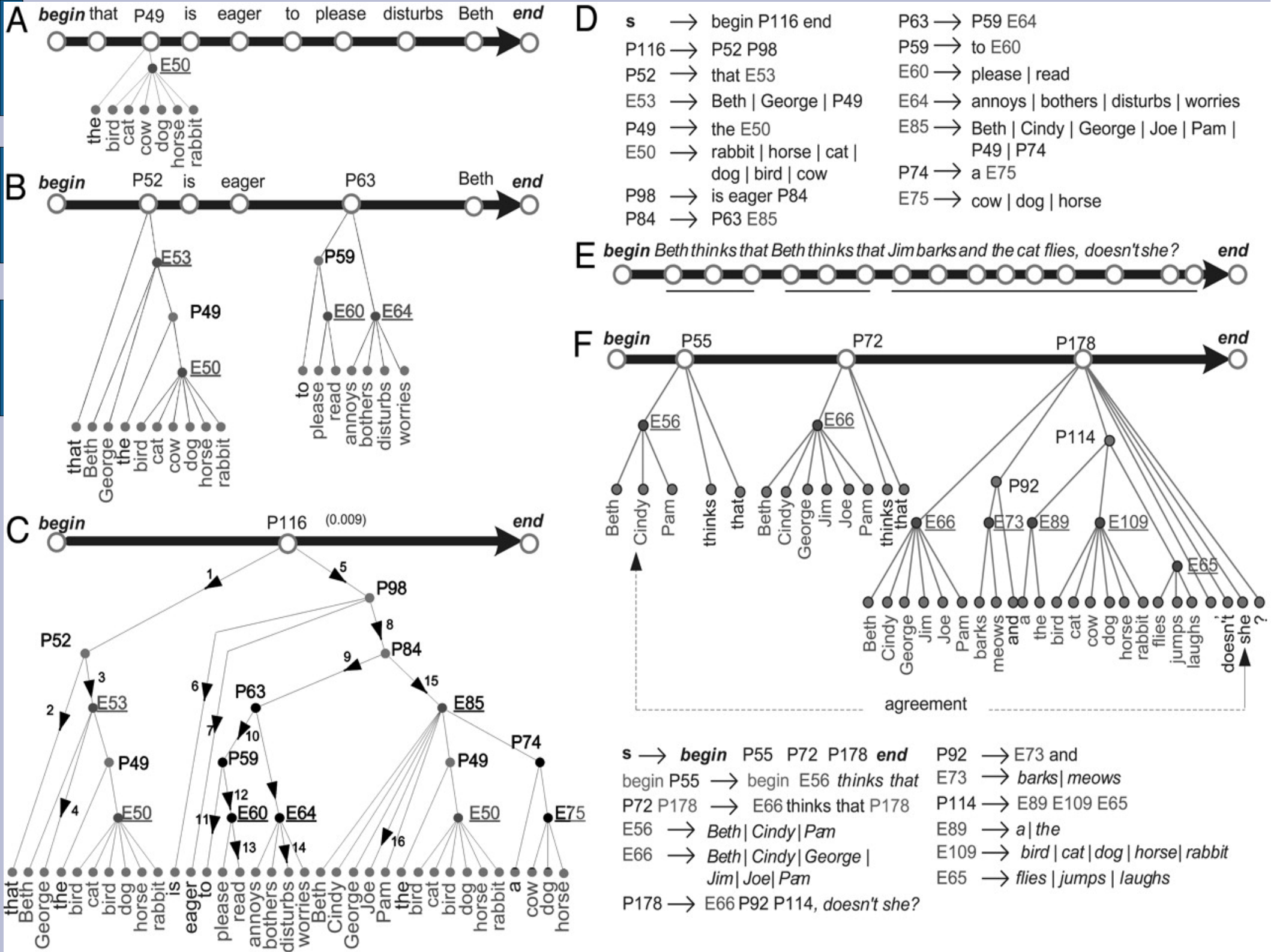
“ADIOS: Automatic Distillation of Structure” (Solan et al. 2005)

- “Linguists traditionally analyze sentences into recursively structured phrasal constituents; at the same time, a distributional analysis of partially aligned sentential contexts reveals in the lexicon clusters that are said to correspond to various syntactic categories (such as nouns or verbs).”
- “We introduce an unsupervised algorithm that discovers hierarchical structure in any sequence data, on the basis of the minimal assumption that the corpus at hand contains partially overlapping strings at multiple levels of organization.”
- http://www.pnas.org/cgi/content/full/102/33/11629



"The definition of a bundle that serves as a candidate pattern, whose beginning and end are signaled by the maxima of PL and PR. It becomes a candidate because of the large drops in these probabilities after the maxima, signifying the divergence of paths at


A
p
1
0.8
0.6
0.4
0.2
0
P_L=1
P_R=1
P_R=1
P_L=5/6
P_R=3/4
P_L=3/5
P_R=4/N_e
P_L=6/N_e
P_R=1/3
begin
end
e1
e2
e3
e4
e5
Significant pattern
B
P
New vertex
C
New vertex
D
i
i+1
i+2
i+3
took
the
table
chair
bed
to
E
L
E
the
blue
red
green
table
chair
bed
is
F
the
blue
red
green
table
chair
bed
is


A
begin that P49 is eager to please disturbs Beth end
E50
the bird cat cow dog horse rabbit
B
begin P52 is eager P63 Beth end
E53
P49
E50
P59
E60 E64
that Beth George the bird cat cow dog horse rabbit
to please read annoys bothers disturbs worries
C
begin P116 (0.009) end
P98
P52
P84
E53
P63
E85
P49
P59
P74
E50
E60 E64
E50
E75
that Beth George the bird cat bird dog horse rabbit is eager to please read annoys bothers disturbs worries Beth Cindy George Joe Pam the bird cat bird dog horse rabbit a cow dog horse
D
s → begin P116 end
P116 → P52 P98
P52 → that E53
E53 → Beth | George | P49
P49 → the E50
E50 → rabbit | horse | cat | dog | bird | cow
P98 → is eager P84
P84 → P63 E85
P63 → P59 E64
P59 → to E60
E60 → please | read
E64 → annoys | bothers | disturbs | worries
E85 → Beth | Cindy | George | Joe | Pam | P49 | P74
P74 → a E75
E75 → cow | dog | horse
E
begin Beth thinks that Beth thinks that Jim barks and the cat flies, doesn't she? end
F
begin P55 P72 P178 end
E56
E66
P114
P92
E66
E73 E89
E109
E65
Beth Cindy Pam thinks that
Beth Cindy George Jim Joe Pam thinks that
Beth Cindy George Jim Joe Pam barks meows and a the bird cat cow dog horse rabbit flies jumps laughs doesn't she ?
agreement
s → begin P55 P72 P178 end
begin P55 → begin E56 thinks that
P72 P178 → E66 thinks that P178
E56 → Beth | Cindy | Pam
E66 → Beth | Cindy | George | Jim | Joe | Pam
P178 → E66 P92 P114, doesn't she?
P92 → E73 and
E73 → barks | meows
P114 → E89 E109 E65
E89 → a | the
E109 → bird | cat | dog | horse | rabbit
E65 → flies | jumps | laughs

Summary

- These automated techniques offer the potential for much more precise and detailed descriptions of the language of audio description.
- Specifically, they can automatically describe a 'local grammar' of audio description.
  - The collocation/recollocation technique has already been applied to corpora of audio description and film scripts by Vassiliou (2006), and to a corpus of subtitles (Lingabavan and Salway 2006).
  - I am currently working to implement the ADIOS algorithm, in order to apply it to audio description**...**