The Role of Terminology and Local Grammar in Video Annotation

Mohammed S. Al-Athel

Submitted for the Award of
Doctor of Philosophy from
University of Surrey

Department of Computing
School of Electronics and Physical Sciences
University of Surrey
Guildford, Surrey GU2 7XH, UK

May 2008

Abstract

The linguistic annotation of video sequences is an intellectually challenging task involving the investigation of how images and words are linked together, a task that is ultimately financially rewarding in that the eventual automatic retrieval of video (sequences) can be much less time consuming, subjective and expensive than when retrieved manually. Much effort has been focused on automatic or semi-automatic annotation.

Computational linguistic methods of video annotation rely on collections of collateral text in the form of keywords and proper nouns. Keywords are often used in a particular order indicating an identifiable pattern which is often limited and can subsequently be used to annotate the portion of a video where such a pattern occurred. Once the relevant keywords and patterns have been stored, they can then be used to annotate the remainder of the video, excluding all collateral text which does not match the keywords or patterns.

A new method of video annotation is presented in this thesis. The method facilitates a) annotation extraction of specialist terms within a corpus of collateral text; b) annotation identification of frequently used linguistic patterns to use in repeating key events within the data-set. The use of the method has led to the development of a system that can automatically assign key words and key patterns to a number of frames that are found in the commentary text approximately contemporaneous to the selected number of frames. The system does not perform video analysis; it only analyses the collateral text.

The method is based on corpus linguistics and is mainly frequency based – frequency of occurrence of a key word or key pattern is taken as the basis of its representation. No assumptions are made about the grammatical structure of the language used in the collateral text, neither is a lexica of key words refined.

Our system has been designed to annotate videos of football matches in English and Arabic, and also cricket videos in English. The system has also been designed to retrieve annotated clips. The system not only provides a simple search method for annotated clips retrieval, it also provides complex, more advanced search methods.

Email: M.Al-Athel@surrey.ac.uk

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

Acknowledgement

I would like to express my deep and sincere gratitude to my supervisor, Professor Khurshid Ahmad for his wide knowledge, support, patience and guidance throughout my studies. His understanding and encouragement have provided the motivation for my dream to come true.

Also, special thanks to the Computing Department secretaries, Lydia and Sophie, for their unlimited support.

I gratefully thank David Cheng for allowing me to use his program "Text Analysis". As well, I gratefully thank Yousif Almas for allowing me to use and evaluate his program "LoLo: A System based on Terminology for Multilingual Extraction". Also, I owe a special debt of gratitude to Professor David Retterer from Ohio Northern University, for his Microsoft Visual Studio teaching during my undergraduate years; this escaped prisoner extends his thanks.

My final and main acknowledgment goes to my family. My parents who made all this possible for me, without them and their support and help, none of this would have happened. Last but not least, my wife, my son, and my daughter who were my inspiration throughout my study.

Contents

List of Figures

List of Tables

List of Equations

Chapter 1

1 Introduction

1.1 Introduction

Much research has been performed on analysing streaming videos. The techniques and methods that are being used have made considerable progress, but more is to be done. In producing summaries of videos or indexing video frames, one has to look at the structure of the events being imaged. The annotation of video (clips) at different levels of collateral linguistic description, including *keywords, phrases, sentences,* and *full texts*, is expected to improve the chances of retrieving still images (frames) from a collection of still images (Srihari and Zhang 2000). In addition, to the collateral use of written language, increasingly one sees the use of collateral audio streams accompanying a video stream in the indexing of news clips and sports events (Xie et al 2004). The annotation is to be used in conjunction with the visual features of the video – again at different levels of visual description – i.e., image, pictures and scenes. There appears to be a consensus emerging in the literature on the general topic of image annotation that a balanced approach to the use of the image-*external* linguistic metadata, expressed in written language as well as spoken (see Figure 1)

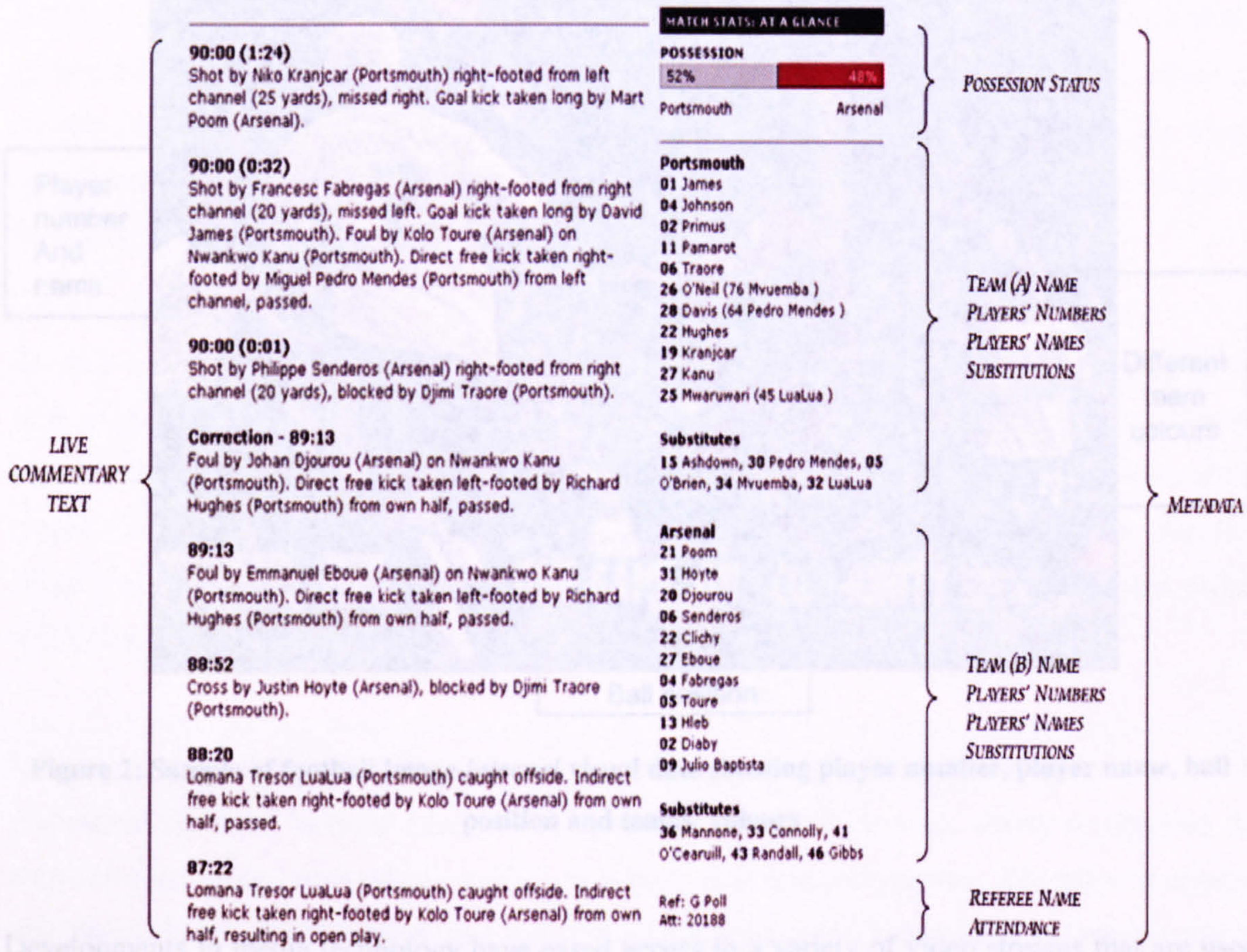


Figure 1: Sample of football image-external linguistic metadata showing the live commentary text, position status, players' names and number, referee name and attendance.

and image-*internal* visual data (scene, ball position, players and their positions in a football video); such visual data is used, for example, in specific object recognition technique (see Figure 2). This will perhaps improve the performance of systems for indexing and retrieving still (Zhou 2002) and moving images (see, for instance, Snoek and Worring 2005).

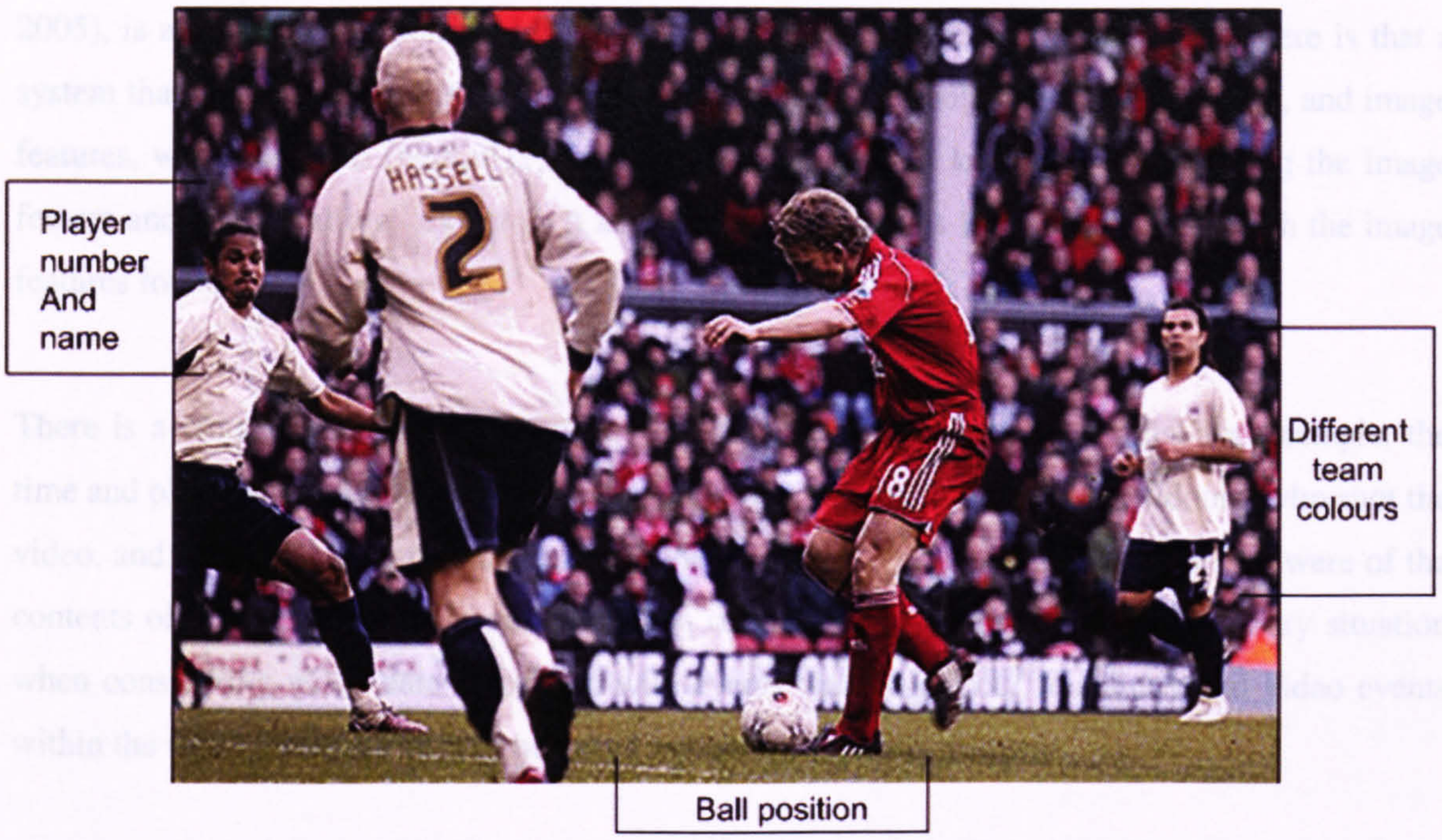


Figure 2: Sample of football image-*internal* visual data showing player number, player name, ball position and teams' colours

Developments in media technology have eased access to a variety of video streams that are used for informing, entertaining, surveillance and a number of other application areas. Video streams are essentially images, sometimes accompanied by a sound track. Within the video there may be text displayed to identify objects as a part of the local scenery. Current video processing techniques and video processing standards focus on indexing these videos for subsequent retrieval, and on the aggregation of video clips from video streams to compile albums of video clips focusing on an event or object. In order to test the hypothesis - that linguistic metadata may improve indexation and retrieval - one has to find such data. The search for, and the choice of, such data should be as objective as possible. The assignment of linguistic descriptors, say keywords or proper names of people or places in the image, on an ad-hoc basis poses two problems: First, how to test the performance of a system if the keywords, for instance, were chosen by the system builders. Second, how is the system to be updated when new sets of images are added, especially with new objects and events depicted therein?

The notion of *collateral texts* was developed to obviate the first problem and the argument was that an item of (written) language found in close proximity to an image, the image caption for example, will comprise keywords related to the object(s) in the image. The development of multi-modal systems that have a learning capability, including Bayesian learning (Barnard et al 2003), statistical learning (Li et al 2004), or neural networks algorithms (Saragiotis, Vrusias and Ahmad

2005), is a potential solution for identifying new objects or events. The argument here is that a system that has been trained to learn the association between collateral keywords, say, and image features, will be able to (partially) annotate an image without keywords by analysing the image feature and then recalling keywords, if any, that the system has learnt to associate with the image features found.

There is a considerable amount of metadata associated with video streaming, for example, the time and place where the video was shot, whether the video was indoors or outdoors, who shot the video, and so on. However, in order to use the metadata, a typical end user must be aware of the contents of the metadata associated with each of the images. This is not a satisfactory situation when considering video data associated with events that may occur frequently and video events within the video streams that may be very frequent or not at all frequent.

We know that experts working in visual domains, ranging from art criticism to sports commentators, from forensic scientists to home decorators, all have the ability to describe the contents of an image, be it still or moving, in a systematic and concise way. Somehow it appears that experts in the visual domain have background knowledge which they bring to bear in order to describe a given image. This background domain knowledge, if it appears in collateral text or audio description, can be used in order to identify and extract the domain special language and further more investigate the existence of a local grammar as we will see later in this thesis.

The research into video annotation includes videos of a variety of types: from video recordings of news broadcasts, complete with the news anchor and field reporters generating collateral keysounds to accompany the news events, to videos of sports events where commentators generate collateral sounds: in some cases, the sounds produced by balls in a ball game are also included as collateral (see, for example Liu et al 2006 for the use of collateral keysounds in a basketball match). For news events especially, the output of the anchors has low variability but that of sports commentators varies throughout one game and across different games. Both these events are generally produced by professional organisations where the emphasis is on excellent visual quality of the images. Then there are images of 'everyday' events where the contents do not vary much including weather reports and arbitrary events, shot by amateurs typically, like the taking off and landing of aeroplanes. We follow Snoek, Worring and Hauptmann (2006) to focus on video sequences that have high production consistency and high content variability – *sports events* (See Figure 3 and Table 1 below).

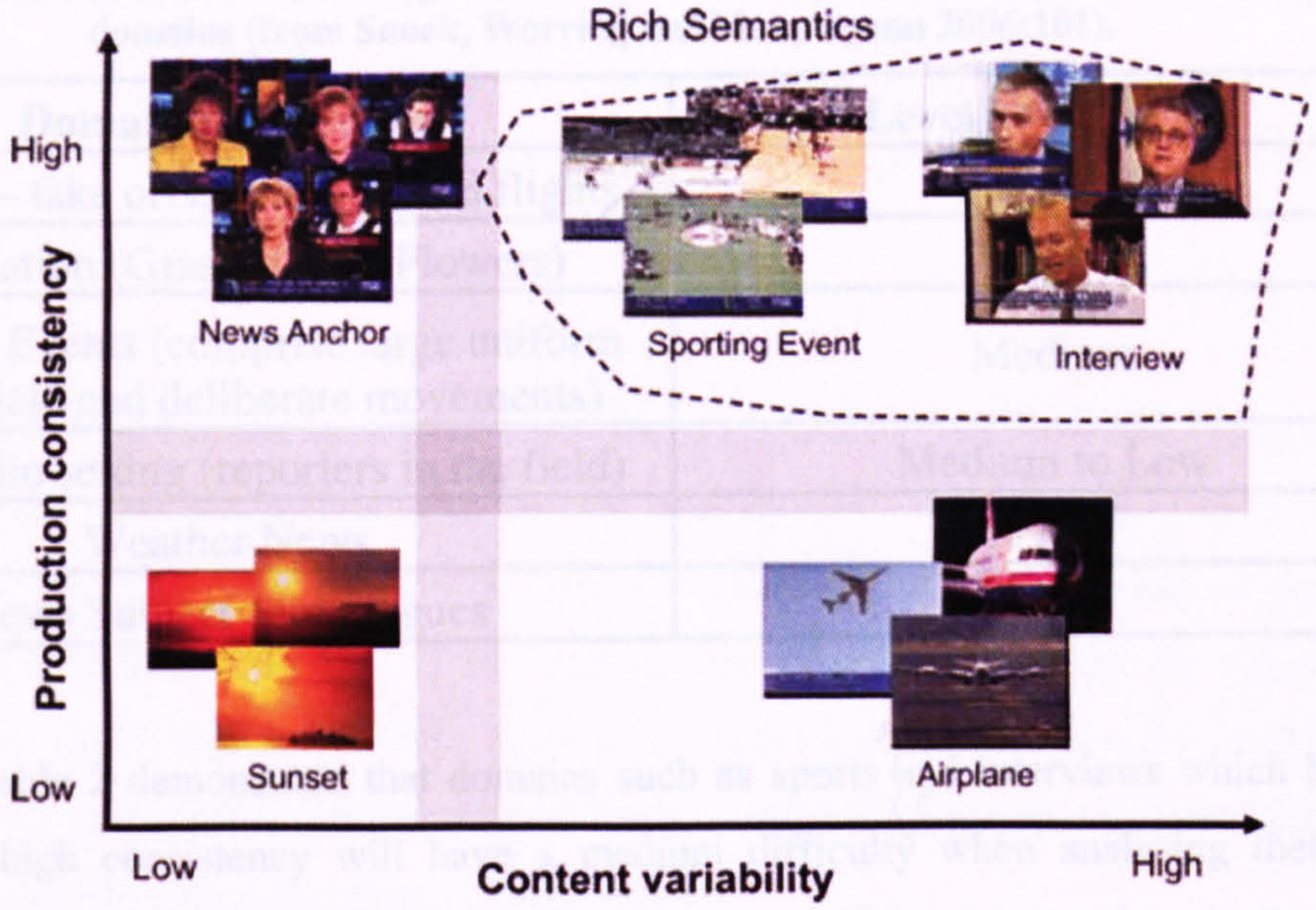


Figure 3: The contingency graph of content variability and production consistency of typical video sequences (based on Snoek, Worring and Hauptmann 2006, pp 91-93)

Table 1: The contingency matrix of content variability and production consistency of typical video sequences (based on Snoek, Worring and Hauptmann 2006, pp 91-93)

		Content Variability	
Production Consistency		Low	High
	Low	Diurnal Events (Sunrise; Sunset)	Arbitrary Events (Aeroplanes taking off)
	High	News Anchors	Sporting Events; Live Interviews

Given that sporting events usually comprise a uniform visual field and comprise deliberate movement of players (and referees/umpires) that have been learnt through training and experience, Snoek et al (2006) argue that such events have a 'medium level of difficulty' of interpretation (see Table 2).

Table 2: The use of domain knowledge and level of difficulty in video sequences from different domains (from Snoek, Worring and Hauptmann 2006:101).

Domain Knowledge	Level of Difficulty
Aircraft – take offs, landings, and flights	Very High
Vegetation (Grass, Trees, Flowers)	High
Sporting Events (comprise large uniform visual field and deliberate movements)	Medium
Non-studio setting (reporters in the field)	Medium to Low
Weather News	Low
News Subject Monologues	Low

Table 1 and Table 2 demonstrate that domains such as sports and interviews which have high visibility and high consistency will have a medium difficulty when analysing their videos. Consistency is a major factor when it comes to video analysis to determine the level of the difficulty. Airplanes, which have the same high level of visibility, have low level of consistency and that results in a high level of difficulty when it comes to video analysis.

For all specialist domains there is a language used by the domain community. One manifestation of such a language, often called sublanguage or special language of the domain, is that it has its own characteristic vocabulary. The vocabulary of a sports commentator dealing with Formula 1 Grand Prix racing will involve ontology based on racing cars and the rules of the competition; the language of the cricket commentator will involve the language of cricket, together with the rules of cricket and relevant background knowledge. In both cases the commentator will have both an historical and working knowledge of the game. Sports commentators commentate on a game with a precise language that uses all the keywords; sometimes these keywords are explained and other times the keywords are assumed to be understood by the audience. What is more remarkable is that keywords are used in a template throughout that is repeated to indicate the onset or the conclusion of key events in the game. In motor racing, for instance, the commentator will often announce at the beginning the positions of various drivers on the grid and the condition of the race track, and towards the end of the race the commentator will describe the relative positions of the various car drivers, the condition of the track and the condition of the motor cars.

Later in this thesis we will look closely at the MUMIS project (Multimedia Index and Searching Environment) (Declerck et al 2001) which is very similar to our system. However, it is important to notice the major difference between the two systems is the manner in which the keywords are

chosen. In MUMIS project, keywords are chosen manually whereas in our system the keywords are chosen automatically by the system itself; and this point is one of the essential points in our contribution. Later on we will answer this question: How can a system choose the keywords automatically?

In this thesis an attempt will be made to provide a mechanism that analyses video streams via their annotations, specifically live commentary written text. The goal is to provide a robust system that can be integrated with other video streaming analysis applications to provide more detailed and useful information; a system that can be used as a front-end for video streaming analysis or that can be used on its own. In either case, the system will show the usefulness of the annotation that is provided. It is true that a picture is worth a thousand words, but at the same time a single word can describe a sequence of video frames.

1.2 Contribution

My main contribution was to explore and exploit the utility of ball-by-ball commentaries that are increasingly being made available with videos of popular sports like football and cricket. These collateral texts, available in English and Arabic and perhaps other languages, describe the various events on the (football or cricket) ground in a special language of the domain of a particular sport. This language comprises a terminology and a set of grammar rules that govern the use of the terms. I have shown that these collateral texts can be used for the automatic identification of single and compound terms and that frequently used phrases comprising the terms present the evidence of a local grammar. These local grammar rules can be used to query a data base of videos containing football (or cricket) in order to look at a compilation of image sequences. These sequences may comprise one player over many matches, many players over one match, or many players over many matches involved in a specific set of actions. The use of collateral texts for identifying terms and for identifying key clauses (governed by a local grammar) is perhaps first in this rather busy field of sports-video indexing where key terms are provided and the end-user is expected to frame his or her own query. My video indexing program is an adjunct contribution in that it helps to test the efficacy of my method and is a practical tool in itself; I intend to see whether this system can be exploited commercially. A related contribution is the multi-lingual approach: much of the sports-video indexing and retrieval work is in English: I initially used Arabic collateral texts for evaluating my approach and found that my method works well with this language that is typologically distinct from English.

The cross-lingual approach is a contribution to the evaluation methodology for a text analysis and information extraction system. I have also evaluated the outcome of my analysis using two very different team sports. Football, which typically lasts for under two hours and where almost all the members of the team are active for two fixed durations – before and after the half-time interval. Cricket is quite different in this respect: the two main protagonists are the bowler and one of two batsmen whilst 9 members of the batting team are waiting their turn. Every 5 or 6 minutes there is a break when 6 balls have been bowled (called an 'over') and the game can last for 1, 3 or 5 days. Cricket and football do not usually have the same spectators. The choice of a totally different sport (cricket) for evaluation tests the evaluation methodology for text analysis and information extraction systems.

I have used a text analysis strategy and associated programs developed by a number of researchers and graduate students at the University of Surrey. The application of these programs to sports commentary analysis has not been undertaken as yet and I have suggested improvements to the existing programs.

1.3 Dissertation Structure

In the next chapter, a review of relevant literature on video annotation and related topics will be undertaken. In chapter 3, the corpus-based method will be presented in detail followed by the description of the implementation and evaluation in chapter 4. In chapter 5, the conclusion will be provided along with suggestions for possible future work in this area of research.

Chapter 2

2 Background and Motivation

The annotation of sports videos usually relies on the spoken word of the commentators – the running commentary that is a good example of creative use of language. The creativity lies in the emotion and excitement that the commentators are trained to put into the description of what they see happening on the ground and at other sporting venues. Liu et al (2006) have classified various sporting events in terms of the keysounds generated by commentators and referees with special reference to *tennis, soccer* and *basketball* (see Table 3).

Table 3: Games and associated keysounds (Liu et al 2006, p 8)

Game	Keysounds			
	Commentator Speech		Referee Sounds	
	Plain	Excited	Whistles	Announcements
Tennis	Beginning/End of a point		Score; fouls; service changes	Beginning and end of game
Soccer	Start and during the game	Goal scored, fouls committed	Free kicks/Penalty Kick; Beginning and end of game	
Basketball	Ditto	Points scored; Fast break, drive	Fouls, Beginning and end of game	

Looking back at Table 1 and Table 2, one can see more reasons why the sports' domains are considered to be at a medium level of difficulty. The additional benefits that come with the video such as commentator speech and other sounds can be a guideline to the occurring events.

On the basis of the keysounds and visual identification of 'court-view scenes' and 'offensive-defensive exchange interval' (ODI), Liu et al have devised a set of heuristics like the one shown in Figure 4 below:

***Input*: Shot classification, ODI information and audio keysound**

***Output*: The event label "foul" and "shot at the basket" for a scene**

IF The current scene is "court-view and non-ODI scene" or (the current scene is ODI scene and its neighbor scenes are non-court view scene) THEN

IF The audio keysound "whistling" has been detected and it does not occur at the beginning of the scene THEN

IF The next "court view scene" is an ODI scene THEN Event "offensive foul" detected

Figure 4: A heuristic based on multi-modal analysis of a video clip and the collateral commentary (Liu et al 2006: 8)

Here we see the use of 'domain' specific terms like *offensive, defensive,* and *foul* together with a compound *offensive foul*; elsewhere in their paper Liu et al talk about *defensive foul* and *shot at the basket.* For basketball enthusiasts such terms do not pose any problem: they belong to a *community* defined by its passion for the game and as a community they have a common language, or at least a common vocabulary. It can be argued communities of different varieties are defined by the existence of a common language within the community. We will say more about special languages later in the Method section of this paper.

In a related work by Tanaka-Ishii et al (1999), where the authors are interested in automatically generating ball-by-ball commentary for a game of robotic football, we see yet more examples of the use of specialist terminology and that of a set of 50 or so rules that comprise a local grammar; Gross (1993) defines local grammar as a way of describing the syntactic behaviour of groups of individual elements which are related but whose similarities cannot easily be expressed using phrase structure rules; to deal with different forms of text organization which occur within otherwise normal text. Researchers also emphasize the local patterns: Local Pattern can be defined as all the words and structures which are regularly associated with the word and contribute to its meaning. A pattern can be identified if a combination of words occurs relatively frequently (Hunston and Francis 2000). The authors distinguish between the so-called local *tags* and global *tags,* and within this distinction between event-based tags and state based tags. The tags are essentially specialist terminology and there are grammar rules of a kind that are used to order the specialist terms (see Table 4):

Table 4: Tanaka-Ishii's use of 'special language' for generating 'ball-by-ball' commentary for a robotic football game (Tanaka-Ishii et al 1998).

Terminology	Rules	
<u>Local *Tags* or terms</u>: *EVENT-BASED*: Kick, Dribble; Pass; *STATE-BASED*: Mark, Problematic Player: Player (goal-scoring) success rate	**Logical consequences:** Infer the consequences from a set of antecedants	**(High Pass- Success Rate player)** **(Pass Pattern player Goal) !(active player)** For inferring the consequence that "player is active" when his pass success rate is high and he has made shots on the goal.
<u>Global Tags or terms</u> *EVENT-BASED*: Change of Form or Side *STATE-BASED*: Team success rate; Score, Avg. distance ball passed	**Logical subsumption.** One proposition is subsumed by another	**(Pass player1 player2)(Kick player1)** **! (Less-important @2);** where Less-important @2 reduces the second antecedent proposition matched by the rule
	State change: State update is carried out in the same way as logical subsumption	**(Form team form1)** **(Form team form2)** **! (Less-important (earlier @1 @2)).**
	Second order relations. For establishing higher-order relations among propositions	**(High Pass Success Rate player)** **(Player On Voronoi Line player)** **!(Reason @1 @2)**

The rules are most frequently occurring significant patterns of play which were hand crafted and subsequently used to generate ball-by-ball commentary for the system developed by Tanaka-Ishii et al (1998).

2.1 Describing Movement

Movement is the act or an instance of moving; a change in place or position. Rittscher has argued that "Rather than understanding the perception of biological motion we intend to construct a machine that is able to recognize certain biological motions" (Rittscher et al. 2003: 475). Their approach to the modelling of movement is to create a "Generative model to recognize the observed biological motion, and recognize the type of motion directly from the spatio-temporal features of the image sequence" (2003: 475). These approaches are based on video analysis. Movement can be categorized into deliberate movement and stylized movement. One has to know the type of movement in order to describe it. For example, in a football game the fans in the stadium provide a movement but generally this is ignored by the researcher. Also, the players in the game are in continuous movement but only some of them including the player who has possession of the ball are relevant to the researcher.

Sports specialist domains contain a wide range of motion entities. Those entities vary from one sub-domain to another. For example, in swimming the main movement will be the movement of the head and arms; in ice-skating the main motions will be the movement of the arms, legs and body-jump. In our domain, football, the main movements will be header, pass, running, and foul kick. Sports movements are limited and governed by their specialist domain; it would be a surprise to see a football player performing a swimming movement during a match for example. Such domains may allow the substitution of one keyword with another but on limited bases. Other domains can be more flexible with keyword substitution. Snoek and Worring (2003) argue that to describe movement in a sports specialist domain, the sports type and the targeted movement have to be known first, in other words, one needs to know what to look for to make the task easier. Sports in general are described as deliberate movement. Someone can anticipate the type of movements to be taken at any given minute based on the domain of the sport, even though a very unusual movement may occur sometimes. A sport like football is, in general, scripted. The team coach chooses the players to participate, the formation of the team and the strategy to be played. However, on the field, how the players move is not and cannot be scripted.

Some sports are categorized as having stylized movement, such as figure skating. The music to be played and the skating to be performed are fully scripted. The participant(s) spend months practising the movements to be performed over and over again seeking perfection. Diving can also be classed as a stylized sport. Although these sports are categorized as stylized movements, they are no different than any other sport when it comes to video analysis or summarization. However, it is worth mentioning the different types of movements in sports in general.

2.2 Analysing Deliberate Movement

In general, sports videos cannot be accessed directly. Modelling the video is a step that one needs to take in order to retrieve any information. Such a step is considered one of the most important tasks in video analysis (see, for example, Navalpakkam et al 2003; Petkovic et al 2001; Petkovic and Jonker 2000; Rittscher et al 2003, and Tahaghoghi et al 2005). Many researchers have been working on enhancing and improving video modality. Rittscher (2003: 476) stated that "the task of recognizing a simple motion like walking, independently from scale and viewing angle, is an enormously challenging problem". The various algorithms that have been introduced to the field of video analysis are too many to list and each has its own characteristics. Looking into sports videos summarization and analysis, many techniques are presented and the majority have taken advantage of the fact that sports videos have predictable events and consistent features (Tahaghoghi et al 2005). Sports video analysis can be categorized as follows:

2.2.1 Video Analysis

This is the most analysed category whose researchers' intuition is to analyse the video without the usage of any other available sources. Many researchers have shown interest in intuitive analysis when analysing sports video. Some researchers are looking to extract a single object (Andrade et al 2003; Kojima et al 2000, and Wolf et al 2002). The techniques they have used for recognizing, indexing or retrieving are built on specific object recognition. Others (Adams et al 2002; Quenot et al 2002; Smeaton and Over 2002, and Snoek and Worring 2005) have shown different techniques which were based on event recognition. Still others (Nam and Tewfik 1999; Ronard and Thuong 2003; Snoek and Worring 2005, and Zhong and Chang 2000) focused on video indexing techniques. Intuitive analysis, analysing what you see, is not limited by all these methods alone; however, those are worth mentioning and recognizing. Regardless of the technique being used, the major problem they all face is the loss of some information when the source video is modelled. The techniques for intuitive analysis involve complex mathematic algorithms for the image processing. Some of these methods are: the moving object and the moving region methods (Lema et al 2001), colour histogram comparison (Tahaghoghi et al 2005), layered video data modelling (Petkovic and Jonker 2000), shot boundary detection (Smeaton and Over 2002), structure parsing (Zhong and Chang 2000), and image segmentation (Petkovic et al 2001). For example, Bertini et al (2006) presented a system for automated video annotation. The argument here is that their system allows effective automatic semantic annotation of video clips with high level concepts by checking their similarity with the visual concepts of the ontology. One point to be mentioned is that their system requires events to be pre-defined. The video analysis methods are not limited by these methods, many more exist. Since this thesis does not deal with image

processing, such algorithms will not be focused on for the time being; however, these algorithms have been looked into and some of them were analysed. Snoek and Worring (2003) find it interesting that most visual and auditory modalities ignore the textual modality that might contain some valuable information. This will be demonstrated later on in this thesis when it is shown that sports video annotation can contain useful information.

2.2.2 Speech Analysis

This section deals with the extraction of image internal features and semantic information from collateral image external sources – sound volume or text caption. The image internal features are analysed by extraction using information and then manually labelling certain features as meaningful. This is, from our point of view, the hardest part to deal with. It combines speech and video analysis. Researchers who undertake elicitation analysis in sport investigate the various events that are detected and mentioned by the commentator. For example, a football commentator will comment on a specific play by how the player should have played the ball. This will involve what actually happened and what should have happened. Also, by the use of speech or voice detection one can try to detect an event via the voice volume. Speech recognition and indexing is the goal for this type of analysis (see Dolbear and Brady 2003; Snoek and Worring 2005; and Wolf et al 2002). Snoek and Worring (2005), for example, explain how the different speech volumes can lead to specific different meanings; sudden loud volume might mean a goal is scored in a football game; continuously increasing volume might mean a team is getting close to scoring a goal. The speech commentary might be converted to a written commentary to search for specific patterns such as goal scoring. Tjondronegoro (2005) has presented a Ph.D. thesis that deals with video detection and annotation using video and audio analysis (see Figure 5 below).

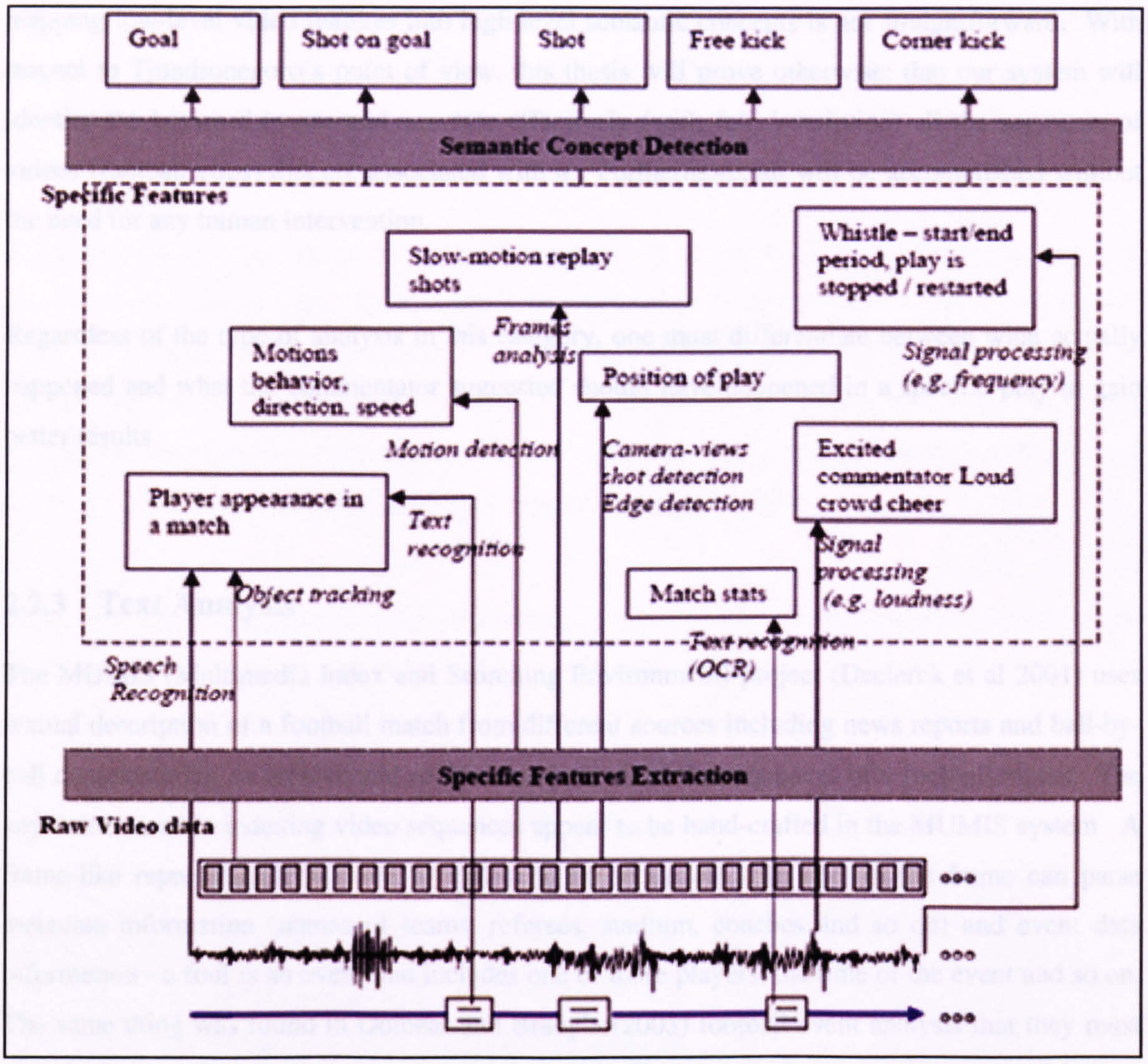


Figure 5: Tjondronegoro system for video detection and annotation using video and audio analysis (Tjondronegoro 2005: 71).

The system illustrates how various data is obtained from different sources. A player's appearance in a match can be obtained from all sources. Position of play can be obtained from video. Excited commentator can be obtained from audio. Match statistics can be obtained from text.

Such a system can work in parallel to the system researched in this thesis and eliminate a part of human interference such as the commentator indicating the start and end of an event. Tjondronegoro (2005: 32) stated that "another alternative for managing video is to annotate the semantics of video segments using key words or free texts". Thus, user queries can be managed using standard query language, such as SQL, and browsing can be based on hierarchical topic (or subject) classification. However, the major limitation of this approach is the fact that it would be perhaps tedious and ineffective to manually annotate every segment of video. The process of

mapping low-level video features into high-level semantic concepts is not straightforward. With respect to Tjondronegoro's point of view, this thesis will prove otherwise: that our system will identify the keyword to use and annotate effectively (with full description) all the segments of videos (football clips) that are associated with it. Furthermore, all will be accomplished without the need for any human intervention.

Regardless of the type of analysis in this category, one must differentiate between what actually happened and what the commentator suggested should have happened in a specific play to gain better results.

2.2.3 Text Analysis

The MUMIS (Multimedia Index and Searching Environment) project (Declerck et al 2001) uses textual description of a football match from different sources including news reports and ball-by-ball commentaries, to archive and retrieve some or all of the sequences of a football match. The key-words used in indexing video sequences appear to be hand-crafted in the MUMIS system. A frame-like representation is used to represent the sequences: the slots in the frame can parse metadata information (names of teams, referees, stadium, coaches and so on) and event data information - a foul is an event that includes one or more players, the time of the event and so on. The same thing was found in Dolbear and Brady's (2003) football event analysis that they must specify the set of semantic terms that can be used. This means the terms are static and human interference exists in order to use their system. One advantage of text-based analysis is that it can recognize events, categorize them into main topics and produce a final summary quicker than image processing algorithms such as those employed by McKeown et al (2003). Yankova and Boytcheva (2003) also agree with what Snoek and Worring (2003) said earlier that "an enormous amount of information exists in natural language text". Also we agree with Yankova and Boytcheva when they state that "it has to be first distilled into more structured form" and later on in this thesis we will discuss the resulting local grammars. Snoek and Worring (2003) also bring up an important point when dealing with extracting information from text: the information needs to be synchronized with the video itself. In our thesis this synchronization is done via the use of the time-stamp which is provided throughout our corpus. Yankova and Boytcheva (2003) take it further and provide five subtasks to consider: lexical analysis, named entity, coreference resolution, syntactic analysis and template pattern matching.

For text-based analysis, for annotation purposes particularly, it is important the term base used for annotation is generated automatically from the text under analysis. Human extraction and annotation is a time consuming and error-prone task.

Researchers in this field focus on the local grammar. It is noticed that with local grammar and patterns, collocation is used. Collocation is the occurrence of two or more words within a short space of each other in the text (Sinclair 1991a: 170). The strength of the local grammar and its derived patterns has opened the doors for research. A variety of systems have been introduced such as Machine Learning (Srikanth et al 2005), video indexing and video annotation (Snoek and Worring 2005), news analysis and summarization (Ahmad et al 2004).

Wang et al (2005) presented a system *Automatic Generation of Personalized Music Sports Video* that is fairly close to the system in this thesis, see Figure 6 below.

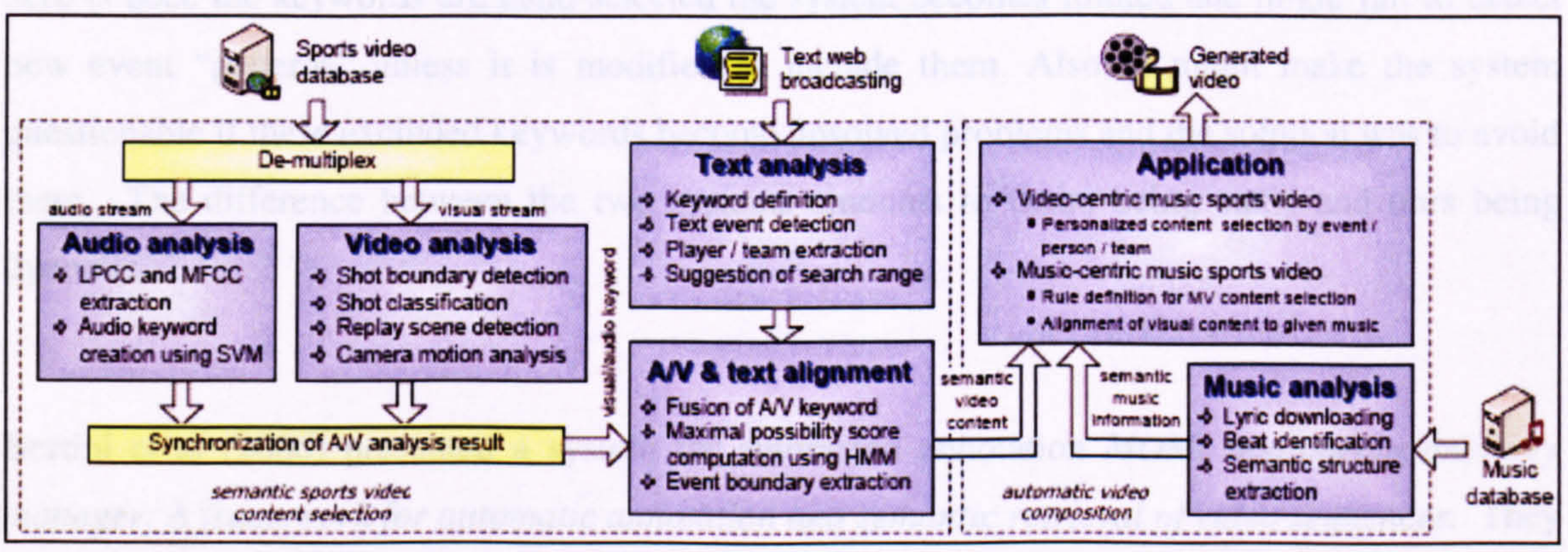


Figure 6: Wang et al's system for Automatic Generation of Personalized Music Sports Video (2005: 737)

This system is made up of two sections: Semantic sports video content selection and automatic video composition. The first section deals with the analysis of the audio stream and the video stream. Then the audio and the video are synchronized. At the same time, text web broadcasting is collected and analysed to be aligned with the audio and video. The second section deals with the music analysis as well receiving the result of the audio, video and text alignment. Then both inputs are combined to produce video-centric music sport video and music-centric video sport video.

Their usage of football videos is interesting because not only is it a globally popular game, but also it presents many difficult challenges for video analysis due to its dynamic structure; the time stamp for the live commentary text doesn't need to be perfect as an estimated time will be efficient. Also, text analysis plays an important role for sports video content selection because the text analysis can greatly increase the event detection performance due to the exact text keywords of the events. Live commentary text information is more freely available as live text commentary or match reports from the Internet; live commentary text possesses very detailed information about the event, related players and approximate times of events. Wang et al's system uses unique nouns: "each type of an event features one or several unique nouns. These nouns are defined as a keyword and by detecting the keywords from live commentary text, the relevant event can be identified" (Wang et al 2005: 739). Furthermore, they stated that for their system to achieve a high detection percentage, "phrases with different meanings should be removed". They went on to state that "these events are chosen because they are either important or difficult to be detected by traditional audio/video analysis techniques. The popular event "shoot" is not selected because the "shoot" event overlaps the combination of "goal" and "save" events". The problem here is once the keywords are hand-selected the system becomes limited and might fail to detect new event "patterns" unless it is modified to include them. Also, it might make the system questionable if these excluded keywords become unsolved problems and the solution was to avoid them. The difference between the two systems amounts to theirs being static and ours being dynamic.

Bertini et al (2006) presented a system for automated annotation *MOM: multimedia ontology manager: A framework for automatic annotation and semantic retrieval of video sequences.* They state that their MOM system "allows effective automatic annotation of video clips with high level concepts". Looking at their system query and retrieval (Figure 7 below) it is noted that their queries are limited and the events are pre-chosen. The events their system retrieves are limited to the choices provided by the GUI. Their way of annotation is to "annotate a sequence of clips with some pre-defined articulated sentence" (pp. 788). That means specific events are chosen and human intervention occurred to annotate these events.

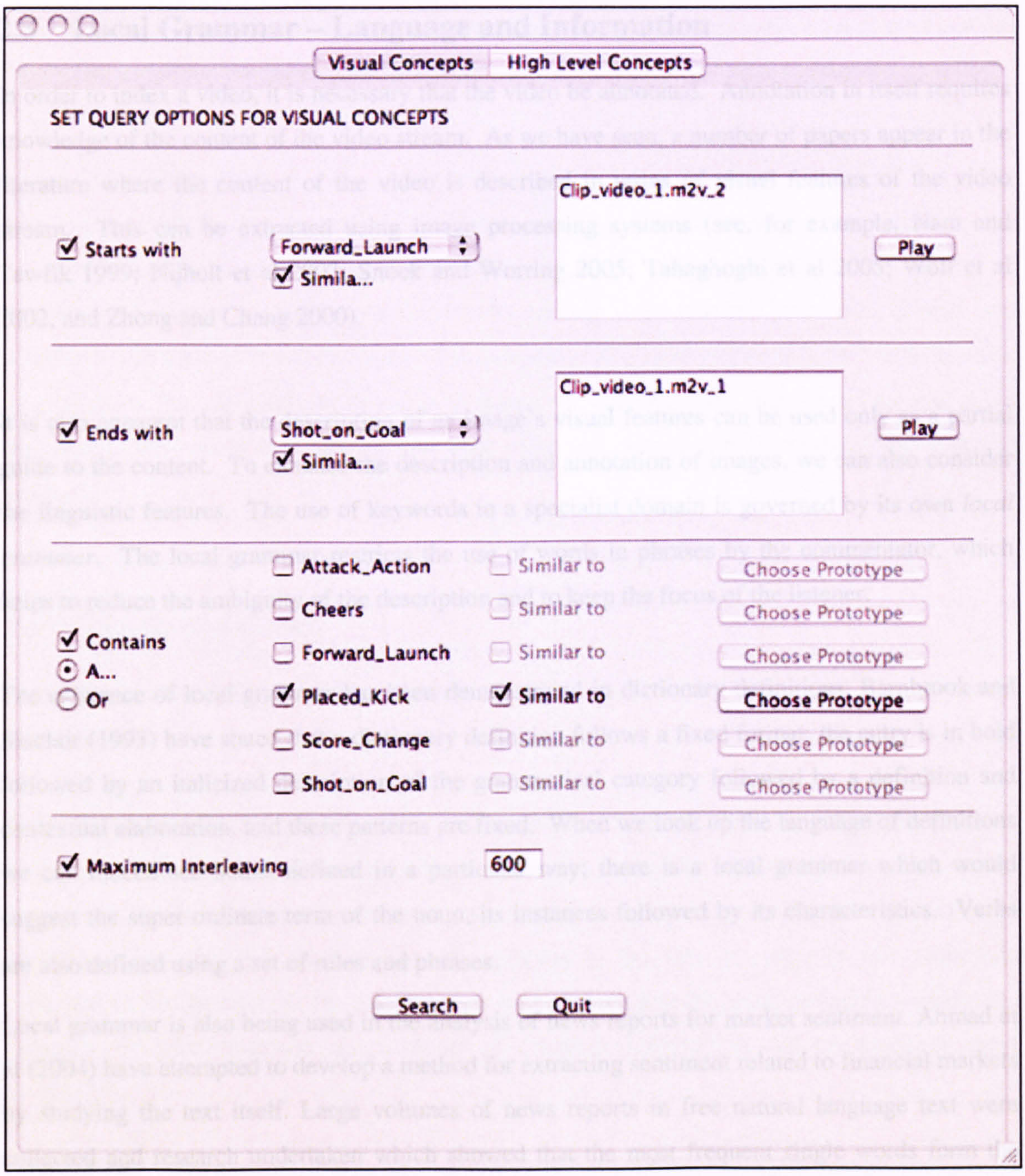


Figure 7: Bertini et al MOM query GUI (2006: 788)

Systems such as these presented by Declerck et al, Wang et al and Bertini et al are more likely to be effective in achieving the goals they were designed for. However, these systems only allow limited queries to be allowed; for a new query to be added a user intervention is required. This is more likely to cause many issues when one's goal is to achieve a dynamically automated system for video indexing and annotation.

2.3 Local Grammar – Language and Information

In order to index a video, it is necessary that the video be annotated. Annotation in itself requires knowledge of the content of the video stream. As we have seen, a number of papers appear in the literature where the content of the video is described in terms of visual features of the video stream. This can be extracted using image processing systems (see, for example, Nam and Tewfik 1999; Nijholt et al 2003; Snoek and Worring 2005; Tahaghoghi et al 2005; Wolf et al 2002, and Zhong and Chang 2000).

It is also apparent that the description of an image's visual features can be used only as a partial guide to the content. To enhance the description and annotation of images, we can also consider the linguistic features. The use of keywords in a specialist domain is governed by its own *local grammar*. The local grammar restricts the use of words in phrases by the commentator, which helps to reduce the ambiguity of the description and to keep the focus of the listener.

The existence of local grammar has been demonstrated in dictionary definitions. Barnbrook and Sinclair (1993) have stated that a dictionary definition follows a fixed format: the entry is in bold followed by an italicized description of the grammatical category followed by a definition and contextual elaboration, and these patterns are fixed. When we look up the language of definitions we can indeed see nouns defined in a particular way; there is a local grammar which would suggest the super ordinate term of the noun, its instances followed by its characteristics. Verbs are also defined using a set of rules and phrases.

Local grammar is also being used in the analysis of news reports for market sentiment. Ahmad et al (2004) have attempted to develop a method for extracting sentiment related to financial markets by studying the text itself. Large volumes of news reports in free natural language text were collected and research undertaken which showed that the most frequent single words form the basis of the most significant collocation patterns; which may extend up to 3 or 4 words either side of the most frequent word. These collocations tend to use and unambiguously extract the sentiment of a report about the movement of a financial instrument. The authors have specifically looked at financial news reporting. They note that the word *percentage* used either as a symbol or as a word is among the most frequent words in their financial corpus running up to tens of millions of words. The strong collocates of the word percentage typically are the words of movements: *rose, fell, rise* and *fall* and their various morphologies indicating time. The collocation also includes the name of a financial instrument. Ahmad et al. have extracted a

number of local grammar patterns which are used to express sentiments about the market and can be used to search for sentiment bearing words in as unambiguous a manner as possible.

Videos in specialist video domains are described using a special language, a language that exists with its own vocabulary governed by a local grammar. These local grammar patterns are repeatedly used in the description of unusual events in the video. This thesis will attempt to determine whether or not one can automatically identify these patterns in a visual domain and whether or not these patterns are robust. The robustness of a local grammar pattern is determined by examining the statistical performance of usage, for example, how often is an unusual event discovered, how often it is missed and how often are normal events misclassified as being unusual.

2.3.1 Harris and Gross

The importance of Zellig Harris and Maurice Gross' work in the field of textual analysis cannot be overstated.

Zellig Sabbetai Harris (October 23, 1909 - May 22, 1992) was an American linguist, mathematical syntactician, and methodologist of science. Originally a Semiticist, he is best known for his work in structural linguistics and discourse analysis and for the discovery of transformational structure in language, all achieved in the first 10 years of his career and published within the first 25. His contributions in the subsequent 35 years, including sublanguage grammar, operator grammar, and a theory of linguistic information, are perhaps even more remarkable.

Maurice Gross (1934-2001) was both a great linguist and a pioneer in natural language processing. As a linguist, Maurice Gross contributed to the revival of formal linguistics in the 1960s, and he created and implemented an efficient methodology for descriptive lexicology. A specialist of natural language processing (NLP), he was also a pioneer of linguistics-based processing. The best-known theory of Maurice Gross is the description of idiosyncratic properties of lexical elements.

Zellig Harris in his book *Language and Information* argued the existence of science sublanguages (1988). Harris stated that "A subset of the sentences of a language constitutes a sublanguage of that language if it is closed under some operations of the language" (pp. 34). Furthermore, Harris

stated that "When we make a separate grammar for a given subject matter, we find not a general dependence on dependence, but specific sets of arguments occurring only under particular sets of operations" (pp. 38). The notion of local grammar was introduced by Harris (1991) where he looked at the use of collocation and 'frozen sentences' in scientific languages, more specifically in biochemistry where he indicated ordinary noun phrases may contain two nouns separated by a verb that require the choice of the nouns preceding the verbs and following the verbs to be restricted. Harris' famous example is "he washed peptides in hydrochloric acid". Harris knows that the sentence "he washed hydrochloric acid in peptides" would not be an acceptable sentence in the grammar of chemistry. The verb *washed* can only take a certain limited class of noun as subject or object. Maurice Gross (1993) advanced the notion of local grammar considerably . He looked at idiomatic expressions, calendrical terms, and calendrical statements and suggested the existence of local grammar based on frequency usage of keywords. Gross (1993:30) suggested that we look at a number of idiomatic expressions, for example: *Bob lost his cool*, *Bob lost his temper*, *Bob lost his cork*, *Bob lost his self-control*, *Bob blew a fuse*, *Bob blew a gasket.* In each of these phrases the determiners are frozen and he described the local grammar as well. In his more elaborated example, Gross describes calendrical terms which include statements like: *the match took place on Tuesday the 3rd of February 2005.* One can also make the following sentence without losing much meaning: *the match took place the 3rd of February 2005.* Or we can say: *the match took place on the 1st Tuesday of May 2005.* These sentences are related and the position to which each of the words is restricted is not by virtue of its grammatical category but by virtue of its frequent use in that sentence. He goes on elaborating his example by saying: *the match took place on Tuesday May the 2nd at noon or at 4 o'clock or at 16:30.* The argument here is that when we say '*the match took place on*', then we describe a date, time or place in a certain order. For instance, '*the match took place in 1969 Tuesday 2nd*' would not be acceptable; the order in which the days are composed is governed by a lineation and is shown by the local grammar. Local grammar will generate and recognise all the acceptable ways of stating calendrical dates and times.

2.3.2 Richard Burton

Richard Burton presented a Ph. D. thesis entitled "Semantic Grammar: A Technique for Efficient Language Understanding in Limited Domains" (1976). Even though his work has no direct relation to this thesis, it is believed that his work has influenced many researchers (for example, Daimi 2002; Zelle and Mooney 1993; Carbonell and Hayes 1994; Ward and Pellom 1999; Slocum 1981; Gabsdil and Lemon 2005, and Nakanishi, et al. 2005) and made significant impact on the Natural Language Processing field and is therefore worthy of brief discussion here.

Burton's thesis addressed the problem of "developing a system which can understand natural language (English) within an educational problem-solving environment" (1976: 1). Burton also stated that there are four requirements. First, efficiency; Burton argued that if a student is at a terminal solving a problem and he/she decides to acquire more information by submitting a query to the system they will have nothing to do while waiting for the search result. During that time, the student is "apt to spend time forgetting pertinent information and losing interest" (*ibid.*: 3). That is the system must understand the query and generate a response within two seconds (or risk the enquirer becoming bored) as some psychological experiments have shown. Second, habitability; that is "one in which the user can make local or minor modifications to an accepted sentence", in other words the system must allow a certain amount of alteration by the user (*ibid.*: 4). Third, self-teaching; that is a student that uses a system should be able to feel the range and limitations of the sublanguage. Fourth, awareness of ambiguity; that is "the program which interprets natural language sentences must be aware that its interpretation is not the only one" (*ibid.*: 6). Burton emphasized "major leverage points that allow these requirements to be met" which are: Limited domain, Limited activities within the domain and Known conceptualizations of the domain .

Burton's thesis presented the development of a technique - "semantic grammars" - for building natural language processors which satisfy these constraints. Burton showed throughout his thesis that the notion of semantic grammar provided a paradigm for organizing the knowledge required in the understanding process which permits efficient parsing. Also, that semantic grammar provided insights into a useful class of dialogue constructs and permitted efficient handling of such phenomena as "pronominalizations" and ellipsis. All that led Burton to introduce the use of Augmented Transition Network (ATN); moreover, the design and implementation of a general ATN complier which increased the speed of execution by translating it into an optimized object program.

2.4 Others Evaluation

In the video analysis field, solid evaluation is essential. Regardless of what the researcher thinks of his/her method or system, evaluating it is the key to measuring success. A well known conference, TRECVID[1], is dedicated to researchers in information retrieval research areas in content based retrieval of video. It is co-sponsored by the National Institute of Standards and Technology (NIST) and Advanced Research and Development Activity (ARDA) centre of the U.S. Department of Defence and was founded in 2003 as an independent evaluation/workshop

[1] http://www-nlpir.nist.gov/projects/trecvid/

from TREC (Text Retrieval Conference). Its goal is to encourage research in information retrieval by providing a large test collection, uniform scoring procedures, and a forum for organizations interested in comparing their results. In 2006, TRECVID targeted: shot boundary determination, high-level feature extraction, search (interactive, manually-assisted, and/or fully automatic) and rushes exploitation (exploratory). And in 2007, TRECVID targeted: shot boundary determination, high-level feature extraction and search (interactive, manually-assisted, and/or fully automatic). Part of TRECVID's evaluation is:

- All transitions: for each file, precision and recall for detection; for each run, the mean precision and recall per reference transition across all files.

- Gradual transitions only: "frame-recall" and "frame precision" will be calculated for each detected gradual reference transition. Averages per detected gradual reference transition will be calculated for each file and for each submitted run.

Precision and Recall are one of the most used measuring tools for evaluation. Precision is a measure of the usefulness of a hitlist; Recall is a measure of the completeness of the hitlist (more details will be provided in the Precision and Recall evaluation section below).

The NTCIR[2] Project (National Institute of Informatics Test Collection for IR Systems Project, based in Tokyo) runs a series of evaluation workshops to enhance the research in information access technologies, including text retrieval, cross-lingual information access, question answering, etc, by providing an infrastructure of evaluation and research including large-scale re-usable test collections, evaluation metrics and methodologies, and a forum of researchers who are interested in exchanging research ideas and evaluation methodologies. They state the following for their evaluation, "Evaluation is a very critical issue for all of the researchers. So please examine how the evaluation is done and how the metrics behave, or whether there are any methods to overcome the limitation of current practice of the evaluation. With your cooperation, we would like to obtain fruitful examination of the evaluation results and metrics".

[2] http://research.nii.ac.jp/ntcir/ntcir-ws6/agenda-en.html

Also, CLEF[3] (Cross-Language Evaluation Forum, which is an activity of the TrebleCLEF Coordination Action under the Seventh Framework Programme of the European Commission) supports global digital library applications by (i) developing an infrastructure for the testing, tuning and evaluation of information retrieval systems operating on European languages in both monolingual and cross-language contexts, and (ii) creating test-suites of reusable data which can be employed by system developers for benchmarking purposes. In their evaluation they stated that all systems are evaluated according to their Mean Average Precision (MAP) as computed by the TREC EVAL software on the pre-existing CLEF relevance-assessments.

As mentioned, these various conferences, workshops and forums emphasize the importance of evaluation in determining acceptance of someone's research. Since the system in this thesis is relative to those who are interested in video analysis and video annotation in concept, the same evaluation (Precision and Recall) will be used to compare the results to those other systems.

2.5 Conclusion

Many researchers have shown interest in analysing sports videos. Varieties of tactics have been introduced and different algorithms have been developed. Some researchers used computerized techniques to analyse the video itself such as the colour histogram, shot detection and image segmentation. Others used computerized techniques to involve speech recognition and combine it with the video analysis. Furthermore, some techniques involved the text commentary that accompanies the sport event; these techniques vary depending on whether the text is obtained from a broadcasting website or from the text caption that is embedded in the video. All these techniques aim to achieve good results with better accuracy in shorter time. The question to be asked here is how are events chosen? It seems, in most cases, that the user has pre-defined the events to be analysed when dealing with video analysis. In speech analysis and text analysis, keywords are pre-selected and, in some cases, some keywords are pre-rejected. This leads these systems to be static. In the case of one parameter being changed, such a system will fail and will require human interference. For example, FIFA (International Federation of Association Football) is considering the use of artificial grass. If this is to be approved then different grass colours might be introduced and systems that use the colour histogram method will fail if the grass is not green. Also, in the text commentary analysis or speech analysis, some pre-selected

[3] http://www.clef-campaign.org/

keywords might get dropped from the source and new keywords introduced. This situation actually exists in the BBC text web broadcasting as new keywords and patterns are introduced in the 2006 commentary. Systems such as Wang et al.'s might fail to recognize them as their system is built on pre-selected keywords.

The system to be introduced in this thesis is, in one sense, dynamic. It builds a corpus of live commentary text; performs several analyses such as tokens frequency, keywords collocations and patterns (events) detection without user interference. Then, the system synchronizes the detected events with its sport video to perform automated video indexing and annotating including complex annotated clips retrieval that allows simple search and advanced search. Despite the fact that this system is tested on the English football domain, it is dynamic and we demonstrate that it can be adapted to other sports' domains.

Chapter 3

3 Method

Videos in specialist video domains are described using a special language, a language that exists with its own vocabulary governed by a local grammar. These local grammar patterns are repeatedly used in the description of unusual events in the video. This thesis will attempt to determine whether or not one can automatically identify these patterns in a visual domain and whether or not these patterns are robust. The robustness of a local grammar pattern is determined by examining the statistical performance of usage, for example, how often is an unusual event discovered, how often it is missed and how often are normal events misclassified as being unusual.

An example of a local grammar in the specialist domain was chosen for this research and it was that of football matches. In the video summarization literature a number of visual domains have been studied, most frequently that of news casts where the anchor person and the reporter(s) describe objects and events in a video stream. Visual features are used in conjunction with the description to index the news story. This visual domain contains a high frequency of unusual events. The anchor and the reporter are familiar with the terminology which is being used and which is generally understood by the public at large. What is being investigated here is the terminology used in a domain where unusual events are infrequent and where there are periods where not much happens although there is a lot of action on the ground. In order to look at video summarization we have chosen the domain of football commentaries in particular ball-by-ball commentary produced by the BBC [online] (see Figure 8 below).

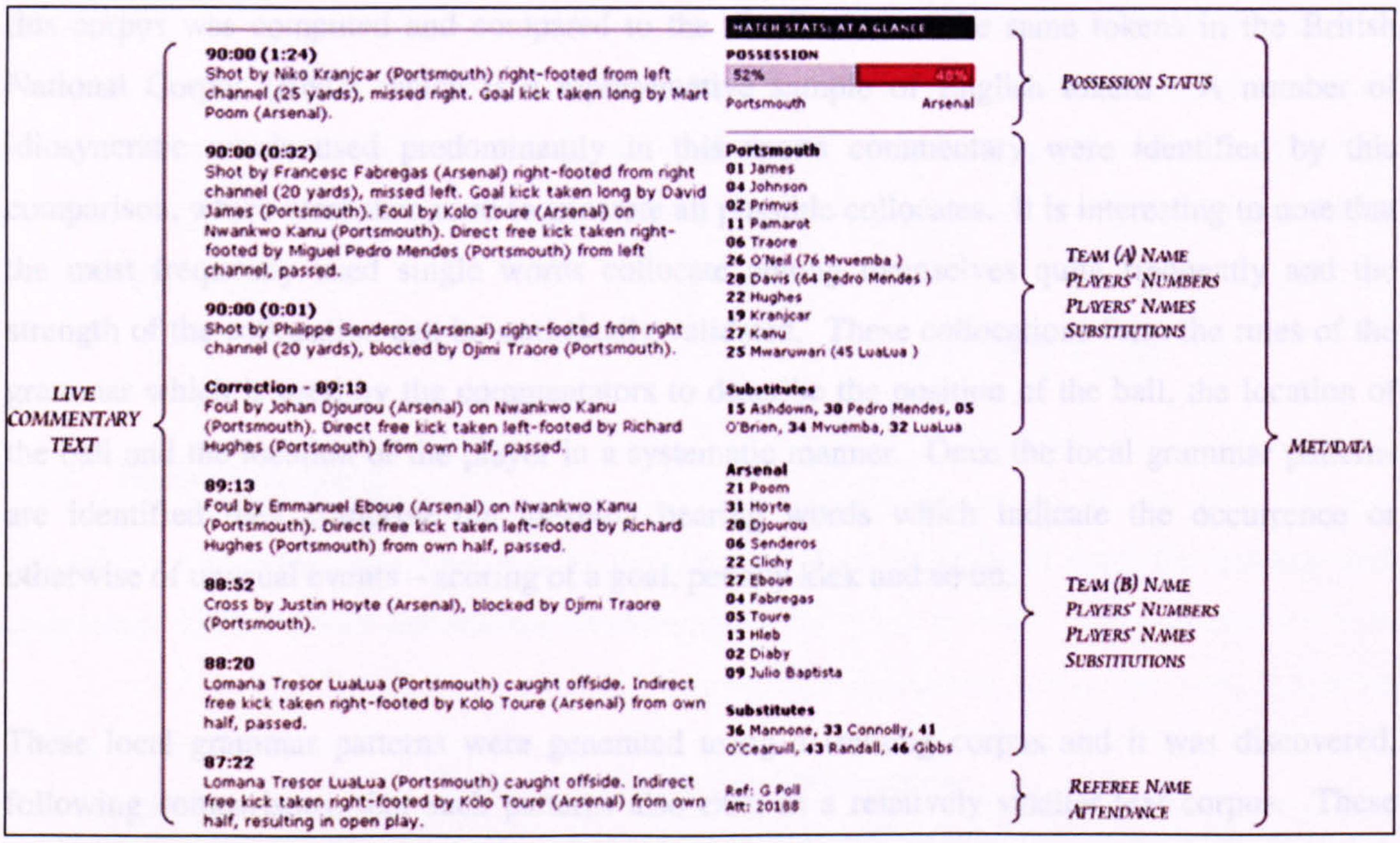


Figure 8: Screenshot of BBC live commentary, http://www.bbc.co.uk , April 2, 2007

The football match live commentary is shown to the left with each commentary accompanied by a time-stamp. To the right, additional metadata are provided: possession status and team information (team name, players' names, players' numbers and the substitutions), the referee name and the total match attendance. Note that this is what is referred to as image-*external* meta-data.

Note the telegraphic nature of the language, each sentence is a meaning unit on its own; there is a time stamp which is metalevel information. The ball-by-ball commentary is produced by a major news organisation to describe events on the football field every few minutes for its viewers who use the internet to be informed about the game. Also, note that the language is not adorned by closed class words which are typically used in newspaper reports: *the*, *of*, *and*, etc. There is no garnishing of the language by introductory clauses, introductory phrases or closing remarks. What the commentator does for every movement of the ball is to describe the position of the players and which player possesses the ball. The vocabulary is limited and, as will be shown, a local grammar employed.

This chapter will demonstrate the notions of a local grammar by an example. Ball-by-ball football commentaries have been collected over a year long period resulting in a corpus of 3 million tokens. Before any real processing can be done on the input text, it needs to be segmented into linguistic units such as words, punctuation, numbers, or alphanumeric, which have been used to describe more than 700 games, each game lasting for 90 minutes. The distribution of tokens in

this corpus was computed and compared to the distribution of the same tokens in the British National Corpus (BNC) which is a representative sample of English tokens. A number of idiosyncratic words used predominantly in this sports commentary were identified by this comparison, which were then used to generate all possible collocates. It is interesting to note that the most frequently used single words collocate among themselves quite frequently and the strength of the collocation can be statistically validated. These collocations form the rules of the grammar which is used by the commentators to describe the position of the ball, the location of the ball and the location of the player in a systematic manner. Once the local grammar patterns are identified they comprise the meaning bearing words which indicate the occurrence or otherwise of unusual events – scoring of a goal, penalty kick and so on.

These local grammar patterns were generated using a training corpus and it was discovered, following comparisons, that such patterns also exist in a relatively smaller text corpus. These patterns have allowed the system to automate the annotation of short sequences of video frames with descriptions which can be identified and analysed by the local grammar. The sequence of frames is then annotated by those grammar patterns containing those words. Subsequently from the annotation the same pattern can be used to search for similar occurrences throughout the video clips within one game or across different games to produce so-called 'albums'.

This thesis will not be dealing with visual features as it will concentrate on whether or not videos can be summarised using linguistic description. However, work is currently being undertaken in close collaboration with a project where visual features will be analysed and used in conjunction with a linguistic description – Reveal Project, the University of Surrey – and it is hoped that our findings will be of some importance to it.

This chapter describes the collection of a corpus of live commentary text. Then an analysis is performed to obtain a variety of results which leads to the extraction of terminology and local grammar. All this leads to automated video indexing and annotation.

3.1 Materials and Method

Any corpus, regardless of its domain, provides information that might not be available via visual analysis. Moreover, a corpus-based method allows flexibility to reorder and manipulate the data.

For example, a corpus-based method for analysing news provides a "Robust system that clusters news into events, categorizes events into broad topics and summarizes multiple articles on each event" (McKeown et al. 2003: 15-16). It will be shown by using the corpus-based method how to construct the most frequent parts-of-speech and the most frequent motion entities and then build the local grammars.

3.1.1 Materials Used

Several corpora have been analysed in this thesis. They have been used for training and evaluation. Table 5 below summarizes these corpora.

Table 5: Corpora general information

Corpus	Matches	Tokens		Year	Source	Duration (hours)
		Total	Avg			
English football - Training	775	3,026,038	3,902	2004	http://www.bbc.co.uk	1,162.5
English football – Testing	57	224,074		2004		85.5
English football – Testing-2	1,048	4,276,938		2006		1,572.0
English cricket - Training	300	4,337,772	14,672	2003-2006	http://content-usa.cricinfo.com	2,400
English cricket - Testing	16	213,899		2007		128
Arabic football - Training	48	53,784	1,162	2006	http://www.alittihad.ae	72
Arabic football - Testing	8	7,849		2006		10.5

Table 5 above shows how many match commentary texts are contained in each corpus together with the total number of tokens. Also, the year the matches were played and the commentary text source.

Tokenization plays an important part in the thesis method as it is the first step. Tokenization is the process of splitting a sentence into its constitute tokens. Four programs were chosen to do the tokenization and POS analysis: the test was run on a Dual Xeon workstation with 4GB memory, with a fresh installation of Windows software:

- **Text Analysis International:** Text Analysis International is a company based in California which has developed an off-the-shelf program called *VisualText*[4]. The

[4] http://www.textanalysis.com/Apps/POS_Tagger/pos_tagger.html - accessed 18 August 2006

tokenization process is fast but it failed to tag 98% of the corpus. In fact, it tagged the 98% as unknown. Such a result is unacceptable as this program needs heavy training.

- **Oliver Mason Tagger (Oliver 2004):** Oliver Mason tagger is well known and Tokenization and POS process is acceptable. Interestingly, it tags *in* as a number. The POS tagger did not recognize 15% of the tokens and it mis-tagged about 10% of the tokens it did tag. Such a result is just about acceptable but better is preferred.

- **GATE (General Architecture for Text Engineering) (Cunningham et al 2002):** This is the project from Sheffield University and requires Java to be installed. The POS part took about 12 seconds which is amazingly fast. However, it failed to show the results as the PC froze up; even when the PC was left to run for 24 hours. Several attempts were tried and all had the same result. The only time it managed to show the result was when the corpus was restricted to 10 matches rather than the 700+ matches. The GATE project then tagged 95% of the corpus correctly. It failed to tag the time stamp as it reported, for example, 01:45 as unknown. Also, it splits a hyphened token into two; throw-in is split into throw and in. The result from the small corpus shows promise but without being able to view the full corpus result, it is unacceptable. GATE also comes with its own sentence splitter. It only worked when the sentence ends with a full stop. When the sentence ends with just an end-line, it does not recognize it and attaches the next sentence to it. This will cause major problems as the time-stamp ends with a line break and not with a period.

- **CLAWS (The Constituent Likelihood Automatic Word-tagging System) (Garside 1987):** This is a project from Lancaster University. Its process speed was acceptable. The result showed 99% success, and the result is saved to a file.

Based on this analysis, CLAWS was chosen as it showed 99% accuracy and the processing time is acceptable.

3.1.2 Method

Ahmad et al. (2005) has presented a 5-step algorithm for time series summarization. The algorithm concept in general seemed to be ideal and with some modification it can be applied for keywords selection and patterns detection (see Figure 9).

1. **INPUT** $CORPUS_{GL}$ /*a general language corpus comprising N_{GL} = 100,000,000 individual words*/
$CORPUS_{SL}$ /*a corpus of specialist texts comprising N_{SL} = 3,026,038 individual words*/
$N_{GL/SL} = N_{GL} / N_{SL} = 33.0465$

2. ***TERM IDETIFICATION***

A) CONTRAST the distribution of words in $CORPUS_{GL}$ and $CORPUS_{SL}$ /*Ahmad and Rogers, 2001*/

I. **COMPUTE** Frequency $n_{SL}(w)$ /*of all words, *w* in $CORPUS_{SL}$*/
Average frequency An_{SL}
Frequency ratio $n_{GL}(w)$ /*of all words, *w* in $CORPUS_{GL}$*/

II. **COMPUTE** Relative frequency $f_{SL}(w) = n_{SL}(w)/N_{SL}$
Average relative frequency Ar_{SL}

III. **COMPUTE** $StDev_n = StDev(\sum(n_{SL}(w)))$ /* StDev of n_{SL} */
$StDev_f = StDev(\sum(f_{SL}(w)))$ /* StDev of f_{SL} */

IV. **COMPUTE** z-score freqency $Z_{SL}(w) = (f_{SL}(w) - Ar_{SL}) / StDev_f$

V. **COMPUTE** $f_{GL}(w) = (1 + f_{SL}(w)) / (n_{GL}(w) / N_{GL})$

VI. **COMPUTE** $D(w) = [(1 + n_{SL}(w)) * N_{GL/SL}] / f_{GL}(w)$
A_D /* Average of D */
$StDev_D = StDev(\sum(D(w)))$ /* StDev of D */

VII. **COMPUTE** z-score weirdness
$Z_D(w) = (D(w) - A_D) / StDev_D$

B) FIRST WORDS SELECTION

a. **LET** S^{SW} be the group of selected words

b. **LET** V be the value of a normalized StDev, V = 0 /* V value can be changed */

c. FOR EACH WORD (w) in $CORPUS_{SL}$
IF $Z_{SL}(w) > V$ AND $Z_D(w) > V$ THEN $w \in S^{SW}$
NEXT (w)

3. ***LOCAL PATTERNS***

A) FIND COLLOCATION patterns for words in S^{SW}
FOR EACH WORD *w* in S^{SW}

a. **COMPUTE** $n_{SL}(w, w^{SW})$ /*frequency of a word *W* co-occurring within *k* words from w^{SW} where $-m \leq k \leq m$ and m=5 (Smadja, 1993)*/

b. **FLITER** Significant collocates (w, w^{SW}, *k*) based on *z-scores* and other moments of $n(w, w^{SW})$

NEXT W^{SW}

B) FIND NEXT COLLOCATION

A. FOR EACH $W \in S^{SW}$
$max(w) = max(n_{SL}(w, w^{SW})) / n_{SL}(w)$ /* max collocation for every word (w) */
NEXT (w)

b. **LET** S2 be the final group for selected words
IF (max(w)) > 85% THEN $w \in S2$

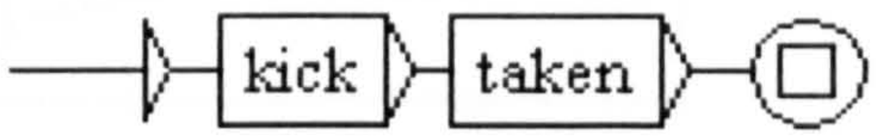
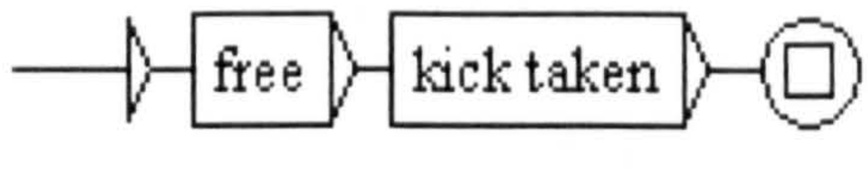
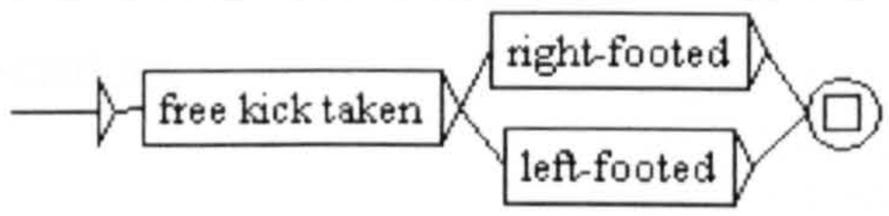
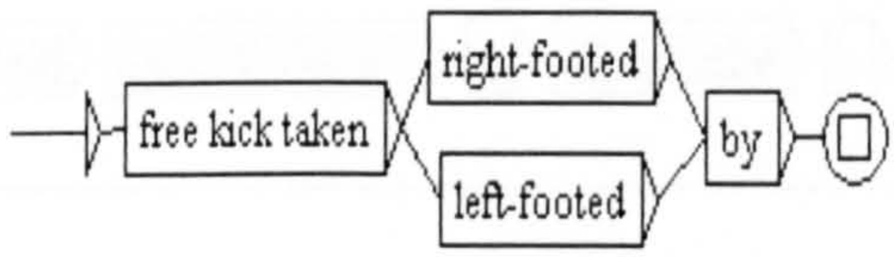
Figure 9: Patterns extraction algorithm

In the first instance, a corpus is built and a ratio of the general corpus to the built corpus is calculated. Then several computations are performed. The purposes of **Term Identification** are to calculate the z-score and the z-score weirdness of the corpus tokens. The z-score for an item indicates how far and in what direction that item deviates from its distribution's mean, expressed

in units of its distribution's standard deviation. Also, term identification finds the keywords with z-score frequency and z-score weirdness that are at least equal to the normalized standard deviation. The **Local Patterns** is to find the words collocation for these keywords using Smadja's method (1994). Then, for each found collocation we determine the next collocation only by accepting strongly collocated keywords; 85% collocation is found to be satisfying. These collocation patterns can eventually span many words but the collocation strength is markedly different in certain patterns. Those chosen patterns are the basis of the local grammar.

According to - John Sinclair, a typical language user has 'available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analyzable into segments' (1991:110). Here collocation plays a significant role, especially the collocation patterns that involve the most frequently used candidate terms. The collocation patterns were extracted based on statistical significance as suggested by Smadja (1994) combined with Ahmad et al.'s 5-step modified algorithm. Table 6 shows how such patterns can be unified.

Table 6: The emergent local grammar of football (Numbers in parentheses are frequencies)

Key Token	Left Collocate	Key Collocate	Right Collocate	Collocation Patterns
kick (*f*=41,013)			taken (*f*=46132)	
	free	*kick taken*		
		free kick taken (*f*=22,449)	left-footed right-footed	
		free kick taken left footed; *free kick taken right footed*	by	

More statistically significant collocates of the keyword *kick* emerge in addition to the one described above:

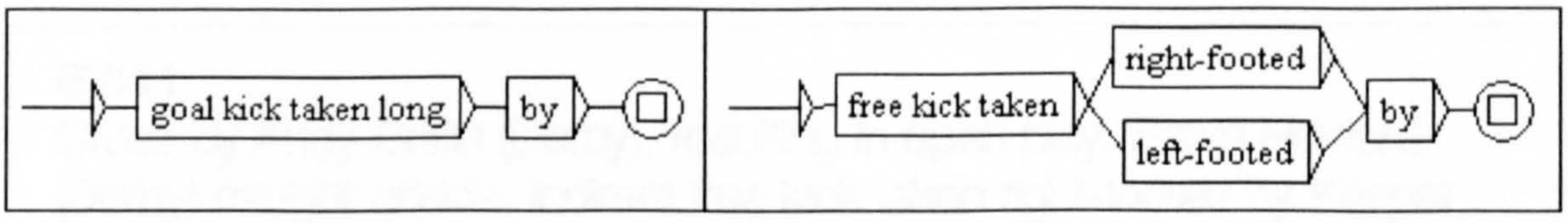


Figure 10: *kick* early collocation

The above two graphs can be joined by a program together at the inputs to the *by* node to produce

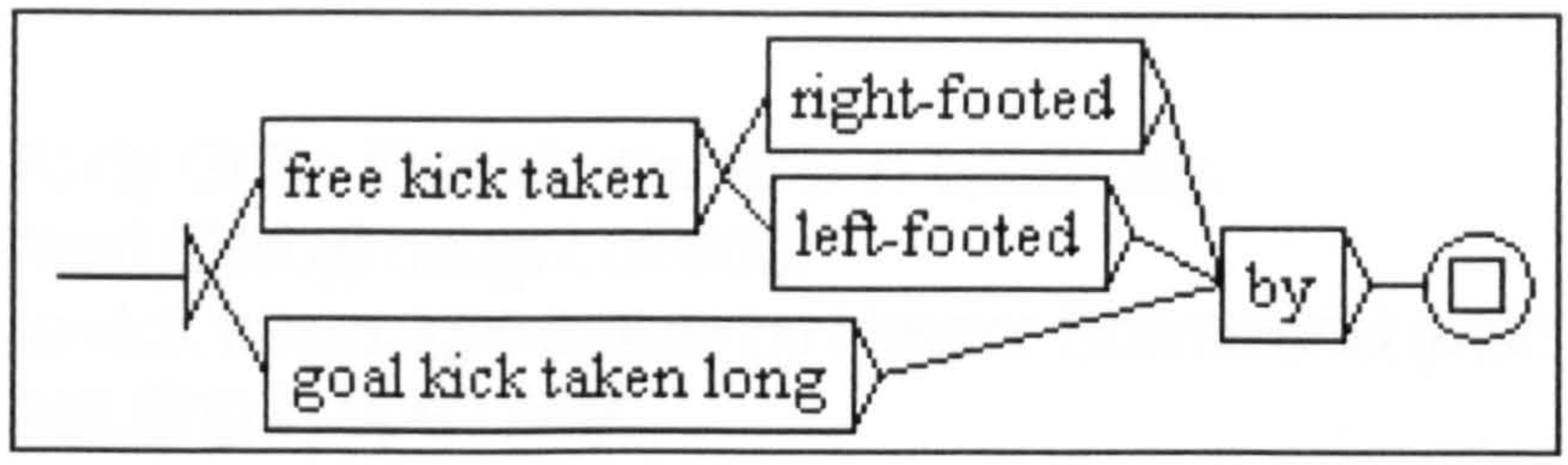


Figure 11: *kick* early collocations combined

Furthermore, an inspection will suggest that the *by* node feeds into nodes comprising names of players and team (in parentheses) and these can be joined up.

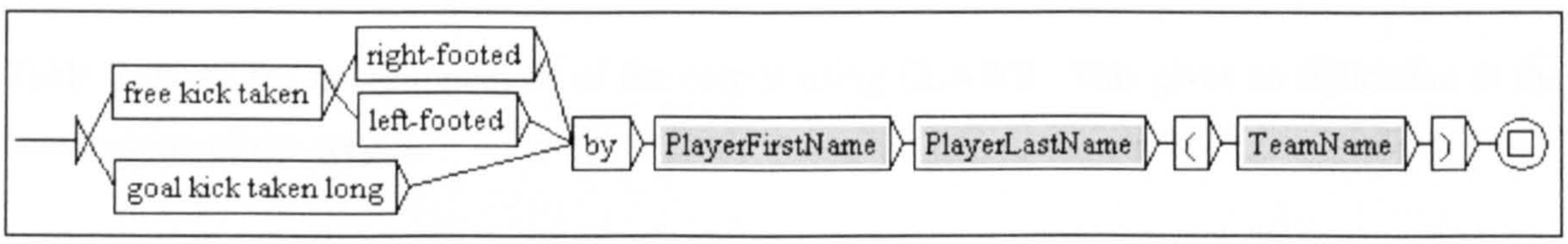


Figure 12: *kick* advanced collocation

3.1.3 Corpus Pre-Analysis

The first step to be taken before analysing the corpus is to separate the corpus sentences. Sometimes the live commentary stamps multiple events with the same time. For example, Figure 13 shows three events: *Cross*, *offside* and *Indirect free kick* and all have the same time stamp.

> 82:41
> Cross by Andy Griffin (Derby), resulting in open play. Steve Howard (Derby) caught offside. Indirect free kick taken right-footed by Kasper Schmeichel (Man City) from own half, resulting in open play.

Figure 13: Sample of the corpus multi-event time stamp (BBC online, 2007)

Our system will convert Figure 13 to Figure 14 where each event starts a new line.

82:41
Cross by Andy Griffin (Derby), resulting in open play.
Steve Howard (Derby) caught offside.
Indirect free kick taken right-footed by Kasper Schmeichel (Man City) from own half, resulting in open play.

Figure 14: Corpus multi-event time stamp separated

This is done simply by finding the full stop which marks the end of an event (sentence) or by finding an end-line.

Table 7 shows the overall analysis of the corpus using CLAWS. This gives an indication of the composition of the corpus.

Table 7: CLAWS analysis for Live Commentary Corpus (N ≅ 3,026,038)

Category	*f*	*f* / N	Description
NP1	861,987	28.49	singular proper noun (London, Jane, Frederick)
NN1	417,975	13.81	singular locative noun (street, Bay)
II	386,418	12.77	preposition
JJ	266,093	8.79	general adjective
)	231,718	7.66	
(	231,717	7.66	
.	215,897	7.13	
MC	191,838	6.34	cardinal number neutral for number (two, three...)
VVN	95,754	3.16	past participle form of lexical verb (given, worked...)
VVG	45,159	1.49	-ing form of lexical verb (giving, working etc.)
VV0	40,588	1.34	base form of lexical verb (give, work etc.)
VVD	38,503	1.27	past tense form of lexical verb (gave, worked etc.)

Proper noun and singular locative noun are the top 2 categories which, when combined together, produce a frequency of 37.52 %. Verbs, which usually present the motion entity, come in the 9th to the 12th position which, when combined, produce a frequency of about 6%. It will be interesting to see how these results differ when the language analysed is Arabic.

3.2 The Evolution of a Local Grammar

3.2.1 Vocabulary Analysis

The first step in reaching our primary objective was to analyse the corpus using System Quirk (Holmes-Higgin et al, 1993)[5]. System Quirk is an application that analyses a text and for each token it provides some statistical information such as: Frequency, Relative Frequency, and Real Frequency. Quirk et al (1985) categorized word classes into five groups: "Closed Classes (Preposition, Pronoun, Determiner, Conjunction, Modal verb, Primary verb); Open Classes (Noun, Adjective, Full verb, Adverb); Numerals; Interjections and Unique function" (1985:67). The top 50 most frequent words returned following analysis using System Quirk are shown in Table 8:

Table 8: QUIRK frequency analysis for live commentary corpus (N ≅ 3,026,038)

Rank	Token	*f*	*f* / N
1	)	231,718	7.66
2	(	231,717	7.66
3	by	179,335	5.93
4	.	162,007	5.35
5	:	123,734	4.09
6	-	113,157	3.74
7	,	68,113	2.25
8	in	65,998	2.18
9	right	53,514	1.77
10	from	52,024	1.72
11	footed	46,781	1.55
12	taken	46,132	1.52
13	throw	42,901	1.42
14	kick	41,013	1.36
15	left	31,838	1.05
16	free	25,958	0.86
17	attacking	25,334	0.84
18	play	22,056	0.73
19	resulting	21,869	0.72
20	goal	21,499	0.71
21	on	20,870	0.69
22	foul	19,875	0.66
23	open	19,798	0.65
24	defending	17,572	0.58
25	yards	16,760	0.55

Rank	Token	*f*	*f* / N
26	half	15,845	0.52
27	own	15,659	0.52
28	cross	14,827	0.49
29	long	14,485	0.48
30	of	14,315	0.47
31	channel	13,893	0.46
32	clearance	11,443	0.38
33	shot	11,378	0.38
34	caught	11,190	0.37
35	to	9,658	0.32
36	line	8,868	0.29
37	corner	8,837	0.29
38	save	8,421	0.28
39	penalty	8,342	0.28
40	area	8,065	0.27
41	paul	7,947	0.26
42	the	7,859	0.26
43	centre	7,503	0.25
44	city	5,417	0.18
45	mark	5,261	0.17
46	david	5,204	0.17
47	offside	5,174	0.17
48	lee	5,172	0.17
49	michael	5,122	0.17
50	side	5,008	0.17

[5] http://www.computing.surrey.ac.uk/SystemQ - accessed 10 May 2005

It is not a surprise that names like Paul, Mark and David appear as they were among the most common names back in the 70s and early 80s, given that the average player's age is late 20s to early 40s[6]. Also, one might question the parentheses and the full-stop being among the top 4. These tokens, as the analysis will show throughout this chapter, are important as they provide specific information and are mentioned in almost every pattern.

For the time being, one-character tokens, proper nouns and closed class words will be ignored as they do not contribute to a pattern meaning. The list then becomes:

Table 9: Most frequent open class words in live commentary corpus (N ≅ 3,026,038)

Rank	Word	*f*	*f* / N
1	right	53,514	1.77
2	footed	46,781	1.55
3	taken	46,132	1.52
4	throw	42,901	1.42
5	kick	41,013	1.36
6	left	31,838	1.05
7	free	25,958	0.86
8	attacking	25,334	0.84
9	play	22,056	0.73
10	resulting	21,869	0.72
11	goal	21,499	0.71
12	foul	19,875	0.66
13	open	19,798	0.65
14	defending	17,572	0.58
15	yards	16,760	0.55
16	half	15,845	0.52

Rank	Word	*f*	*f* / N
17	own	15,659	0.52
18	cross	14,827	0.49
19	long	14,485	0.48
20	channel	13,893	0.46
21	clearance	11,443	0.38
22	shot	11,378	0.38
23	caught	11,190	0.37
24	line	8,868	0.29
25	corner	8,837	0.29
26	save	8,421	0.28
27	penalty	8,342	0.28
28	area	8,065	0.27
29	centre	7,503	0.25
30	city	5,417	0.18
31	offside	5,174	0.17
32	side	5,008	0.17

Choosing the appropriate keywords plays a major part in an information extraction system. Allowing the keywords to be chosen manually is subjective and prevents automation. New keywords may exist but until they are added manually, the system will ignore them. Choosing keywords based on their frequency, as has been done so far, can be sufficient and some may consider it the best way. However, special characters are among the most frequent tokens and there is an argument for dismissing them.

[6] http://names.mongabay.com

A more meaningful method can be followed by choosing appropriate keywords. Ahmad et al.'s algorithm has been extended to compute the weirdness coefficient of each of the tokens, that is to compare the relative frequency of the tokens in a specialist corpus (the Commentary corpus) and the general language corpus (BNC). The normalized score, or *z*-score, of the frequency and weirdness of each token is then computed. Those tokens that have statistically significant differences in distribution, measured by whether or not the z-scores were positive, will be chosen. This can be done automatically and will allow the system to monitor any changes as they happen. The first 10 most frequent tokens (excluding the one character tokens) that have statistically significant distributions are shown in Table 10.

Table 10: Top Ten Keywords based on the weirdness calculation excluding the one character tokens (Special Corpus = 3,026,038 tokens, BNC = 100,000,000 tokens)

Specialist Corpus (*N*=3,026,038)				General Language Corpus (*N*=100,000,000)		Weirdness
Rank	Token	Abs. Freq	Real Freq	Abs. Freq	Real Freq	
1	footed	46,781	1.55%	38	0.00%	40,726
2	throw	42,901	1.42%	3317	0.00%	427
3	kick	41,013	1.36%	2269	0.00%	597
4	attacking	25,334	0.84%	1160	0.00%	722
5	foul	19,875	0.66%	1001	0.00%	656
6	defending	17,572	0.58%	1188	0.00%	489
7	clearance	11,443	0.38%	875	0.00%	432
8	line	8,868	0.29%	23716	0.02%	12
9	corner	8,837	0.29%	7096	0.01%	41
10	save	8,421	0.28%	7350	0.01%	38

Looking back at Ahmad et al's modified algorithm, more precisely where the average of all tokens' weirdness and the StDev (standard deviation) of all tokens' weirdness are calculated, it is noticed that *footed* has the highest weirdness level. With further analysis, Table 11 shows the list of words with at least zero weirdness level and the number of patterns that are associated with listed words.

Table 11: Keywords with minimum zero weirdness level and the patterns associated with them

(N=3,026,038, Total Corpus Patterns = 170,282)

Word	*f*	*f* / N	Total Patterns not Included in Higher Level Keywords	%	Total Distinguished Patterns	%	Weirdness Level
footed	46,781	27.47	46,781	27.473	46,781	27.47	12
line	8,868	5.21	490	0.288	47,271	27.76	
corner	8,837	5.19	270	0.159	47,541	27.92	
save	8,421	4.95	2,680	1.574	50,221	29.49	3
penalty	8,342	4.90	1,048	0.615	51,269	30.11	
area	8,065	4.74	103	0.060	51,372	30.17	
centre	7,503	4.41	37	0.022	51,409	30.19	2
offside	5,174	3.04	53	0.031	51,462	30.22	
side	5,008	2.94	368	0.216	51,830	30.44	
inswinging	4,835	2.84	0	0.000	51,830	30.44	
missed	4,801	2.82	88	0.052	51,918	30.49	
passed	4,779	2.81	3	0.002	51,921	30.49	
post	4,734	2.78	125	0.073	52,046	30.56	
drilled	4,466	2.62	2	0.001	52,048	30.57	1
wing	4,043	2.37	216	0.127	52,264	30.69	
substitution	3,628	2.13	39	0.023	52,303	30.72	
tactical	3,045	1.79	1	0.001	52,304	30.72	
over	2,928	1.72	983	0.577	53,287	31.29	
header	2,898	1.70	60	0.035	53,347	31.33	
blocked	2,837	1.67	499	0.293	53,846	31.62	
kick	41,013	24.09	11,116	6.528	64,962	38.15	
attacking	25,334	14.88	24,153	14.184	89,115	52.33	
foul	19,875	11.67	43	0.025	89,158	52.36	
defending	17,572	10.32	16,794	9.862	105,952	62.22	
clearance	11,443	6.72	4,974	2.921	110,926	65.14	
outswinging	2,691	1.58	0	0.000	110,926	65.14	
out	2,547	1.50	1,262	0.741	112,188	65.88	
bar	2,488	1.46	16	0.009	112,204	65.89	
far	2,175	1.28	27	0.016	112,231	65.91	
behaviour	1,817	1.07	162	0.095	112,393	66.00	
unsporting	1,817	1.07	0	0.000	112,393	66.00	0
assist	1,268	0.74	2	0.001	112,395	66.01	
unknown	722	0.42	0	0.000	112,395	66.01	
hit	579	0.34	78	0.046	112,473	66.05	
headed	465	0.27	5	0.003	112,478	66.05	
football	444	0.26	26	0.015	112,504	66.07	
given	417	0.24	7	0.004	112,511	66.07	
round	404	0.24	35	0.021	112,546	66.09	
dissent	329	0.19	91	0.053	112,637	66.15	
every	248	0.15	249	0.146	112,886	66.29	
victory	242	0.14	14	0.008	112,900	66.30	

Table 11 shows, for each keyword, the total number of patterns that are associated with it, the total number of patterns percentage compared to the total patterns in the corpus, the total number of patterns that other higher level keywords have reported already and the percentage of that comparing to the total patterns in the corpus; for example, the pattern *Penalty kick taken right-footed* will be counted with *footed* since *footed* has the highest weirdness level and will not be counted as a pattern for *kick* or *taken*. The progressive total is the total count of the unique patterns and the percentage of those patterns compared to the patterns in the total corpus. *footed* alone would allow the system to catch 27% of the corpus patterns. When adding the patterns of the keywords with level 3 weirdness, the system will be able to catch 30% of the corpus total patterns. Catching all the patterns from all the keywords with minimum weirdness of level 0 will allow the system to catch 66.30% of the total corpus patterns. This indicates that, using this system, automation should be readily achievable.

3.2.2 Collocation Analysis

At this point, the output from System Quirk is taken and used as an input for COLLOCATE, which is part of the System Quirk workbench[7]. The purpose of this application is to find the concordance for those chosen motion words throughout the corpus. It applies Smadja's method and allows re-collocation to be performed. Collocations are word (single or compound) pairs. The pair can exist next to each other or within 5 neighbours left or right with interspersing words in between. We have used the outlined method by Smadja (1994) for retrieving collocations from text. Smadja defines a peak containing a high frequency word *w* as a tuple (w_i, *distance*, *strength*, *spread*, *j*) that verifies the following criteria:

$strength = \frac{freq_i - \overline{f}}{\sigma} \geq k_0$	(1)
$spread \geq U_0$	(2)
$p^i_j \geq \overline{p_i} + (k_1 \times \sqrt{U_i})$	(3)

Equation 1: Smadja high frequency criteria

Where

(1) is used to eliminate low frequency collocates. The $freq_i$ is the frequency of the collocation of w_i with *w*; $\overline{f}$ is the average frequency, σ is the standard deviation and k_0 is

[7] http://www.computing.surrey.ac.uk/SystemQ - accessed 13 July 2005

the strength threshold. This threshold usually has a value of one for the task of language generation.

(2) requires that the histogram of the ten relative frequencies of appearance of w_i within five words of w to have at least one spike. The histograms are rejected if the variance threshold $U_0 < 10$. The variance is usually computed using the equation below.

$$U_i = \frac{\sum_{j=1}^{10}\left(p_i^j - \overline{p_i}\right)^2}{10}$$

where p_j^i (j in the above tuple) and $\overline{p_i}$ are the frequency of one collocate at a certain distance from w and their average respectively.

(3) pulls out the significant relative positions of two words. Thus, this inequality eliminates columns whereas (a) and (b) select rows. It states that the frequency threshold of one collocate at a certain distance from w has to be at least one standard deviation above the average frequency of one row collocates ($k_1 = 1$).

Smadja suggests, and we can confirm, that the parameters for (k_0, k_1, U_0) = (1, 1, 10) give good results for our collocation. The collocate of the word *taken* was chosen because it is one of the most frequently used open class words in the collection.

Table 12 shows the first four collocations for *taken.* An explanation will be provided of how local grammar is extracted and finite automation is built.

Table 12: *taken* Collocation (N ≅ 3,026,038)

Step	-1	keyword	1	2	*f*	U-Score	K-Score	Strength
1	kick	taken			37,395	125,518,000	29.35	3.00
2	free	kick taken			22,449	45,356,200	20.99	3.00
	goal	kick taken			14,895	19,967,500	13.88	3.00
		kick taken	right-footed		17,524	27,638,200	16.36	3.00
		kick taken	long		14,380	18,610,600	13.40	3.00
		kick taken	left-footed		4,924	2,182,120	4.50	3.00
		kick taken		by	36,811	124,656,000	36.24	3.12
3		free kick taken	right-footed		17,524	27,638,200	16.36	3.00
		free kick taken	left-footed		4,924	2,182,120	4.50	3.00
		free kick taken		by	22,448	45,350,189	20.93	3.00
		goal kick taken	long		14,383	18,606,900	21.27	3.00
		goal kick taken		by	14,363	19,865,700	22.53	3.07
4		free kick taken right-footed	by		17,524	27,358,100	34.49	3.18
		free kick taken left-footed	by		4,924	2,159,150	30.32	3.19
		goal kick taken long	by		14,363	19,865,700	22.53	3.07

In Step (1), both *kick* and *taken* are among the 10 most frequent words in the corpus. The word *taken* comprises 1.52% of the corpus (46,132 words) and the word *kick* 1.36% of the corpus (41,013 words). The collocation of *kick* and *taken* occurs 37,395 times. That is, if it is accepted that *kick* is used in conjunction with *taken* 91% of the time and *taken,* being more frequent, is used with *kick* 80% of the time. This is a very strong collocation which indicates the nature of the language that is being used. This will give us a finite state automaton, shown in Figure 15:

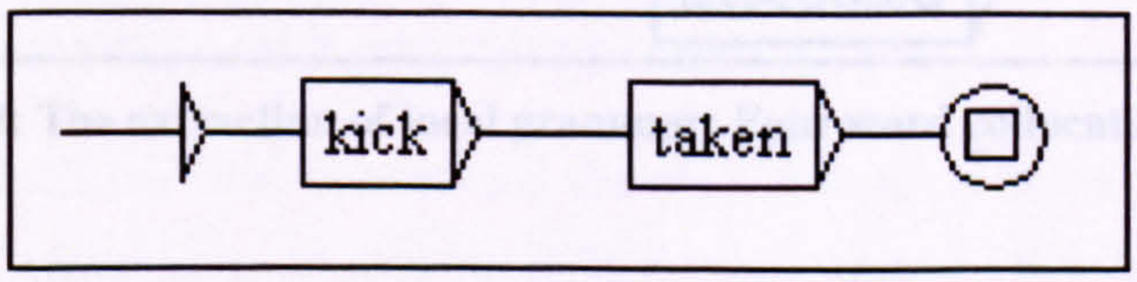


Figure 15: The extraction of local grammar: Two word collocation

In Step (2), the collocation of *kick taken* is examined and it can be seen that it has a number of other collocates that are significant. Looking at three words collocates we find: *free kick taken* (22,449), *goal kick taken* (14,895), *kick taken right-footed* (17,524), *kick taken long* (14,380) and *kick taken left-footed* (4,924). These are the most frequent collocates of *kick taken*. We can then extend the finite automaton in Figure 15 to Figure 16 and Figure 17:

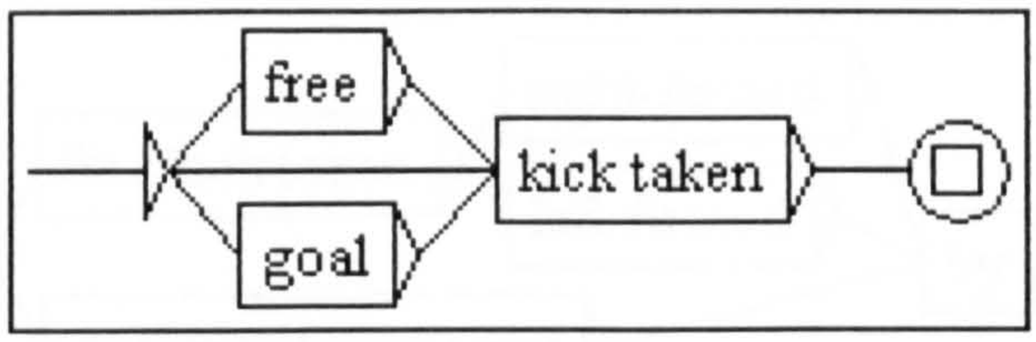


Figure 16: The extraction of local grammar: Three word collocation Part-1

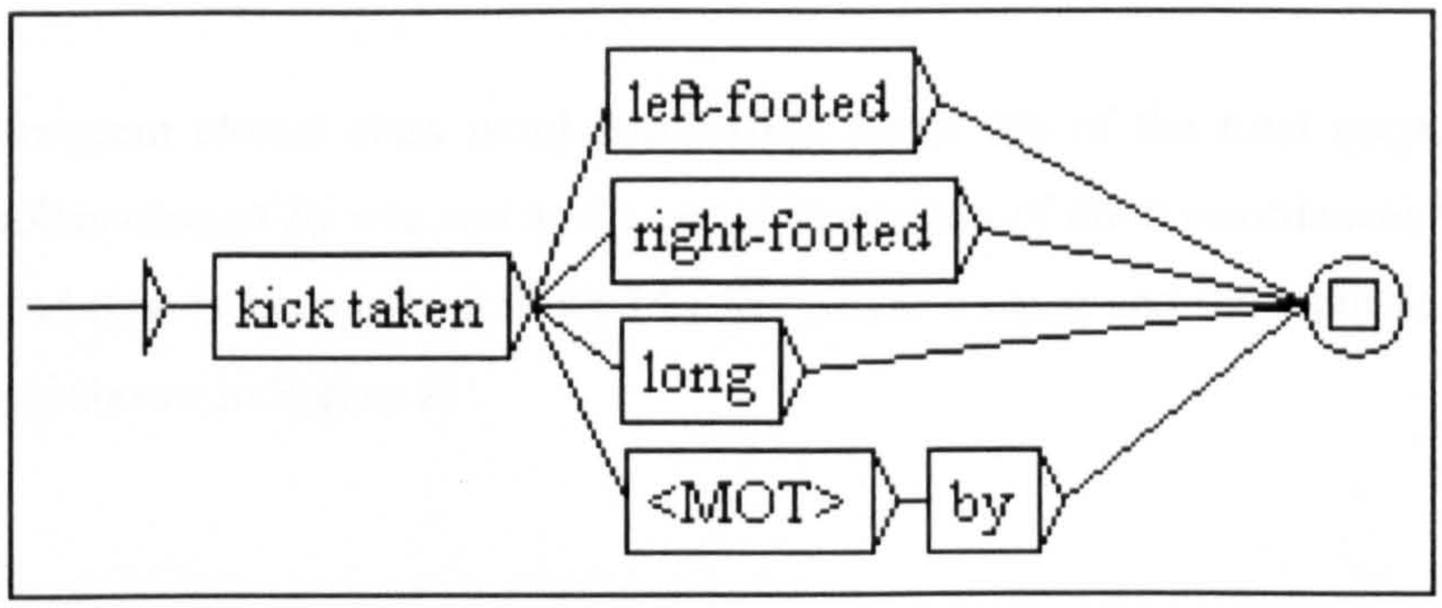


Figure 17: The extraction of local grammar: Three word collocation Part-2

In Step (3), when we take the collocation *kick taken right-footed* or *kick taken left-footed* the key collocate is *free*; at the same time, in the collocation *kick taken long* the key collocate is *goal*. This leads to Figure 18 and Figure 19:

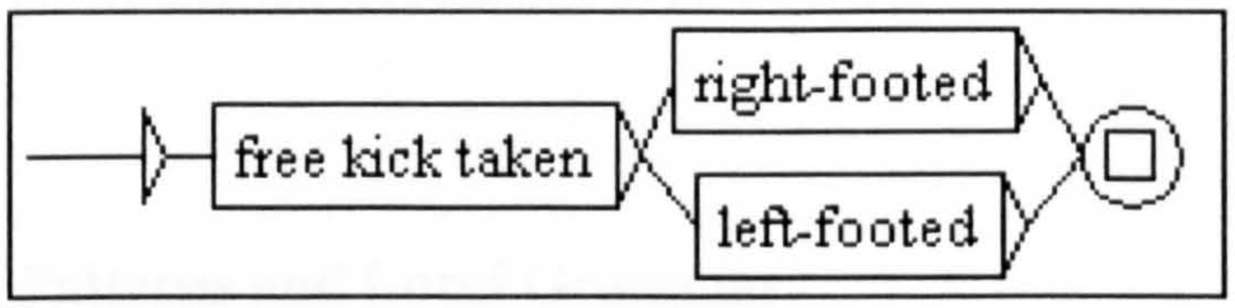


Figure 18: The extraction of local grammar: Four word collocation Part-1

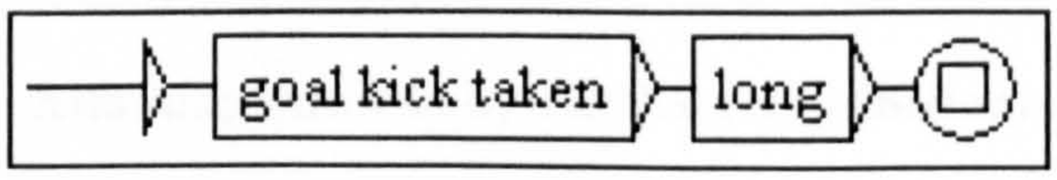


Figure 19: The extraction of local grammar: Four word collocation Part-2

For Step (4), it can be seen that by doing 3 words collocation, there are now certain fixed frozen phrases which can be used. If this analysis is extended, we see the key *by* is added to give new patterns, as in Figure 20:

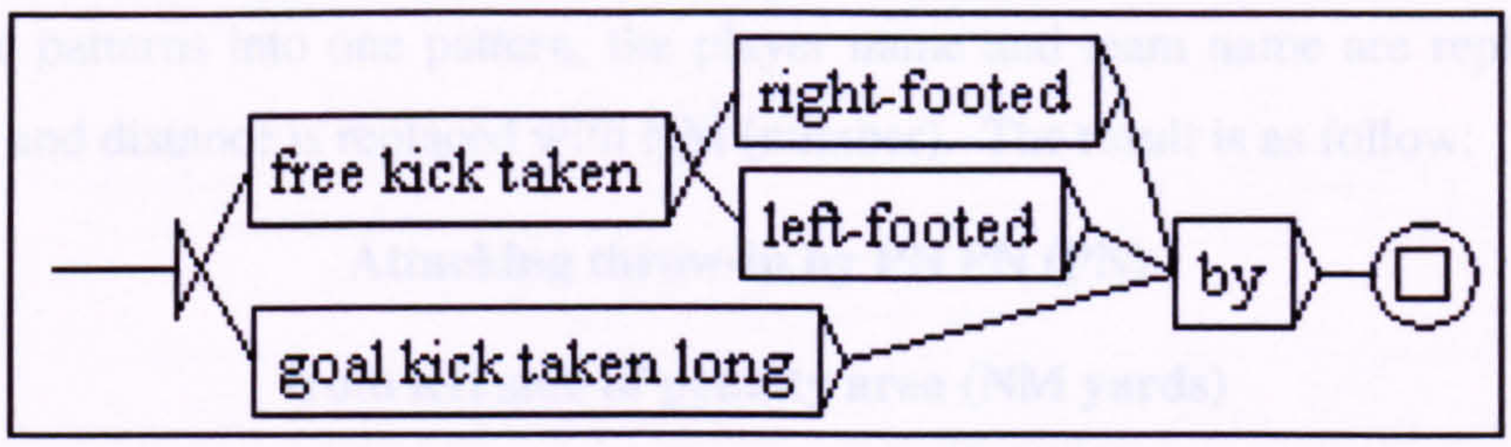


Figure 20: The extraction of local grammar: 5-6 word collocation

By is a very frequent closed class word comprising about 6% of the total corpus. When the frequency of collocation of *by* was run and a visual inspection of the concordances was produced, it was discovered that *by* is usually followed by the player's name and the team name. The final local grammar is shown in Figure 21:

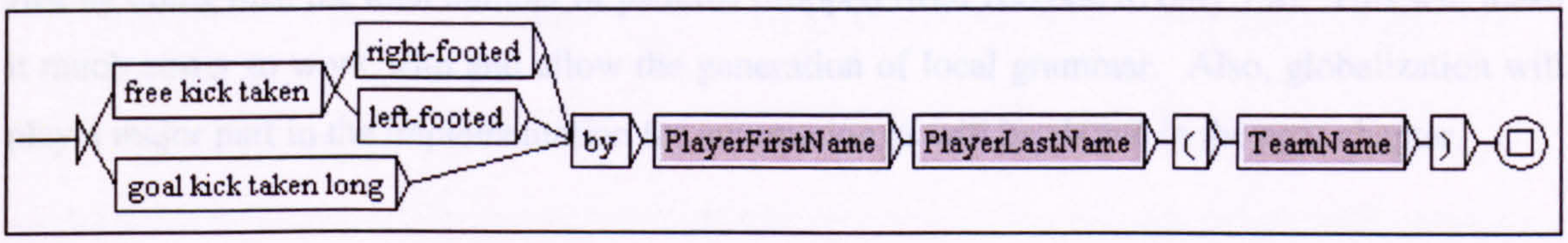


Figure 21: A local grammar for ball – by – ball commentary

Appendix A contains a set of further, less frequent, grammar patterns which describe usual events in more detail, including the pattern *free kick taken* which shows that more keys can be added to the local grammar if more than five collocations steps are taken.

3.2.3 Unifying Patterns and Local Grammar

It is noted that many patterns look similar and the only difference is the player's name and the team name. Also, in some patterns only the distance is changed. For example, the following patterns are the same with exception to the players' names and teams' names:

Attacking throw-in by Tal Ben Haim (Bolton)

Attacking throw-in by Ivan Campo (Bolton)

Also, the following sub-patterns only have different distances.

from left side of penalty area (18 yards)

from left side of penalty area (12 yards)

To unite these patterns into one pattern, the player name and team name are replaced with PN (proper noun) and distance is replaced with NM (number). The result is as follow:

Attacking throw-in by PN PN (PN).

from left side of penalty area (NM yards)

That led to the introduction of a procedure, which we termed *globalization*:

1. Every capitalized word in a sentence is to be replaced with PN with exception to the first word in the sentence.
2. Every number in a sentence is to be replaced with NM.

Just by doing this, the total number of patterns dropped from 100,000 to only 190. This will make it much easier to work with and allow the generation of local grammar. Also, globalization will play a major part in the implementation for automation as will be shown in the next chapter.

A point that can be argued is whether or not it is beneficial to treat repeated PN as one. For example:

Table 13: Sample of actual patterns and their global patterns

Actual Pattern	Globalized
Attacking throw-in by Leighton Baines (Wigan).	Attacking throw-in by PN PN (PN)
Attacking throw-in by Phil Jagielka (Sheff Utd).	Attacking throw-in by PN PN (PN PN)

Wigan, the team name, is changed to PN; whereas Sheffield Utd is changed to PN PN. Should this be followed for all patterns? The system will not count them as one; instead it will use them as different unique patterns. In this thesis, repeated PN will not be counted as one.

For the system to automate video indexing, it needs a local grammar so that it can validate patterns from the live commentary. Looking back at the *kick taken* collocation, the following is one of its patterns

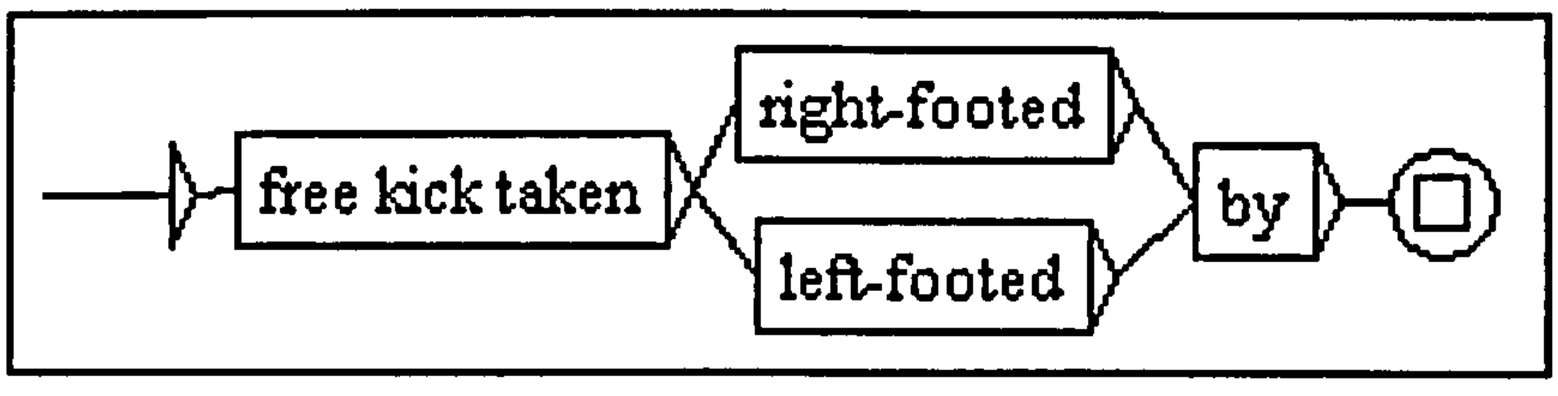


Figure 22: *kick taken* early collocation

Let this pattern be identified as ***Kick-1.*** Table 14 shows the collocation of this sub-pattern.

Table 14: *Kick-1* next phase collocation

Key Token	Right Phrase Collocate	
KICK-1	PN PN (PN PN)	
	PN PN PN (PN PN)	
	PN PN PN PN (PN PN)	

The above table can be summarized as follows:

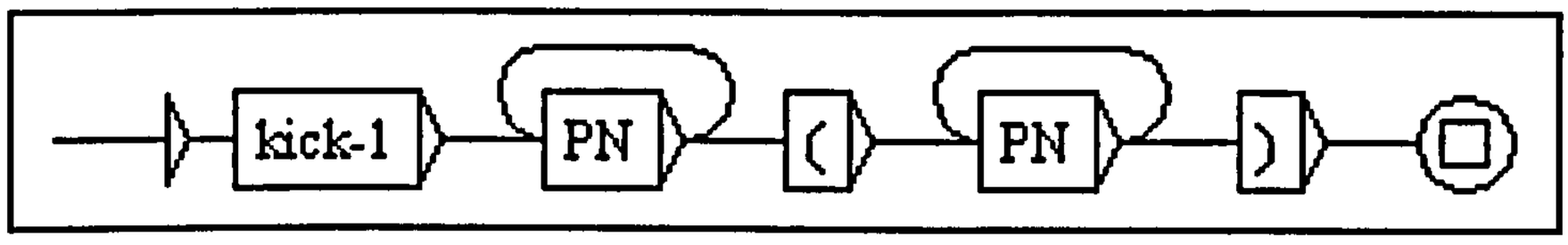


Figure 23: *Kick-1* next phase collocation

The kick pattern, up to now, can now be showing as follow:

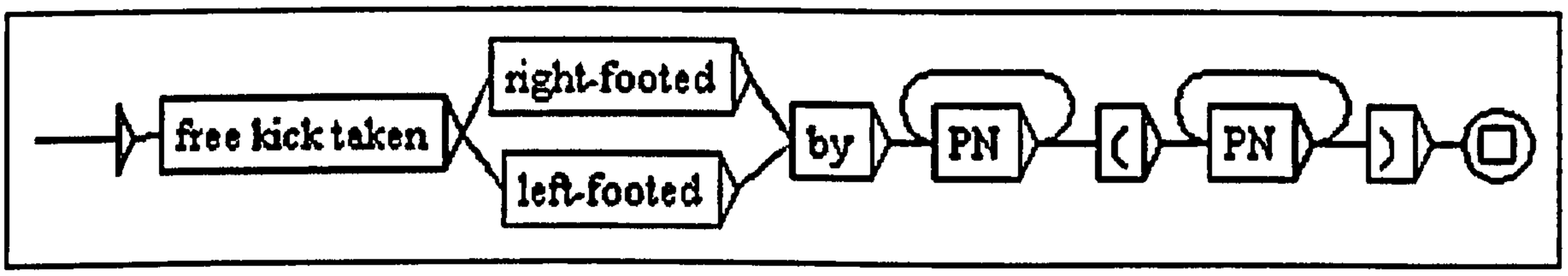


Figure 24: *kick* collocation summarized

Again, let the above free kick taken pattern be ***FK.***

And let this PN pattern be named ***PNP***

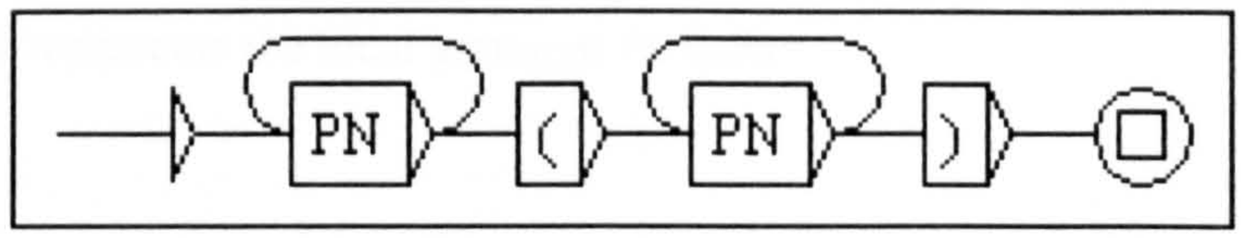


Figure 25: *PNP* collocation

As before, let ***FK*** be our key collocate and run the next collocation. Table 15 shows the complete pattern with all its possibilities.

Table 15: FK phrase collocation

Key Token	Key Collocate	Right Collocate	Right Phrase(s) Collocate	
FK (f = 23,468)	from (f = 17,616)	centre	of penalty area	passed Resulting in open play
		own	half (NM yards)	missed right missed left over the bar save (caught) by PNP blocked by PNP
		right left	by-line	passed resulting in open play
			channel	(NM yards) blocked by PNP
			side of penalty area	passed Resulting in open play (NM yards) save (caught) by PNP
			side of six-yard box	passed resulting in open play
			wing	(NM yards) missed right (NM yards) missed left save (tipped round post) by PNP clearance by PNP save (caught) by PNP save (tipped over) by PNP

Such analysis was also completed for the other high-weirdness keywords collocation. The following FS (Finite State) diagram represents all their patterns.

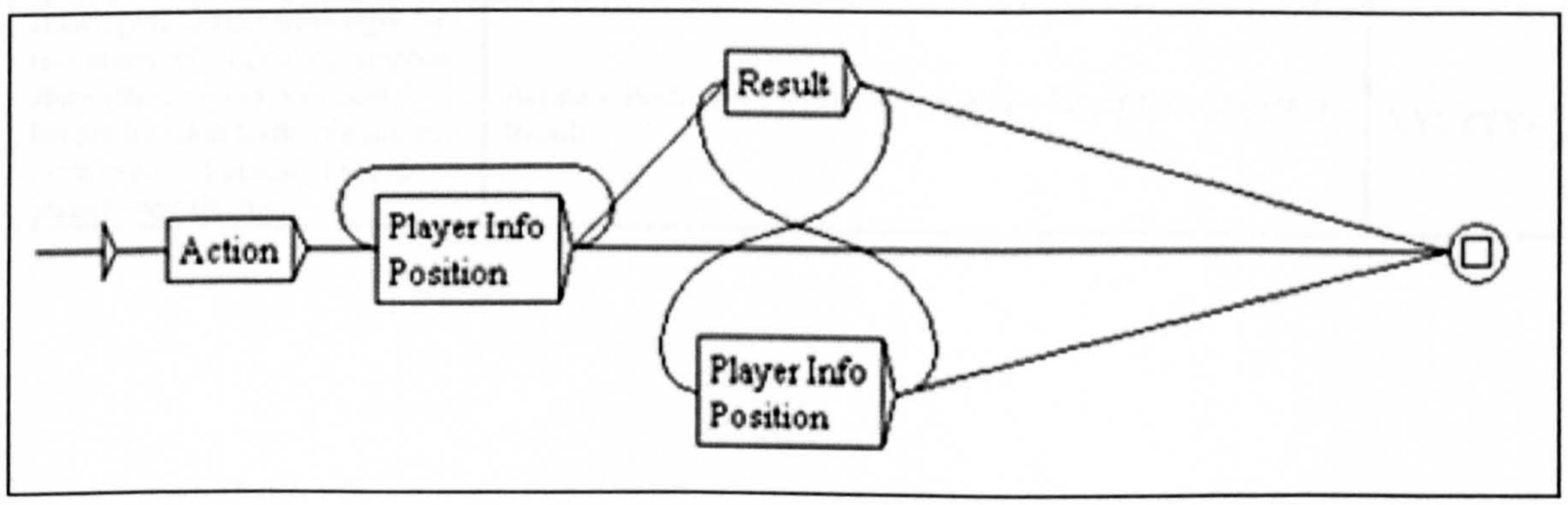


Figure 26: The Local Grammar Finite Automata

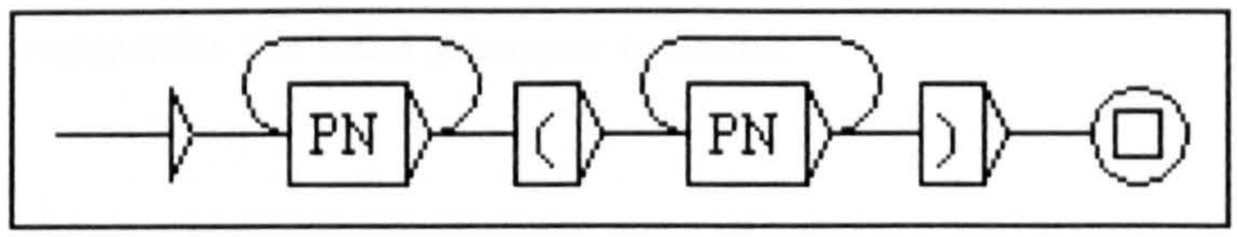


Figure 25: *PNP* collocation

As before, let ***FK*** be our key collocate and run the next collocation. Table 15 shows the complete pattern with all its possibilities.

Table 15: FK phrase collocation

Key Token	Key Collocate	Right Collocate	Right Phrase(s) Collocate	
FK (f = 23,468)	from (f = 17,616)	centre	of penalty area	passed Resulting in open play
		own	half (NM yards)	missed right missed left over the bar save (caught) by PNP blocked by PNP
		right left	by-line	passed resulting in open play
			channel	(NM yards) blocked by PNP
			side of penalty area	passed Resulting in open play (NM yards) save (caught) by PNP
			side of six-yard box	passed resulting in open play
			wing	(NM yards) missed right (NM yards) missed left save (tipped round post) by PNP clearance by PNP save (caught) by PNP save (tipped over) by PNP

Such analysis was also completed for the other high-weirdness keywords collocation. The following FS (Finite State) diagram represents all their patterns.

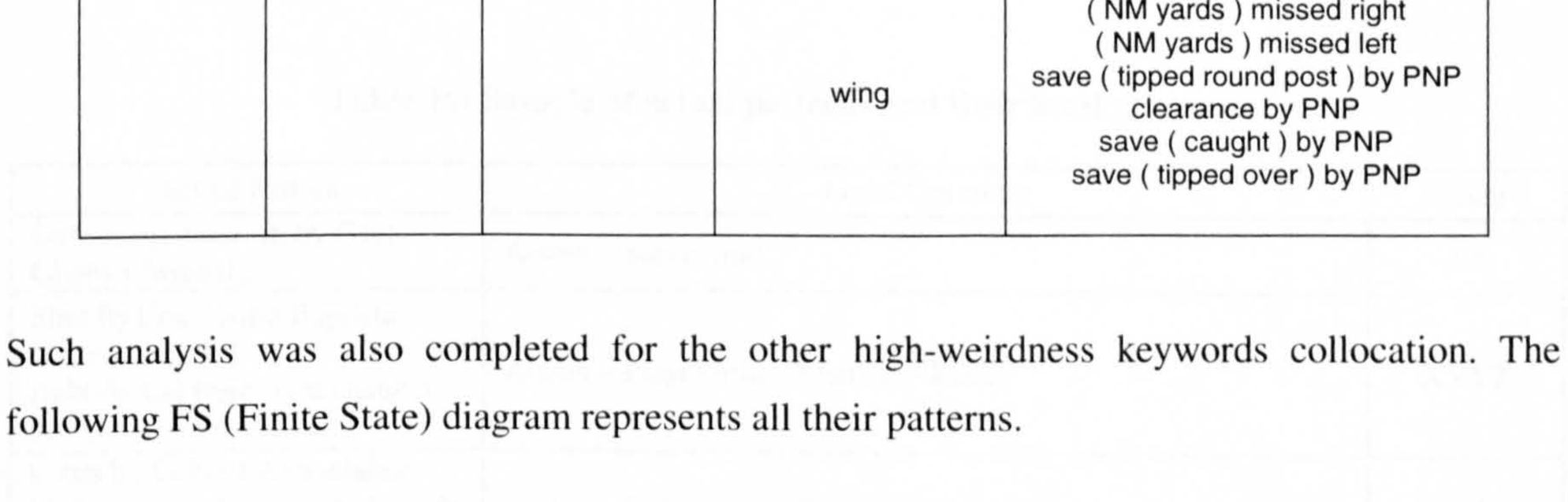


Figure 26: The Local Grammar Finite Automata

Equation 2 below represents the local grammar formula:

$$\sum_{t-start-time}^{end-time} [X_t Y_t^+ (Z_t \oplus Y_t)^*]$$

Equation 2: The local grammar formula

Where:

- X = Action
- Y = Player Info or Position
- Z = Result.
- $\oplus$ = OR
- $+$ = At least one repetition
- $*$ = Possible repetition

The following are examples of possible occurrence (See Table 16 below):

Table 16: Sample of actual patterns and their local grammar

Actual Pattern	Local Grammar	String
Defending throw-in by Gael Clichy (Arsenal).	Action – player info	XY
Shot by Cesar Julio Baptista (Arsenal) curled right-footed from right channel (20 yards), over the bar.	Action – Player info - Position – Result	XYYZ
Cross by Gabriel Agbonlahor (Aston Villa), blocked by Franck Queudrue (Fulham).	Action – Player Info – Result – Player Info	XYZY
Inswinging corner from right by-line taken left-footed by Stephen Hunt (Reading) to near post, header by Glen Little (Reading) from centre of penalty area (6 yards), over the bar.	Action – Position – Player Info - Result – Player Info – Position - Result	XYYZYYZ

3.3 System Design

Figure 27 graphically depicts the overall system process. A video and its text are the input for the system. First, they go through the training section. Second, the patterns are synchronized with the video. Finally, video indexing and video annotation is applied.

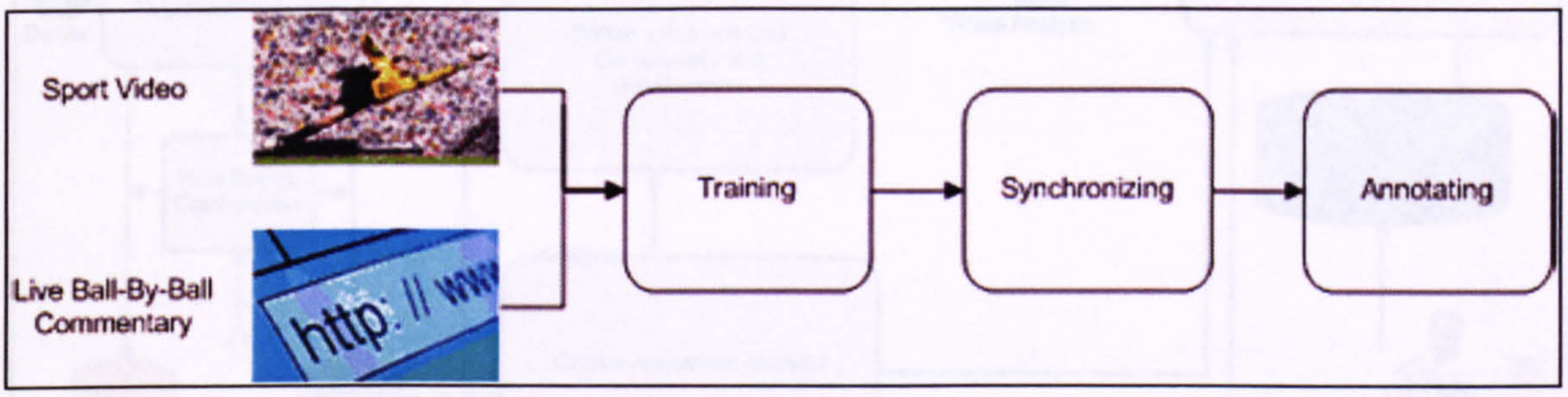


Figure 27: Overall system design

Figure 28 below shows the system in more detail. The system is divided into 4 sections: System Server, Synchronizer, Video Clipping and CPU.

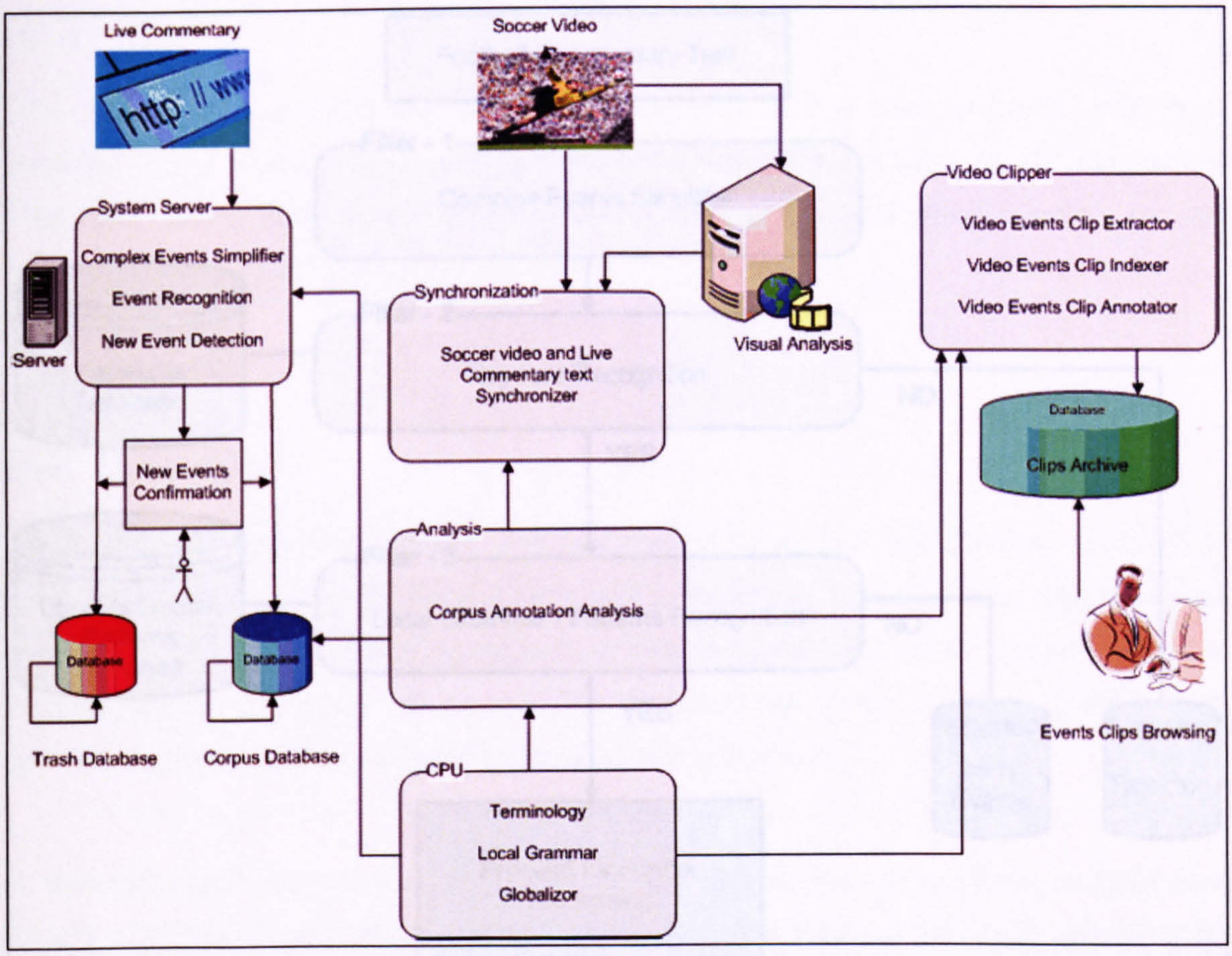


Figure 28: Complete system in detail

3.3.1 System Server

System server is divided into 3 sections: Complex Events Simplifier, Event Recognition and New Event Detection. When the commentary text file is received by the System Server, it simplifies any complex events. Then the commentary text is processed to filter its events into recognized events and new events. The recognized events are sent directly to either the Corpus Database if it passed the local grammar or to the Trash Database if it matches a pattern that was rejected before. New events, once detected, are validated and confirmed to which database it belongs to. Figure 29 below shows a proposed algorithm of automation of events detection in football commentary text.

.

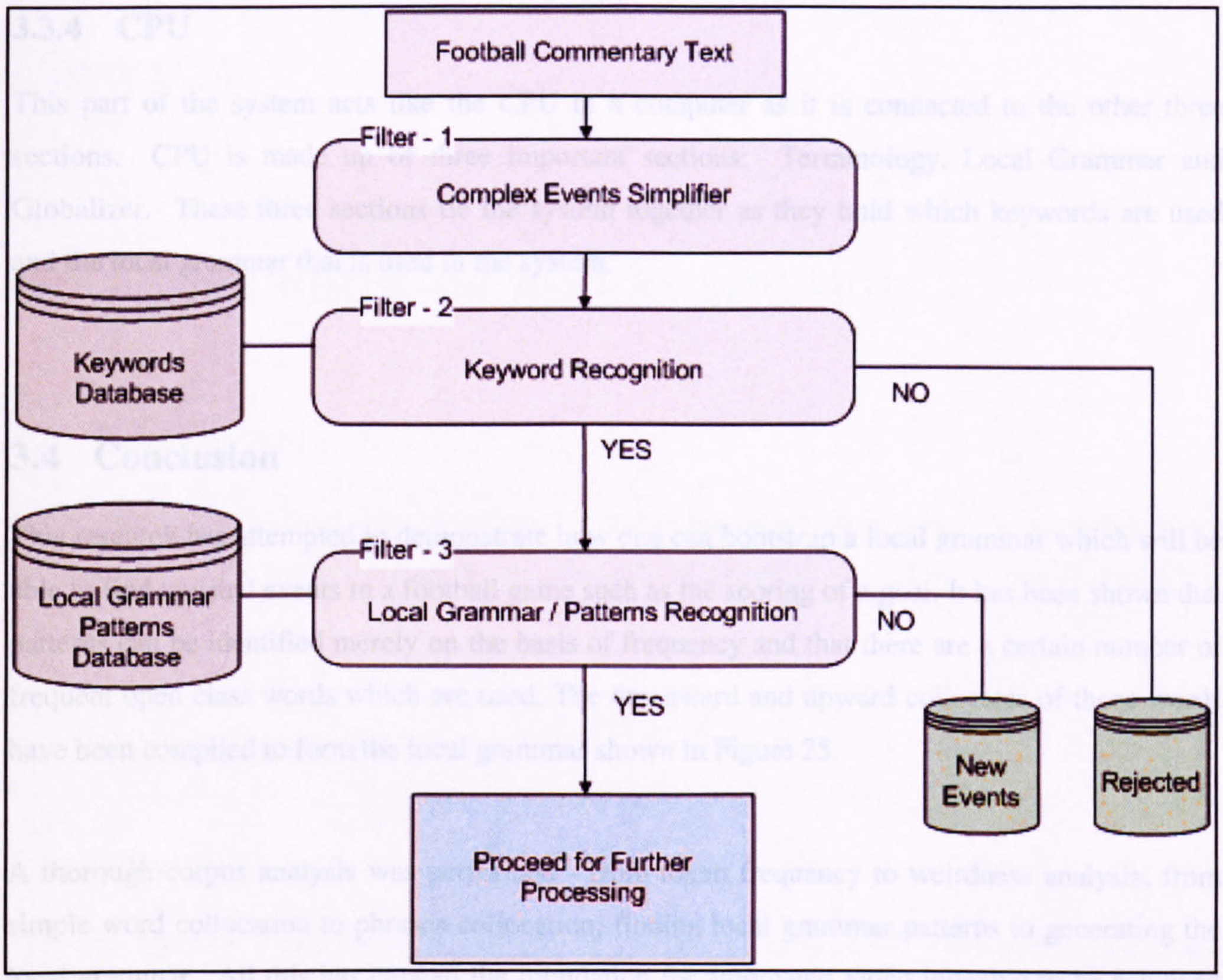


Figure 29: Proposed algorithm for filtering events in football commentary text

3.3.2 Synchronizer

This is the part where the system synchronizes the events that are detected in the commentary text with their segments in the video. At this point, Visual Analysis is needed just to train the system and get an average interval time for the detected events.

3.3.3 Video Clipper

This part of the system receives the synchronization results and does the decoding and the encoding. For each event, it extracts its segments from the video, indexes it, annotates it and then archives it. An end-user has the ability to perform simple and complex searches on these archived clips.

3.3.4 CPU

This part of the system acts like the CPU in a computer as it is connected to the other three sections. CPU is made up of three important sections: Terminology, Local Grammar and Globalizer. These three sections tie the system together as they hold which keywords are used and the local grammar that is used in the system.

3.4 Conclusion

This research has attempted to demonstrate how one can bootstrap a local grammar which will be able to find unusual events in a football game such as the scoring of a goal. It has been shown that patterns can be identified merely on the basis of frequency and that there are a certain number of frequent open class words which are used. The downward and upward collocates of those words have been compiled to form the local grammar shown in Figure 25.

A thorough corpus analysis was performed - from token frequency to weirdness analysis; from simple word collocation to phrases collocation; finding local grammar patterns to generating the local grammar. All this has created the foundation for automated video indexing to be achieved as will be shown in the next chapter.

It should be noted that the system started without any pre-specified keywords. No assumption was made about any word or about patterns which were identified by looking at the corpus itself.

.

Chapter 4

4 Implementation and Evaluation

4.1 Introduction

In this chapter the automated video indexing implementation will be introduced. A walk-through of how it works and the options it includes will be provided. Also, a corpus evaluation is performed.

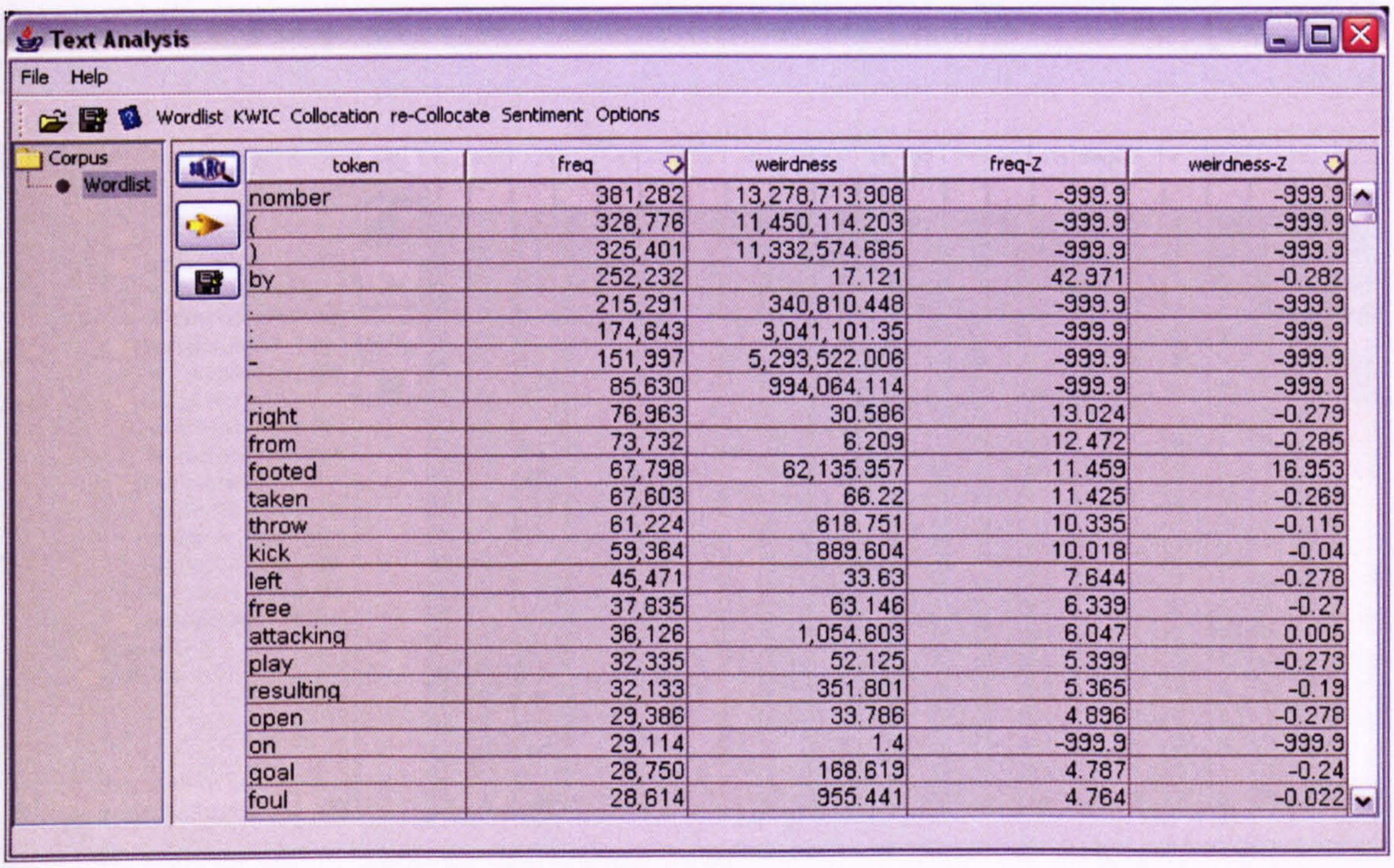
4.2 Implementation

4.2.1 Text Analysis

The first application to be used is *Text Analysis System* (Cheng 2007). This application is the result of Cheng's Ph.D. thesis which was based on System Quirk and COLLOCATE applications that we have previously used in this thesis. As a result, *Text Analysis System* is an advanced version of System Quirk and COLLOCATE and has the ability to do the following:

- Tokens frequency, tokens weirdness, frequency Z-score and weirdness Z-score
- Tokens concordance
- Tokens collocation and re-collocation
- The ability to change the settings for collocation and concordance to get different results
- The ability to include and exclude specific words

This application has eliminated the need to use System Quirk and COLLOCATE individually. Also, calculations are done much faster now, see Figure 30.



token	freq	weirdness	freq-z	weirdness-z
nomber	381,282	13,278,713.908	-999.9	-999.9
(	328,776	11,450,114.203	-999.9	-999.9
)	325,401	11,332,574.685	-999.9	-999.9
by	252,232	17.121	42.971	-0.282
.	215,291	340,810.448	-999.9	-999.9
:	174,643	3,041,101.35	-999.9	-999.9
-	151,997	5,293,522.006	-999.9	-999.9
,	85,630	994,064.114	-999.9	-999.9
right	76,963	30.586	13.024	-0.279
from	73,732	6.209	12.472	-0.285
footed	67,798	62,135.957	11.459	16.953
taken	67,603	66.22	11.425	-0.269
throw	61,224	618.751	10.335	-0.115
kick	59,364	889.604	10.018	-0.04
left	45,471	33.63	7.644	-0.278
free	37,835	63.146	6.339	-0.27
attacking	36,126	1,054.603	6.047	0.005
play	32,335	52.125	5.399	-0.273
resulting	32,133	351.801	5.365	-0.19
open	29,386	33.786	4.896	-0.278
on	29,114	1.4	-999.9	-999.9
goal	28,750	168.619	4.787	-0.24
foul	28,614	955.441	4.764	-0.022

Figure 30: Cheng's Text Analysis showing tokens, frequency, weirdness, frequency z-score and weirdness z-score (Cheng 2007)[8]

Figure 31 below shows the concordance and the collocation of *kick*

[8] Text Analysis uses **nomber** for all the number.

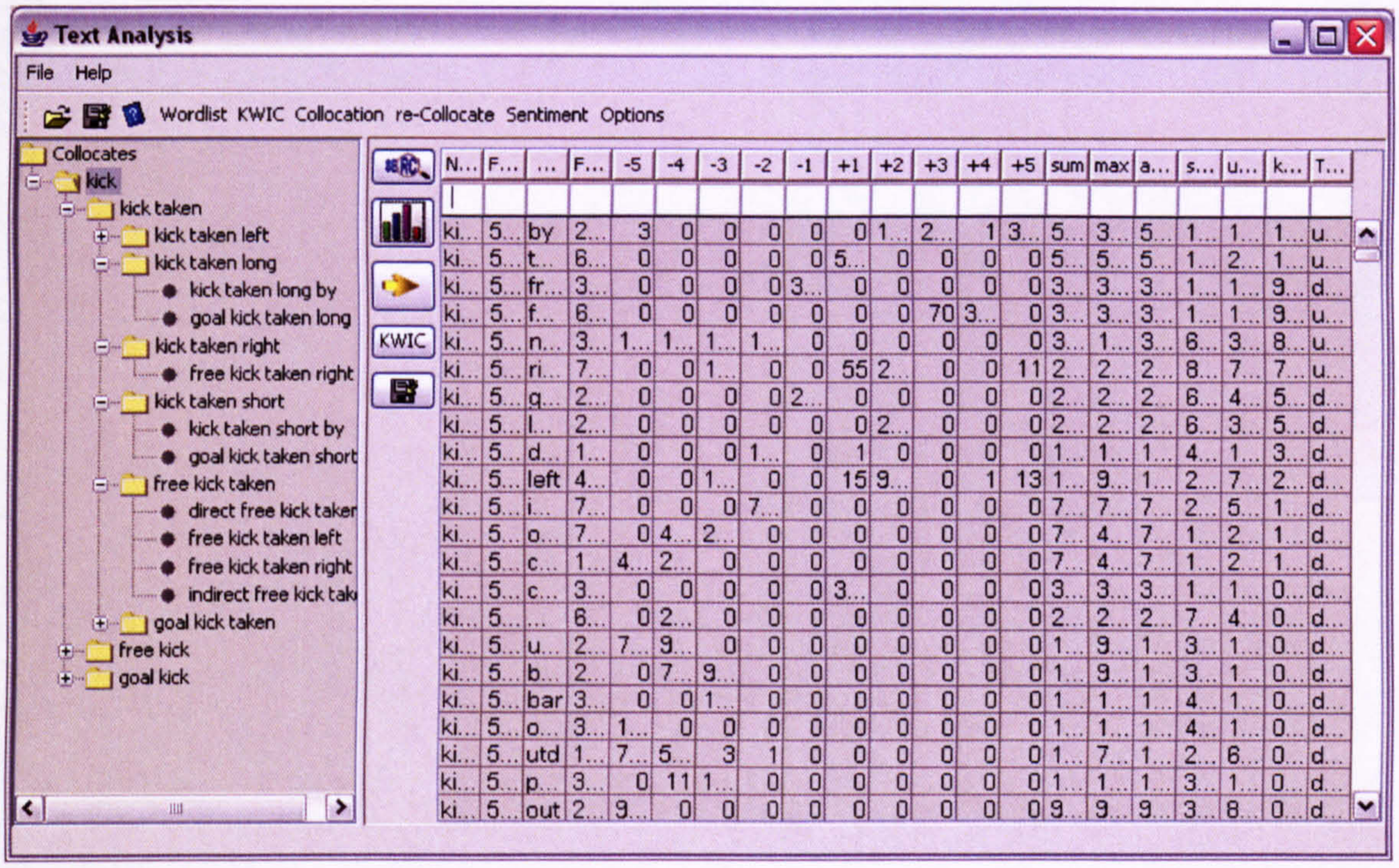


Figure 31: Cheng's Text Analysis showing a sample of *kick* collocation (Cheng 2007)

The saved result of the Text Analysis application is the input of another application *Protégé*[9]. This input file needs no modification or editing; it is setup by the Text Analysis application to be a *Protégé* input file.

Protégé is an application that shows the variant patterns that are detected from a single token and its collocation(s), see Figure 32.

[9] http://protege.stanford.edu

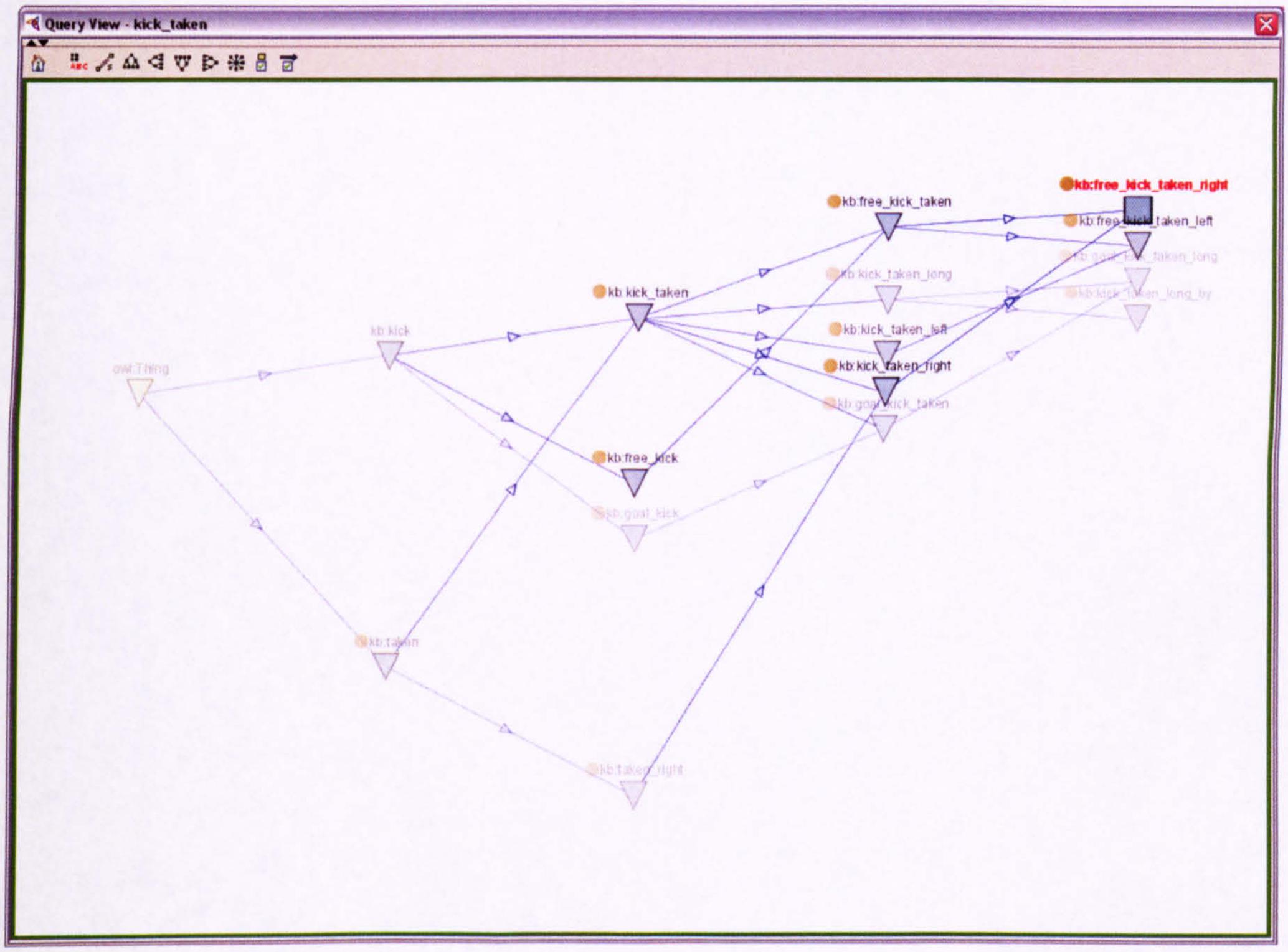


Figure 32: *Protégé* showing the patterns that are detected for *kick*

Protégé makes it easier to detect the patterns, especially if a new pattern exists. *Protégé* can also provide a collocation list as shown in Figure 33.

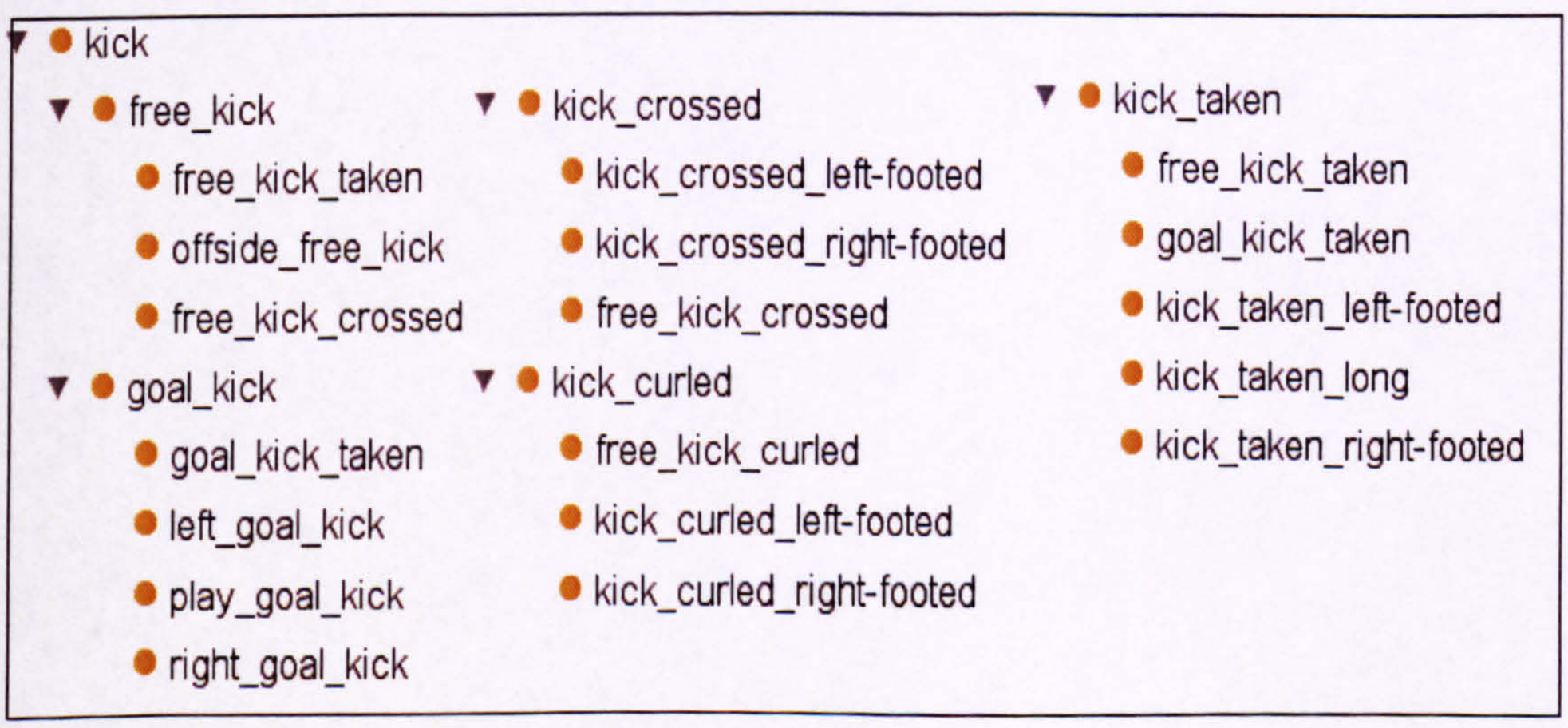


Figure 33: *Protégé* showing significant collocates of kick and associated collocates

Note that fast and direct results are obtained from these two applications without the need to go through several applications to achieve the same results.

4.2.2 Text and Video Synchronization

As noted before, events in the live commentary text are time-stamped and this time-stamp is essential in allowing the system to annotate the video. Two issues need to be considered in order to achieve an efficient video annotation: (i) Multiple events sharing the same time-stamp which usually occurs when there is a fast-moving game with actions happening quickly over a very short period of time, commentated upon at speed and (ii) time-stamp not being correctly reported by the commentary text.

The time-stamp can be dedicated to one event as shown in Figure 34

9:31
Goal kick taken long by Matthew Glennon (Huddersfield).

Figure 34: Sample of commentary text 1-Event time-stamp

.

Similarly, a time-stamp can be shared by two events as shown in Figure 35

11:56
Foul by Liam Ridgewell (Birmingham) on Andy Holdsworth (Huddersfield). Direct free kick taken right-footed by Frank Sinclair (Huddersfield) from own half, resulting in open play.

Figure 35: Sample of commentary text 2-Events time-stamp

Or even by three events as shown in Figure 36

33:30
Direct free kick taken left-footed by Robbie Williams (Huddersfield) from right channel, header by Chris Brandon (Huddersfield) from left side of six-yard box (6 yards), save (parried) by Maik Taylor (Birmingham). Foul by Chris Brandon (Huddersfield) on Stephen Kelly (Birmingham). Direct free kick taken right-footed by Maik Taylor (Birmingham) from own half, resulting in open play.

Figure 36: Sample of commentary text 3-Events time-stamp

In the case of multiple events sharing the same time-stamp, it is clear that it is impossible for these events to occur at the same time. For example, Figure 35 shows *Foul* and *Direct free kick* events sharing the same time-stamp. Also, in Figure 36, *Direct free kick*, *foul*, and another *Direct free kick* event are sharing the same time-stamp. Analysis is needed to be done to solve this problem and help in assigning new modified time-stamps for the other events.

It is also noted that most of the 1-event time-stamps are not perfectly synchronized with the event in the video file within an acceptable range (five seconds before or after). This is often due to the live commentator speaking either before the event or just after, rather than coinciding exactly.

Any event has starting point and an ending point. Some events have a shorter interval than others. For example, *foul* and *shot* events may have a maximum of one second interval. Other events such as *throw-in* and *corner* might have a longer interval which could be up to two minutes in some cases. For the purpose of this thesis, the ending point will be the focus point. To illustrate the issue of the commentary text time-stamp not being reported correctly with conjunction of the event's interval, let's examine the following example closely.

7:16
Goal kick taken long by Roy Carroll (West Ham).

Figure 37: Goal kick event from Manchester City and West Ham match, 2006

Figure 37 shows a commentary text *goal kick* event that happened 7 minutes and 16 seconds from the start of the match. To see what actually happened, Figure 38 below shows the starting point of the *goal kick* event. Note that the starting point is reported at 7:17

Figure 38: Goal kick event starting point in Manchester City and West Ham, 2006

However, Figure 39 below shows the ending point of this *goal kick* event.

Figure 39: Goal kick event ending point in Manchester City and West Ham, 2006

Notice that ending point is actually 7:42. This indicates that the interval length for this event is 25 seconds, which is the length of time between the commentator announcing that the goal kick has been awarded and it being taken. For the system to correctly annotate this event's video, the time-stamp needs to be adjusted to (7:16 + 0:25) 7:41.

One way to investigate this issue is by watching the football match video and comparing the event times with the live commentary time-stamp; this is the only time we had to analyse the video itself. Ten random football video matches and their live commentary texts were analysed. Table 17 and Table 18 show the relative information obtained as a result of this analysis (negative number means the event happened before the commentary text time-stamp)

Table 17: Events analysis (Live commentary text versus actual video) Part-1

		Defending throw-in	Attacking throw-in	Direct free kick	Indirect free kick
Games	Smallest Offset (sec)	-10	-10	0	0
10	Largest Offset (sec)	58	68	160	46
	Total Events	231	272	225	62
	Avg. Event Per Game	23.1	27.2	22.5	6.2
	Avg. Time (sec)	16.95	16.61	24.45	22.66
	Mean (sec)	14	13	24	22.5
	StDev (sec)	14.07	14.14	19.15	12.46
	AveDev (sec)	11.33	11.50	12.75	10.53

Table 18: Events analysis (Live commentary text versus actual video) Part-2

		Goal Kick	Foul	corner	shot	cross	goal	offside	free kick
Games	Smallest Offset (sec)	-6	-50	-6	-10	-6	0	-20	9
10	Largest Offset (sec)	130	70	60	35	5	29	4	60
	Total Events	187	278	146	153	202	24	62	60
	Avg. Time (sec)	23.31	0.23	24.23	0.33	0.21	4.42	-1.45	4.7
	Avg. Event Per Game	18.7	27.8	13.3	15.3	20.2	2.4	6.2	31.06
	Mean (sec)	22	0	25	0	0	0	0	30
	StDev (sec)	16.68	7.59	15.27	4.71	1.20	9.51	3.98	11.25
	AveDev (sec)	11.77	2.24	12.63	1.41	0.67	6.86	2.47	8.46

Defending throw-in occurred 231 times in the 10 matches. As explained above, our goal is to find the ending point. For *defending throw-in,* the smallest offset, with respect to its time stamp from the live commentary text, is -10 seconds and the largest offset is 58 seconds. It has a mean value of 14 seconds, StDev of 14.07 seconds and AveDev (Average Deviation) of 11.33 seconds. For the time being, we will use the mean value to adjust the live commentary time stamp. Figure 40 below illustrates the 231 *defending throw-in* events with their actual time while Y=0 indicates the live commentary text stamp-line. The events are plotted as they were detected.

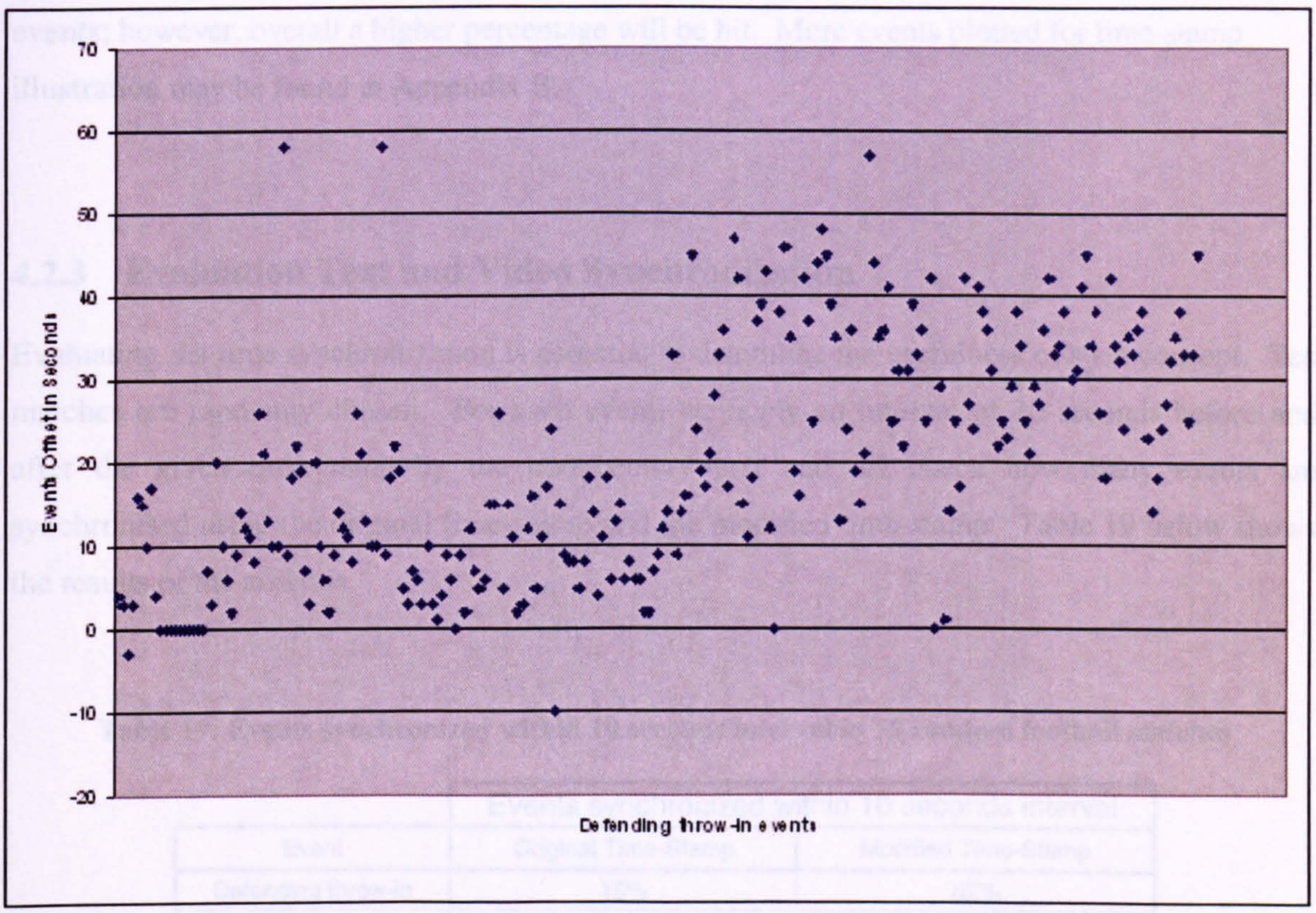


Figure 40: Defending throw-in events with their offset time (sec) with respect to the Live Commentary time stamp (mean = 14 seconds)

As shown in Figure 40, only a few *Defending throw-in* events video times matched the commentary time-stamp. The majority are between 5 seconds after to 50 seconds.

To extend this explanation further, Figure 41 shows how the live commentary time stamps an event which in reality is actually offset by seconds or minutes.

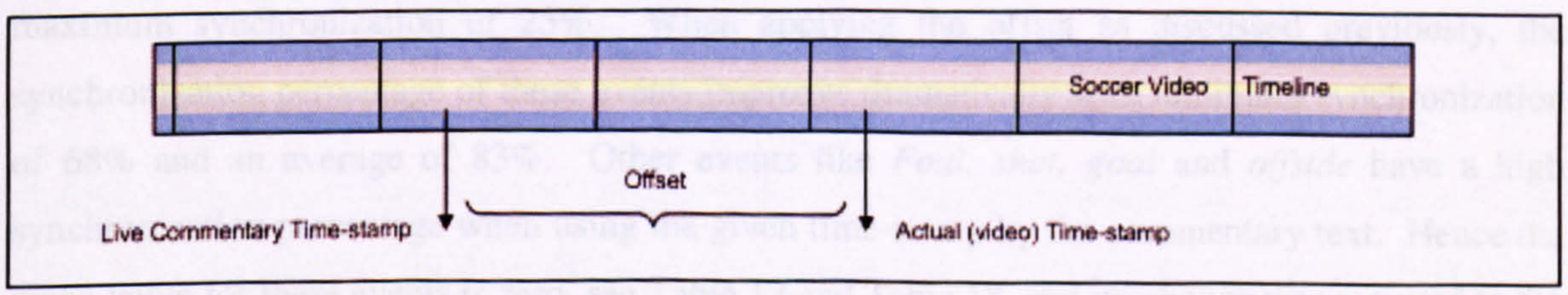


Figure 41: Illustrating the offset through the timeline

Based on the results given in Table 17 and Table 18, the system will use the mean value to adjust the live commentary time-stamp. It is expected that some new time-stamps may miss their

events; however, overall a higher percentage will be hit. More events plotted for time-stamp illustration may be found in Appendix B.

4.2.3 Evaluation Text and Video Synchronization

Evaluating the time synchronization is essential to determine the usefulness of such concept. Ten matches are randomly chosen. For each event, we apply an interval of 10 seconds before and after the given time-stamp by the commentary text and we check how many events are synchronised using the original time-stamp and the modified time-stamp. Table 19 below shows the results of the analysis.

Table 19: Events synchronized within 10 seconds interval in 10 random football matches

	Events synchronized within 10 seconds interval	
Event	Original Time-Stamp	Modified Time-Stamp
Defending throw-in	15%	87%
Attacking throw-in	25%	92%
Direct free kick	18%	88%
Indirect free kick	9%	68%
Goal Kick	27%	93%
Foul	91%	91%
corner	20%	83%
shot	96%	96%
cross	98%	98%
goal	83%	83%
offside	91%	91%
free kick	3%	84%

Events such *Defending throw-in, Attacking throw-in, Direct free kick* and *Indirect free kick* have low synchronization percentages when using the time-stamp provided by the commentary text; maximum synchronization of 25%. When applying the offset as discussed previously, the synchronization percentage of these events improves dramatically with minimum synchronization of 68% and an average of 83%. Other events like *Foul, shot, goal* and *offside* have a high synchronization percentage when using the given time-stamp by the commentary text. Hence the mean value for these events is zero, see Table 17 and Table 18, and no change was required to the time-stamp. This shows that the concept of time-stamp offset only applies to those events with low synchronization percentages and does not affect those events with high synchronization percentages. This will improve the overall events synchronization which is the main goal of this analysis. Figure 42 below illustrates the results obtained from Table 19.

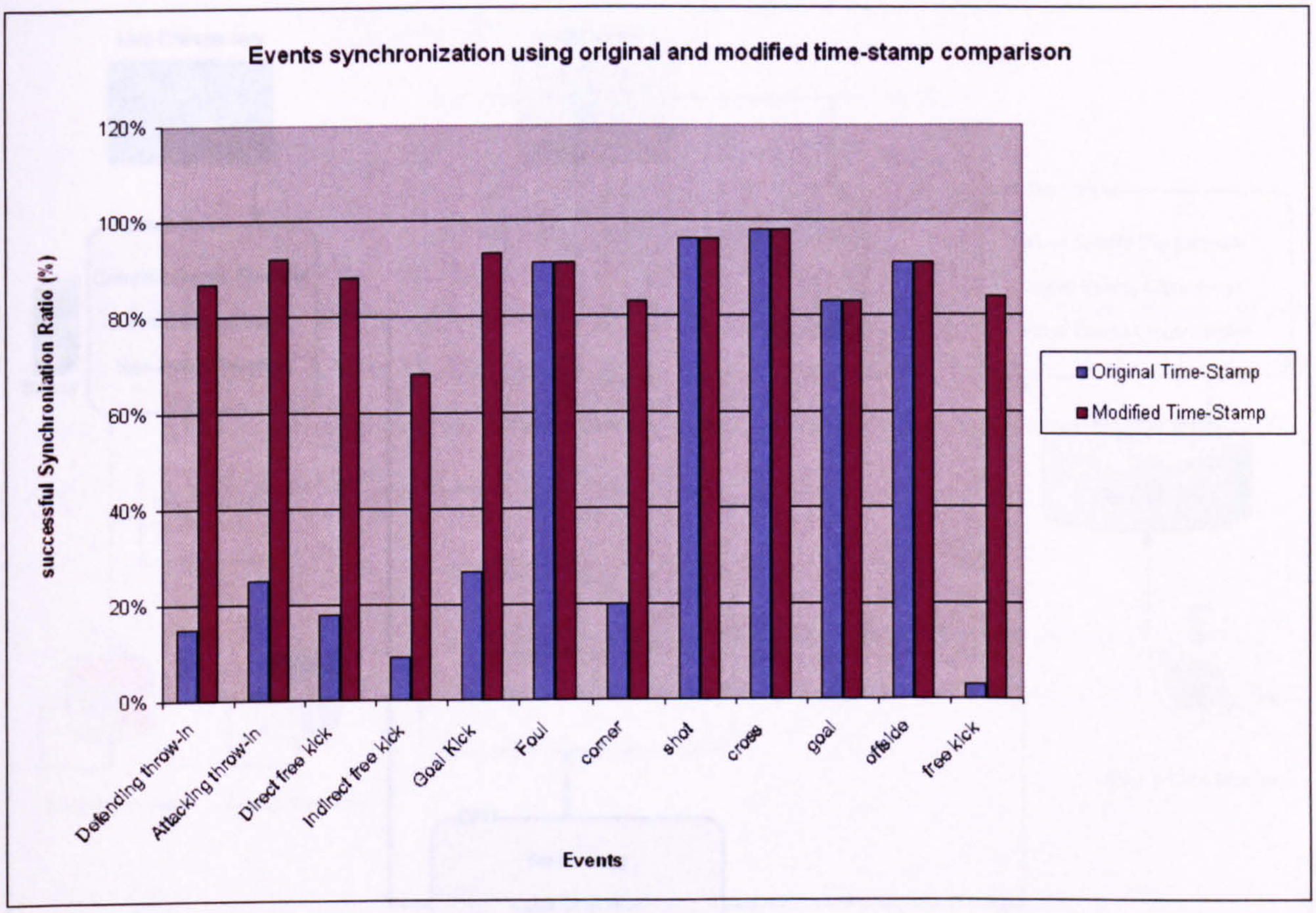


Figure 42: Events synchronization using original and modified time-stamp comparison

4.3 Automated Video Indexing and Annotation

The system is now ready to extract the events segments and automatically annotate them. The synchronization information along with the football video and the commentary text are passed to the video cutting and annotating section in the system.

4.3.1 Terminology, Local Grammar and Globalization

Terminology, Local Grammar and Globalization filtrations are immediately applied when a new event is to be processed by the system. These filters are what differentiate this system from others that might appear to be very similar but which utilise methods based only on patterns analysis. Figure 43 below shows the Terminology, Local Grammar and Globalization part of the system (being focused)

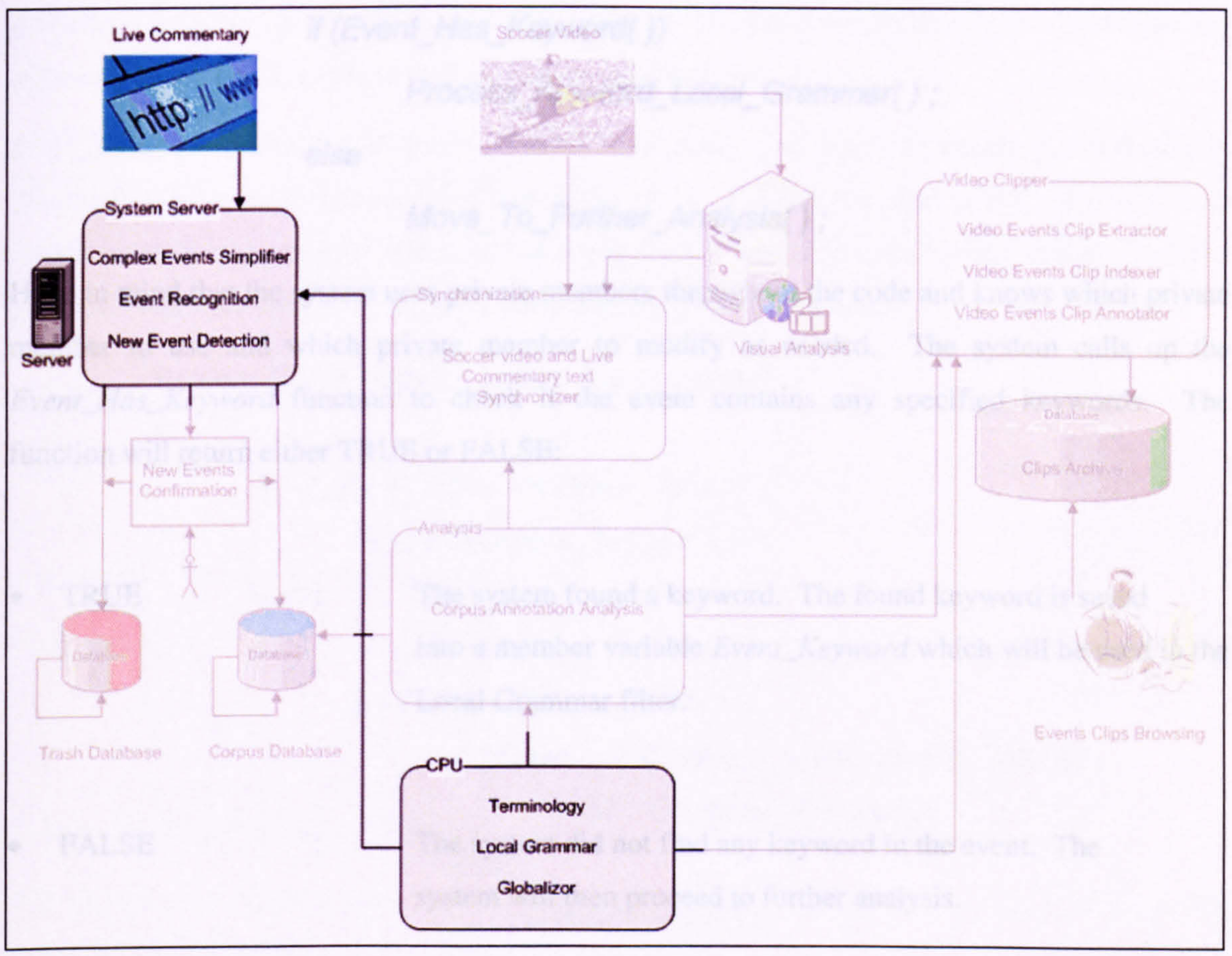


Figure 43: Terminology, Local Grammar and Globalization filtrations

Recall the proposed algorithm of automation of events detection in football commentary text, see Figure 29, chapter 3. The following is part of the system code that is written to apply the proposed algorithm.

The following code is run once when the system starts:

```
Load_Keywords( ) ;
```

The system then loads the automatically analyzed keywords, as discussed in chapter 3 (Table 22 specifically), into a dynamic array with respect to their weirdness level.

When the system receives an event to process (after event simplification if needed), the Terminology filter is applied:

```
if (Event_Has_Keyword( ))
        Process_Keyword_Local_Grammar( ) ;
else
        Move_To_Further_Analysis( ) ;
```

Have in mind that the system uses private members throughout the code and knows which private member to use and which private member to modify as needed. The system calls up the *Event_Has_Keyword* function to check if the event contains any specified keywords. The function will return either TRUE or FALSE:

- TRUE : The system found a keyword. The found keyword is saved into a member variable *Event_Keyword* which will be used in the Local Grammar filter.

- FALSE : The system did not find any keyword in the event. The system will then proceed to further analysis.

The Terminology filter helps in speeding up processing the events without wasting time applying further analysis to events that can be found rejected at a much earlier stage. Without the use of Terminology, similar systems might take more time to achieve the same results.

Once the event is filtered as accepted by the Terminology filter, the Local Grammar filter is applied. The system will use the keyword that is stored in the *Event_Keyword* to determine the patterns and the local grammar that are to be applied.

```
Load_Event_Keyword_Globalized_Pattems( ) ;
If (Map_Event_To_Globalized_Pattems( ))
        Add_Event_To_Accepted_DataBase( ) ;
else
        Add_Event_To_Rejected_DataBase( ) ;
```

The system first looks up the private member *Event_Keyword* and loads its globalized patterns. This will save more time than applying all available globalized patterns. The system then calls up the *Map_Event_To_Globalized_Patterns* which, as its name suggests, maps the *Event* to the

globalized patterns. If there is a match, the system moves the Event to the accepted database and if there is no match, *Event* is moved to the rejected database.

Map_Event_To_Globalized_Patterns function runs some local grammar checking, for example, that the first word in the event cannot be a proper noun. The function then calls *Compare_Tokens* function which is a recursive function that compares the tokens in both *Event* and the globalized patterns. The following code is part of the *Compare_Tokens* function:

```
if ( Get_Token(Pattern_Token, Event_Token)
{
        If ( Pattern_Token == Event_Token)
                return Compare_Tokens( ) ;
        else
        {
                if ( Pattern_Token == "PN")
                        if ( Check_PN( Pattern_Token, Event_Token))
                                return Compare_Tokens( ) ;
                        else
                                return FALSE ;

                else  if ( Pattern_Token == "NM")
                        if (Check_NM(Pattern_Token, Event_Token))
                                return Compare_Tokens( ) ;
                        else
                                return FALSE ;
                else
                        return FASLE ;
        }
}
else
        return TRUE ;
```

The code checks if there are any more tokens to be checked. If there are none, it means all tokens' comparisons have passed successfully and *Event* has matched one of the patterns. If there is token to be processed, then the system checks if the tokens are the same and if not, the system checks if they are either proper nouns or numbers with the use of *CLAWS_POS(token)* function

which denotes the part of speech category for the token. As long as the tokens' POS categories are matching, the function will compare the next token (if any). If no more tokens are to be compared, it means that all tokens' POS comparison are a match and the function will return TRUE indicating that the event pattern has matched one of the patterns in the system database. If a token POS comparison does not match at any given position in the event pattern, the function will return FALSE indicating the event pattern did not match the chosen pattern in the system database. The system will then move on to the next "keyword" pattern that is stored in the database to compare it to the detect event pattern from the incoming commentary text.

The point to bear in mind, when programming in Visual Studio.Net, is that the code will be spread all over. Some functions are idle and waiting to be called while other functions are a thread waiting for specific events to happen.

Next we will look at the system GUIs when applying these codes

4.3.2 Processing Key Patterns

The system starts by processing the incoming text and analysing its patterns. Figure 44 shows the main GUI for the system.

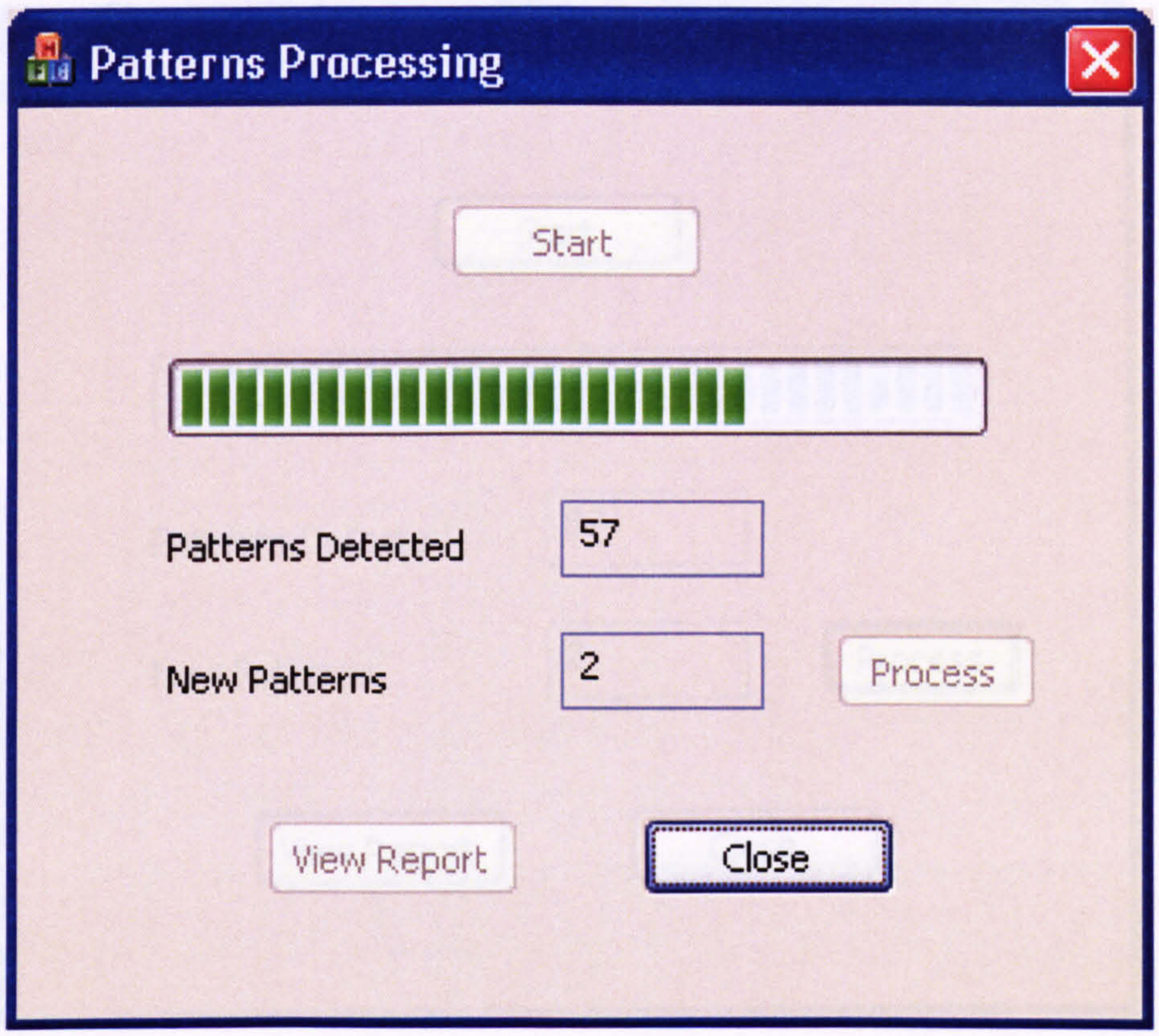


Figure 44: System training GUI

The GUI does not show much information as the system does everything in the background and is therefore suitable for a user who is training the system. The GUI is written in C++ using Microsoft Visual Studio Dot Net 2005. When the Start button is clicked, a thread will be run to detect new inputs. Once a video and its text are detected, the system starts the processing. When detecting new patterns, the user gets to check them and train the system. Instantly, the video indexing and annotating sections process the video file and make clips of the patterns. Once the whole process is complete (see Figure 45) the user can process the new patterns as they may exist (see Figure 46) or view the full report, (see Figure 47).

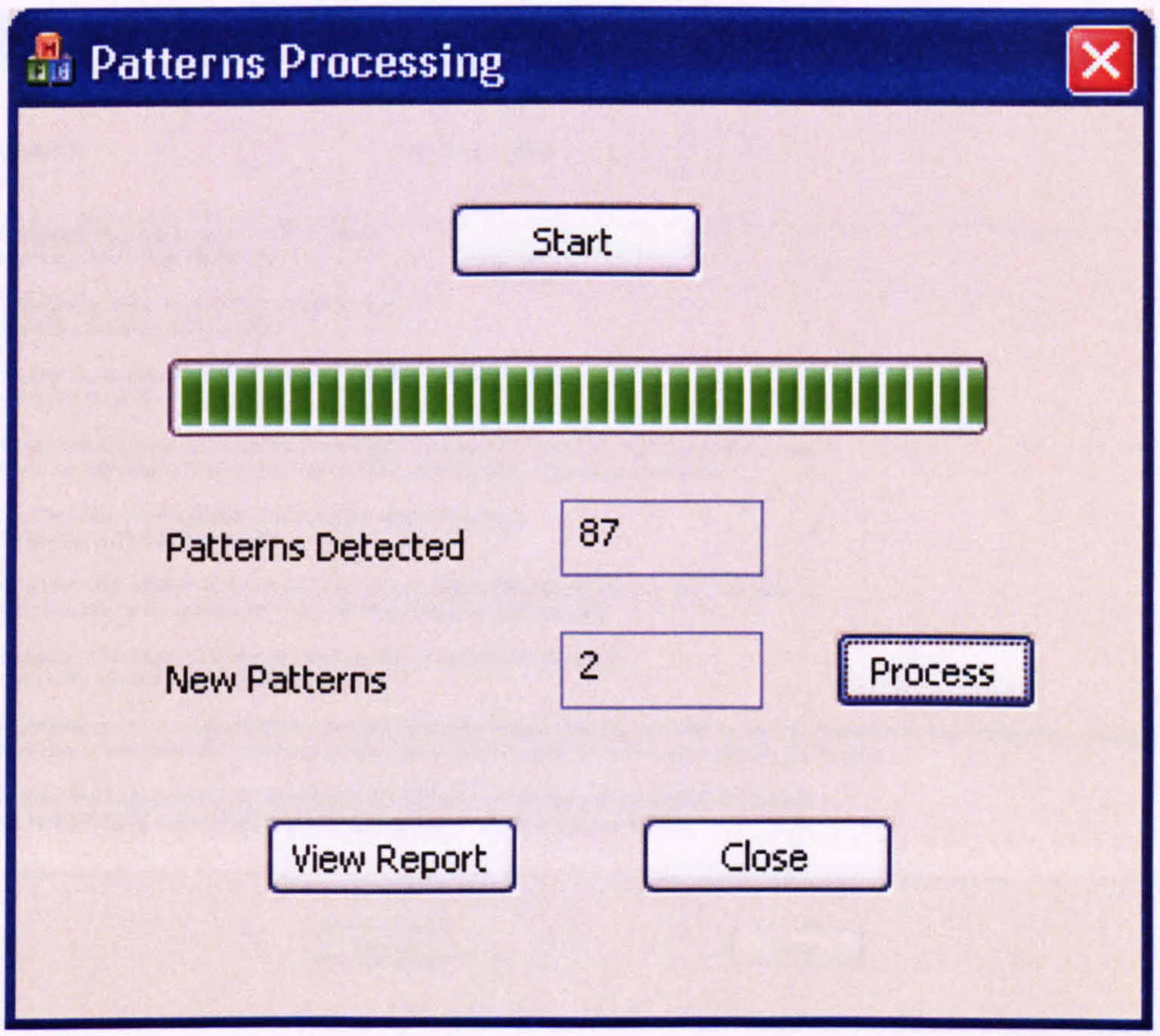


Figure 45: Patterns processing GUI

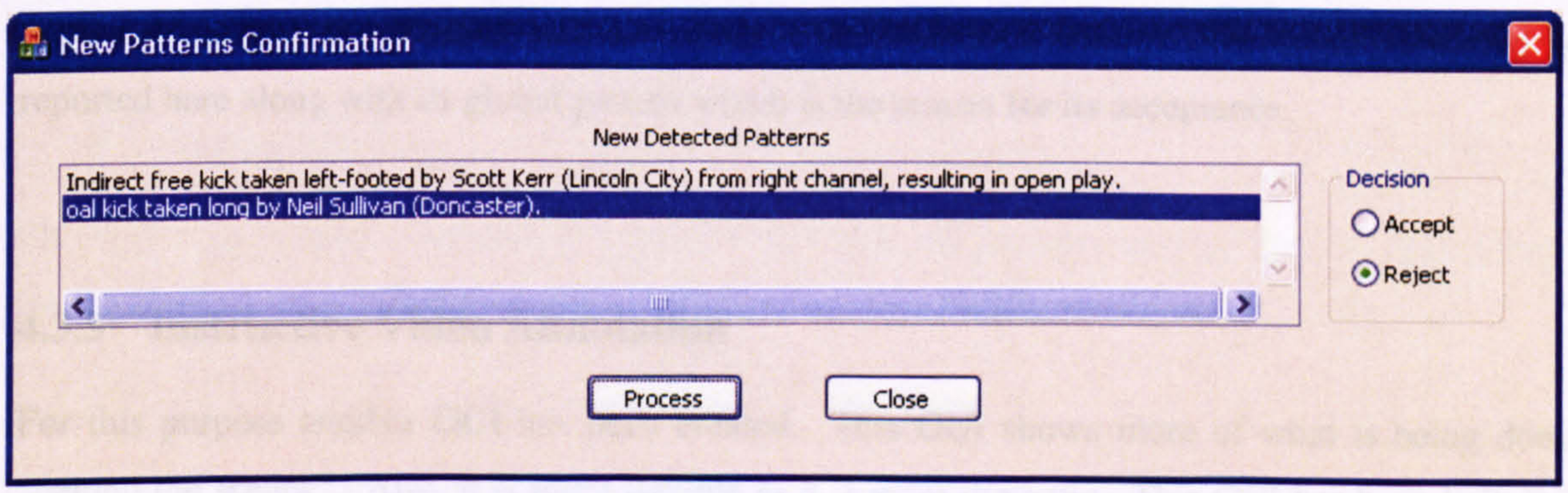


Figure 46: Processing new patterns GUI

When processing the new patterns, each unknown pattern will have the option of being accepted or rejected. In Figure 46, a new pattern of *free kick* is detected and awaits confirmation. This exists due to the system spotting *Indirect* keyword for the first time. The other new pattern is a mis-type from the source that was meant to say *Goal kick* instead of *oal kick.*

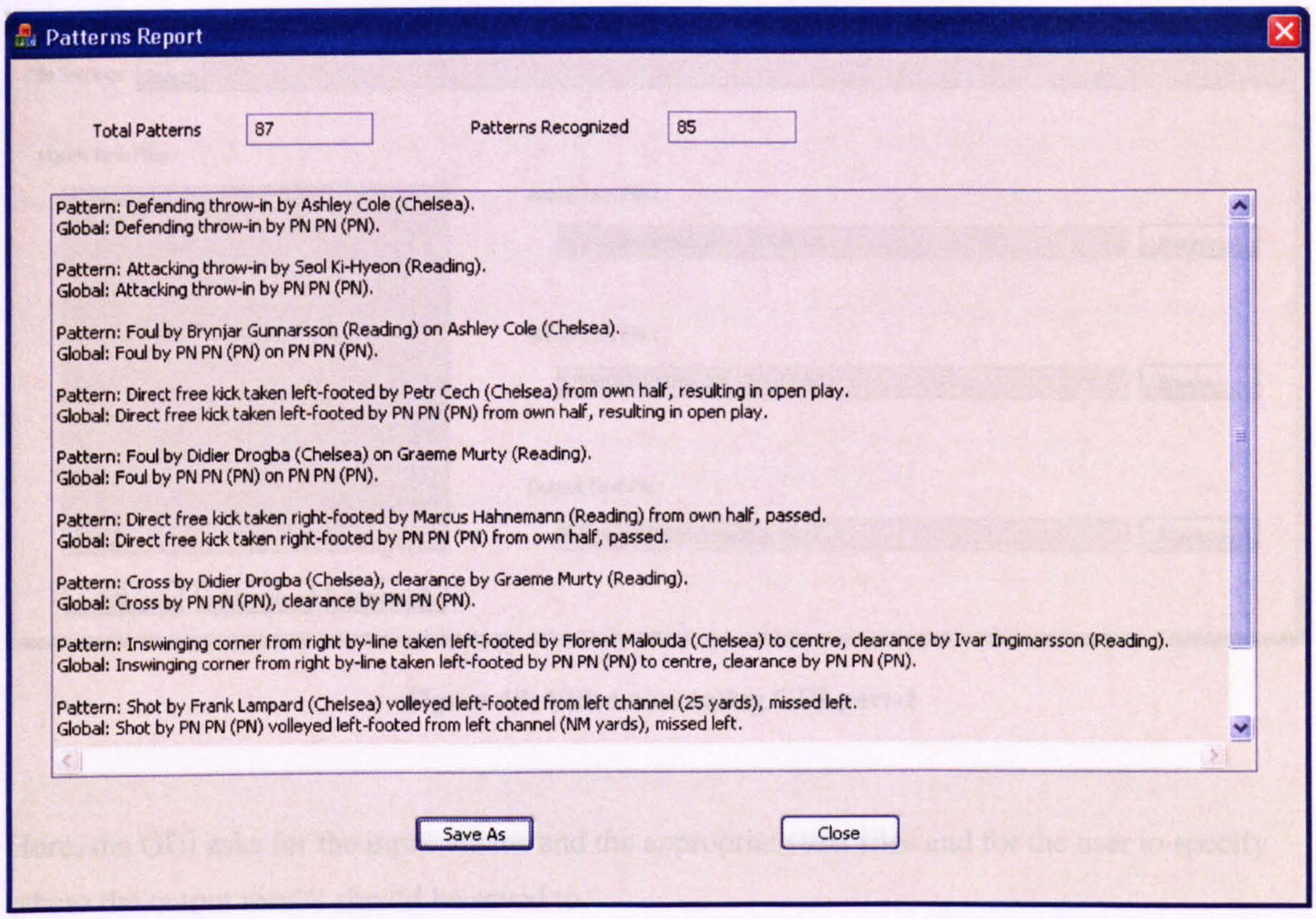


Figure 47: Patterns process report GUI

Figure 47 shows the patterns process report. Each detected pattern that is auto-accepted is reported here along with its global pattern which is the reason for its acceptance.

4.3.3 Instructive Video Annotation

For this purpose another GUI has been created. This GUI shows more of what is being done "behind the scenes". Also, it is more flexible as it is more dynamic. This GUI has been built for demonstration purposes and is comprised of two separate GUIs. The first GUI is for video indexing and annotation, and the second GUI is for video cutting, searching and playing. Both GUIs are written in Microsoft Visual Studio Dot Net 2005 using the C# language. The reason for choosing C# language instead of C++ is to make these GUIs server-side ready if they are to be used on the internet. The first of these GUIs is made up of two sections. Figure 48 shows the first section of this GUI, the File Settings.

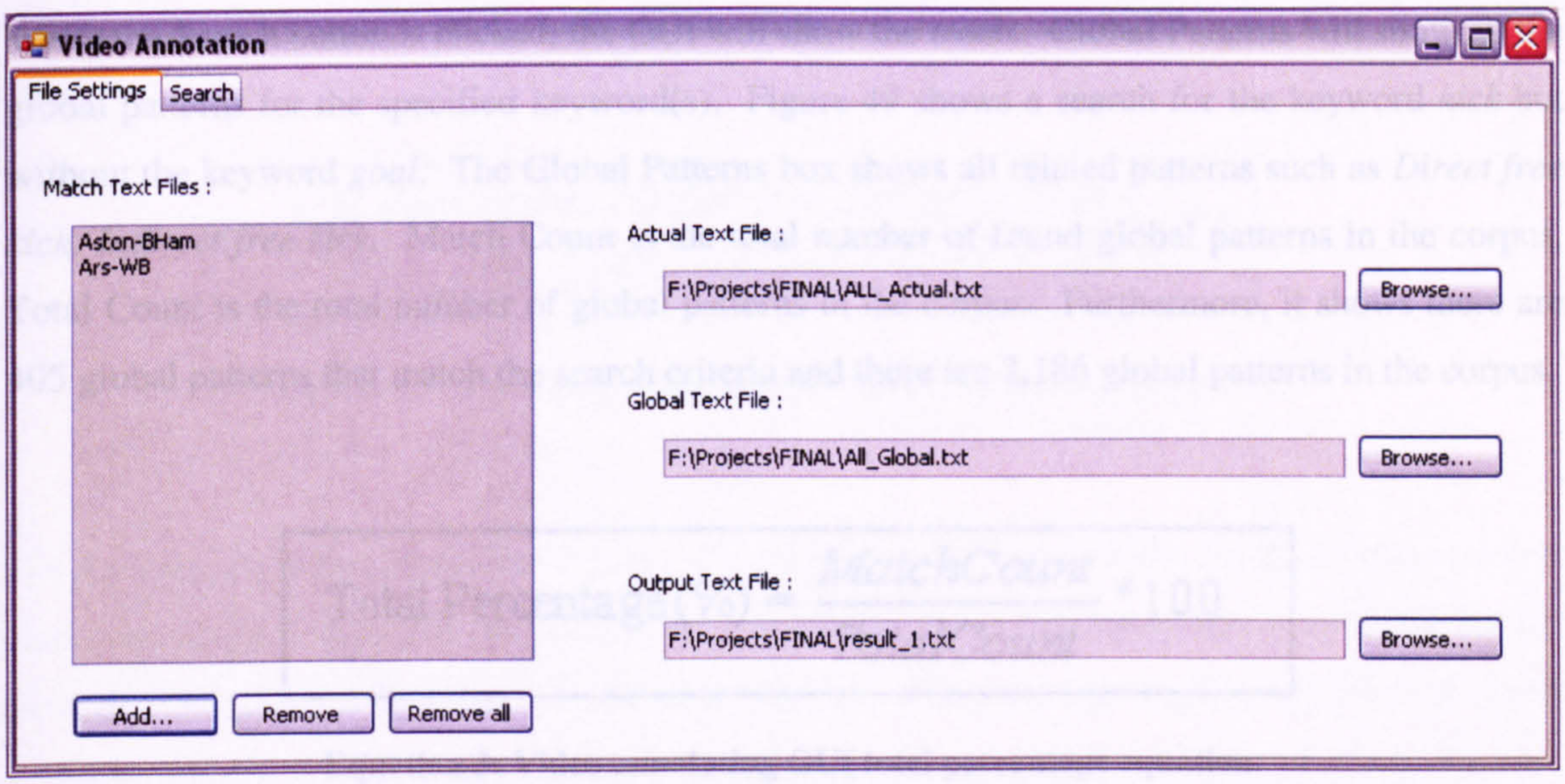


Figure 48: Video annotating GUI part-1

Here, the GUI asks for the input videos and the appropriate text files and for the user to specify where the output results should be saved to.

- **Match Text Files:** The live commentary text file(s).
- **Actual Text File:** The live commentary text files combined into one file.
- **Global Text File:** The actual text file globalized.
- **Output Text File:** The file to save the search result(s) to.

The system only needs the match text files. Combining and globalization are done by the system itself. However, this is shown in this GUI so the user becomes aware of what is being used.

The second part of the GUI deals with searching (see Figure 49 below). The search function gives the user the choice of simple search and advanced search. Simple search allows the search for a single keyword or phrase. Advanced search allows for single keyword search, multiple keywords (contiguous or non-contiguous) as well as allowing exclusion of specific keywords. When searching for non-contiguous keywords a filter becomes available to specify a search for all/any keywords.

The video clip options provide three static dynamic parameters: the Offset, which has been explained previously; a Roll-back option which enables the clip to start X seconds before its offset point; and Length which tells the system how long this clip is to run/cut for in seconds.

When the Search button is clicked, the GUI will show the result. Global Patterns will show all the global patterns for the specified keyword(s). Figure 49 shows a search for the keyword *kick* but without the keyword *goal*. The Global Patterns box shows all related patterns such as *Direct free kick, Indirect free kick.* Match Count is the total number of found global patterns in the corpus. Total Count is the total number of global patterns in the corpus. Furthermore, it shows there are 405 global patterns that match the search criteria and there are 2,186 global patterns in the corpus.

$$\text{Total Percentage}(\%) = \frac{MatchCount}{TotalCount} * 100$$

Equation 3: Video annotating GUI total percentage equation

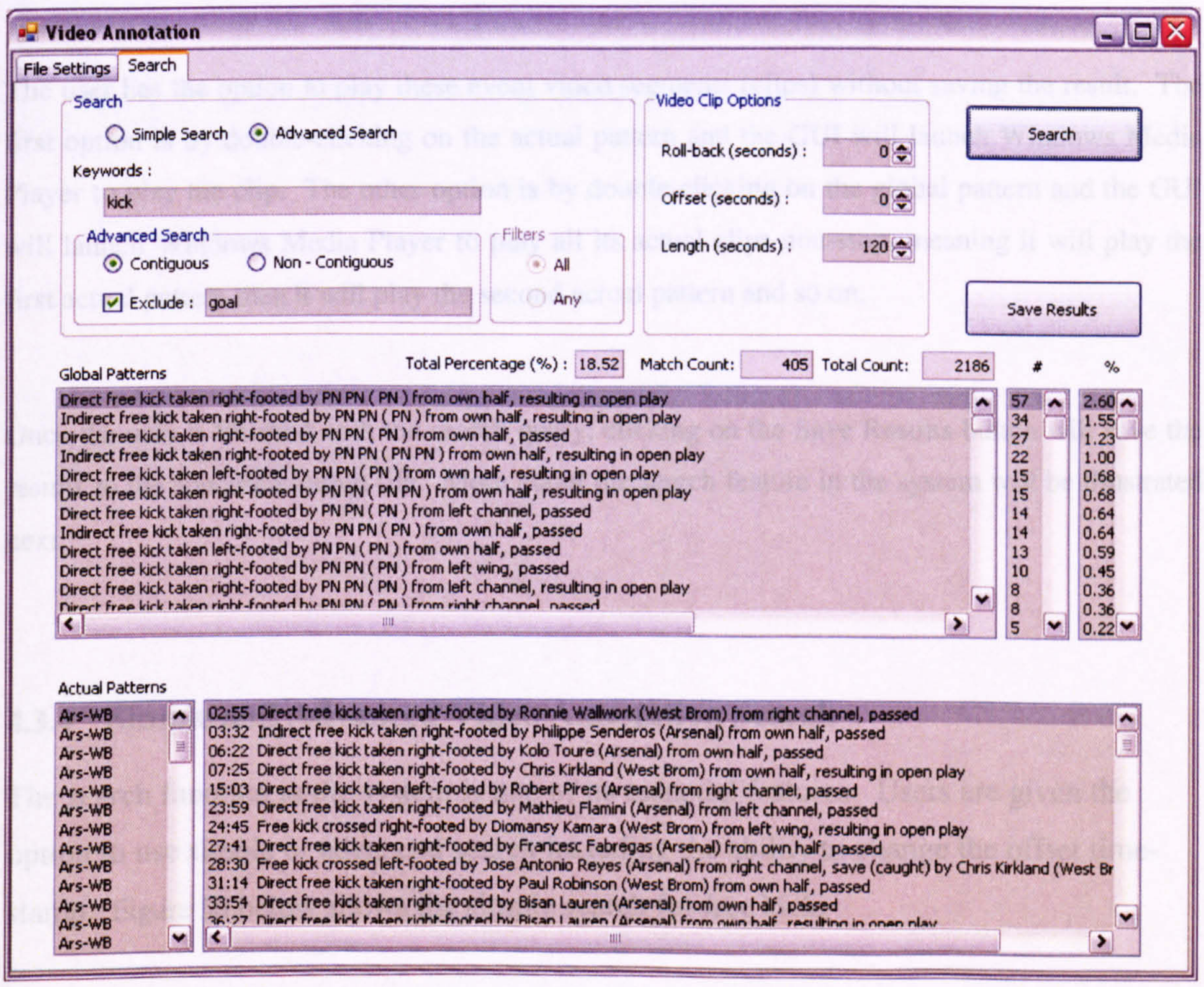


Figure 49: Video annotating GUI part-2

The # column shows the total count of this pattern in the input text file(s). The % column calculates:

$$\frac{\sum TheGlobalPatternInTheInputFile}{\sum TotalGlobalPatternsInTheCorpus} * 100$$

Equation 4: Percentage column equation in the system GUI part-2

The Actual Pattern box shows the actual (original) patterns that match the search criteria and that have been found in the input text file(s). Note that it shows the two teams who played that match, the time-stamp for the event and the full event pattern. When the user highlights (left click once) a global pattern, the GUI will highlight the first matching actual pattern, and when the user highlights an actual pattern, the GUI will highlight its global pattern.

The user has the option to play these event video segments (clips) without saving the result. The first option is by double-clicking on the actual pattern and the GUI will launch Windows Media Player to play the clip. The other option is by double-clicking on the global pattern and the GUI will launch Windows Media Player to play all its actual clips non-stop, meaning it will play the first actual pattern then it will play the second actual pattern and so on.

Once the user is finished with the search query, clicking on the Save Results button will save the results to the specified output file. More about the search feature in the system will be illustrated next.

4.3.4 Simple and Advanced Video Annotation Search

The search function in the system is one of its essential features. Users are given the option to use simple or advanced search including the ability to change the offset time-stamp. Figure 50 below shows the simple search for free kick.

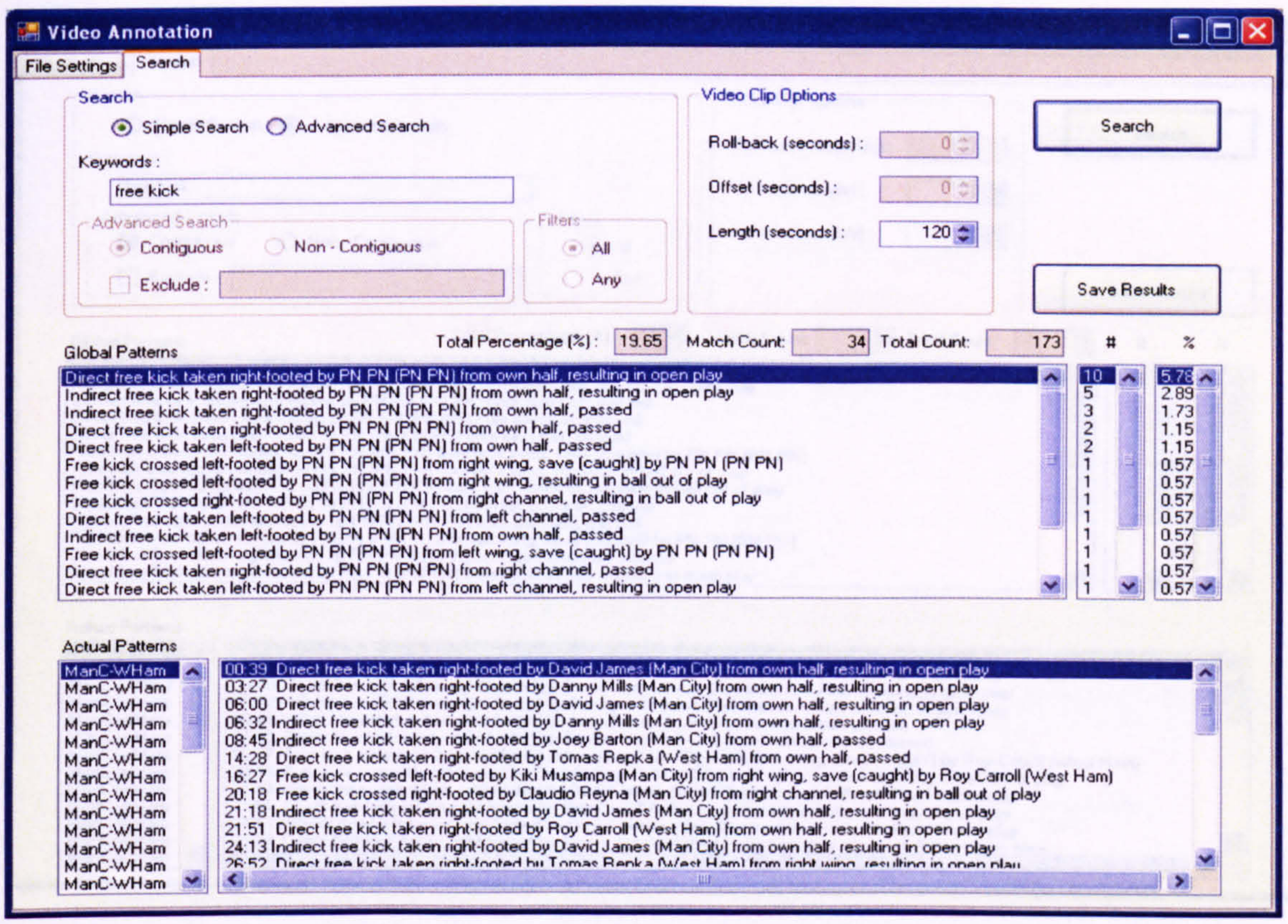


Figure 50: *free kick* simple search

Figure 51 below shows the same search for free kick using the advanced search option with Contiguous being selected.

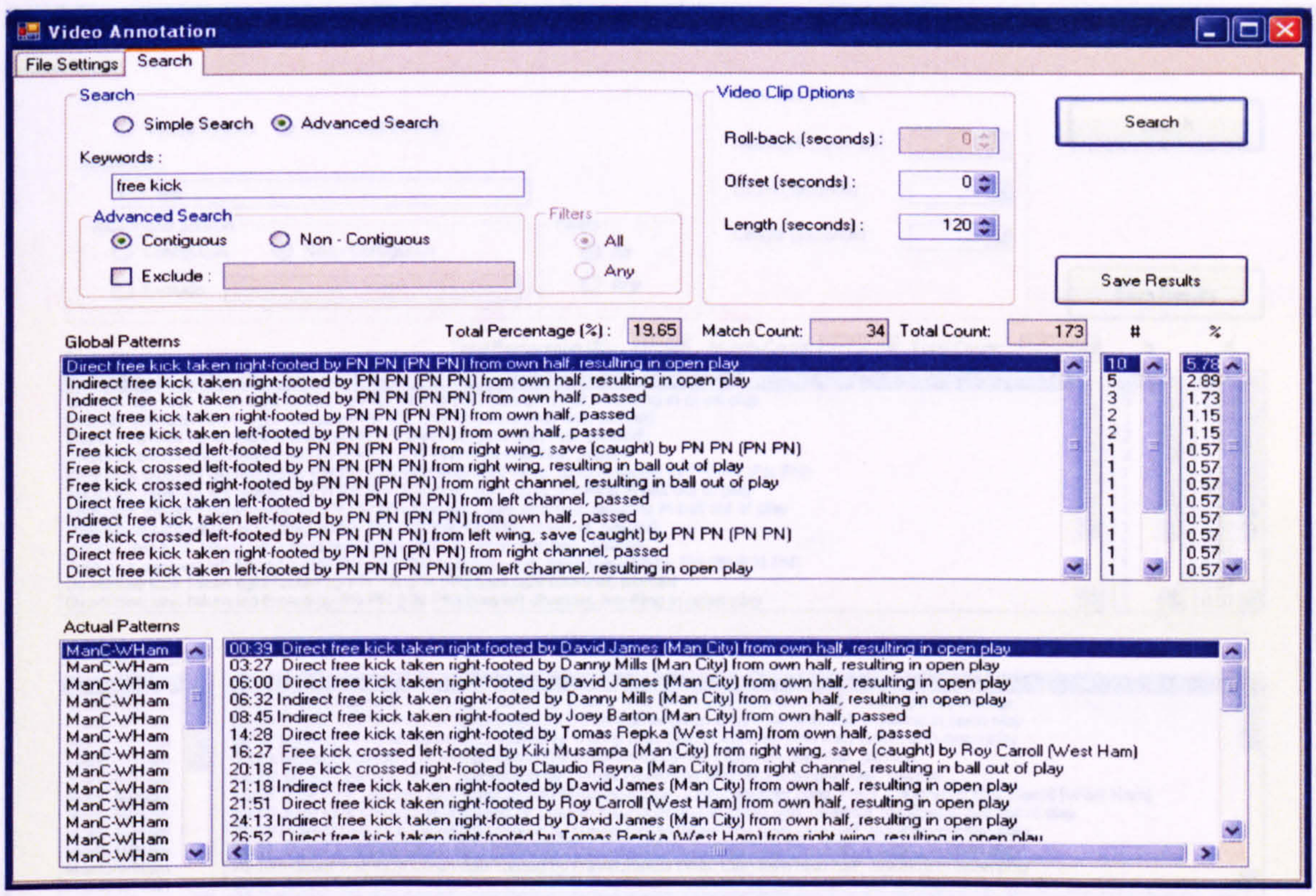


Figure 51: *free kick* advanced search using Contiguous option

Figure 52 below shows the kick free advanced search option with Non-Contiguous being selected. Pay attention to the order of the searched word as they are reversed and the system is still able to detect the appropriate events.

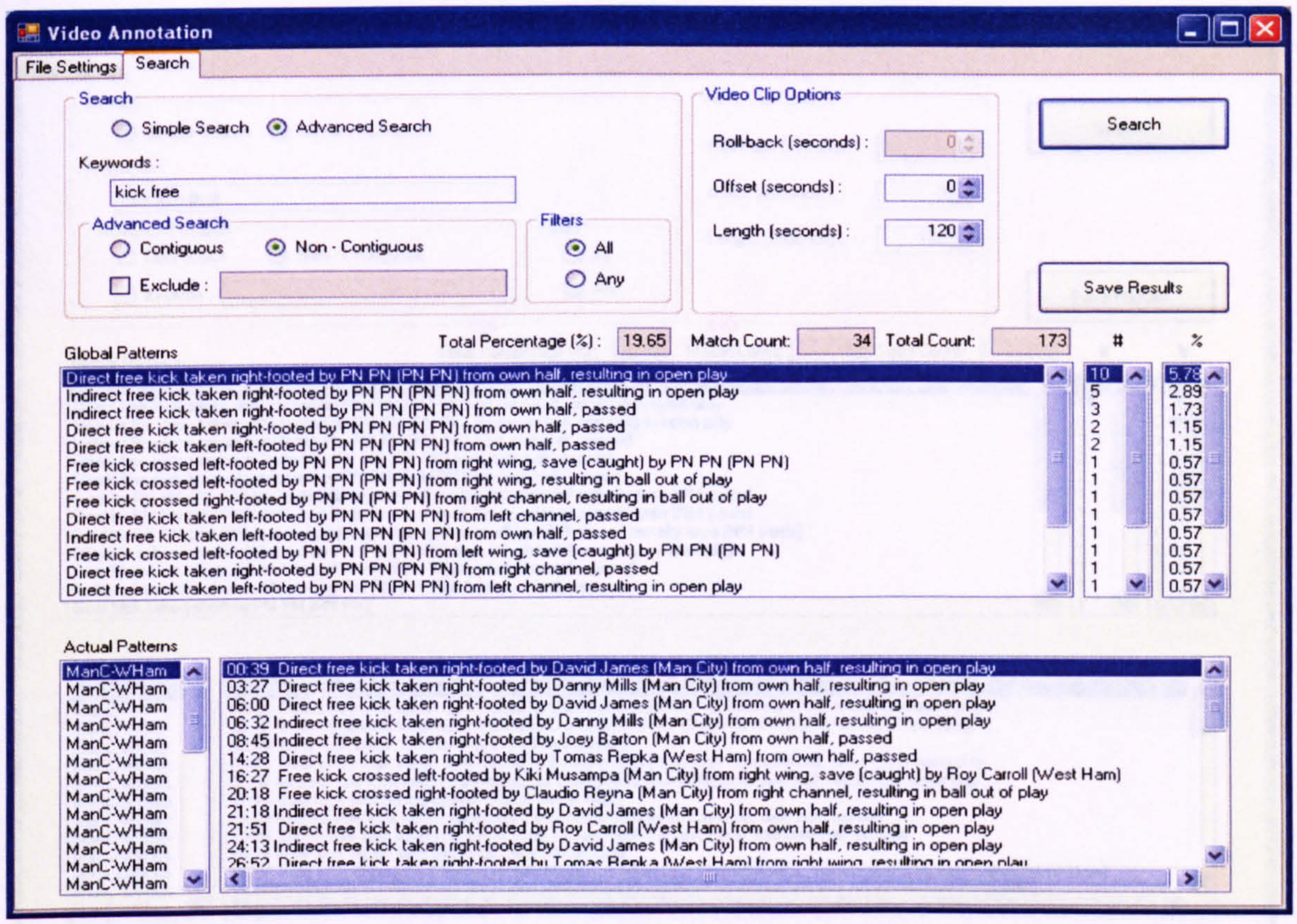


Figure 52: *kick free* advanced search using Non-Contiguous option

Figure 53 below shows kick free goal advanced search with Non-Contiguous option selected and the keywords choice changed to Any instead of All. Notice that some additional events have appeared such as *goal kick*. This shows the strength of the Non-Contiguous option especially when looking back at Figure 52 where *goal* was not among the keywords to search for.

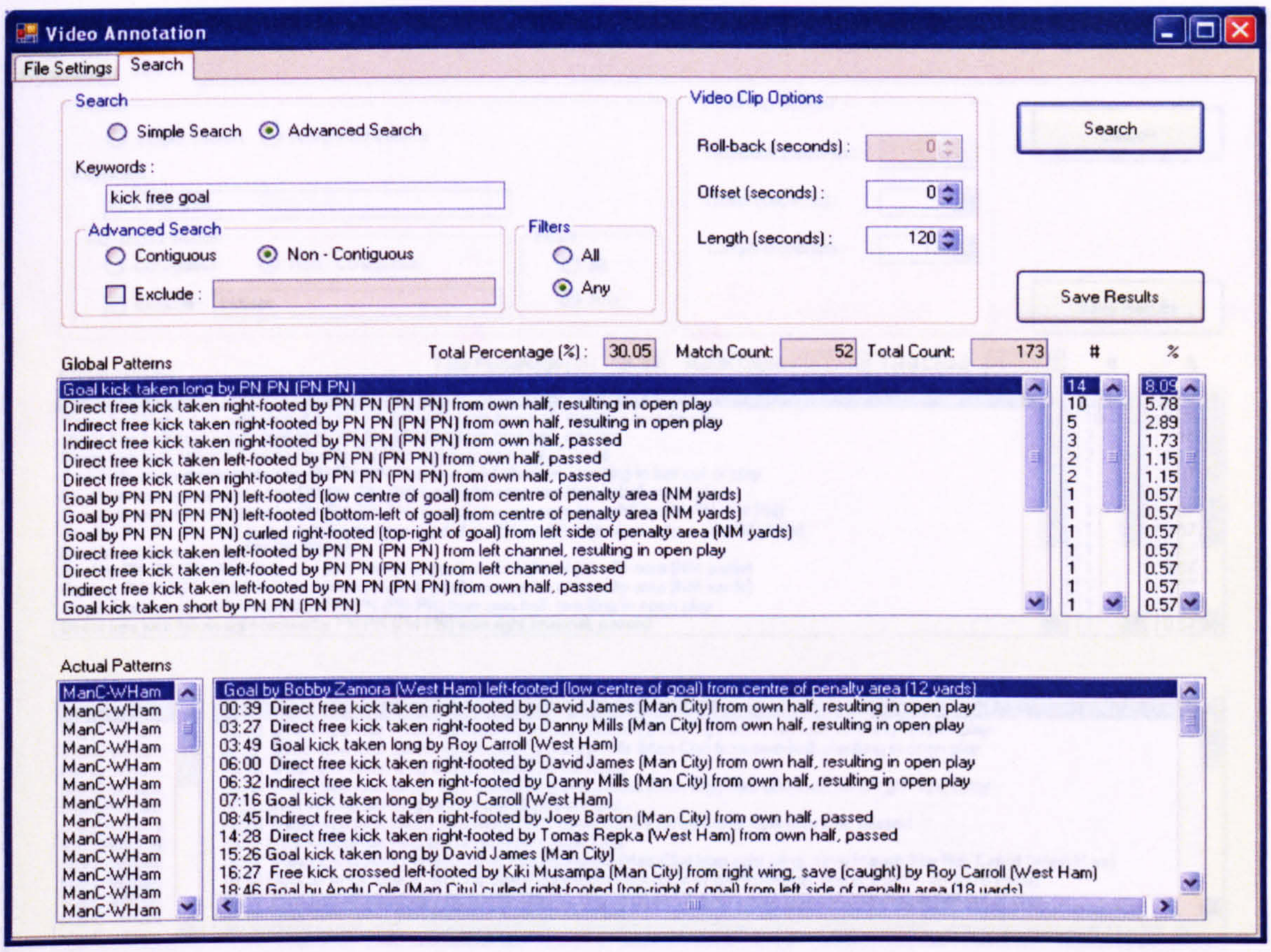


Figure 53: *kick free goal* advanced search using Non-Contiguous option

Figure 54 below shows a more complicated search. The same advanced search we have seen in Figure 53 but enabling the Exclude feature. In this advanced search we have chosen to exclude *indirect* and as a result all the indirect events have been ignored.

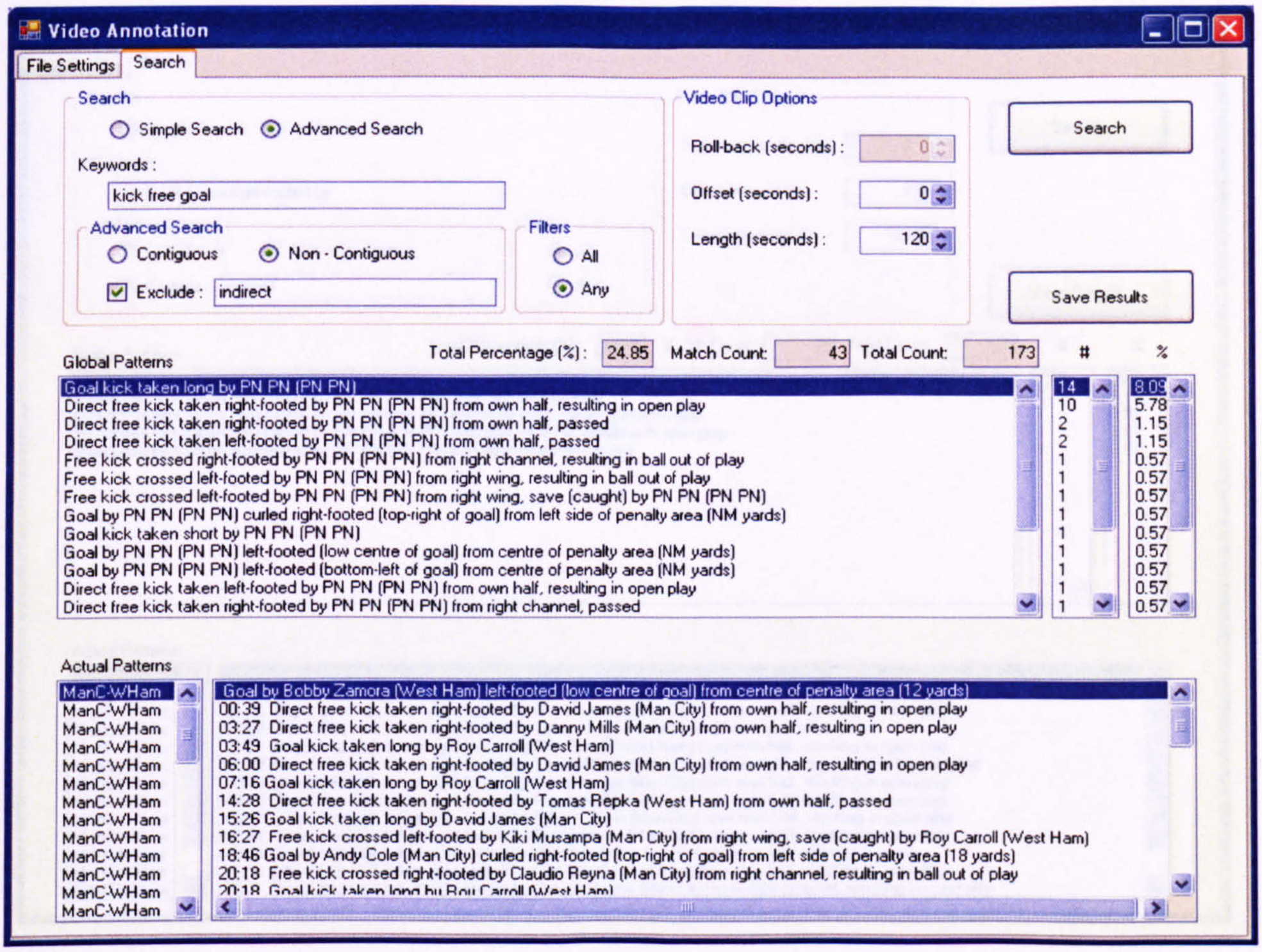


Figure 54: *kick free goal* advanced search using Non-Contiguous option with Exclude being enabled

Figure 55 below shows a more precise advanced search for *free kick taken right-footed by*. Non-Contiguous is selected and All keywords is selected. As a result, the system returned the events that only matched the given query.

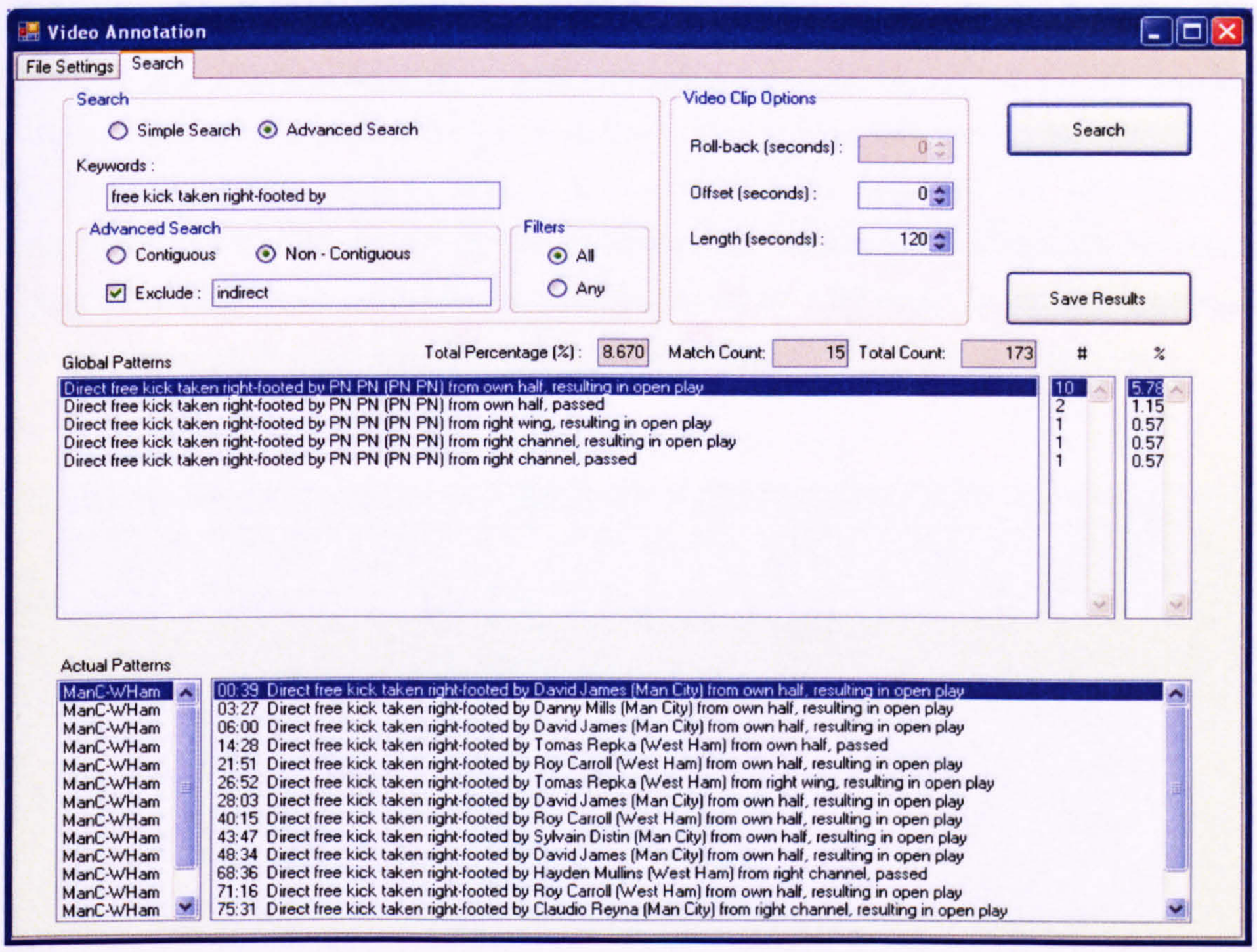


Figure 55: ***free kick taken right-footed by*** **advanced search**

Figure 56 shows that the same search and result can be obtained without having to know the exact event pattern as we have seen in Figure 55 above. The search keywords are scrambled and yet the system returned the correct results. This specific feature is essential. The user does not need to know the exact pattern to find it; by entering the keywords as they decide what they want to search for, the system will still return the correct results.

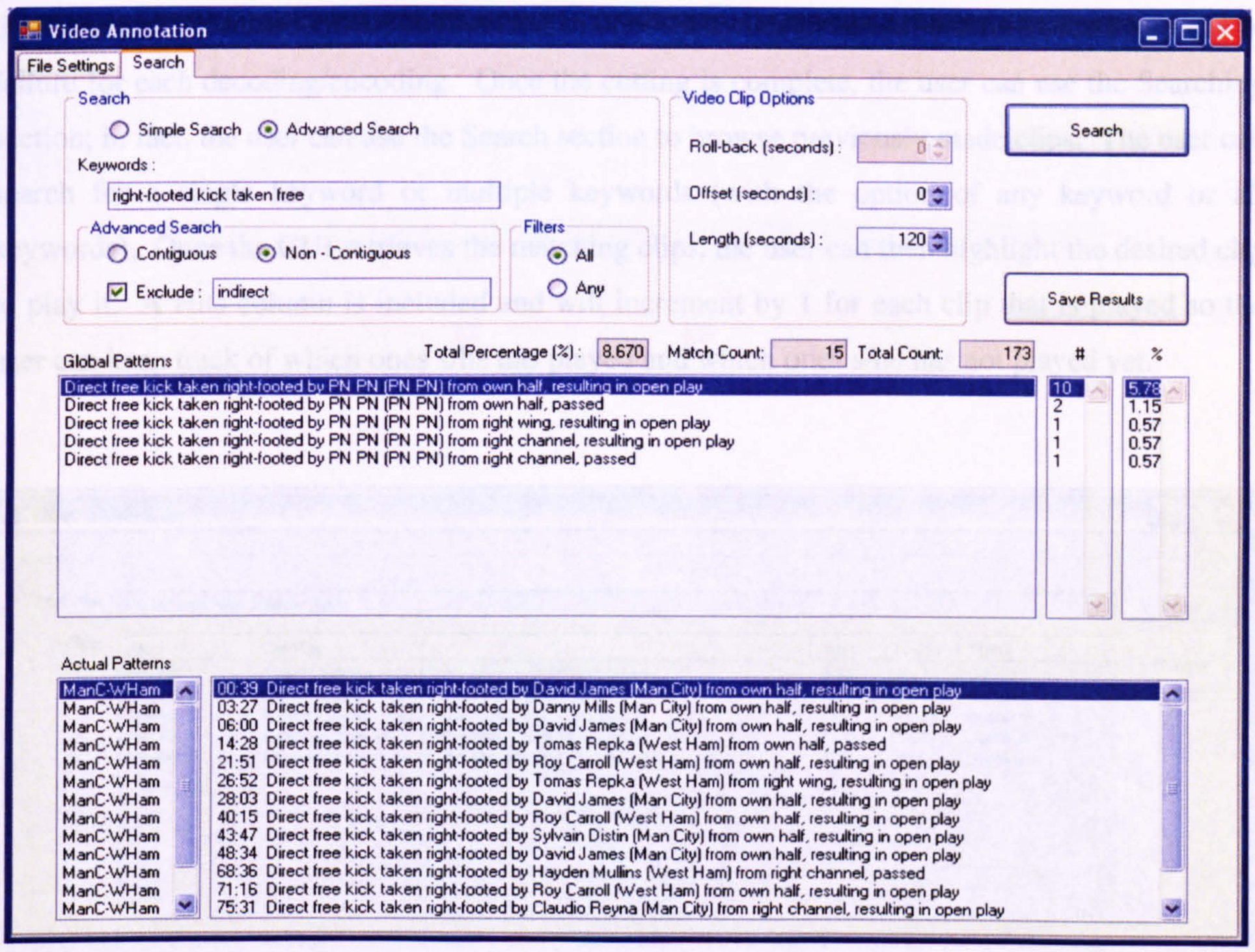


Figure 56: ***right-footed kick taken free*** **advanced search**

4.3.5 Additional Features

In addition to the basic and advanced search features in the system GUI (see above), the search section also allows the user to view/play the clips before annotating them. The user has the option to double-click on the actual event in the GUI to play that specific event. More beauty, the user has the optional to double-click on the global event and the system will play all related actual events sequentially.

4.3.6 Video Cutting, Browsing and Playing

The second GUI of the demo GUI is for video cutting, browsing and playing as may be seen in Figure 41. The first GUI provides the search result text file. When loading the saved result text file, the GUI lists all the patterns that are to be clipped along with the teams' names, the start time and end time (based on the user entered values for Roll-back, Offset and Length from part-1). Once the Process button is clicked, the GUI will start the decoding and encoding to make clips of

the patterns. A status column is included to show the GUI progress and to indicate success or failure for each decoding/encoding. Once the cutting is complete, the user can use the Searching section; in fact, the user can use the Search section to browse previously made clips. The user can search for a single keyword or multiple keywords (with the option of any keyword or all keywords). Once the GUI retrieves the matching clips, the user can then highlight the desired clip to play it. A Hits column is included and will increment by 1 for each clip that is played so the user can keep track of which ones s/he has played and which ones s/he has not played yet.

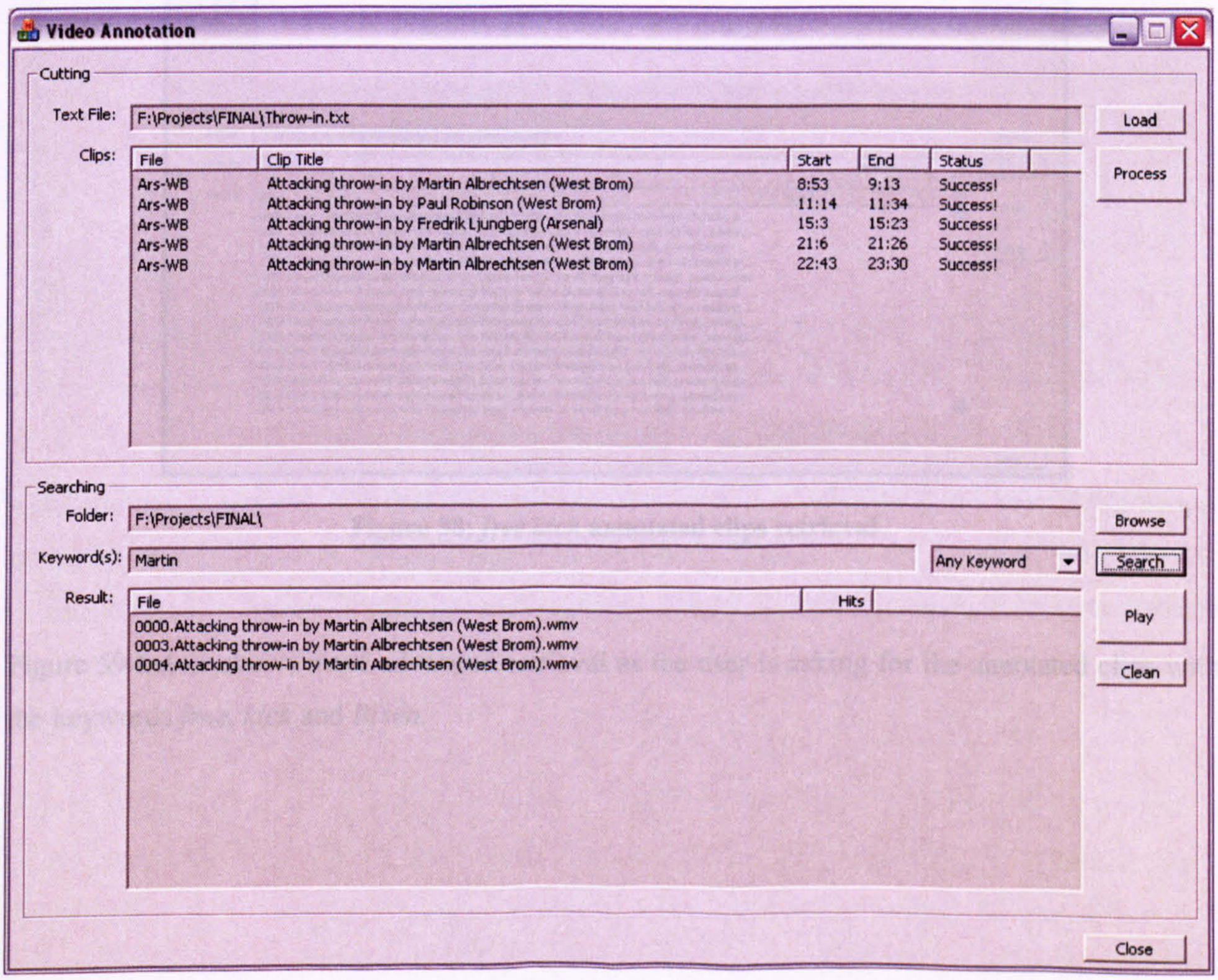


Figure 57: Video cutting, browsing and playing GUI

As shown in Figure 57 above, *Martin* was the selected keyword and the system retrieved those matching clips.

4.3.7 Simple and Advanced Annotated Clips Retrieval

The system provides the user with simple and advanced annotated clips retrieval methods. Figure 58 below shows a simple retrieval of the annotated clips with the keywords *free* or *kick.*

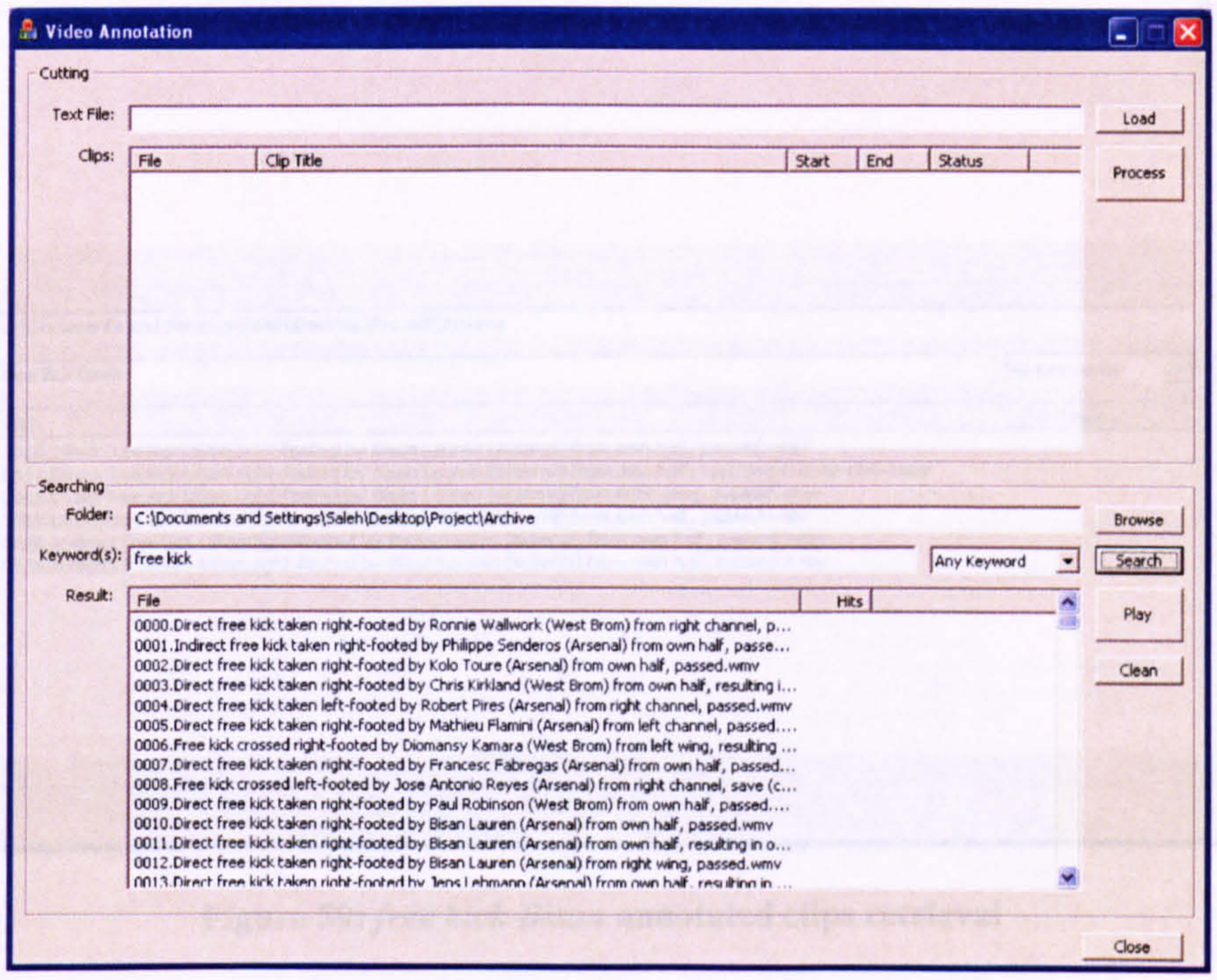


Figure 58: *free kick* annotated clips retrieval

Figure 59 below shows more advanced retrieval as the user is asking for the annotated clips with the keywords *free*, *kick* and *Bisen.*

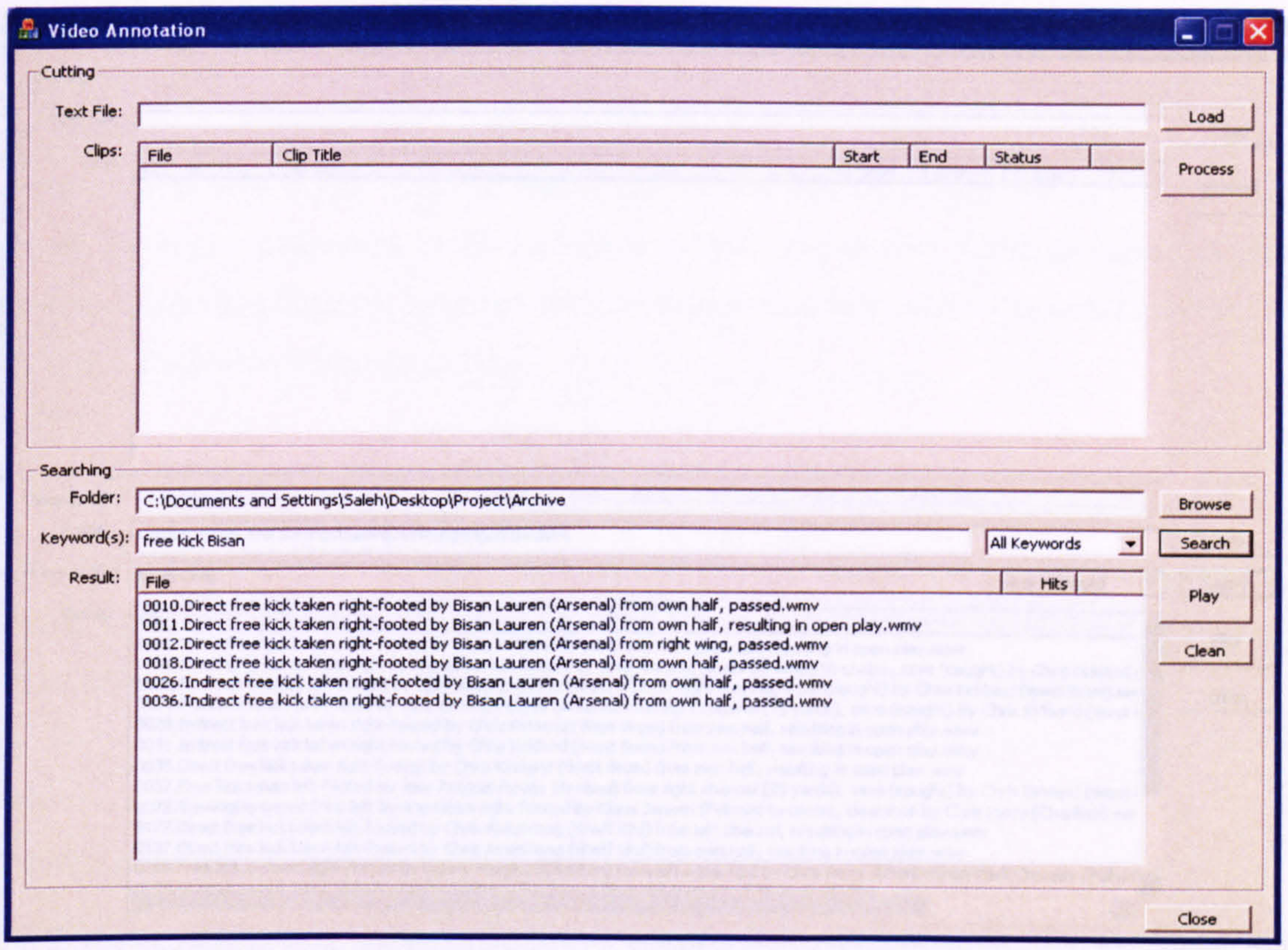


Figure 59: *free kick Bisan* annotated clips retrieval

Figure 60 below shows a different way of retrieving the annotated clips as the user is requesting all the annotated clips with the keywords *Bisen* or *Chris.* This shows the ability of the system to be able to retrieve annotated clips not only by using special football keywords but also to retrieve annotated clips by searching for player names or team names.

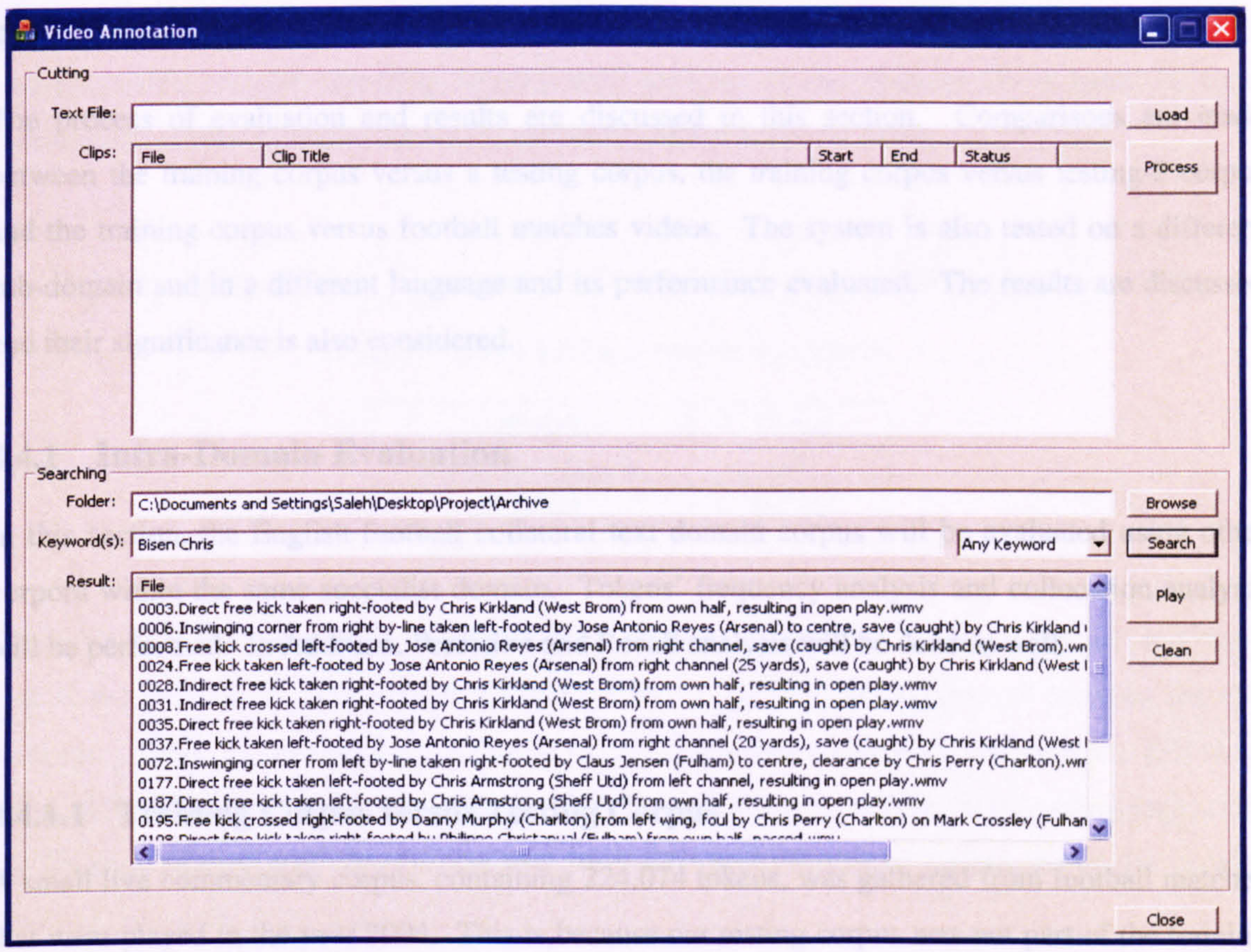


Figure 60: *Bisen Chris* annotated clips retrieval

4.3.8 Conclusion

Implementation has shown that it is possible to eliminate human input in many areas to achieve automated video annotation. Text Analysis has provided the needed calculations (token list, token frequency, token weirdness, token z-score, token weirdness z-score, concordance and collocations). *Protégé* illustrated the collocation result as a graph and as a list. The video annotation GUIs demonstrated one way to automate the video annotation with the use of text and video synchronization. All this is developed from finding significant keywords that helped build patterns; the system was then able to locate the segments in the video file and then to automatically cut and annotate them.

.

4.4 Evaluation

The process of evaluation and results are discussed in this section. Comparisons are made between the training corpus versus a testing corpus, the training corpus versus testing-2 corpus and the training corpus versus football matches videos. The system is also tested on a different sub-domain and in a different language and its performance evaluated. The results are discussed and their significance is also considered.

4.4.1 Intra-Domain Evaluation

In this section, the English football collateral text domain corpus will be evaluated using other corpora within the same specialist domain. Tokens' frequency analysis and collocation analysis will be performed. In Addition, Precision and Recall analysis will be done as well.

4.4.1.1 Training Corpus versus Testing Corpus

A small live commentary corpus, containing 224,074 tokens, was gathered from football matches that were played in the year 2004. This is because our testing corpus was not part of the training corpus. First, a part-of-speech comparison analysis was undertaken using CLAWS. This will compare the structure of both corpora based on their words' categories. The results are shown in Table 20:

Table 20: Comparing Part-of-Speech analysis using CLAWS ($N_T = 224,074$ and $N_C = 3,026,038$)

	Testing Corpus		Training Corpus	
Category	f_T	$R_T = f_T / N_T$	f_C	$R_C = f_C / N_C$
NP1	47,473	25.96	861,987	28.49
II	20,379	11.14	386,418	12.77
NN1	19,980	10.92	417,975	13.81
(	17,472	9.55	231,717	7.66
)	17,471	9.55	231,718	7.66
JJ	13,680	7.48	266,093	8.79
MC	10,696	5.85	191,838	6.34
.	9,301	5.09	215,897	7.13
VVN	4,899	2.68	95,754	3.16
VVG	2,471	1.35	45,159	1.49
VV0	1,925	1.05	40,588	1.34
VVD	1,725	0.94	38,503	1.27

Mann-Whitney U-test is used to explain the results from Table 20. Mann-Whitney U-test is a non-parametric test for assessing whether two samples of observations come from the same distribution. Table 21 below shows the results fro Mann-Whitney U-test.

Table 21: Mann-Whitney U-test result for CLAWS POS analysis between (English football) training and testing commentary corpora

Variable	Value
Mann-Whitney U	66
p-Value	0.729

The p-Value is the value that answers the question: are they similar? Mann-Whitney U-test states that if the p-Value is greater than 0.5, then there is insignificant difference between the two samples. Table 21 shows the p-Value = 0.729 which means there is insignificant difference between the POS frequency ratio in the English football live commentary text training corpus and the English football live commentary text testing corpus. Figure 61 below shows the plotting graph of the results from Table 20.

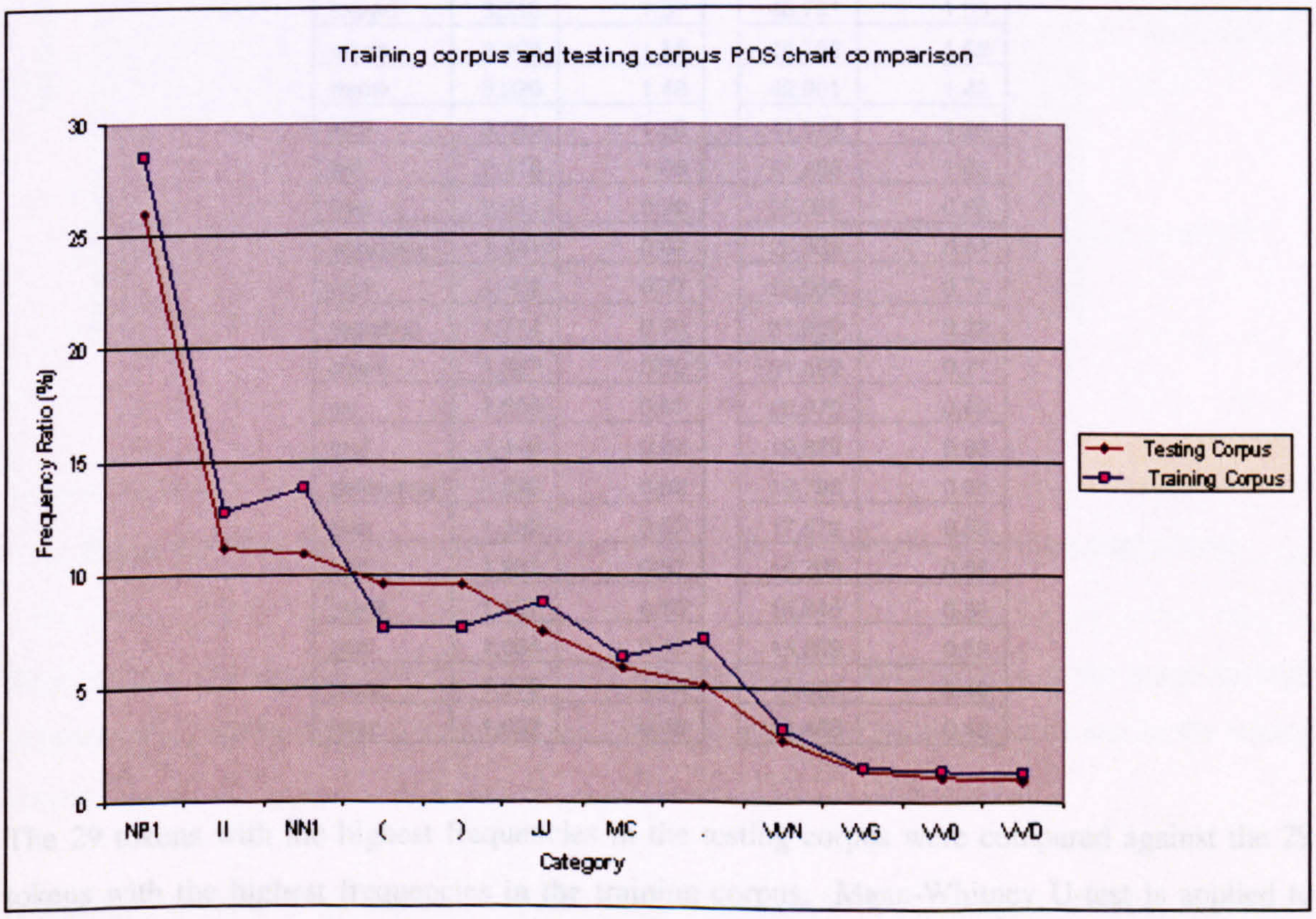


Figure 61: Training corpus versus testing corpus POS chart comparison

The graph above shows that despite the difference in the corpora sizes, their structures are almost the same. NP1 (singular proper nouns) are the most frequent; followed by NN1 (singular locative nouns) and II (preposition). Then parentheses come next.

The next test was to check the frequency of the most frequent tokens in the testing corpus and compare them with the training corpus using System Quirk. Table 22 demonstrates the comparison:

Table 22: Comparing frequencies using Quirk (N_T = 224,074 and N_C = 3,026,038)

	Testing Corpus		Training Corpus	
Token	f_T	$R_T = f_T / N_T$	f_C	$R_C = f_C / N_C$
)	17,472	7.80	231,718	7.66
(	17,472	7.80	231,717	7.66
by	13,482	6.02	179,335	5.93
-	12,957	5.78	162,007	5.35
.	11,648	5.20	123,734	4.09
:	9,275	4.14	113,157	3.74
in	5,043	2.25	68,113	2.25
,	4,614	2.06	65,998	2.18
right	3,828	1.71	53,514	1.77
from	3,772	1.68	52,024	1.72
footed	3,513	1.57	46,781	1.55
taken	3,468	1.55	46,132	1.52
throw	3,326	1.48	42,901	1.42
kick	3,089	1.38	41,013	1.36
left	2,415	1.08	31,838	1.05
free	2,011	0.90	25,958	0.86
attacking	1,941	0.87	25,334	0.84
play	1,722	0.77	22,056	0.73
resulting	1,714	0.76	21,869	0.72
open	1,567	0.70	21,499	0.71
on	1,506	0.67	20,870	0.69
foul	1,438	0.64	19,875	0.66
defending	1,382	0.62	19,798	0.65
own	1,282	0.57	17,572	0.58
half	1,281	0.57	16,760	0.55
yards	1,127	0.50	15,845	0.52
goal	1,084	0.48	15,659	0.52
cross	1,070	0.48	14,827	0.49
long	1,032	0.46	14,485	0.48

The 29 tokens with the highest frequencies in the testing corpus were compared against the 29 tokens with the highest frequencies in the training corpus. Mann-Whitney U-test is applied to compare the result from Table 22, see Table 23 below.

Table 23: Mann-Whitney U-test result for System Quirk analysis between (English football) training and testing commentary corpora

Variable	Value
Mann-Whitney U	419
p-Value	0.981

Table 23 shows the p-Value = 0.981 which means there is insignificant difference between the tokens frequency ratio in the English football live commentary text training corpus and the English football live commentary text testing corpus. Figure 62 below illustrates the similarity of the tokens frequency ratio in both corpora.

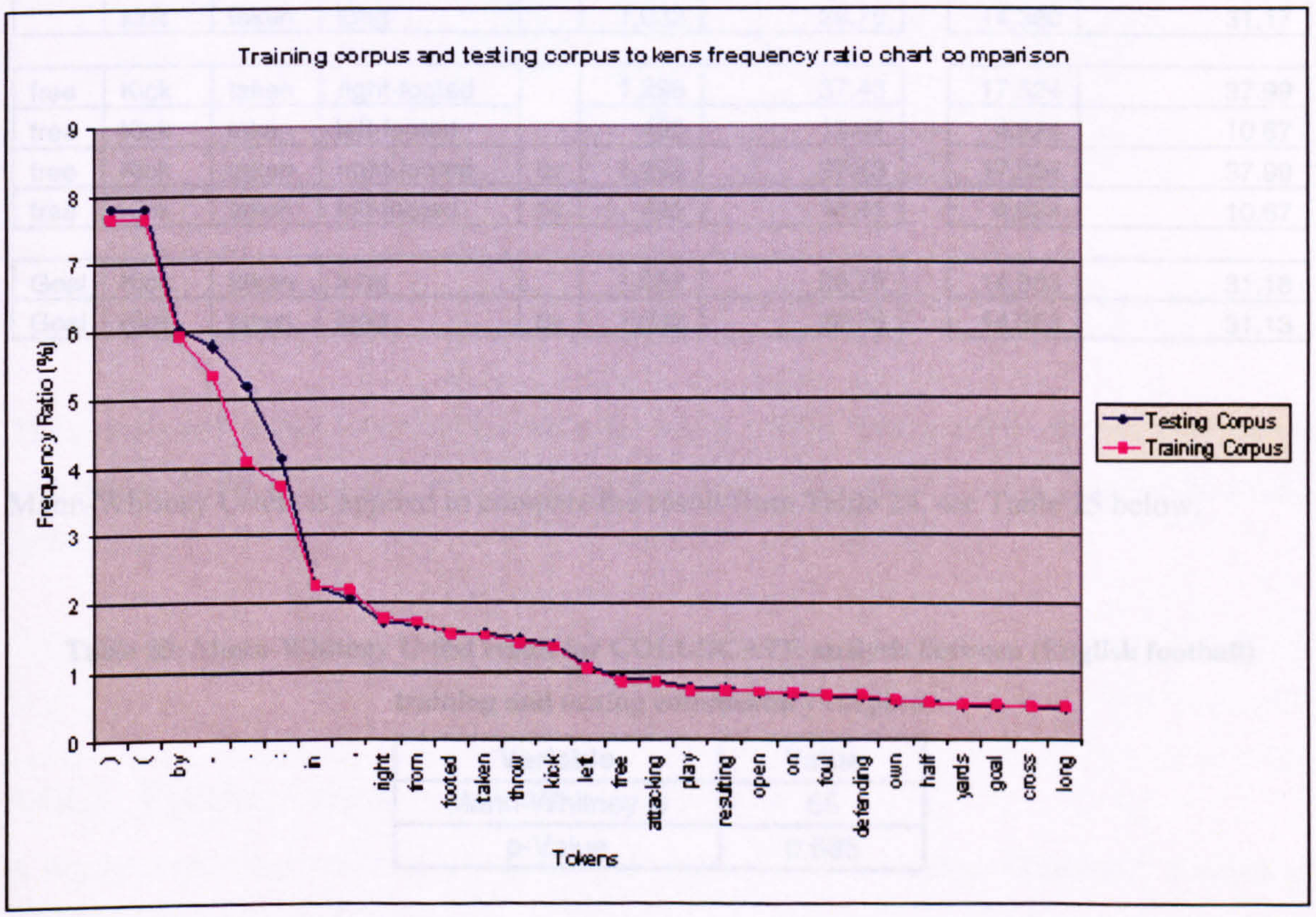


Figure 62: Training corpus versus testing corpus tokens frequency ratio chart comparison

One can see that the frequency ratio of the most frequent tokens is almost the same in both corpora. This shows the consistency of the tokens usage in the training corpus and in the testing corpus

It was then decided to go one step further and use COLLOCATE to compare 2-word, 3-word, 4-word and 5-word collocations using the same word chosen previously, *taken*. The goal here is to

see if the collocations ratio change or are still within a range that allows us to say that the training corpus results still stand. Table 24 shows the comparison:

Table 24: Comparing *taken* collocation frequencies using COLLOCATE

(f_{Ttaken}= 3,468, f_{Ctaken} = 46,132 , N_T = 224,074 and N_C = 3,026,038)

					Testing		Training	
					f_T	$R_t = f_T / f_{Ttaken}$	f_C	$R_C = f_C / f_{Ctaken}$
	kick	taken			2,838	81.83	37,395	81.06
free	kick	Taken			1,763	50.84	22,449	48.66
goal	kick	taken			1,075	31.00	14,895	32.29
	kick	taken	right-footed		1,298	37.43	17,524	37.99
	kick	taken	left-footed		465	13.41	4,924	10.67
	kick	taken	long		1,032	29.76	14,380	31.17
free	Kick	taken	right-footed		1,298	37.43	17,524	37.99
free	Kick	taken	left-footed		465	13.41	4,924	10.67
free	Kick	taken	right-footed	by	1,298	37.43	17,524	37.99
free	Kick	taken	left-footed	by	465	13.41	4,924	10.67
Goal	Kick	taken	long		1,032	29.76	14,383	31.18
Goal	Kick	taken	long	by	1,032	29.76	14,363	31.13

Mann-Whitney U-test is applied to compare the result from Table 24, see Table 25 below.

Table 25: Mann-Whitney U-test result for COLLOCATE analysis between (English football) training and testing commentary corpora

Variable	Value
Mann-Whitney U	65
p-Value	0.685

Table 25 shows the p-Value = 0.685 which means there is insignificant difference between *taken* collocation frequency ratio in the English football live commentary text training corpus and the English football live commentary text testing corpus. Figure 63 below shows the plotting graph of *taken* collocation frequencies ratio.

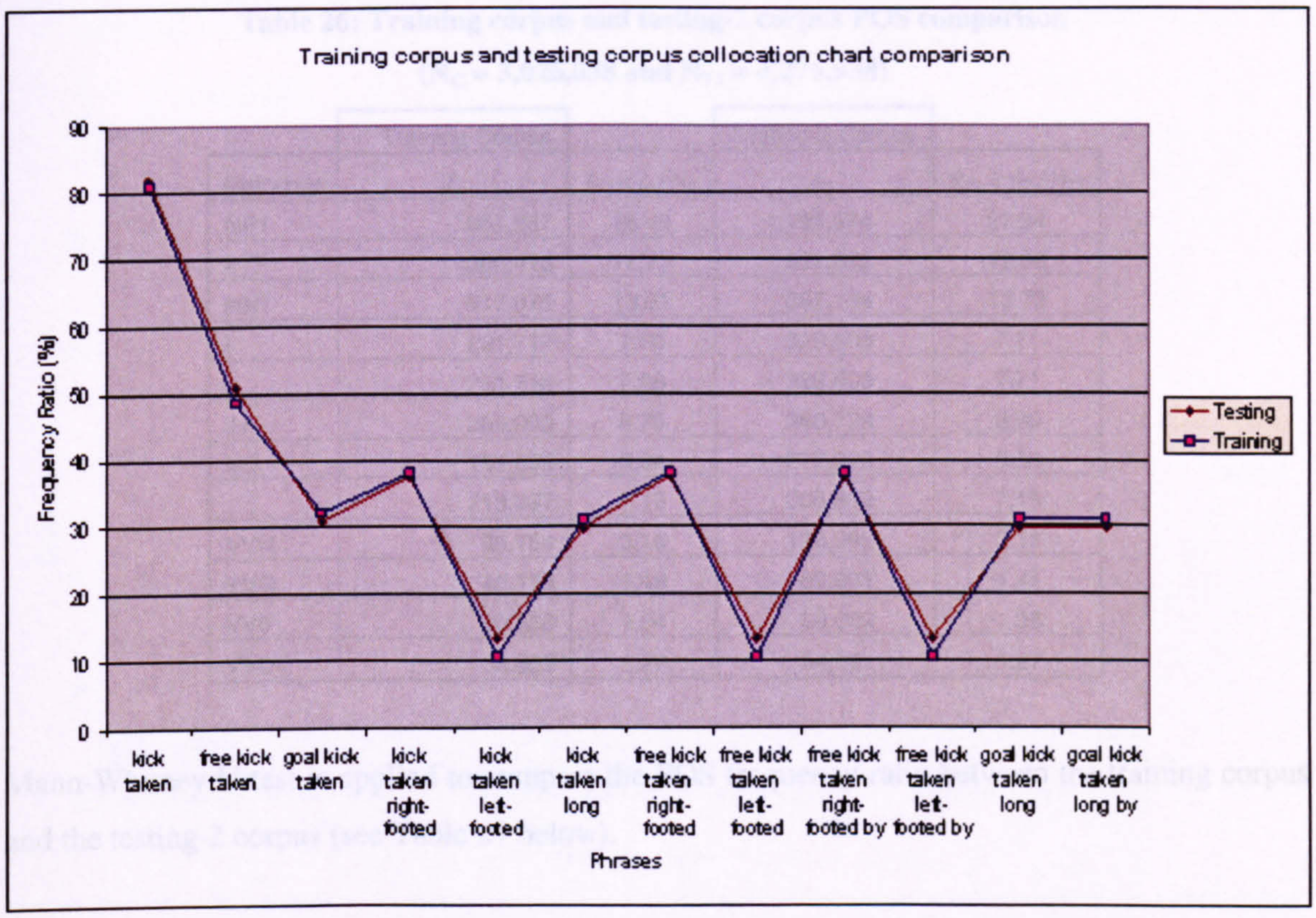


Figure 63: *taken* collocation frequencies ratio chart comparison (training versus testing corpora)

The comparison analysis between the training corpus and the testing corpus showed that there is consistency in the tokens list and their frequency ratio. Also, there is consistency in the tokens collocations.

4.4.1.2 Training Corpus versus Testing-2 Corpus

A bigger corpus was collected during the year 2006. This new corpus (Testing-2) consists of 4,276,938 tokens (over 1,000 matches). The concept here is to find out if the corpus analysis is still the same using tokens from a different year. A part-of-speech comparison analysis was undertaken using CLAWS. This will compare the structure of both corpora based on their words' categories. The results are shown in Table 26:

Table 26: Training corpus and testing-2 corpus POS comparison

(N_C = 3,026,038 and N_{T2} = 4,276,938)

	Training Corpus		Testing-2 Corpus	
Category	f_c	$R_C = f_c / N_C$	f_{T2}	$R_{T2} = f_{T2} / N_{T2}$
NP1	861,987	28.49	1,194,974	27.94
II	386,418	12.77	551,082	12.88
NN1	417,975	13.81	587,194	13.73
(	231,717	7.66	329,803	7.71
)	231,718	7.66	329,803	7.71
JJ	266,093	8.79	380,738	8.90
MC	191,838	6.34	272,098	6.36
.	215,897	7.13	305,892	7.15
VVN	95,754	3.16	136,092	3.18
VVG	45,159	1.49	62,981	1.47
VV0	40,588	1.34	58,032	1.36
VVD	38,503	1.27	54,501	1.27

Mann-Whitney U-test is applied to compare the POS frequency ratio between the training corpus and the testing-2 corpus (see Table 27 below).

Table 27: Mann-Whitney U-test result for POS CLAWS analysis between (English football) training and testing-2 commentary corpora

Variable	Value
Mann-Whitney U	68.5
p-Value	0.840

Table 27 shows the p-Value = 0.840 which means there is insignificant difference between the POS frequency ratio in the English football live commentary text training corpus and the English football live commentary text testing-2 corpus. Figure 64 below illustrates the similarity of the POS frequency ratio in both corpora.

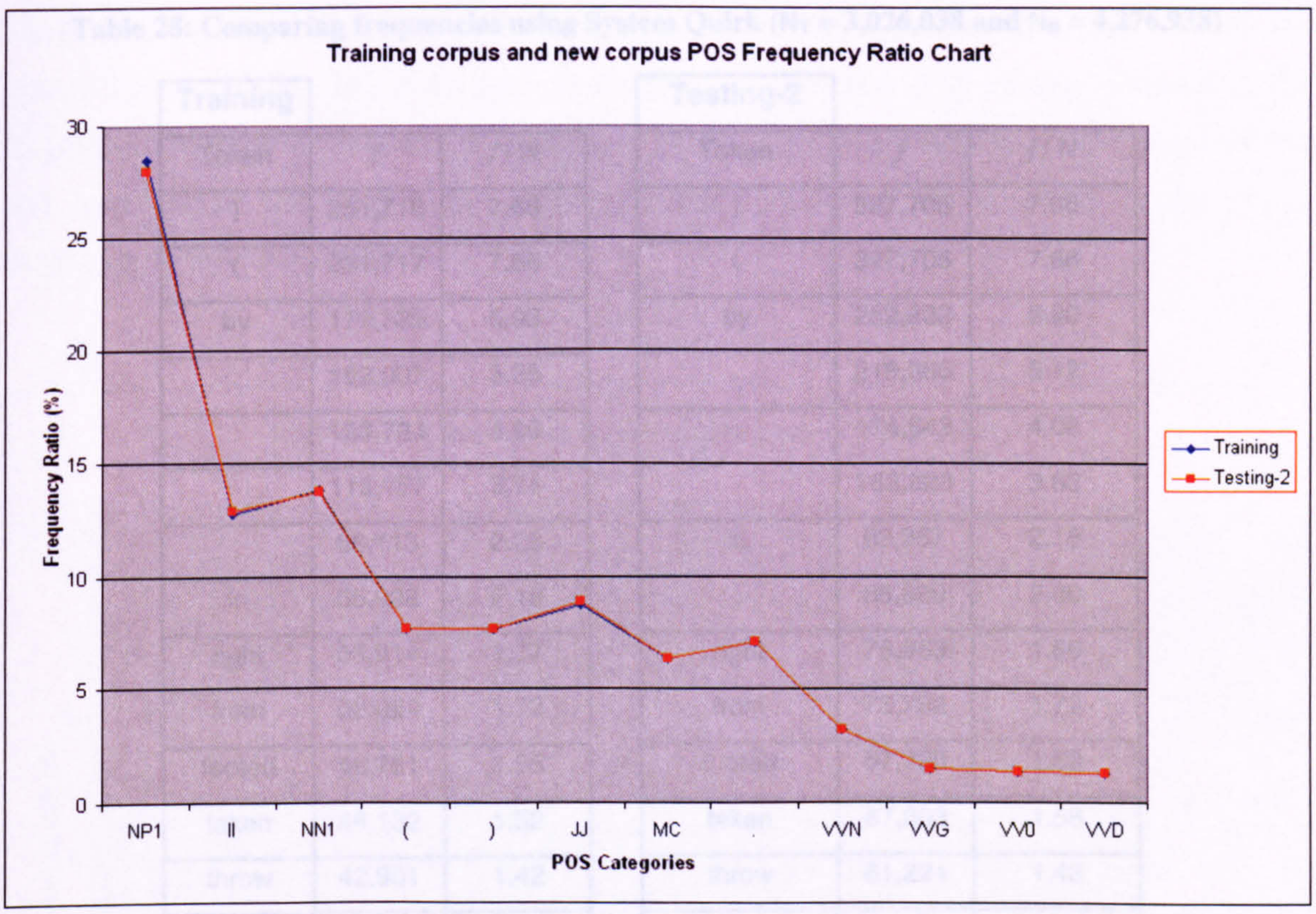


Figure 64: Training corpus and testing-2 corpus POS comparison chart

The POS categories frequency ratios are almost overlapping. This shows that both corpora have almost the same structure and there is consistency throughout the years.

Then, the most frequent tokens are analysed and compared using System Quirk and the results are shown in Table 28:

Table 28: Comparing frequencies using System Quirk (N_T = 3,026,038 and N_N = 4,276,938)

Training		
Token	*f*	*f* / N
)	231,718	7.66
(	231,717	7.66
by	179,335	5.93
.	162,007	5.35
:	123,734	4.09
-	113,157	3.74
,	68,113	2.25
in	65,998	2.18
right	53,514	1.77
from	52,024	1.72
footed	46,781	1.55
taken	46,132	1.52
throw	42,901	1.42
kick	41,013	1.36
left	31,838	1.05
free	25,958	0.86
attacking	25,334	0.84
play	22,056	0.73
resulting	21,869	0.72
goal	21,499	0.71
on	20,870	0.69
foul	19,875	0.66
open	19,798	0.65
defending	17,572	0.58
yards	16,760	0.55

Testing-2		
Token	*f*	*f* / N
)	327,705	7.66
(	327,705	7.66
by	252,232	5.90
.	219,066	5.12
:	174,643	4.08
-	163,823	3.83
in	93,357	2.18
,	85,630	2.00
right	76,963	1.80
from	73,732	1.72
footed	67,798	1.59
taken	67,603	1.58
throw	61,224	1.43
kick	59,364	1.39
left	45,471	1.06
free	37,836	0.88
attacking	36,241	0.85
play	32,335	0.76
resulting	32,133	0.75
open	29,386	0.69
on	29,114	0.68
goal	28,779	0.67
foul	28,673	0.67
defending	24,942	0.58
yards	23,193	0.54

Mann-Whitney U-test is applied to compare the result from Table 28, see Table 29 below.

Table 29: Mann-Whitney U-test result for Quirk analysis between (English football) training and testing-2 commentary corpora

Variable	Value
Mann-Whitney U	308
p-Value	0.930

Table 29 shows the p-Value = 0.930 which means there is insignificant difference between the tokens frequency ratio in the English football live commentary text training corpus and the English football live commentary text testing-2 corpus. This shows that even after 2 years, tokens still have similar frequency ratio. Figure 65 below illustrates the similarity of the tokens frequency ratio in both corpora.

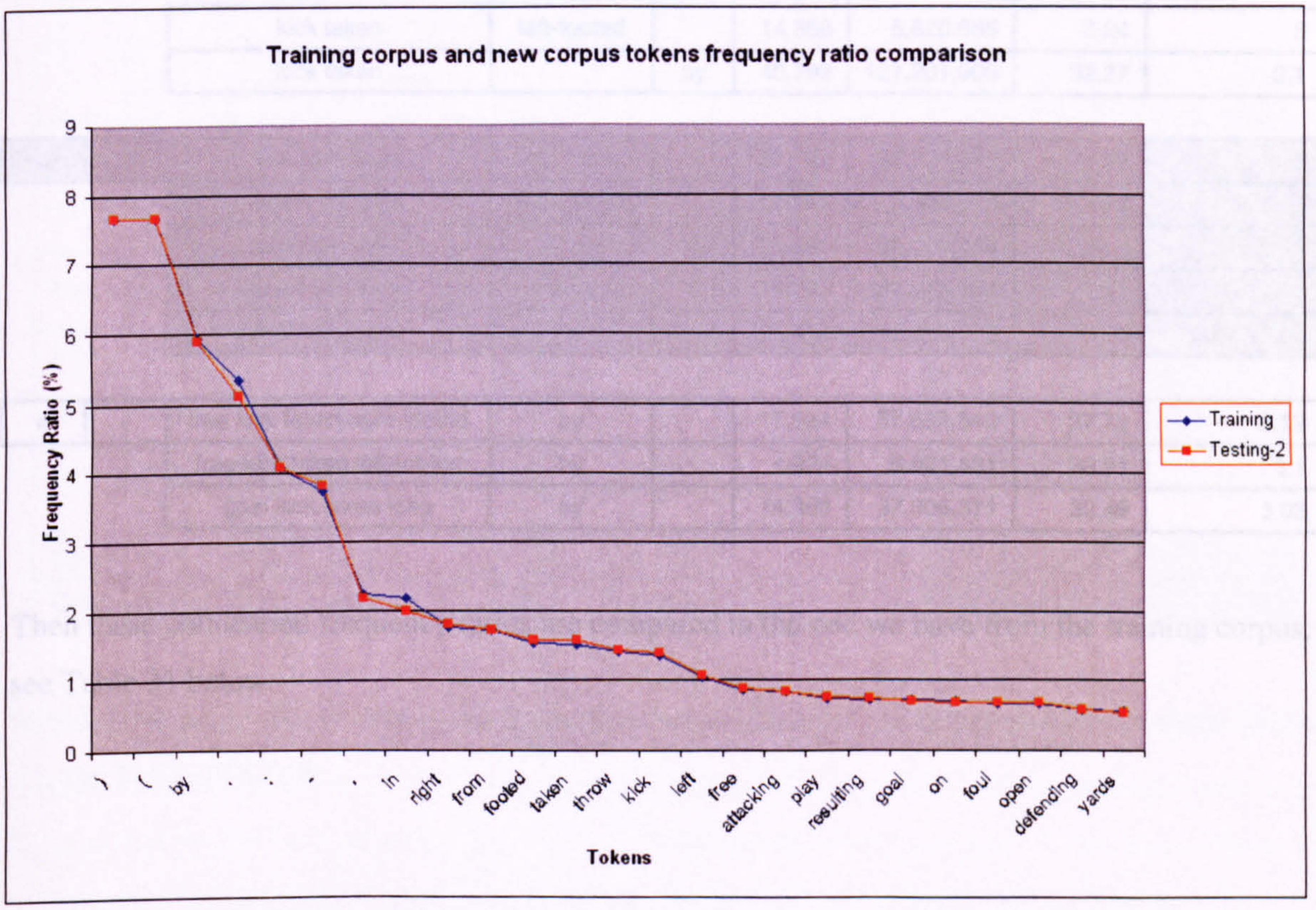


Figure 65: Training corpus and testing-2 corpus tokens frequency comparison chart

The live commentary text training corpus that was collected in 2004 (770 matches) and the live commentary testing-2 corpus that was collected in 2006 (1,000 matches) have the same most frequent tokens. These tokens frequency ratio are almost identical in both corpora and this shows a consistency in the live commentary corpus.

The next step is to look at how *taken* collocates in the testing-2 corpus, see Table 30 below.

Table 30: *taken* Collocation (N = 4,276,938)

Step	-1	keyword	1	2	*f*	U-Score	K-Score	Strength
1	kick	taken			48,851	270,776,898	13.93	3
2	free	kick taken			29,392	100,352,310	8.04	3
	goal	kick taken			19,430	41,332,041	12.23	3
		kick taken	right-footed		17,524	57,724,890	9.78	3
		kick taken	long		4,924	37,948,064	7.51	3
		kick taken	left-footed		14,380	5,620,585	3.24	3
		kick taken		by	40,792	127,201,908	32.27	3.1
3		free kick taken	right-footed		17,524	57,724,890	9.78	3
		free kick taken	left-footed		4,924	5,620,585	3.24	3
		free kick taken		by	25,380	100,268,854	26.1	3
		goal kick taken	long		14,383	37,948,064	7.51	3
		goal kick taken		by	21,412	41,242,270	13.97	3.07
4		free kick taken right-footed	by		17,524	57,652,343	37.73	3.12
		free kick taken left-footed	by		4,924	5,631,431	33.91	3.1
		goal kick taken long	by		14,363	37,808,371	32.46	3.03

Then these collocation frequency ratios are compared to the one we have from the training corpus, see Table 31 below.

Table 31: *taken* collocation frequency ratio comparison

(f_{Ttaken}=46,132, f_{Ntaken} = 59,364 , N_T = 3,026,038 and N_N = 4,276,938)

					Training		Testing-2	
					f_T	$R_T = f_T / f_{Ttaken}$	f_N	$R_N = f_N / f_{Ntaken}$
	kick	taken			37,395	81.06	48,851	82.29
free	kick	taken			22,449	48.66	29,392	49.51
goal	kick	taken			14,895	32.29	19,430	32.73
	kick	taken	right-footed		17,524	37.99	22,431	37.79
	kick	taken	left-footed		4,924	10.67	6,949	11.71
	kick	taken	long		14,380	31.17	20,534	34.59
free	kick	taken	right-footed		17,524	37.99	22,431	37.79
free	kick	taken	left-footed		4,924	10.67	6,949	11.71
free	kick	taken	right-footed	by	17,524	37.99	22,431	37.79
free	kick	taken	left-footed	by	4,924	10.67	6,949	11.71
Goal	kick	taken	long		14,383	31.18	19,534	32.91
Goal	kick	taken	long	by	14,363	31.13	19,412	32.70

Mann-Whitney U-test is applied to compare *taken* collocation frequency ratio result from Table 31, see Table 32 below.

Table 32: Mann-Whitney U-test result for *taken* collocation analysis between (English football) training and testing-2 commentary corpora

Variable	Value
Mann-Whitney U	63
p-Value	0.602

Table 32 shows the p-Value = 0.602 which means there is insignificant difference between *taken* collocation frequency ratio in the English football live commentary text training corpus and the English football live commentary text testing-2 corpus. Figure 66 below shows the plotting graph of *taken* collocation frequencies ratio.

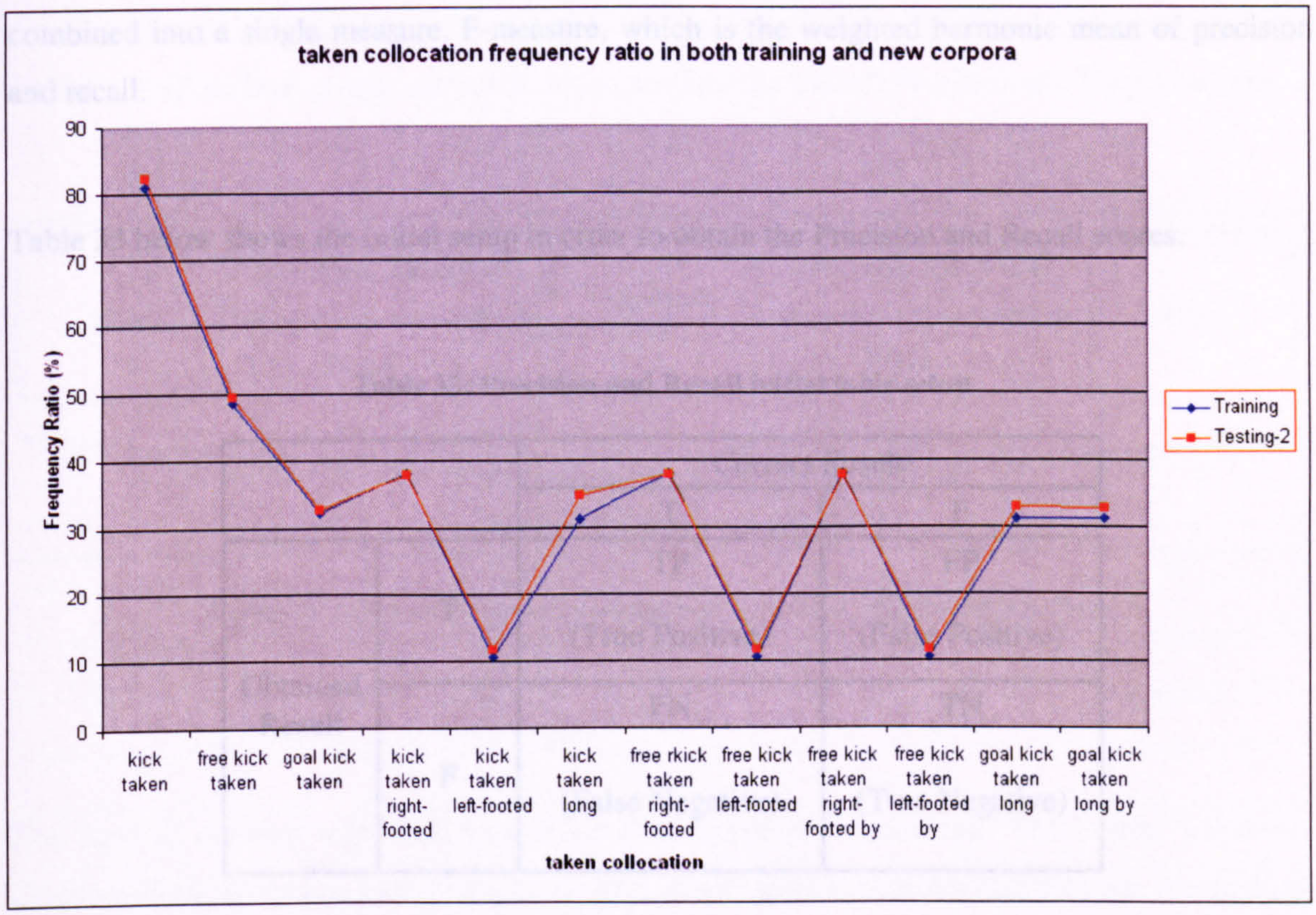


Figure 66: Training corpus and testing-2 corpus *taken* collocation comparison chart

The comparisons between the training corpus and the testing-2 corpus show that the difference is minimal. This shows that the collocation analysis result we obtained from the testing-2 corpus is the same as the collocation analysis result we have from the training corpus. Even though there is a two years difference between both corpora, the live commentary text is still consistent.

4.4.1.3 Precision and Recall

Precision and Recall are two widely used measures for evaluating the quality of results in domains such as Information Retrieval. Precision is defined as the *number of relevant documents* retrieved by a search *divided by the total number of documents retrieved* by that search, and Recall is defined as the *number of relevant documents* retrieved by a search *divided by the total number of existing relevant documents* (which should have been retrieved). In Information Retrieval, a perfect Precision score of 1.0 means that every result retrieved by a search was relevant (but says nothing about whether all relevant documents were retrieved) whereas a perfect Recall score of 1.0 means that all relevant documents were retrieved by the search (but says nothing about how many irrelevant documents were also retrieved). Usually Precision and Recall scores are

combined into a single measure, F-measure, which is the weighted harmonic mean of precision and recall.

Table 33 below shows the initial setup in order to obtain the Precision and Recall scores.

Table 33: Precision and Recall initial table setup

		Correct Result	
		T	F
Obtained Result	T	TP (True Positive)	FP (False Positive)
	F	FN (False Negative)	TN (True Negative)

Explanation of TP, FP, FN and TN will be provided shortly when we look at an actual analysis table.

Equation 5 below shows the equations used to calculate the scores for Precision, Recall and F-measure.

$$\text{Recall} = \frac{TP}{TP + FN}, \quad \text{Precision} = \frac{TP}{TP + FP}$$

$$F = \frac{2 * \text{Recall} * \text{Precision}}{\text{Recall} + \text{Precision}}$$

Equation 5: Precision and Recall Equations

Since the system does not access the video files directly to detect events, Precision and Recall has to be done at two stages, as shown in Figure 67. First, Precision and Recall analysis occurs

between the video and the live commentary text, then again between the live commentary text and the system. Then both results are combined to produce the overall Precision and Recall analysis.

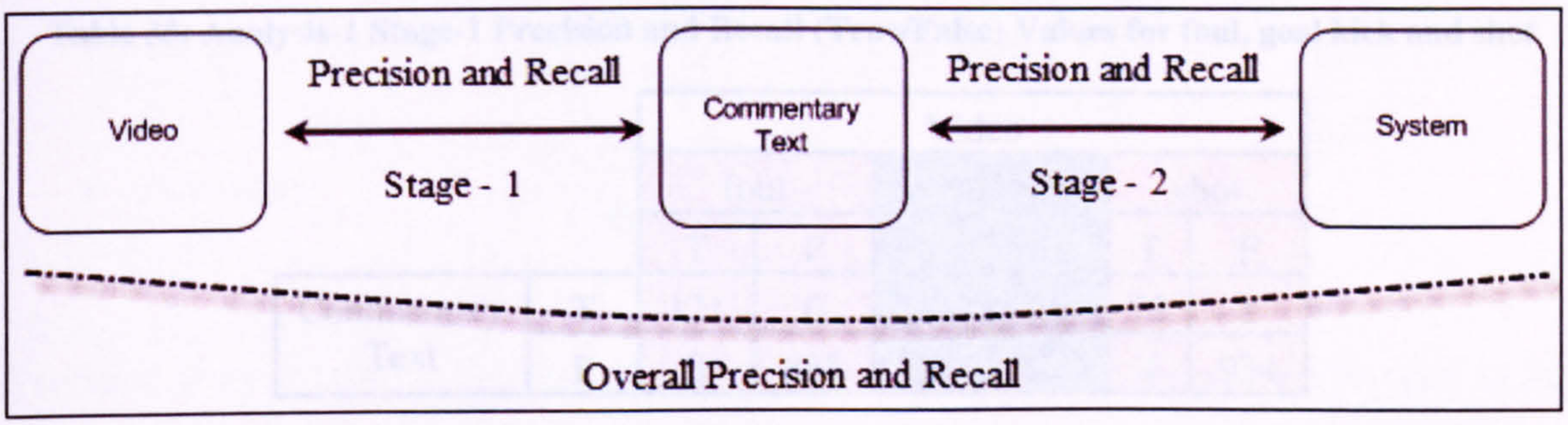


Figure 67: Overview of the two stages of Precision and Recall Analysis

Five football matches were randomly selected and manually annotated. The events' keywords (throw-in, foul, goal kick, kick taken, cross, shot and corner) are chosen based on the terminology analysis, as previously discussed and analysed in Chapter 3, of keywords with high weirdness. To avoid ambiguity and to confirm analysis results, the Precision and Recall analysis will be done twice using different sets of randomly selected matches and the results will be compared.

4.4.1.3.1 Precision and Recall Analysis-1

Table 34 below shows the Precision and Recall (True/False) values for throw-in, kick taken, cross and corner for stage-1 which is between the video and the commentary text.

Table 34: Analysis-1 Stage-1 Precision and Recall (True/False) Values for throw-in, kick taken, cross and corner

		Video							
		throw-in		kick taken		cross		corner	
		T	F	T	F	T	F	T	F
Commentary Text	T	271	1	247	0	93	1	82	0
	F	6	735	3	759	4	913	3	924

Table 35 below shows the Precision and Recall (True/False) values for foul, goal kick and shot for stage-1 which is between the video and the commentary text.

Table 35: Analysis-1 Stage-1 Precision and Recall (True/False) Values for foul, goal kick and shot

		Video					
		foul		goal kick		shot	
		T	F	T	F	T	F
Commentary Text	T	121	0	120	2	72	1
	F	3	885	9	886	4	934

To explain the meaning of these numbers let us consider the results for *goal kick* shown in Table 35. The video and the commentary text detected the same *goal kick* event 120 times (True Positive). The video detected a *goal kick* event nine times where the commentary text failed to do so (False Negative). The commentary text detected two *goal kick* events where the video did not confirm it (False Positive). The video and the commentary text detected 886 other matching events that are not a *goal kick* event (True Negative).

Table 36 shows the total values of the Precision and Recall (True/False) values; the sum of all similar values from Table 34 and Table 35:

Table 36: Analysis-1 Stage-1 Precision and Recall (True/False) total values

		VIDEO	
		T	F
Commentary Text	T	1,006	5
	F	32	6,036

The video and commentary text both reported a matching event 1,006 times. The commentary text failed to report an existing event 32 times. The commentary text reported a false event five times. Both the video and the commentary text reported 6,036 matching events that did not belong to the search query for the specified event.

Applying Equation 5, the scores for Precision and Recall for stage-1 can now be obtained.

- Precision $= \frac{1{,}006}{1{,}006+5} = 0.995$
- Recall $= \frac{1{,}006}{1{,}006+32} = 0.969$
- F-measure $= \frac{2\times0.995\times0.969}{0.995+0.969} = 0.982$

Precision score of 0.995 means that 99.5% of the events returned to the user are relevant to the query. Recall score of 0.969 means that 96.9% of the relevant events are retrieved. F-measure score of 0.982 shows the weighted harmonic mean. Having all scores above 0.970 indicates that the information retrieval accuracy between the video and the commentary text is well above the acceptable level.

The second part of this analysis calculates the Precision and Recall for stage-2 which is between the commentary text and the system.

Table 37 below shows the Precision and Recall (True/False) values for throw-in, kick taken, cross and corner for stage-2 which is between the commentary text and the system.

Table 37: Analysis-1 Stage-2 Precision and Recall (True/False) Values for throw-in, kick taken, cross and corner

		Commentary Text							
		throw-in		kick taken		cross		corner	
		T	F	T	F	T	F	T	F
System	T	271	0	246	0	93	0	81	0
	F	1	733	1	758	1	911	1	923

Table 38 below shows the Precision and Recall (True/False) values for foul, goal kick and shot for stage-2 which is between the video and the commentary text.

Table 38: Analysis-1 Stage-2 Precision and Recall (True/False) Values for foul, goal kick and shot

		Commentary Text					
		foul		goal kick		shot	
		T	F	T	F	T	F
System	T	120	0	121	0	72	0
	F	1	884	1	883	1	932

Table 39 shows the total values of the Precision and Recall (True/False) values; the sum of all similar values from Table 37 and Table 38.

Table 39: Analysis-1 Stage-2 Precision and Recall (True/False) total values

		Commentary Text	
		T	F
System	T	1,004	0
	F	7	6,024

Applying Equation 5, the scores for Precision and Recall for stage-2 can now be obtained.

- Precision $= \frac{1{,}004}{1{,}004+0} = 1.0$
- Recall $= \frac{1{,}004}{1{,}004+7} = 0.993$
- F-measure $= \frac{2\times1.0\times0.993}{1.0+0.993} = 0.996$

Precision score of 1.0 means that 100% of the events returned to the user are relevant to the query. In normal statistical usage, a score of 1.0 would indicate that a system is perfect. Video annotation and retrieval field is still being investigated and systems are far from achieving perfection. However, our case here is an exception and a score of 1.0 is expected. Anything less than 1.0 will mean our system has failed to detect the events that it filtered and a system reinvestigation will be required to determine the reason of such failure. Have in mind that this is only between the commentary text and the system when the Precision score is 1.0. Recall score of 0.993 means that 99.3% of the relevant events are retrieved. F-measure score of 0.996 shows the

weighted harmonic mean. Having all scores above 0.990 indicates that the information retrieval accuracy between the commentary text and the system is well above the acceptable level.

Now that stage-1 and stage-2 Precision and Recall scores are calculated, the overall Precision and Recall can be calculated, see Table 40 below.

Table 40: Analysis-1 Overall Precision and Recall Scores

	Stage-1	Stage-2	Overall
Precision	0.995	1.0	0.995
Recall	0.971	0.993	0.964
F-measure	0.982	0.996	0.979

Precision score of 0.995 means that when a user uses the system and submits a search for specific event(s), 99.5% of the returned results are relevant. Recall score of 0.964 means that 96.4% of the relevant events are retrieved. F-measure score of 0.979 shows the weighted harmonic mean. Having Precision, Recall and F-measure above 0.95 indicates that the information retrieval accuracy between the video and the system, going through the commentary text, is well above the acceptable level.

4.4.1.3.2 Precision and Recall Analysis-2

For the purpose of evaluating the results of analysis-1, and to avoid ambiguity, another five football matches were randomly chosen to calculate the Precision and Recall for the same keywords that were chosen in Analysis-1. Table 41 below shows the Precision and Recall (True/False) values for throw-in, kick taken, cross and corner for stage-1 which is between the commentary text and the system.

Table 41: Analysis-2 Stage-1 Precision and Recall (True/False) Values for throw-in, kick taken, cross and corner

		Video							
		throw-in		kick taken		cross		corner	
		T	F	T	F	T	F	T	F
Commentary Text	T	271	1	247	0	93	1	82	0
	F	6	735	3	759	4	913	3	924

Table 42 below shows the Precision and Recall (True/False) values for foul, goal kick and shot for stage-1.

Table 42: Analysis-2 Stage-1 Precision and Recall (True/False) Values for foul, goal kick and shot

		Video					
		foul		goal kick		shot	
		T	F	T	F	T	F
Commentary Text	T	121	0	120	2	72	1
	F	3	885	6	886	4	934

Table 43 shows the total values of the Precision and Recall (True/False) values; the sum of all similar values from Table 41 and Table 42.

Table 43: Analysis-2 Stage-1 Precision and Recall (True/False) total values

		VIDEO	
		T	F
Commentary Text	T	1,031	8
	F	32	6,186

Applying Equation 5, the scores for Precision and Recall for stage-1 can now be obtained.

- Precision $= \frac{1{,}031}{1{,}031+8} = 0.992$
- Recall $= \frac{1{,}031}{1{,}031+32} = 0.970$
- F-measure $= \frac{2\times0.992\times0.970}{0.992+0.970} = 0.981$

Table 44 shows analysis-1 and analysis-2 Precision, Recall and F-measure scores.

Table 44: Stage-1 (Analysis-1 and Analysis-2) Precison and Recall scores comparison

	Analysis-1	Analysis-2
	Stage-1	Stage-1
Precision	0.995	0.992
Recall	0.969	0.97
F-measure	0.982	0.981

One can notice the similarity in the Precision and Recall scores in both analyses. The similarity can be shown more clearly when looking at Figure 68 below:

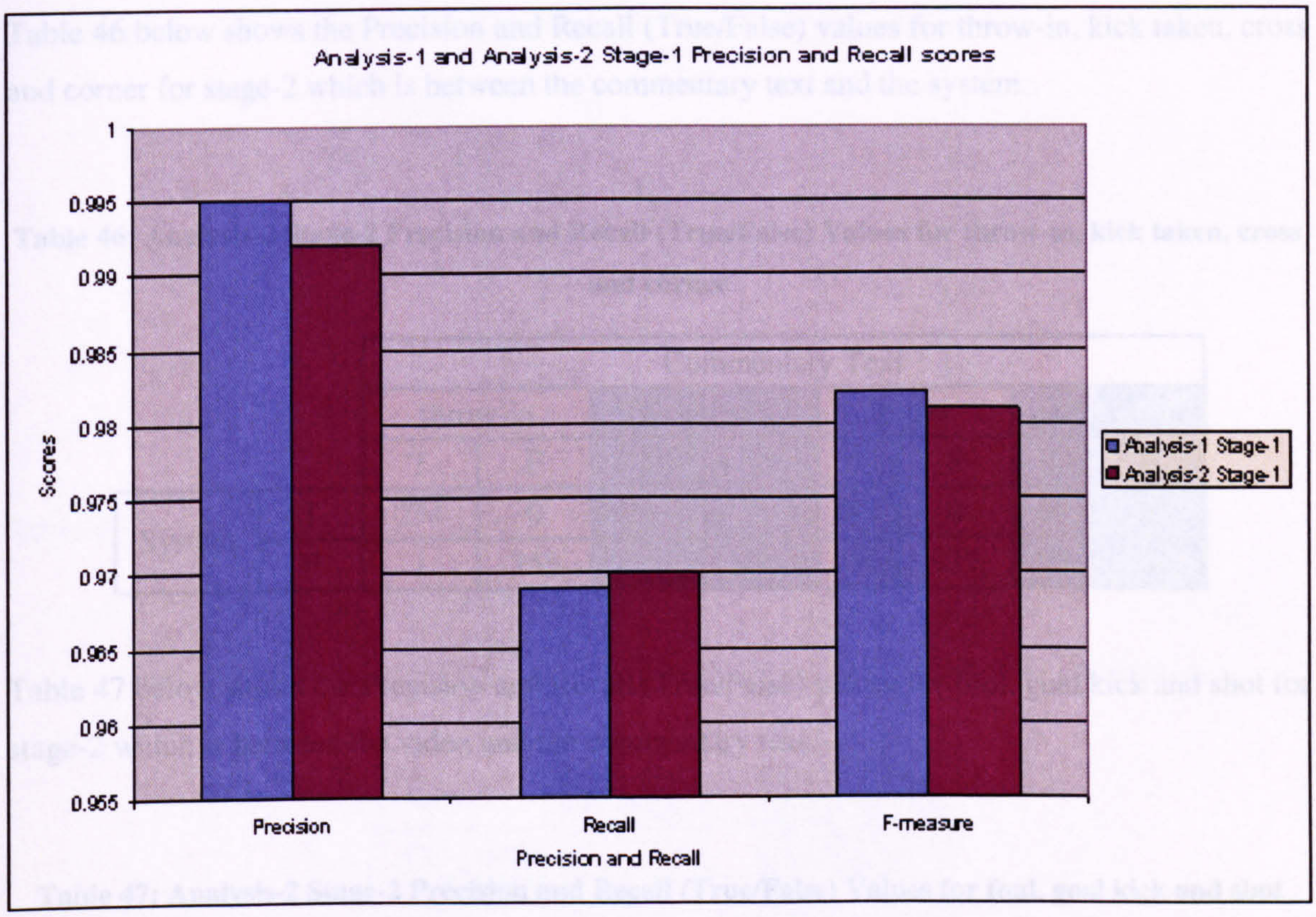


Figure 68: Stage-1 (Analysis-1 and Analysis-2) Precision and Recall scores comparison chart

However, the best way to compare both results is by using a test like Mann-Whitney U-test, see Table 45 below:

Table 45: Mann-Whitney U-test comparison result between analysis-1 and analysis-2 stage-1 Precision and Recall scores.

Variable	Value
Mann-Whitney U	4
p-Value	0.827

The result shows a p-Value of 0.827 which indicates there is insignificant difference between the Precision and Recall scores at stage-1 for both analysis-1 and analysis-2.

The second part of this analysis is to calculate the Precision and Recall for stage-2 which is between the commentary text and the system and compare the results to those from analysis-1.

Table 46 below shows the Precision and Recall (True/False) values for throw-in, kick taken, cross and corner for stage-2 which is between the commentary text and the system.

Table 46: Analysis-2 Stage-2 Precision and Recall (True/False) Values for throw-in, kick taken, cross and corner

		Commentary Text							
		throw-in		kick taken		cross		corner	
		T	F	T	F	T	F	T	F
System	T	274	0	251	0	95	0	80	0
	F	1	758	1	781	1	937	1	952

Table 47 below shows the Precision and Recall (True/False) values for foul, goal kick and shot for stage-2 which is between the video and the commentary text.

Table 47: Analysis-2 Stage-2 Precision and Recall (True/False) Values for foul, goal kick and shot

		Commentary Text					
		foul		goal kick		shot	
		T	F	T	F	T	F
System	T	122	0	133	0	77	0
	F	1	910	1	910	1	955

Table 48 shows the total values of the Precision and Recall (True/False) values; which are the sum of all similar values from Table 46 and Table 47.

Table 48: Analysis-2 Stage-2 Precision and Recall (True/False) total values

		Commentary Text	
		T	F
System	T	1,032	0
	F	7	6,192

Applying Equation 5, the scores for Precision and Recall for stage-2 can now be obtained.

- Precision $= \frac{1{,}032}{1{,}032+0} = 1.0$
- Recall $= \frac{1{,}032}{1{,}032+7} = 0.993$
- F-measure $= \frac{2\times1.0\times0.993}{1.0+0.993} = 0.996$

Table 49 shows analysis-1 and analysis-2 Precision, Recall and F-measure scores.

Table 49: Stage-2 (Analysis-1 and Analysis-2) Precison and Recall scores comparison

	Analysis-1	Analysis-2
	Stage-2	Stage-2
Precision	1	1
Recall	0.993	0.993
F-measure	0.996	0.996

Stage-2 Precision and Recall scores in both analysis-1 and analysis-2 are almost identical, see Figure 69 below.

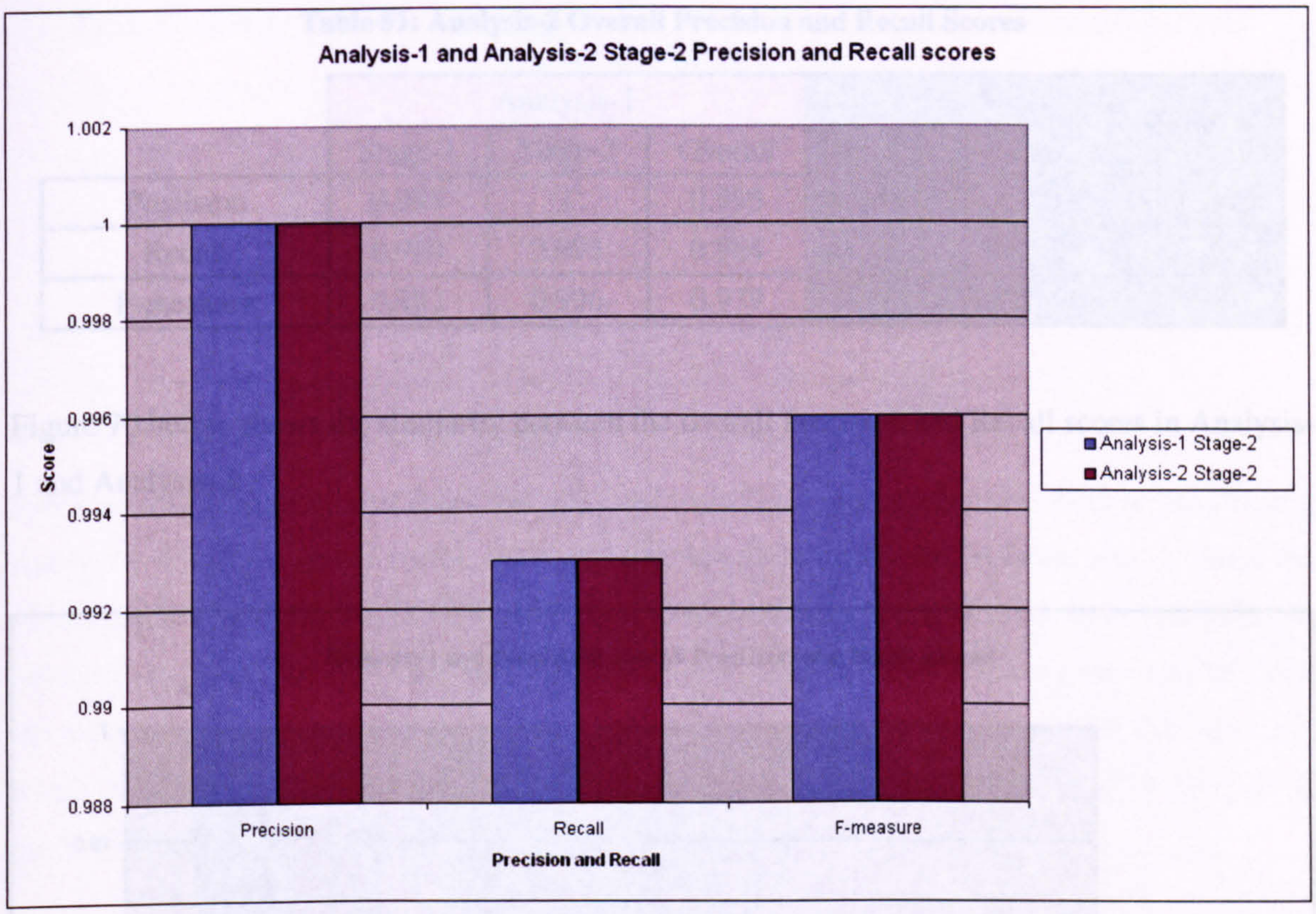


Figure 69: Stage-2 (Analysis-1 and Analysis-2) Precision and Recall scores comparison chart

Table 50 shows the Mann-Whitney U-test result.

Table 50: Mann-Whitney U-test comparison result between analysis-1 and analysis-2 stage-2 Precision and Recall scores.

Variable	Value
Mann-Whitney U	4.5
p-Value	1

The p-Value of 1.0 not only shows that there is insignificant difference between the Precision and Recall scores at stage-2 for both analysis-1 and analysis-2; it also shows that Precision and Recall scores are almost identical.

The last comparison to look at is the analysis-1 and analysis-2 overall Precision and Recall scores. See Table 51 below.

Table 51: Analysis-2 Overall Precision and Recall Scores

	Analysis-1			Analysis-2		
	Stage-1	Stage-2	Overall	Stage-1	Stage-2	Overall
Precision	0.995	1	0.995	0.992	1	0.992
Recall	0.969	0.993	0.964	0.97	0.993	0.963
F-measure	0.982	0.996	0.979	0.981	0.996	0.977

Figure 70 below shows the similarity between the overall Precision and Recall scores in Analysis-1 and Analysis-2.

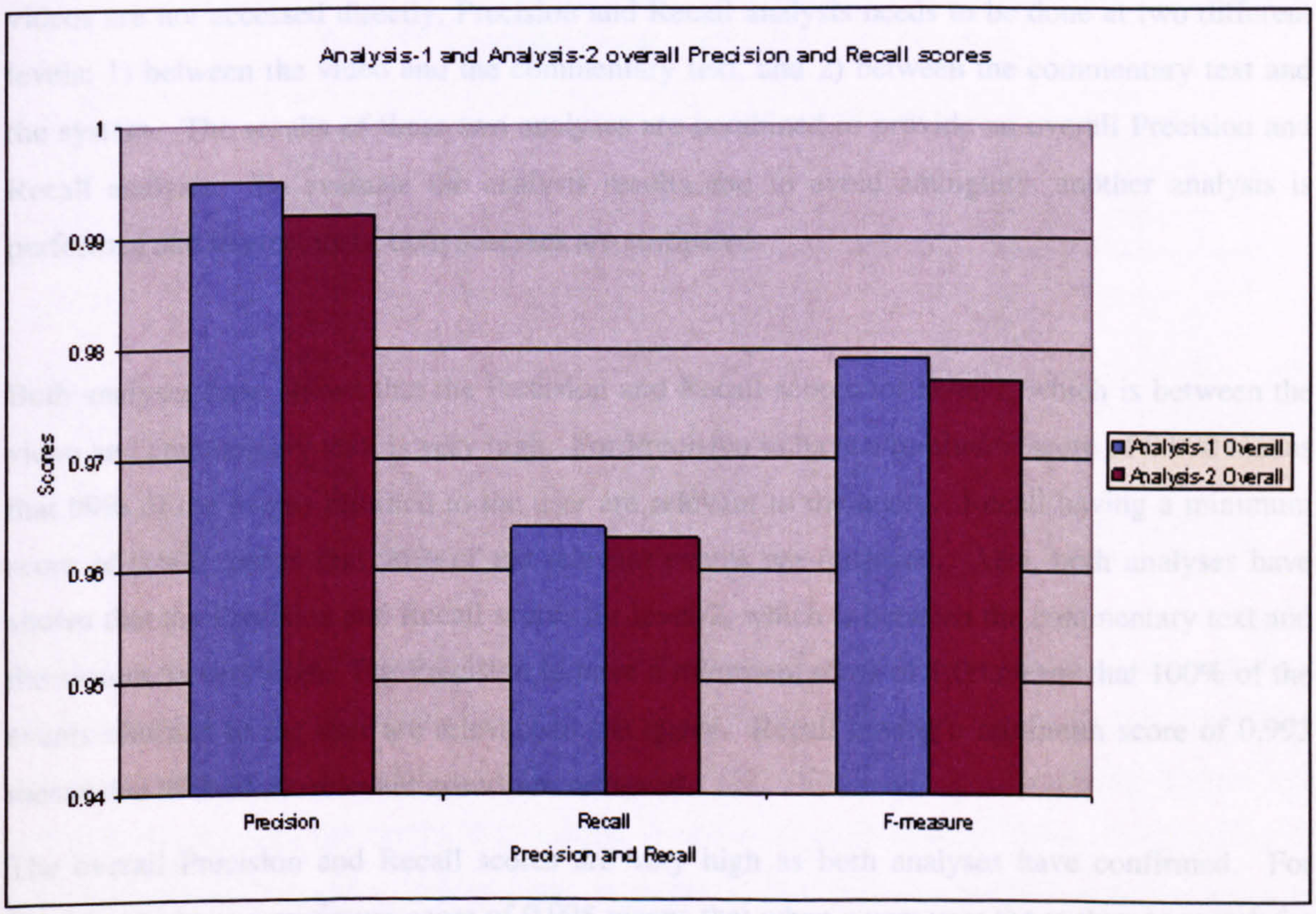


Figure 70: Overall (Analysis-1 and Analysis-2) Precision and Recall scores comparison chart

When applying Mann-Whitney U-test, Table 52 below, the p-Value of 0.513 shows that there is insignificant difference between the analysis-1 and analysis-2 overall Precision and Recall scores.

Table 52: Mann-Whitney U-test comparison result between analysis-1 and analysis-2 overall Precision and Recall scores.

Variable	Value
Mann-Whitney U	3
p-Value	0.513

4.4.1.3.3 Precision and Recall Analysis Conclusion

Precision and Recall are measures that are used to evaluate the quality of results in Information Retrieval. Precision measures the exactness and Recall measures the completeness. Since the videos are not accessed directly, Precision and Recall analysis needs to be done at two different levels: 1) between the video and the commentary text, and 2) between the commentary text and the system. The results of these two analyses are combined to provide an overall Precision and Recall analysis. To evaluate the analysis results and to avoid ambiguity, another analysis is performed and the results of both analyses are compared.

Both analyses have shown that the Precision and Recall scores for level-1, which is between the video and commentary text, is very high. For Precision to have a minimum score of 0.990 means that 99% of the events returned to the user are relevant to the query. Recall having a minimum score of 0.960 means that 96% of the relevant events are retrieved. Also, both analyses have shown that the Precision and Recall scores for level-2, which is between the commentary text and the system, is very high. For Precision to have a minimum score of 1.00 means that 100% of the events returned to the user are relevant to the query. Recall having a minimum score of 0.993 means that 99% of the relevant events are retrieved.

The overall Precision and Recall scores are very high as both analyses have confirmed. For Precision to have a minimum score of 0.995 means that when a user uses the system to search for a specific event, 99.5% of events the search returns are relevant. Recall having a minimum score of 0.964 means that when a user uses the system to search for a specific event, 96.4% of the relevant events are returned to the user.

The overall observation shows strong communication between the system and the actual video(s). Having Precision and Recall scoring above 0.9 on all levels confirms that the information retrieval percentage is high.

4.4.1.4 System Strength - Commentary Text versus Football Video

The live commentary corpus has proven its consistency throughout the previous comparisons and shown it is consistent. The next question to be answered is how strong is the corpus, how much information does the corpus have compared to the football video(s)? Despite the corpus consistency, if the corpus only contains minimal information, then one can argue against its usefulness. To find out and try to answer this question, five football matches have been analysed visually. Events were written down as they happened in the video with their timing and then, for each match, the events were compared to the match live commentary text. Table 53 shows the result of this analysis.

Table 53: 5-matches events analysis for miss and catch

	throw-in	foul	kick taken	cross	shot	corner
Match 1	3	0	0	1	1	0
Match 2	2	0	1	0	0	1
Match 3	0	3	0	0	0	0
Match 4	0	2	2	0	0	2
Match 5	2	0	0	1	0	0
Total misses	7	5	3	2	1	3
Average text miss per game	1.4	1	0.6	0.4	0.2	0.6
Events detected in the training corpus	42,927	19,283	37,394	13,944	9,811	12,381
Average text event per game	55.74	25.04	48.56	18.11	12.7	17.15
Average text missing percentage per game	2.45	3.84	1.22	2.16	1.55	3.38
Average text catching Percentage per game	97.55	96.16	98.78	97.84	98.5	96.62

Table 53 shows the main patterns, with high frequency ratio in the corpus, and how many of them the live commentary text missed. This is not to be confused with the system catch and miss which will be evaluated next. This analysis compares the events that are mentioned in the live commentary text and the events that happen in actual football match. *Throw-in* (attacking or defending) event was missed three times in match-1, two times in match-2 and two times in match-5 when the matches' commentary text was checked against the matches' video visual analysis. The table then shows the average text missed per game. From the live commentary analysis, the total number occurrence of each pattern is counted and an average event per game is estimated. The corpus Average missing percentage per game is calculated as:

$$\frac{AverageMissPerGame}{(AverageMissPerGame + AverageEventPerGame)} * 100$$

Equation 6: Corpus average missing percentage per Game

The average event catching per game percentage is 97%. To obtain such a result just from text analysis instead of video analysis is very promising.

4.4.1.5 System Strength - The System versus the Corpus

Now that the live commentary text strength is evaluated, the next step is to evaluate the strength of our system. Considering the most frequent patterns, Table 54 shows the total of each pattern in the corpus and how many of them our system was able to detect.

Table 54: Patterns detected by our system versus patterns in the training corpus

Patterns	Attacking throw-in	Defending throw-in	Inswinging corner	Outswinging corner	Foul by	Kick taken	Goal kick	cross by	shot by
Corpus	25,365	17,573	4,844	2,692	20,186	37,346	14,905	14,963	12,074
System	25,361	17,565	4,820	2,682	19,283	37,334	14,283	13,944	11,811
%	99.98	99.95	99.50	99.63	95.53	99.97	95.83	93.19	97.82

Attacking throw-in pattern occurred 25,365 times in the training corpus. Our system managed to detect 25,361; that is 99.98%. Our system also managed to detect more than 99.50% of the *Defending throw-in, Inswinging corner, Outswinging corner and Kick taken* patterns in the training corpus. Also, our system managed to detect more than 95% of the *shot by, Foul by* and *Goal kick* patterns. *Cross by* pattern detection was about 93%.

4.4.1.6 System Strength - The System versus Football video

Now that our system catching percentage is evaluated with respect to the corpus, and the corpus catching percentage is evaluated with respect to the football video, a conclusion of our system catching percentage with respect to the football video can now be presented. For our system's actual catching percentage to be calculated, the following formula is applied:

$Corpus_{F}$ = *Corpus Catching Percentage with Respect to the Football Video*

$System_{C}$ = *Our System's Catching Percentage with Respect to the Corpus*

$System_{A}$ = *Our System's Actual Catching Percentage with Respect to the Football Video*

$$System_{A} = Corpus_{F} \times System_{C}$$

Equation 7: System actual catching percentage equation

Table 55 shows the conclusion that is drawn from Table 53 and Table 54.

Table 55: Overall system catching percentage

Patterns	Attacking throw-in	Defending throw-in	Inswinging corner	Outswinging corner	Foul by	Kick taken	cross by	shot by
Corpus catching %	97.55	97.55	96.62	96.62	96.16	98.78	97.84	98.5
System catching %	99.98	99.95	99.5	99.63	95.53	99.97	93.19	97.82
System actual catching %	97.53	97.51	96.14	96.26	91.86	98.75	91.18	96.35

The result above shows our system's actual catching percentage with respect to the events in the football video. Our system's actual catching percentage of *Attacking throw-in* event is $(97.55\%, Corpus_V)$ X $(99.98\%, System_C) = 97.53\%$.

The results obtained from are still satisfying. Our system's actual catching percentage of most frequent events is above 90% and in some cases the actual catching percentage is above 95%.

4.4.1.7 Corpus Size

A further area to consider is the corpus size, and a question to be answered is what is the minimum corpus size that is effective? To answer this question two analyses have been performed. For both analyses, a set of corpora were constructed from the training corpus. Corpus of one match live commentary text, corpus of five matches live commentary text, corpus of fifty matches live commentary text, corpus of one hundred matches live commentary text and so on; incrementing the next corpus by fifty matches until the full size corpus is reached. For each corpus, the most frequent tokens list was analysed. To avoid ambiguity, 20 random samples of corpus of one match live commentary text, corpus of five matches live commentary text and corpus of ten matches live commentary text were used. For each of these three corpora, the average tokens frequency were calculated and used as its final tokens frequency list. This is done to avoid any unusual results that may occur. For both analyses, rank correlation analysis was performed. Rank correlation is a measure of the degree of relationship between variables listed in order or rank.

The first analysis is to calculate the rank correlation between each corpus and its previous corpus as the corpus is building up. Table 56 below shows the correlation analysis result with Figure 71 showing its graph.

Table 56: Token-list correlation as the corpus is building up

Corpus size (in match)	Correlation
5	0.007
10	0.023
50	0.530
100	0.883
150	0.973
200	0.986
250	0.994
300	0.993
350	0.990
400	0.996
450	0.999
500	0.999
550	1.000
600	0.999
650	0.999
700	1.000

The rank correlation between corpus-5 and corpus-1 is 0.007 which means their tokens frequent list is very different. As the corpus is building up, the degree of relationship is getting stronger. This can be seen between corpus-100 and corpus-50, corpus-150 and corpus-100 and corpus-200 and corpus-150.

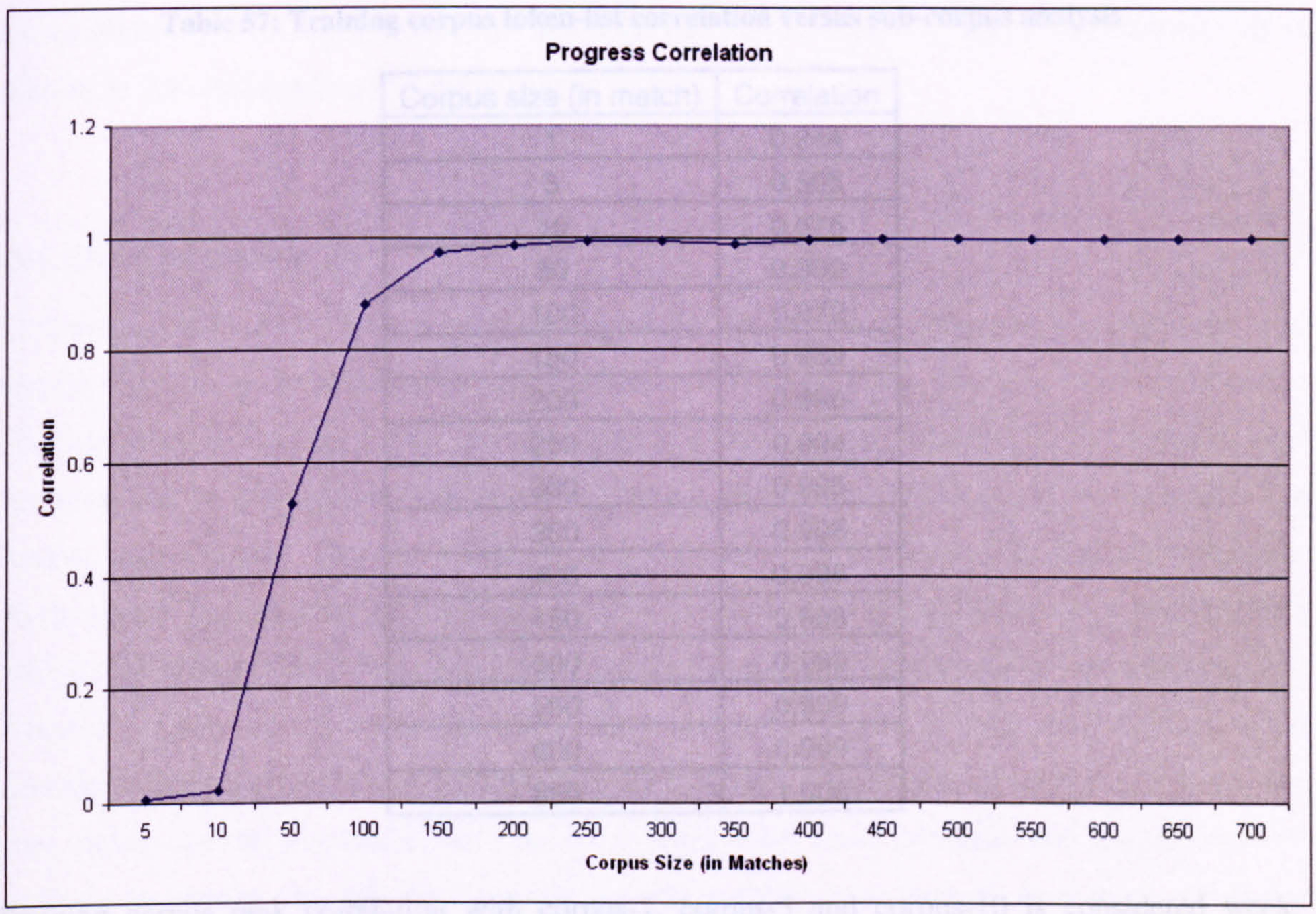


Figure 71: Chart showing token-list correlation as the corpus is building up

It can be seen that once the corpus reaches 100 live commentary text matches, the correlation is at 88% match. Once the corpus size is 150 live commentary text matches, the correlation is stable at 97% and then fluctuates at the 99% level.

For the second analysis, the rank correlation is calculated between the training corpus and each sub-corpus. Table 57 shows the correlation analysis result with Figure 72 showing its graph.

Table 57: Training corpus token-list correlation versus sub-corpus analysis

Corpus size (in match)	Correlation
1	0.244
5	0.505
10	0.675
50	0.899
100	0.972
150	0.989
200	0.996
250	0.994
300	0.995
350	0.996
400	0.998
450	0.998
500	0.999
550	0.999
600	0.999
650	1.000

Training corpus rank correlation with corpus-1, corpus-5 and corpus-10 is considered weak. When the corpus size is 50 matches an acceptable correlation was found (0.899%). Once the corpus size is 200 matches, a consistent correlation was found (0.99%).

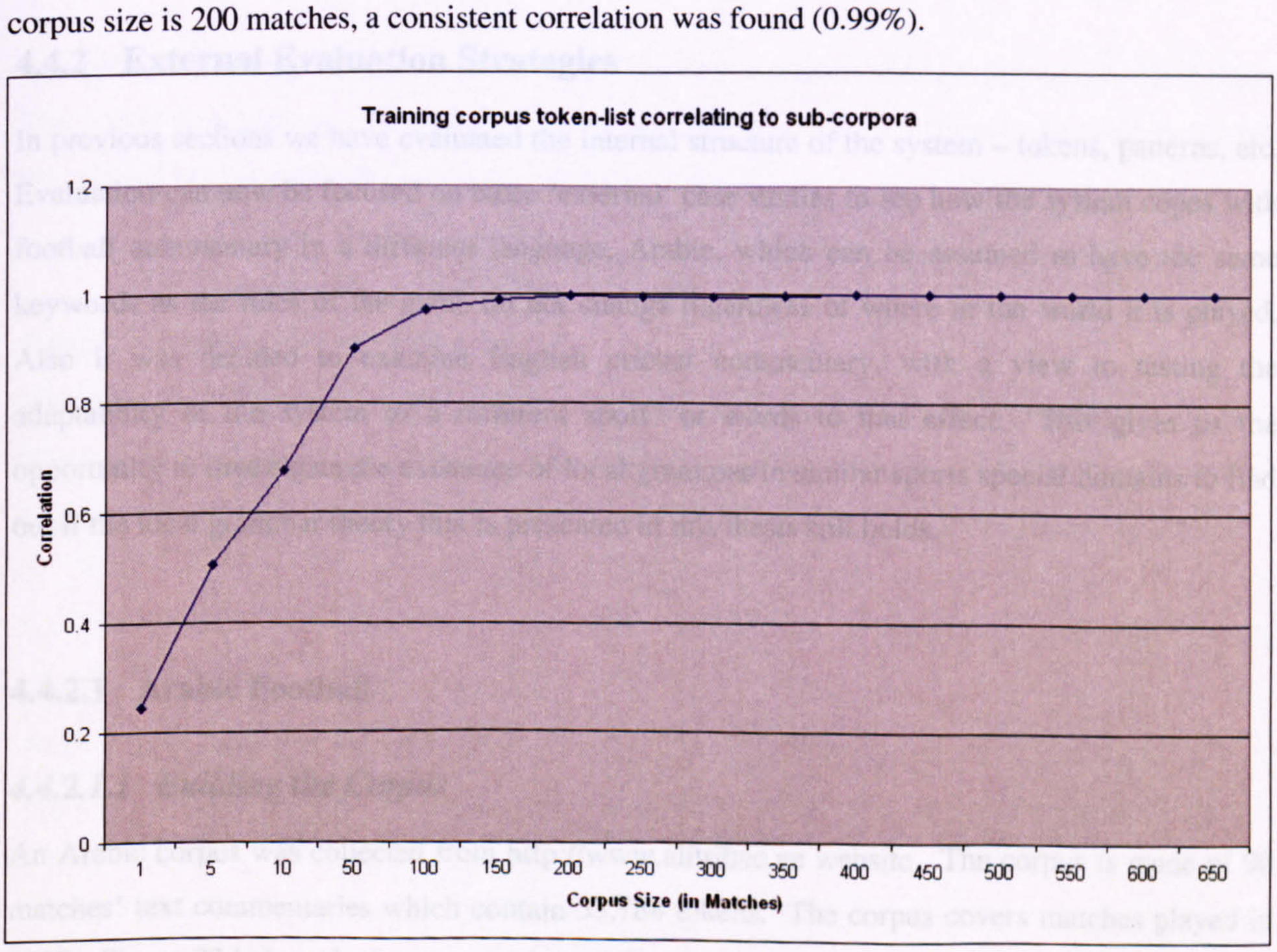


Figure 72: Chart showing training corpus token-list correlation with sub-corpora

From Figure 71 and Figure 72, it can be concluded that the optimal choice for minimal corpus size to be 200 matches live commentary text.

4.4.1.8 Conclusion

Our method is based on the argument that specialist domains have their own local grammar and terminology as we indicated in the Introduction. The comparison that was made between the English football training corpus and the English football testing corpus showed that the local grammar and terminology analysis from the training corpus also existed in the testing corpus and results were a match. This confirmed the structure solidity of the English football sub-language. Furthermore, we analysed the training corpus which was collected in 2004 and testing-2 corpus which was collected in 2006. The tokens' analysis and the collocation analysis confirmed the structure solidity of our special domain. That is, that even after 2 years the English football domain still has the same local grammar and terminology. It is probably therefore safe to assume that, unless drastic new game rules are introduced, such local grammar and terminology are unlikely to change and will consequently provide a stable base through time upon which our system can develop

4.4.2 External Evaluation Strategies

In previous sections we have evaluated the internal structure of the system – tokens, patterns, etc. Evaluation can now be focused on more 'external' case studies to see how the system copes with football commentary in a different language, Arabic, which can be assumed to have the same keywords as the rules of the game do not change regardless of where in the world it is played. Also it was decided to examine English cricket commentary, with a view to testing the adaptability of the system to a different sport" or words to that effect. This gives us the opportunity to investigate the existence of local grammar in similar sports special domains to find out if the local grammar theory that is presented in this thesis still holds.

4.4.2.1 Arabic Football

4.4.2.1.1 Building the Corpus

An Arabic corpus was collected from http://www.alittihad.ae website. The corpus is made of 90 matches' text commentaries which contain 53,784 tokens. The corpus covers matches played in 2007. Figure 73 below shows a screenshot of the commentary of a match between Germany and

Romania. Germany and Romania are highlighted in red (top right) and a translation of some features on the page are written in red (top left).

Figure 73: Screenshot of Arabic football commentary from http://www.alittihad.ae website

The picture caption in the screen shot reads: The German team reached Romania's net 3 times. The article first line says: Germany 2nd team beat Romania 3-1 in a friendly football match that was played in Cologne with 44,000 attendances.

4.4.2.1.2 Corpus Pre-Analysis

The first step to be taken before analysing the corpus is to segment the corpus sentences. However, the Arabic language differs from English. Sentences can be joined with conjunctions. Also, the time-stamp in the Arabic articles is more challenging to extract, see Figure 74 below.

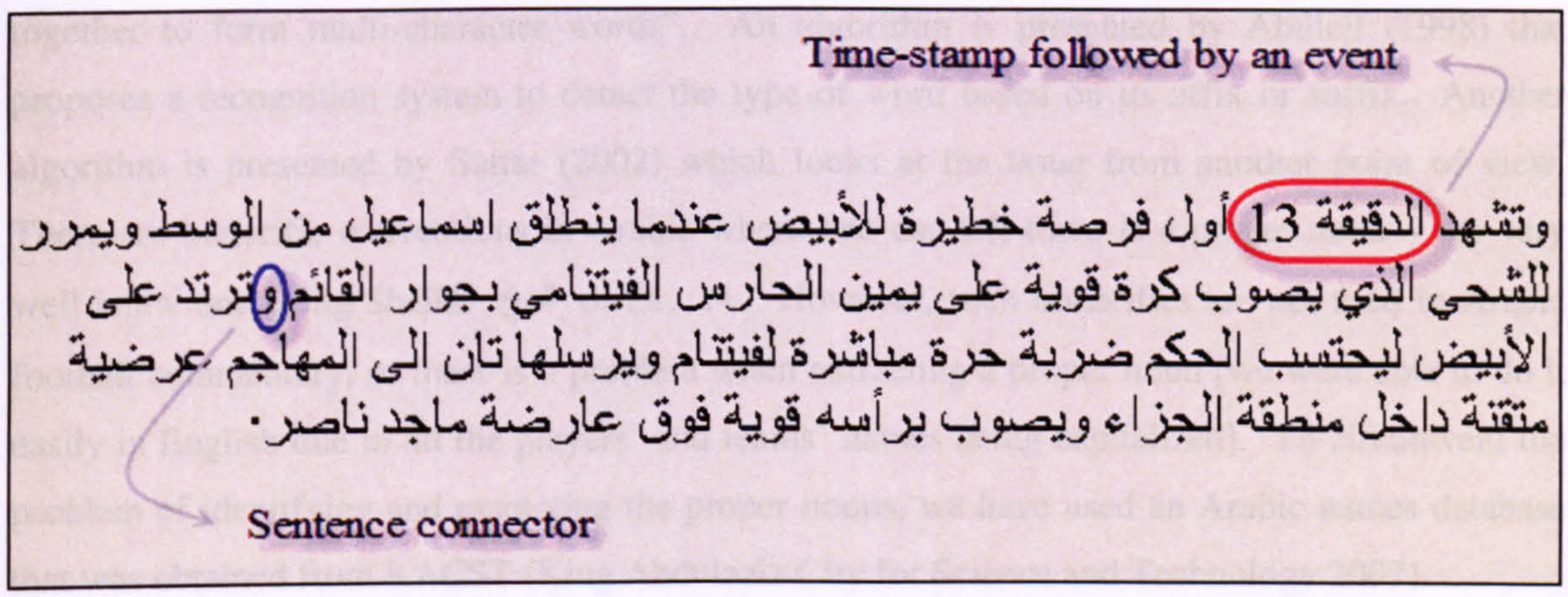


Figure 74: Sample of Arabic football commentary

Using the same method we applied in the English football commentary and subsequently in the cricket commentary will only work on part of the corpus. Furthermore, Table 58 shows the available tools for Arabic and English languages.

Table 58: Generally available tools for Arabic language text analysis

Generally Available Tools	Arabic	English
Sentence segmentation	X	√
Proper noun detection	X	√
Frequency analysis	√	√
Weirdness analysis	√	√
Collocation	√	√
Local Grammar	X	√

Sentence segmentation is a topic that is still being investigated quite extensively (see Madnani 2007; Sattar 2002). However, some statistically-based programs appear to perform reasonably well and can identify the end of sentences. In terms of sentence boundaries it is worth pointing out the role conjunctions play in the Arabic language, in particular the role of letter و when used as a part of a word or when used as a part of a sentence. For instance, the time in Arabic is وقت where و is actually part of the word itself. However, it could also be used to make a conjunction between two sentences as shown in Figure 74. Another challenge in Arabic is the lack of capitalization which helps in the detection, especially in English, of proper nouns. Madnani (2007: 10) states that with "Languages such as Chinese and Arabic, the task [Tokenization] is more difficult since there are no explicit boundaries. Furthermore, almost all characters in such non-segmented languages can exist as one-character words by themselves, and can also join

together to form multi-character words". An algorithm is presented by Abuleil (1998) that proposes a recognition system to detect the type of word based on its affix or suffix. Another algorithm is presented by Sattar (2002) which looks at the issue from another point of view. There are honorific conventions in Arabic where one can tell there is a proper noun – the very well know one being Sheikh شيخ or Dr. د. However, such honorifics are not used in Arabic football commentary, so there is a problem when extracting a proper noun [we were able to do it easily in English due to all the players' and teams' names being capitalized]. To circumvent the problem of identifying and extracting the proper nouns, we have used an Arabic names database that was obtained from KACST (King Abdulaziz City for Science and Technology 2007).

4.4.2.1.3 Vocabulary and Collocation Analysis

We performed a frequency analysis of the 53,784 tokens corpus of Arabic football commentary and the results roughly match what we found in the English football commentaries, see Table 59 below.

Table 59: Arabic football commentary text tokens analysis based on frequency (N ≅ 53,784)

Token	Gloss	Freq	Freq Z-score	Weirdness	Weird Z-score
في	in	1,780	54.71	1	-0.27
من	from	1,537	47.21	1	-0.27
على	on	830	25.41	1	-0.27
الدقيقة	the minute	503	15.33	190	1.34
اللذي	which	418	12.71	1	-0.27
بعد	after	401	12.18	2	-0.26
المباراة	the game	377	11.44	238	1.75
كرة	ball	369	11.2	239	1.76
الى	to	342	10.36	1	-0.27
الشوط	the half	339	10.27	889	7.29
الكرة	the ball	321	9.72	134	0.87
علي	Ali	292	8.82	3	-0.25
الوحدة	Alwehdah	286	8.64	40	0.07
محمد	Mohammed	267	8.05	4	-0.24
عن	about	263	7.93	1	-0.27
إلى	to	262	7.9	0	-0.27
الثاني	the second	252	7.59	7	-0.21
عندما	when	243	7.31	4	-0.24
أن	that	238	7.16	0	-0.27
الهدف	the goal	228	6.85	22	-0.09

For example, there are high frequency words like فى (in) , اللذى (which) , عندما (when) , المباراة (the match) , There are open class words, especially proper nouns like the very common محمد (Mohammed) , على (Ali). Also there is a team name الوحده (Al-Wehdah) which also appears among the top words. The list also shows some other nouns like الكره (the ball) and numerous others like الدقيقه (minute) and الهدف (the goal). There are also some closed class words that can't be as easily categorized as in English.

One could have looked for the word *footed* in the Arabic language corpus but it is not one of the most frequent, in fact, it only shows in one instance. However, when we look at the selective words according to the criteria we outline in the English language corpus analysis for football for specialist text, we select the words with positive z-score for frequency and weirdness and those words are the candidate terms of the domain, see Table 60 below.

Table 60: Arabic football commentary text tokens analysis based on weirdness z-score (N ≅ 53,784)

Token	Gloss	Freq	Freq Z-score	Weirdness	Weird Z-score
ركنية	corner	84	2.41	4186	35.33
تمريرة	kick	58	1.61	2891	24.31
المرمى	the goal	198	5.92	2467	20.71
العنابي	the red	47	1.27	2342	19.64
الوحداوي	Alwehdawi	42	1.11	2093	17.53
الوهيبي	Alweiheibi	41	1.08	2043	17.1
محرزا	scoring	38	0.99	1894	15.83
المنهالي	Almenhali	33	0.83	1645	13.72
سددها	shot	33	0.83	1645	13.72
موريتو	Moreno	33	0.83	1645	13.72
"	"	32	0.8	1595	13.29
بستروفيتش	Pestrovetch	32	0.8	1595	13.29
عرضية	corss	123	3.61	1533	12.76
ميتروفيتش	Metrovetch	30	0.74	1495	12.44
بالمرمى	in the goal	26	0.62	1296	10.75
فرهاد	Ferhad	26	0.62	1296	10.75
وسدد	shot	50	1.36	1246	10.32
هوار	Haowar	24	0.56	1196	9.9
يلعبها	played	24	0.56	1196	9.9
النوبي	Alnoobi	23	0.53	1146	9.47

When we look at those words, we discover about five keywords. The first keyword is ركنيه which means corner. Incidentally the frequency z-score and weirdness z-score for corner in the

English football commentary corpus are (3.4 , 0). Among these five keywords is the word تمريره (kick) which is quite an interesting one beside being the second highest. Also, it is one of the marginally ambiguous words which is equivalent to *kick* in the English language. We will look at the variations in the use of word تمريرة . In Arabic the inflection of words does not happen at the end of the word. In English, for instance, the word *kick* which starts as noun, can be used as a verb (to kick) and the inflection on the word could be kicks, kicked, kicking and so forth. One can find various patterns of تمريرة (kick) which is very common as Table 61 shows that تمريره infixed, postfixed, prefixed and other operations that are not common in English.

Table 61: The variation of the word تمريرة in Arabic

Sentence			Variation	Detected in Corpus
Kick			تمريرة	
2 kicks			تمريرتان	
3+ kicks			تمريرات	
The	kick		التمريرة	√
He	kicked	the ball	مرر	√
He	kicks	the bal	يمرر	√
She	kicked	the ball	مررت	X
She	kicks	the ball	تمرر	X
They (2)	kicked	the ball	مررا	X
They (2)	kick	the ball	يمرران	X
They (3+)	kicked	the ball	مرروا	√
They (3+)	kick	the ball	يمروون	√

For instance, in 'he kicked the ball and she kicked the ball', the word مرر is changed to agree with the feminine pronoun. Also, in 'he kicks the ball and she kicks the ball', the word يمرر is changed (prefixing) يمرر for 'he kicks the ball' andتمرر for 'she kicks the ball'.

Following our method that we select the most frequent candidate terms and their collocation patterns, it is very clear that the collocation patterns of تمريرة are quite frequent. The key

collocate of تمريره is proper noun which comes before it. The proceeding words are عرضيه (cross) and سحريه (great), see Table 62 below.

Table 62: "kick" تمريرة collocation in Arabic football commentary (N ≅ 53,784 and تمريرة freq = 58)

Step	-2	-1	Keyword	1	2	3	*f*	*f* / N	U-Score
1		pn	تمريرة				19	32.76	23.8
1			تمريرة	عرضية			15	25.86	20.25
1			تمريرة	سحرية			7	12.07	4.41
2	أرسل	pn	تمريرة				13	22.41	1
2	أرسل	pn	تمريرة	عرضية			5	8.62	1
2	أرسل	pn	تمريرة	سحرية			8	13.79	1
3	أرسل	pn	تمريرة	عرضية	داخل		5	8.62	1
3	أرسل	pn	تمريرة	سحرية	داخل		8	13.79	1
4	أرسل	pn	تمريرة	عرضية	داخل	منطقة	5	8.62	1
4	أرسل	pn	تمريرة	سحرية	داخل	منطقة	8	13.79	1

4.4.2.1.4 Evaluation

A smaller live commentary corpus, containing 7,649 tokens, was gathered from football matches that were played in the year 2007. First, tokens frequency was analysed and compared with the tokens frequency from the cricket training corpus (see Table 63 below).

Table 63: Token frequency comparison in Arabic corpora

(Training = 53,784 tokens, Testing = 7,849 tokens)

	Testing Corpus		Training Corpus
Token	Freq	Rel. Freq.	Rel. Freq.
في	253	3.22	3.31
من	200	2.55	2.86
على	135	1.72	1.54
الدقيقة	82	1.04	0.94
الشوط	57	0.73	0.63
بعد	57	0.73	0.75
العين	56	0.71	0.39
المباراة	55	0.70	0.70
الثاني	54	0.69	0.47
كرة	54	0.69	0.69
الذي	51	0.65	0.78
إلى	51	0.65	0.49
الكرة	46	0.59	0.60
الى	44	0.56	0.64
عندما	37	0.47	0.45
مع	37	0.47	0.35
حيث	36	0.46	0.30

Table 42 shows the distribution of the words and we see the first 3 words have the same rank, and because of the small corpus we see more of the proper nouns ranked much higher. Table 64 below shows the result of Mann-Whitney U-test.

Table 64: Mann-Whitney U-test result for token frequency analysis between (Arabic football) training and testing commentary corpora

Variable	Value
Mann-Whitney U	126
p-Value	0.523

The p-value is 0.523 which indicates there is insignificant difference between the Arabic commentary training and testing corpora.

Figure 75 below shows the plotted chart of the tokens frequency from both corpora.

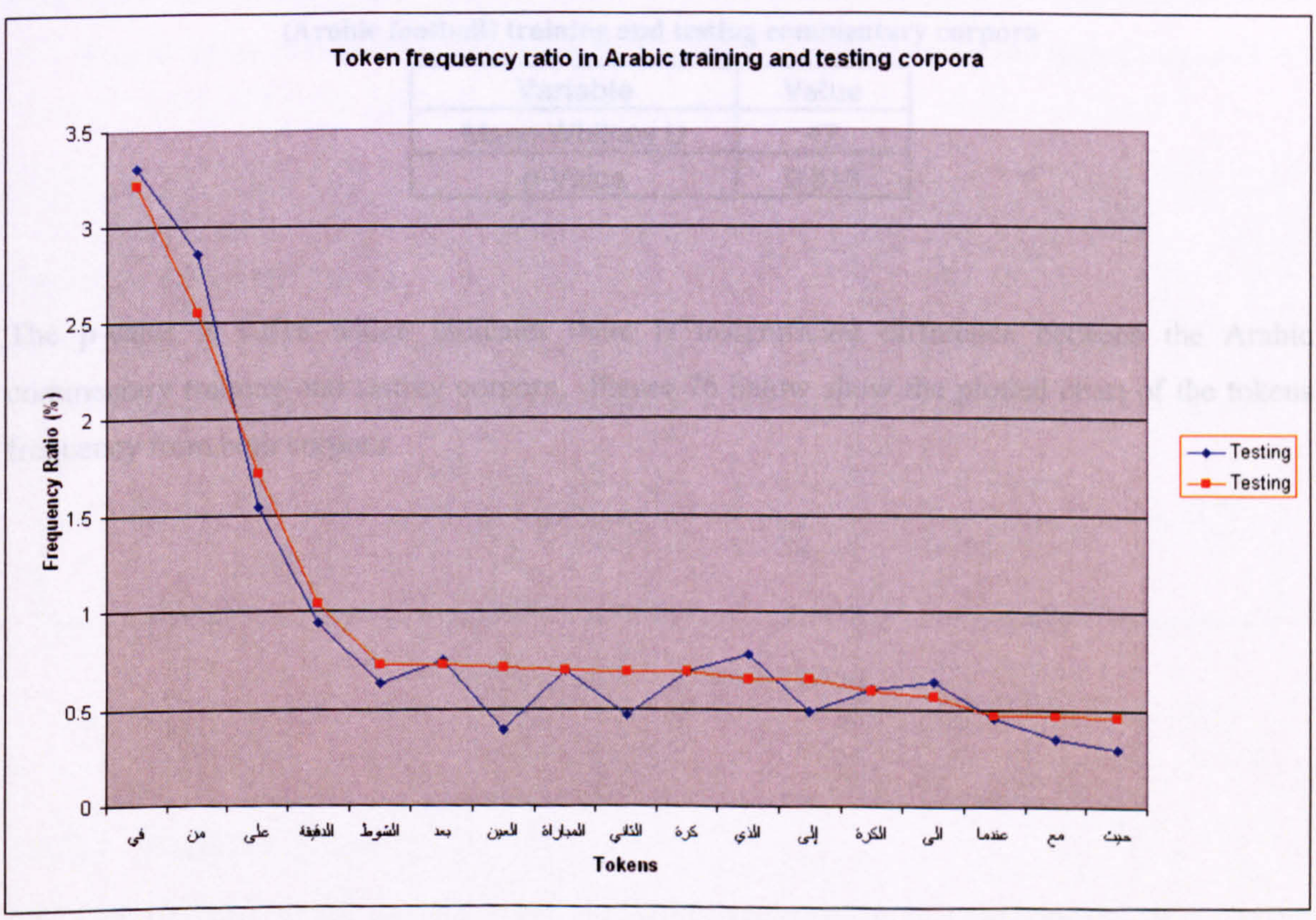


Figure 75: Tokens frequency ratio in (Arabic) commentary training and testing corpora chart

Table 65 below show the "kick" (تمريرة) frequency comparison.

Table 65: "kick" (تمريرة) frequency comparison chart (تمريرة freq : training = 58, testing = 9)

	Testing Corpus								Training Corpus
Step	-2	-1	Keyword	1	2	3	Freq	Freq. Ratio	Freq. Ratio
1		pn	تمريرة				4	44.44	32.76
1			تمريرة	عرضية			2	22.22	25.86
1			تمريرة	سحرية			1	11.11	12.07
2	أرسل	pn	تمريرة				2	22.22	22.41
2	أرسل	pn	تمريرة	عرضية			2	22.22	8.62
2	أرسل	pn	تمريرة	سحرية			1	11.11	13.79
3	أرسل	pn	تمريرة	عرضية	داخل		2	22.22	8.62
3	أرسل	pn	تمريرة	سحرية	داخل		1	11.11	13.79
4	أرسل	pn	تمريرة	عرضية	داخل	منطقة	1	11.11	8.62
4	أرسل	pn	تمريرة	سحرية	داخل	منطقة	1	11.11	13.79

Table 66 below shows the result of Mann-Whitney U-test.

Table 66: Mann-Whitney U-test result for تمريرة collocation frequency analysis between (Arabic football) training and testing commentary corpora

Variable	Value
Mann-Whitney U	47
p-Value	0.818

The p-value is 0.818 which indicates there is insignificant difference between the Arabic commentary training and testing corpora. Figure 76 below show the plotted chart of the tokens frequency from both corpora.

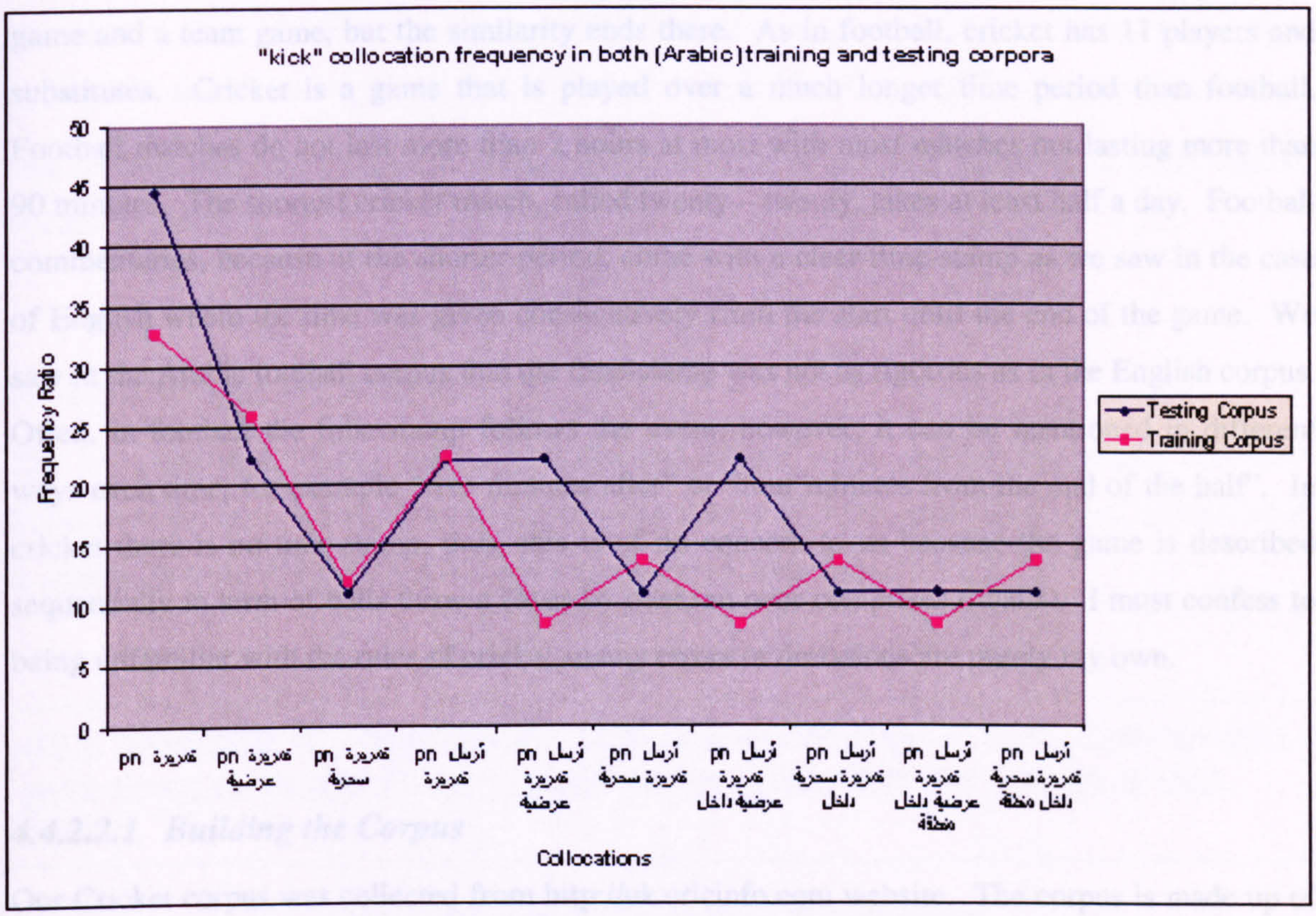


Figure 76: تمريره collocation frequency ratio in both (Arabic) training and testing corpora

4.4.2.1.5 Local Grammar

From previous analysis, the proposed local grammar for تمريره would be as follow:

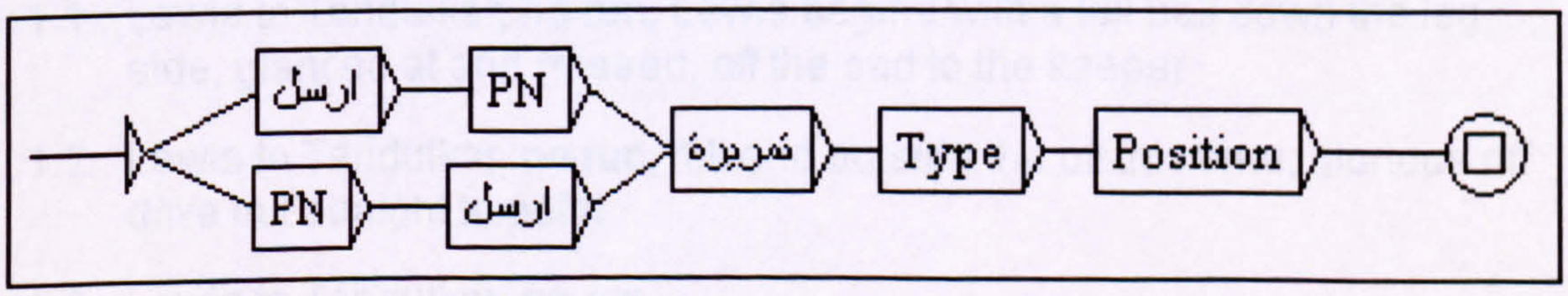


Figure 77: تمريرة local grammar

4.4.2.2 English Cricket

We chose another sport at random to test the finding of our claim that each special language has its own special grammar, in addition, it has its own vocabulary that has key patterns which are categorized by frequency which can be used subsequently to mount queries and which can be of use as a utility for the sport. In this sense we chose the game cricket which, like football, is a ball

game and a team game, but the similarity ends there. As in football, cricket has 11 players and substitutes. Cricket is a game that is played over a much longer time period than football. Football matches do not last more than 2 hours at most with most matches not lasting more than 90 minutes. The shortest cricket match, called twenty – twenty, takes at least half a day. Football commentaries, because of the shorter period, come with a clear time-stamp as we saw in the case of English where the time was given consecutively from the start until the end of the game. We saw in the Arabic football corpus that the time-stamp was not as rigorous as in the English corpus. Often, in football the time-stamp follows the event, however, it can be mentioned in different ways each time, for example "five minutes after" or "four minutes from the end of the half". In cricket there is no time-stamp, thus, this is of no concern to us because the game is described sequentially in term of balls thrown (over by over, an over comprises 6 balls). I must confess to being unfamiliar with the rules of cricket so any errors or omissions are purely my own.

4.4.2.2.1 Building the Corpus

Our Cricket corpus was collected from http://uk.cricinfo.com website. The corpus is made up of 300 matches' text commentaries which contain 4,337,772 tokens. The corpus covers matches played in 2004, 2005 and 2006. Figure 78 below shows a screenshot of this commentary

End of over 1 (maiden) - India 0/0

SC Ganguly 0* (6b) **JM Anderson** 1-1-0-0
SR Tendulkar 0* (0b)

1.1 Lewis to Tendulkar, no run, Lewis begins with a full ball down the leg side, glanced at and missed, off the pad to the keeper

1.2 Lewis to Tendulkar, no run, full and outside the off this time, glorious off drive but straight to point

1.3 Lewis to Tendulkar, no run

1.4 Lewis to Tendulkar, no run, fuller and on the pads this time, Tendulkar walks across and dabs the ball to the leg side, still no runs on the board

1.5 Lewis to Tendulkar, no run, again fullish and on the off, Tendulkar comes forward to play a stroke, the ball goes off a thickish inside edge to the leg side

1.6 Lewis to Tendulkar, no run, very full and quite straight, Tendulkar drives with a straight bat and finds mid-on for the second maiden over on the trot

Figure 78: Screenshot of cricket commentary from content-uk.cricinfo.com site

Note that the cricket commentary does not use a time-stamp, instead the commentary is marked by the over and the inning (i.e. '1.1' = 1st over, 1st ball, '1.2' = 1st over, 2nd ball and so on).

The method that was applied in analyzing the football corpus was followed in analyzing the cricket corpus.

4.4.2.2.2 Corpus Pre-Analysis

The first step to be taken before analysing the corpus is to segment the corpus sentences. Sometimes the live commentary stamps multiple events with the same time. Also, globalization is applied as well. For example Figure 79 shows three sentences.

30.2 Tendulkar to Collingwood. That is a smoking good shot. Collingwood gets down on one knee and launches Tendulkar into the stands at mid-wicket.

Figure 79: Sample of the cricket corpus multi-event

Our system will convert Figure 79 to Figure 80 where each sentence starts a new line.

30.2 PN to PN.
That is a smoking good shot.
PN gets down on one knee and launches PN into the stands at midwicket.

Figure 80: Cricket corpus multi-event separated

4.4.2.2.3 Vocabulary and Collocation Analysis

Tokens frequency analysis was performed using System Quirk. Table 67 shows the 25 most frequent tokens of the original corpus.

Table 67: QUIRK Frequency analysis for cricket Commentary Corpus (N ≅ 4,337,772)

Rank	Token	*f*	*f* / N	Freq. z-score
1	nomber	577,582	13.32	-999.90
2	to	262,744	6.06	18.03
3	,	248,026	5.72	-999.90
4	.	196,180	4.52	-999.90
5	run	150,450	3.47	10.30
6	no	122,659	2.83	8.38
7	)	122,389	2.82	-999.90
8	(	122,386	2.82	-999.90
9	the	113,372	2.61	-999.90
10	-	111,518	2.57	-999.90
11	*	58,378	1.35	-999.90
12	off	53,849	1.24	3.65
13	of	47,661	1.1	-999.90
14	runs	44,919	1.04	3.03
15	and	44,577	1.03	-999.90
16	on	38,570	0.89	-999.90
17	/	35,309	0.81	-999.90
18	over	32,762	0.76	2.20
19	a	28,044	0.65	-999.90
20	leg	25,013	0.58	1.66
21	it	24,648	0.57	-999.90
22	outside	24,278	0.56	1.61
23	ball	22,844	0.53	1.52
24	back	20,843	0.48	1.38
25	short	19,291	0.44	1.27

Table 68 below shows the most frequent open class tokens after filtering out both the closed class words and the words with low frequency z-score.

Table 68: Most Frequent Open Class Words in cricket Commentary Corpus (N ≅ 4,337,772)

Rank	Token	*f*	*f* / N	Freq. z-score
1	run	150,450	3.47	10.30
2	runs	44,919	1.04	3.03
3	over	32,762	0.76	2.20
4	leg	25,013	0.58	1.66
5	outside	24,278	0.56	1.61
6	ball	22,844	0.53	1.52
7	back	20,843	0.48	1.38
8	short	19,291	0.44	1.27
9	mid	18,763	0.43	1.23

Run and *runs* are chosen to perform the collocations as *taken* was chosen in the football live commentary collocation analysis. Table 69 below shows *runs* initial four collocations. At this point globalization is applied; NM stands for number and PN stands for Proper Noun.

Table 69: *runs* collocation (N ≅ 4,337,772)

Step	-1	keyword	*f*	U-Score	K-Score	Strength
1	nomber	runs	40,106	132,242,539	12,122	41.33
2	nomber	nomber runs	19,760	48,913,237	7,372	39.33
2	pn	nomber runs	8,511	52,191,782	7,615	30.89
3	over	nomber nomber runs	19,492	34,097,331	6,155	9.78
3	to	pn nomber runs	7,627	4,845,175	2,320	5.77
4	pn	to pn nomber runs	7,612	4,623,940	2,284	6.39

Figure 81 below shows the initial pattern for *runs* and let this pattern be identified as *Runs-1*

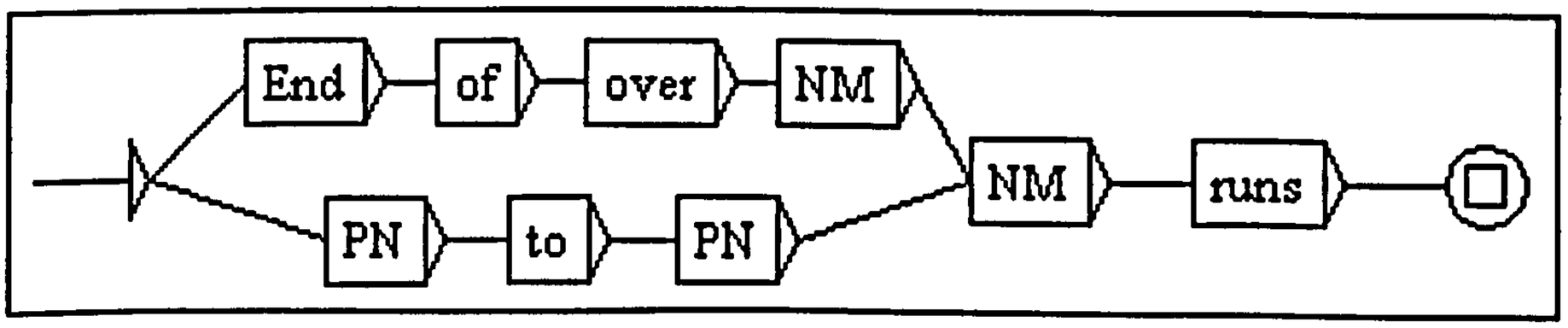


Figure 81: *runs* initial pattern

Table 70 below shows *run* initial four collocations

Table 70: *run* collocation (N ≅ 4,337,772)

Step	-1	keyword	*f*	U-Score	K-Score	Strength
1	no	run	91,042	377,102,018	24,187	47.82
1	nomber	run	34,489	142,468,064	12,582	12.62
2	pn	no run	90,826	192,472,389	21,472	31.49
2	pn	nomber run	28,339	96,050,460	10,331	37.16
2	nomber	nomber run	4,400	40,044,149	6,670	17.35
3	to	pn no run	90,743	163,812,089	20,983	28.18
3	to	pn nomber run	25,219	54,133,909	7,756	25.54
3	over	nomber nomber run	4,398	1,720,574	1,383	4.08
4	pn	to pn no run	90,718	152,372,183	16,354	25.03
4	pn	to pn nomber run	25,194	52,687,792	7,651	33.72

Figure 82 below shows the initial patterns for *run* and let this pattern be identified as *Run-1*

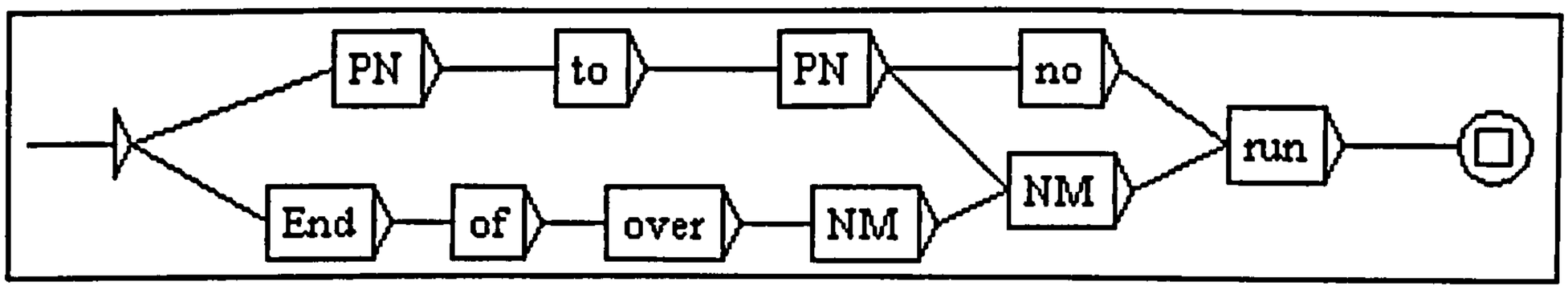


Figure 82: *run* initial patterns

4.4.2.2.4 Local Grammar

Run and *runs* collocations are taken a step further. Using Run-1 and Runs-1 sub-patterns a further collocation was performed. Table 71 shows most of the *runs* complete patterns.

Table 71: *Runs-1* phrase collocation

Key Token	Key Collocate	Right Collocate	Right Phrase(s) Collocate	U-Score
Runs-1 *(f =27,104)*	PN *(f=19,489)*	#	# NM runs required from NM overs	52,435,493
	short *(f=678)*	and	a bit of width outside the off just a touch wide outside off PN outs wide	40,182
		of a length	and on the off and outside the off outside leg stump outside off on the middle and leg on the stumps around the off stump line around the off stump outside the off stump	
	full *(f=567)*	and	on the middle on the stumps outside the off	28,031
		on	leg stump off stumps the stumps	
		delivery		
	on *(f=321)*	a length	and just outside off stumps around the off and middle around the off stump line	15,675
		the	leg and middle middle and leg pads stumps	
	good *(f=181)*	length	on middle and leg on middle and off outside off stump	2,911
	tossed *(f=177)*	up	and on the stumps and outside the off on leg stumps on the middle on middle and leg outside leg stump outside off stump	2,828

Table 72 below shows most of the *run* complete patterns

Table 72: *Run-1* phrase collocation

Key Token	Key Collocate	Right Collocate	Right Phrase(s) Collocate	U-Score
Run-1 *(f = 125,531)*	(none) *(f = 115,912)*			161,656,137
	PN *(f = 2,006)*	NM	NM delivery NM on leg stump NM on middle and leg NM on off stump NM on the stump NM outside leg stump NM outside off stumps NM on a length NM on middle	140,541,375
	good *(f =1,641)*	delivery finish		2,490,263
		length	delivery ball again on leg stump on middle on off stump on off stumps	
	short *(f =1,360)*	and	wide	3,685,384
		of	a driving length a good length a length	
	back *(f =1,300)*	of	a length again a length and on the off a length and outside the off	1,505,473
	full *(f =1,220)* fuller *(f =659)*	and	on the legs on the middle and leg on the off and middle on the off on the pads on the stumps outside the off stump straight	2,948,115
	tossed *(f =887)*	on	leg middle and leg middle	386,523
		up	on leg stump on middle and leg on off stumps outside off stump	
	fullish *(f =486)*	and	on the off on the stumps outside the off	59,091

From Table 71 and

Table 72 a general local grammar for *run* and *runs* can be drawn, see Figure 83 below.

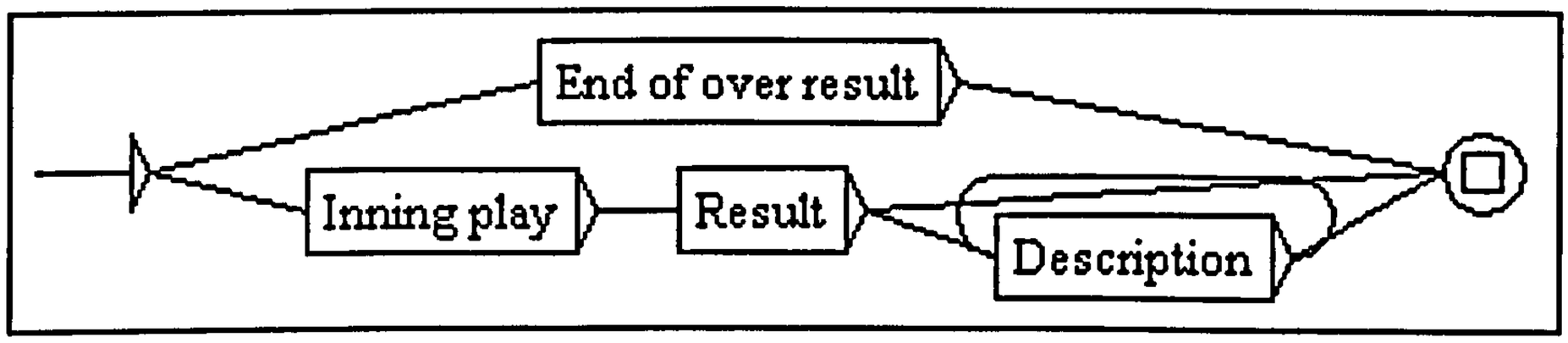


Figure 83: *run* and *runs* local grammar

4.4.2.2.5 Evaluation

A smaller live commentary corpus, containing 213,899 tokens, was gathered from football matches that were played in the year 2007. First, tokens frequency was analysed and compared with the tokens frequency from the cricket training corpus (see Table 73 below)

Table 73: Token frequency comparison between (English Cricket)

(Training = 4,337,772 tokens and Testing = 213,899 tokens)

Token	*f*	Testing Freq. Ratio	Training Freq. Ratio
nomber	24,660	11.53	13.32
,	14,241	6.66	5.72
to	11,131	5.2	6.06
.	10,438	4.88	4.52
the	8,698	4.07	2.61
-	6,792	3.18	2.57
and	5,809	2.72	1.03
run	5,371	2.51	3.47
no	4,192	1.96	2.83
(	3,660	1.71	2.82
)	3,659	1.71	2.82
a	3,400	1.59	0.65
off	3,399	1.59	1.24
on	3,204	1.5	0.89
of	2,797	1.31	1.1
it	2,419	1.13	0.57
*	2,018	0.94	1.35
leg	1,693	0.79	0.58
outside	1,636	0.76	0.56
over	1,554	0.73	0.76
runs	1,442	0.67	1.04

Figure 84 below shows the comparison result from Table 52 above.

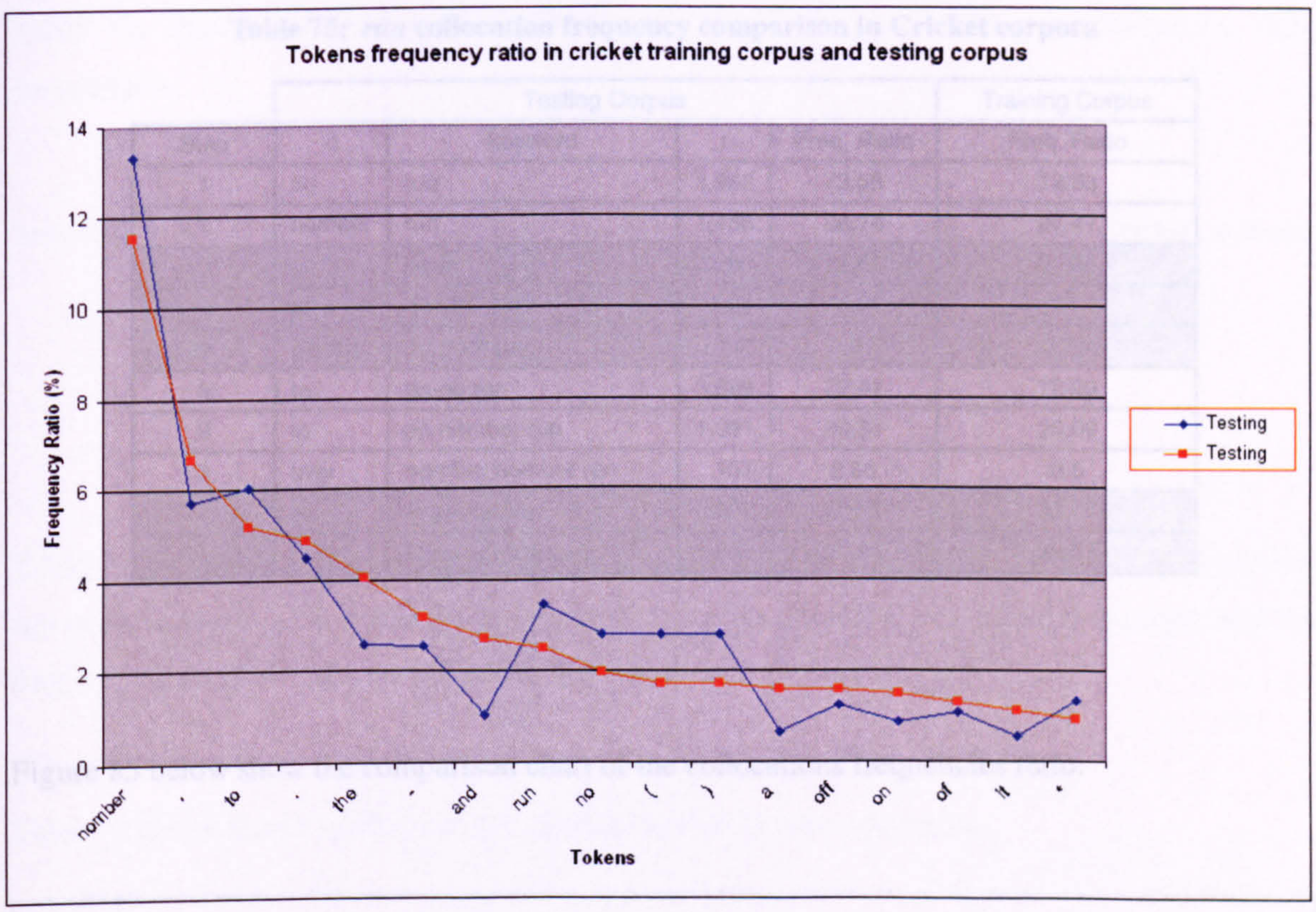


Figure 84: Tokens frequency comparison (Cricket) training and testing corpora chart

Table 74 below shows the result of

Table 74: Mann-Whitney U-test result for tokens analysis between (English-Cricket) training and testing commentary corpora

Variable	Value
Mann-Whitney U	199.5
p-Value	0.597

Table 74 shows the p-Value = 0.597 which means there is insignificant difference between the tokens frequency ratio in the English cricket live commentary text training corpus and the English cricket live commentary text testing corpus

Table 75: *run* collocation frequency comparison in Cricket corpora

	Testing Corpus				Training Corpus
Step	-1	keyword	*f*	Freq. Ratio	Freq. Ratio
1	no	run	3,952	73.58	72.53
1	nomber	run	1,436	26.74	27.47
2	pn	no run	3,889	72.41	72.35
2	pn	nomber run	1,091	20.31	22.58
2	nomber	nomber run	140	2.61	3.51
3	to	pn no run	3,889	72.41	72.29
3	to	pn nomber run	1,091	20.31	20.09
3	over	nomber nomber run	137	2.55	3.5
4	pn	to pn no run	3,889	72.41	72.27
4	pn	to pn nomber run	1,091	20.31	20.07

Figure 85 below show the comparison chart of the collocations frequencies ratio.

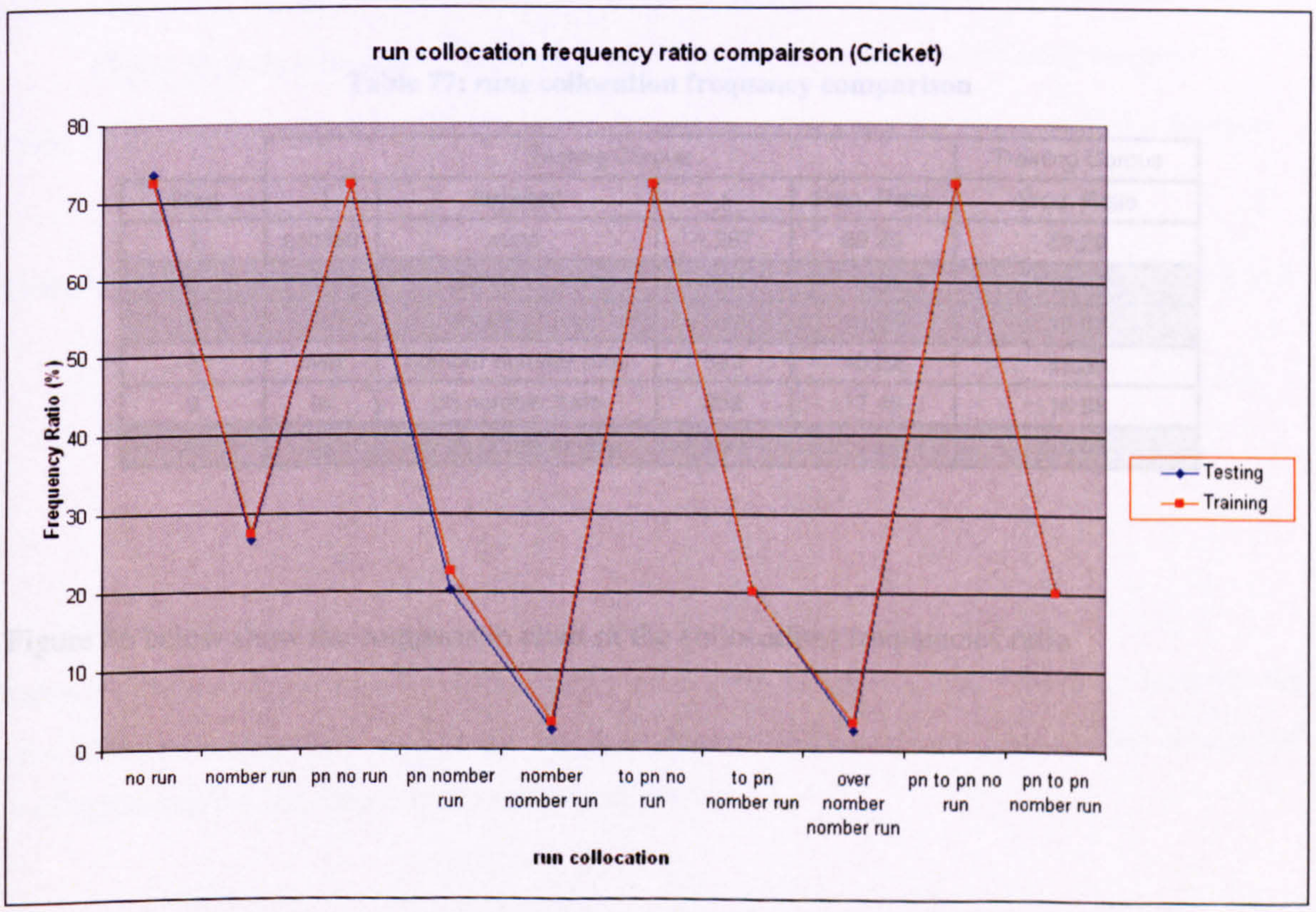


Figure 85: : *run* collocation frequency ratio (Cricket) comparison chart

Table 76 shows the result of Mann-Whitney U-test for run collocation frequency ratio comparison.

Table 76: Mann-Whitney U-test result for *run* collocation analysis between (English-Cricket) training and testing commentary corpora

Variable	Value
Mann-Whitney U	46
p-Value	0.761

Table 76 shows the p-Value = 0.761 which means there is insignificant difference between the *run* collocation frequency ratio in the English cricket live commentary text training corpus and the English cricket live commentary text testing corpus.

Table 77 below shows the frequency ratio comparison of *runs* collocation.

Table 77: *runs* collocation frequency comparison

	Testing Corpus				Training Corpus
Step	-1	keyword	*f*	Freq. Ratio	Freq. Ratio
1	nomber	runs	1,287	89.25	89.29
2	nomber	nomber runs	589	40.85	43.99
2	pn	nomber runs	263	18.24	18.95
3	over	nomber nomber runs	580	40.22	43.39
3	to	pn nomber runs	252	17.48	16.98
4	pn	to pn nomber runs	250	17.34	16.95

Figure 86 below show the comparison chart of the collocations frequencies ratio

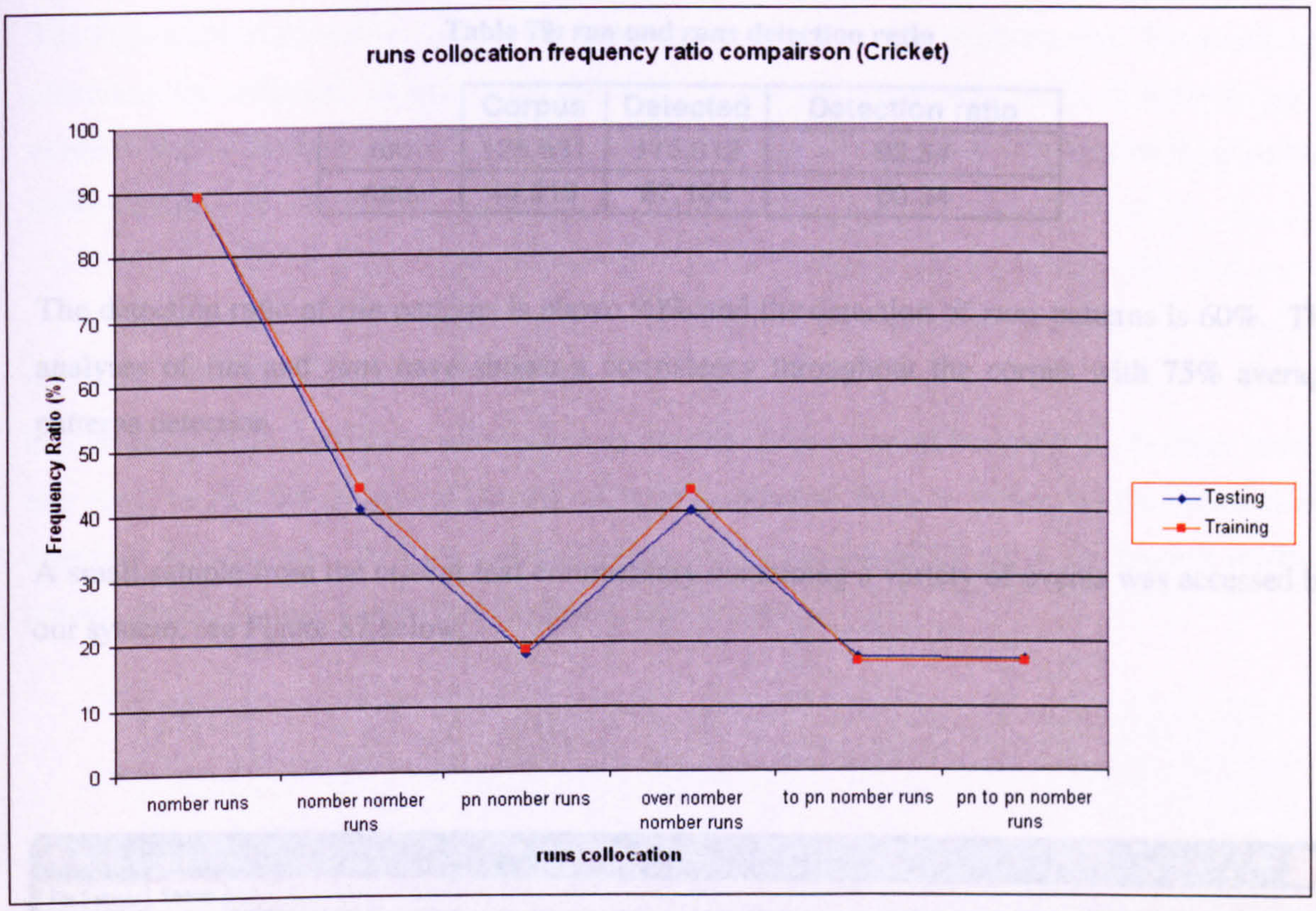


Figure 86: *runs* collocation frequency ratio chart (cricket training and testing corpora)

Table 78 shows the Mann-Whitney U-test result for runs collocation frequency ratio comparison

Table 78: Mann-Whitney U-test result for *runs* collocation analysis in (English-Cricket) training and testing commentary corpora

Variable	Value
Mann-Whitney U	17
p-Value	0.872

Table 78 shows the p-Value = 0.872 which means there is insignificant difference between the *runs* collocation frequency ratio in the English cricket live commentary text training corpus and the English cricket live commentary text testing corpus.

Table 79 below shows the strength of *run* and *runs* patterns detection using the method that our system applied

Table 79: *run* and *runs* detection ratio

	Corpus	Detected	Detection ratio
run	125,531	115,912	92.34
runs	44,919	27,104	60.34

The detection ratio of *run* patterns is above 90% and the detection of *runs* patterns is 60%. The analyses of *run* and *runs* have shown a consistency throughout the corpus with 75% average patterns detection.

A small sample from the cricket text commentary containing a variety of events was accessed by our system, see Figure 87 below.

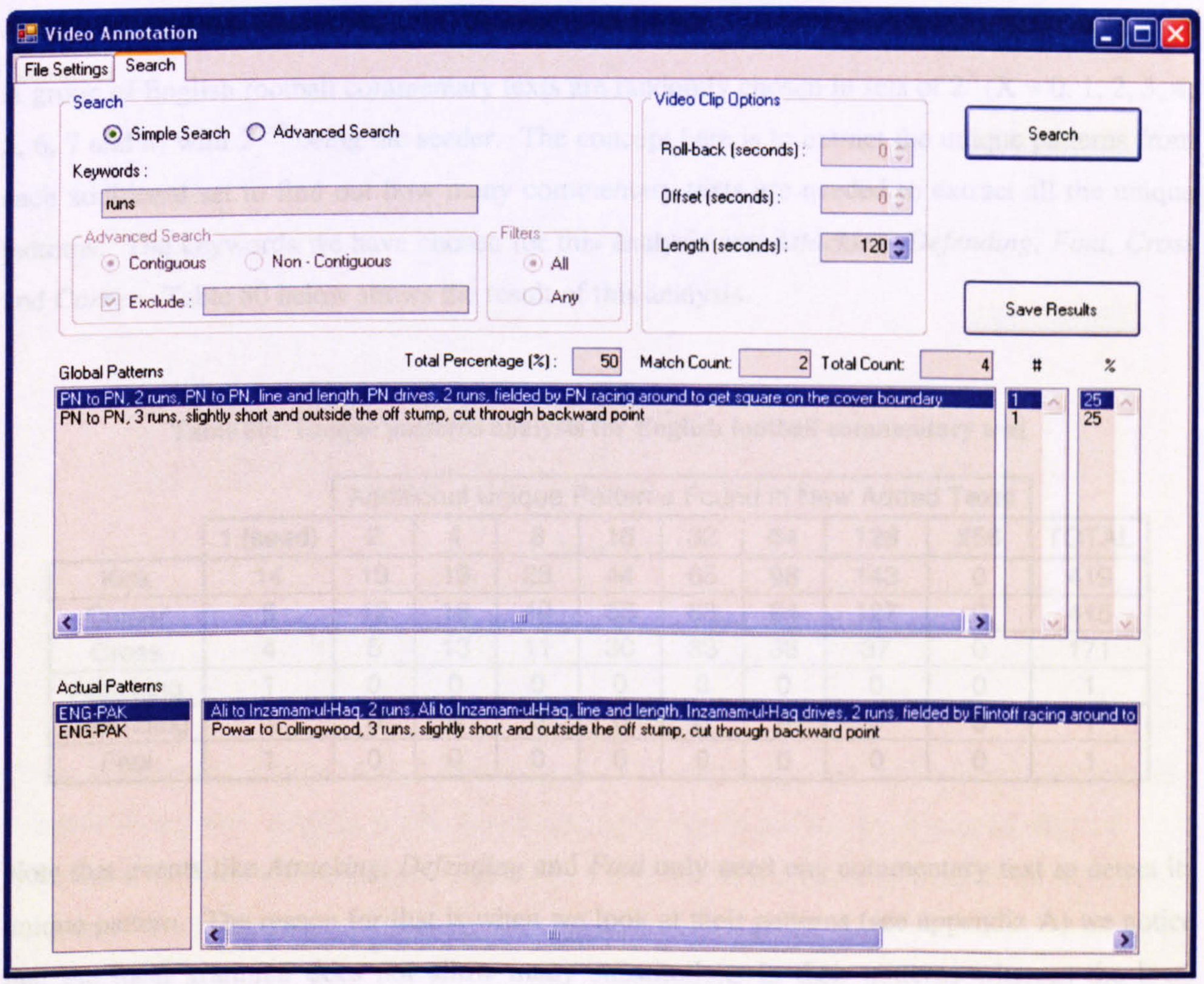


Figure 87: Screen shot of the System Video Annotation processing a sample from the cricket commentary

The Figure above shows our system's ability to process cricket text commentary. The result was examined for accuracy. It was found that our system performed very well. If time-stamping exists in the corpus under investigation, our system would be able to extract the events' clips and archive them.

4.4.2.3 Patterns Existence Analysis

One might question how many matches are needed in order to see all the patterns. In other words, one needs to find the maximum number of needed collateral texts to present all the events to be detected, at least once. A study and evaluation will be carried out to cover the English football corpus, Arabic football corpus and English cricket corpus.

4.4.2.3.1 English Corpus Unique Patterns Analysis

A group of English football commentary texts are randomly chosen in sets of 2^X (X = 0, 1, 2, 3, 4, 5, 6, 7 and 8) with $2^{X=0}$ being the seeder. The concept here is to extract the unique patterns from each additional set to find out how many commentary texts are needed to extract all the unique patterns. The keywords we have chosen for this analysis are: *Attacking*, *Defending*, *Foul*, *Cross* and *Corner*. Table 80 below shows the result of this analysis.

Table 80: Unique patterns analysis for English football commentary text

	Additional Unique Patterns Found in New Added Texts								
	1 (seed)	2	4	8	16	32	64	128	256	TOTAL
Kick	14	13	19	23	44	65	98	143	0	419
Corner	8	12	18	40	58	68	84	127	0	415
Cross	4	5	13	11	30	33	38	37	0	171
Attacking	1	0	0	0	0	0	0	0	0	1
Defending	1	0	0	0	0	0	0	0	0	1
Foul	1	0	0	0	0	0	0	0	0	1

Note that events like *Attacking*, *Defending* and *Foul* only need one commentary text to detect its unique pattern. The reason for that is when we look at their patterns (see appendix A) we notice that the local grammar does not allow many substitutions in their patterns whereas the local grammar allows many substitutions in *Kick*, *Corner* and *Corner* patterns. From one commentary text we can extract about 14 unique *kick* patterns. When adding 2 more commentary texts, we

find 13 more unique *kick* patterns. When 4 commentary texts are added, we find 19 more unique *kick patterns*. An additional 23 unique *kick* patterns are found when adding 8 commentary texts. Notice that the *kick* unique patterns exist up to 128 commentary matches. When adding 2^8 (256) matches, no more unique patterns are found for any of the keywords we have selected. To avoid ambiguity, the same analysis has been done three more times with the commentary text corpus being shuffled to avoid repetition. Figure 88 below shows the four analysis unique patterns comparison.

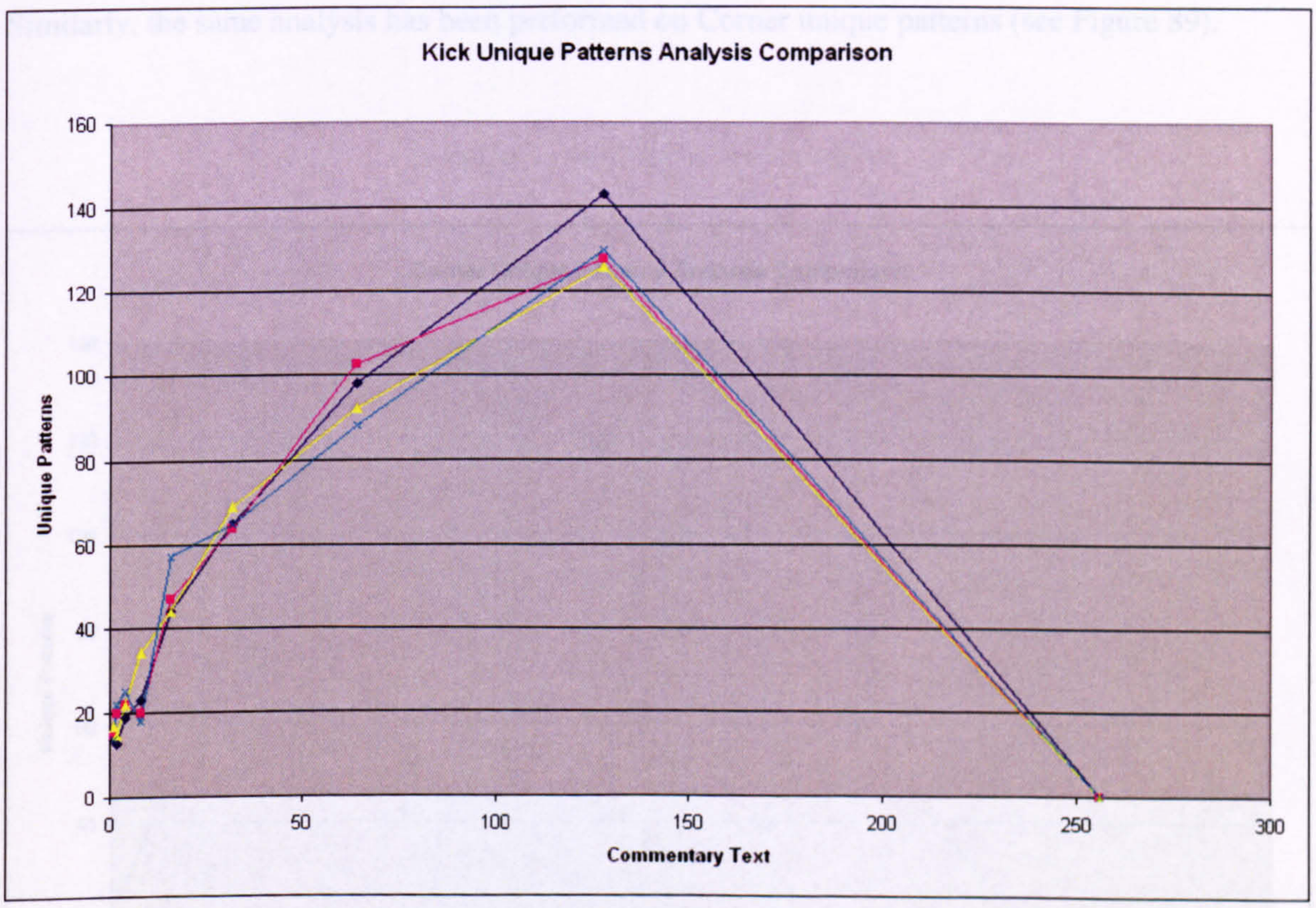


Figure 88: Kick unique patterns analysis comparison

Since we are comparing more than 2 samples, we will use the Kruskal Wallis Test instead of the Mann-Whitney test. Kruskal Wallis Test is a non-parametric method for testing equality of population medians among groups. Intuitively, it is identical to a one-way analysis of variance with the data replaced by their ranks. It is an extension of the Mann-Whitney U test to 3 or more groups.

Table 81: Kruskal Wallis Test result for kick (4-groups) unique patterns analysis in (English-Football) Commentary texts

Variable	Value
Kruskal Wallis Test	3
p-Value	0.991

Table 81 shows the p-Value = 0.991 which means there is insignificant difference between the *kick* unique patterns detection in the 4-groups. This indicates that there is solidity in the *kick* unique patterns appearance.

Similarly, the same analysis has been performed on Corner unique patterns (see Figure 89).

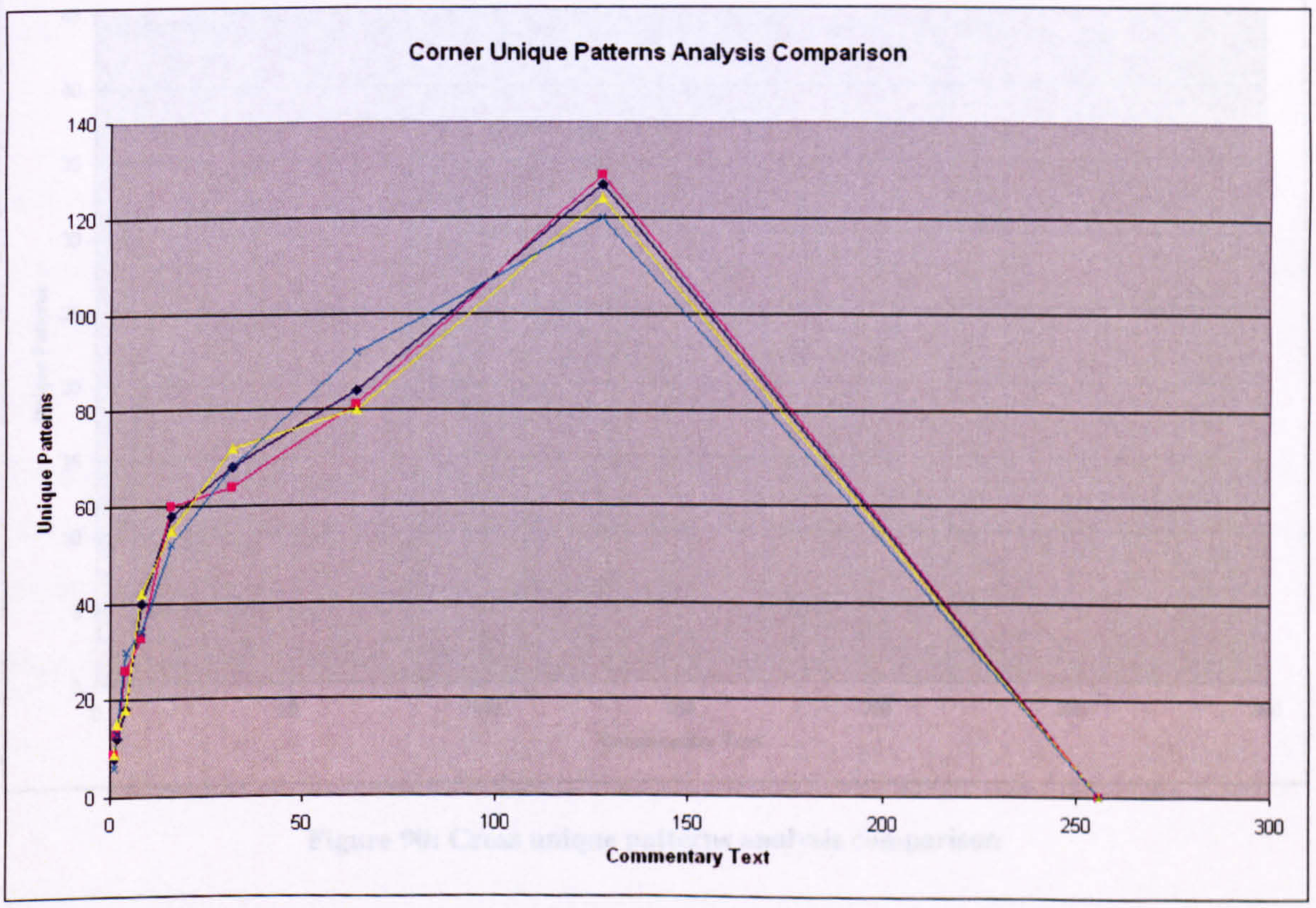


Figure 89: Corner unique patterns analysis comparison

Table 82 below shows the comparison analysis of Kruskal Wallis Test.

Table 82: Kruskal Wallis Test result for *Corner* (4-groups) unique patterns analysis in

(English-Football) Commentary texts

Variable	Value
Kruskal Wallis Test	3
p-Value	0.998

Table 82 shows the p-Value = 0.998 which means there is insignificant difference between the *kick* unique patterns detection in the 4-groups. This indicates that there is solidity in the *kick* unique patterns appearance.

Figure 90 below shows the Cross unique patterns analysis comparison chart.

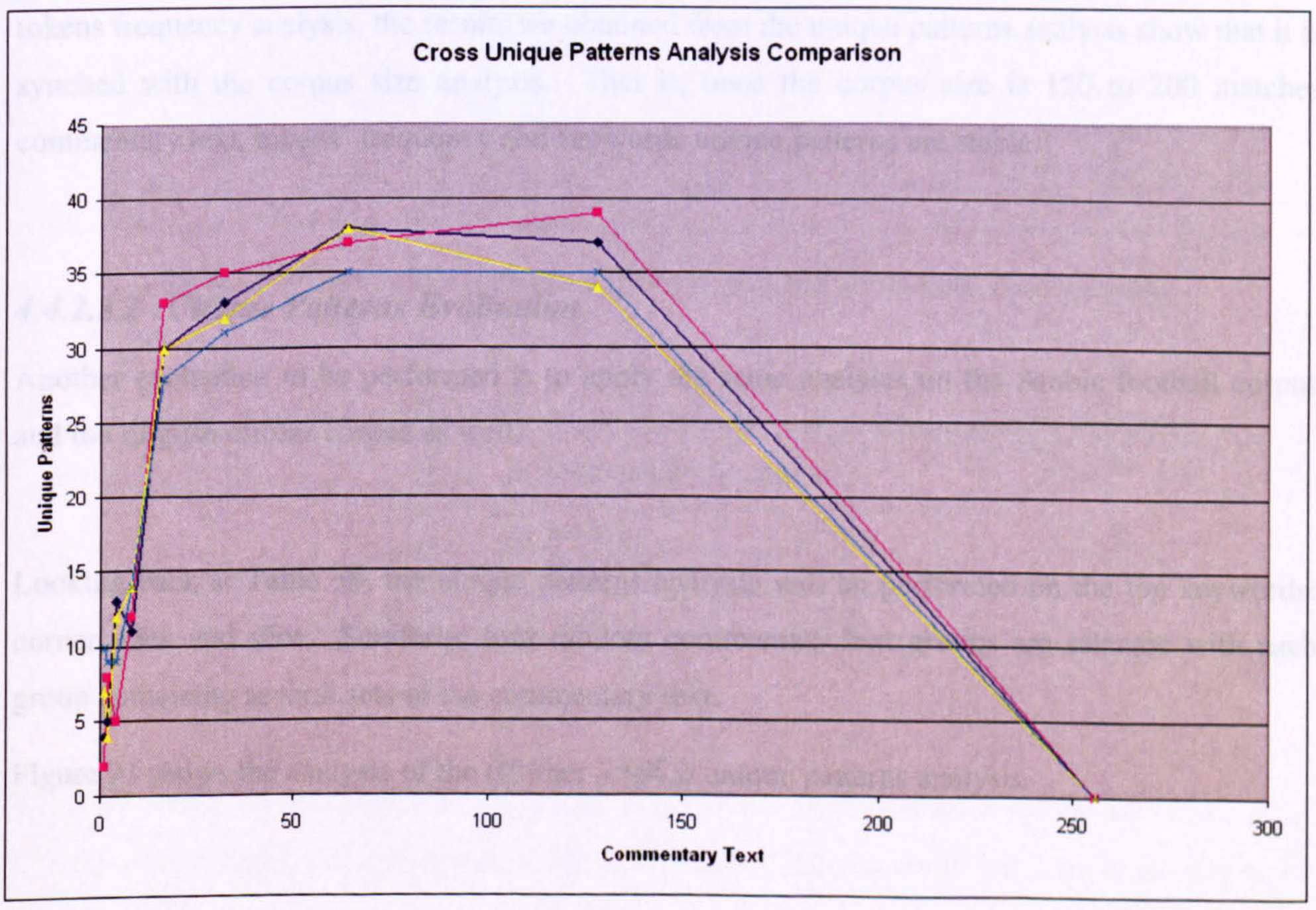


Figure 90: Cross unique patterns analysis comparison

Table 83 below shows the comparison analysis of Kruskal Wallis Test.

Table 83: Kruskal Wallis Test result for Cross (4-groups) unique patterns analysis in (English-Football) Commentary texts

Variable	Value
Kruskal Wallis Test	3
p-Value	0.997

Table 82 shows the p-Value = 0.997 which means there is insignificant difference between the *kick* unique patterns detection in the 4-groups. This indicates that there is solidity in the *kick* unique patterns appearance.

The analyses that have been done show that there is a consistency in detecting unique patterns. Looking back at Figure 71 and Figure 72 when analysing the suggested corpus size based on tokens frequency analysis, the results we obtained from the unique patterns analysis show that it is synched with the corpus size analysis. That is, once the corpus size is 150 to 200 matches commentary text, tokens' frequency and keywords unique patterns are stable.

4.4.2.3.2 Unique Patterns Evaluation

Another evaluation to be performed is to apply the same analysis on the Arabic football corpus and the English cricket corpus as well.

Looking back at Table 38, the unique patterns analysis will be performed on the top keywords: corner, kick and shot. Similarly, four random commentary text groups are selected with each group containing several sets of the commentary text.

Figure 91 shows the analysis of the (Corner – ركنيه) unique patterns analysis.

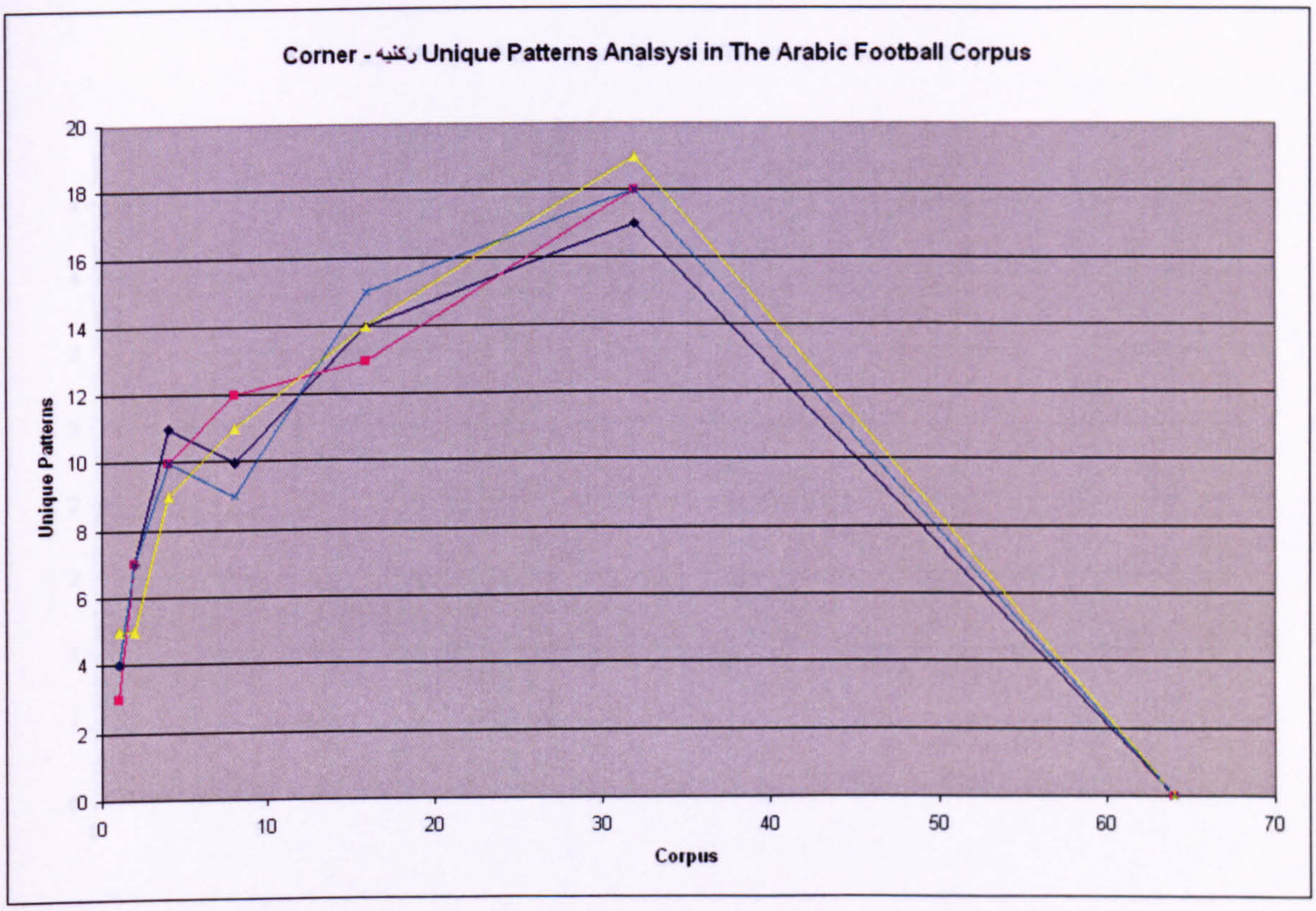


Figure 91: Corner - ركنيه unique patterns analysis in the Arabic football corpus

Table 84: Kruskal Wallis Test result for Corner - ركنيه unique patterns analysis in Arabic football corpus

Variable	Value
Kruskal Wallis Test	3
p-Value	0.993

Table 84 shows the p-Value = 0.993 which means there is insignificant difference between the Corner - ركنيه unique patterns detection in the 4-groups. This indicates that there is solidity in the Corner - ركنيه unique patterns appearance.

Figure 92 below show the Kick - تمريره unique patterns analysis in the Arabic football corpus.

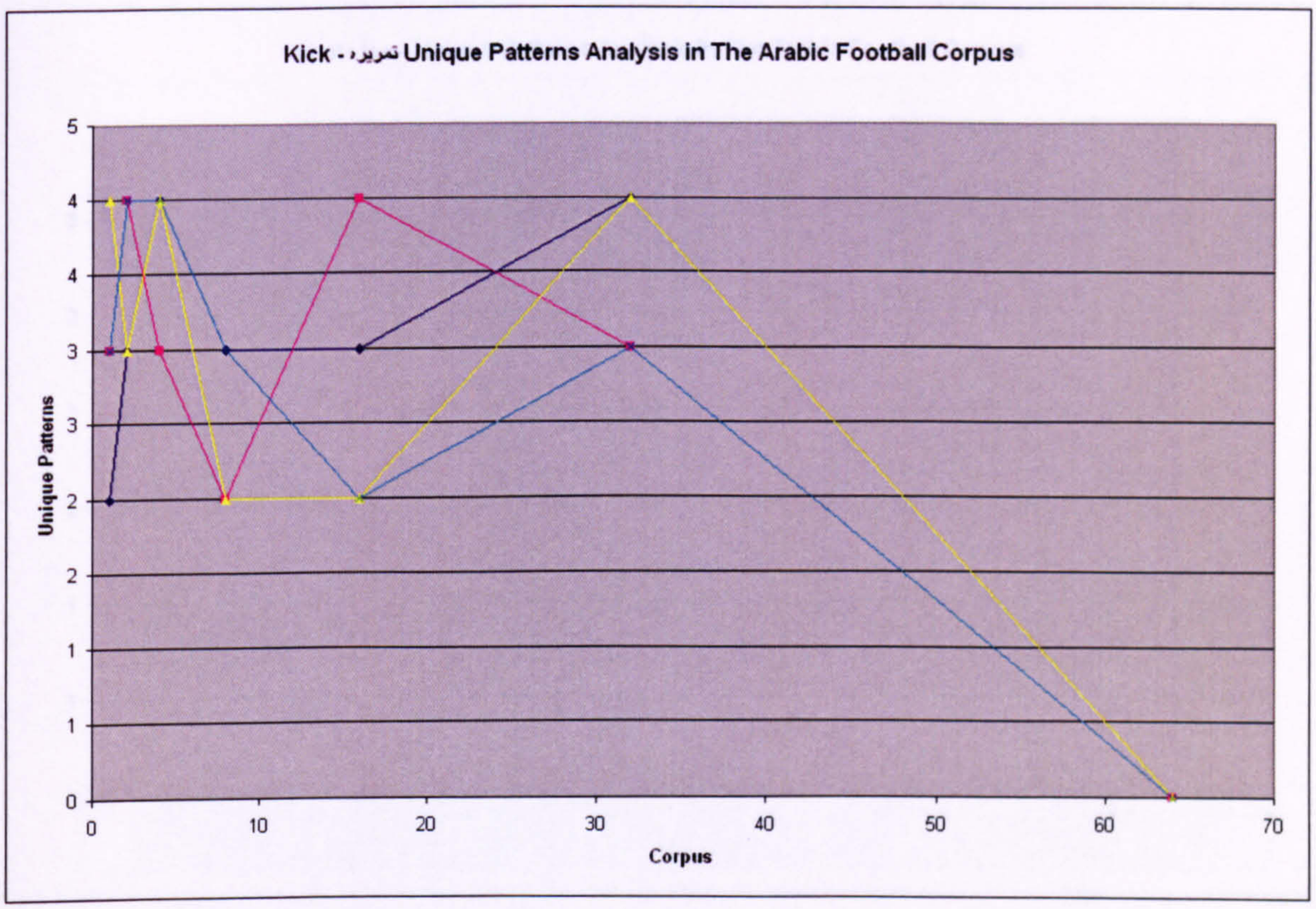


Figure 92: Kick - تمريره unique patterns analysis in the Arabic football corpus

Table 85 below shows the result of the Kruskal Wallis Test result for Kick - تمريره unique patterns analysis in Arabic football corpus

Table 85: Kruskal Wallis Test result for Kick - تمريره unique patterns analysis in the Arabic football corpus

Variable	Value
Kruskal Wallis Test	3
p-Value	0.618

Table 85 shows the p-Value = 0.618 which means there is insignificant difference between the Kick - تمريره unique patterns detection in the 4-groups. This indicates that there is solidity in the Kick - تمريره unique patterns appearance.

Figure 93 below shows Shot - تسديده unique patterns analysis in the Arabic football corpus.

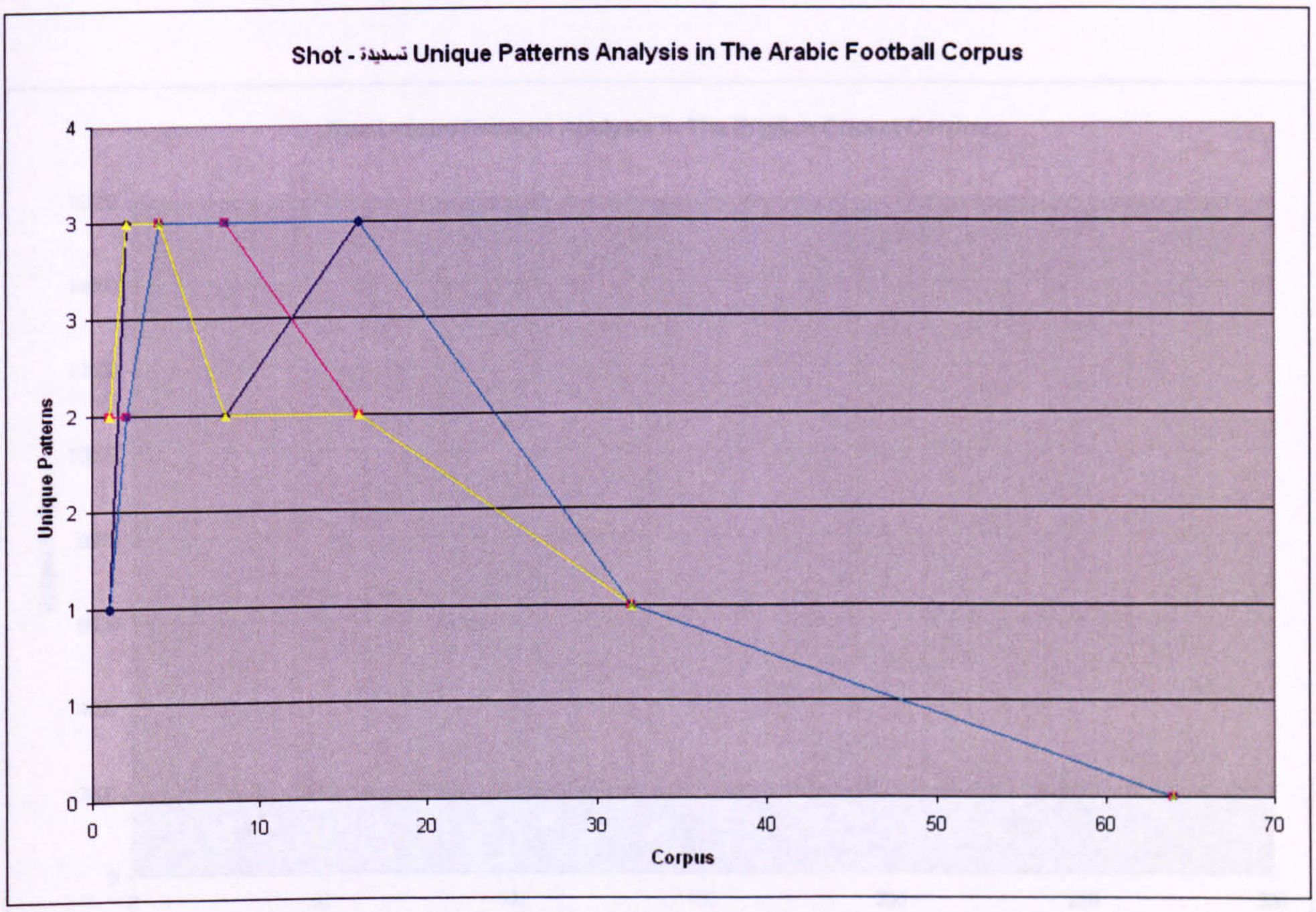


Figure 93: Shot - تسديده unique patterns analysis in the Arabic football corpus

Table 86 below shows the Kruskal Wallis Test result for Shot - تسديده unique patterns analysis in the Arabic football corpus.

Table 86: Kruskal Wallis Test result for Shot - تسديده unique patterns analysis in the Arabic football corpus

Variable	Value
Kruskal Wallis Test	3
p-Value	0.525

Table 86 shows the p-Value = 0.525 which means there is insignificant difference between the *kick* unique patterns detection in the 4-groups. This indicates that there is solidity in the *kick* unique patterns appearance.

Similar analysis was done in the English cricket corpus. The chosen keywords are: run and runs.

Figure 94 below shows Run unique patterns analysis in the English cricket corpus.

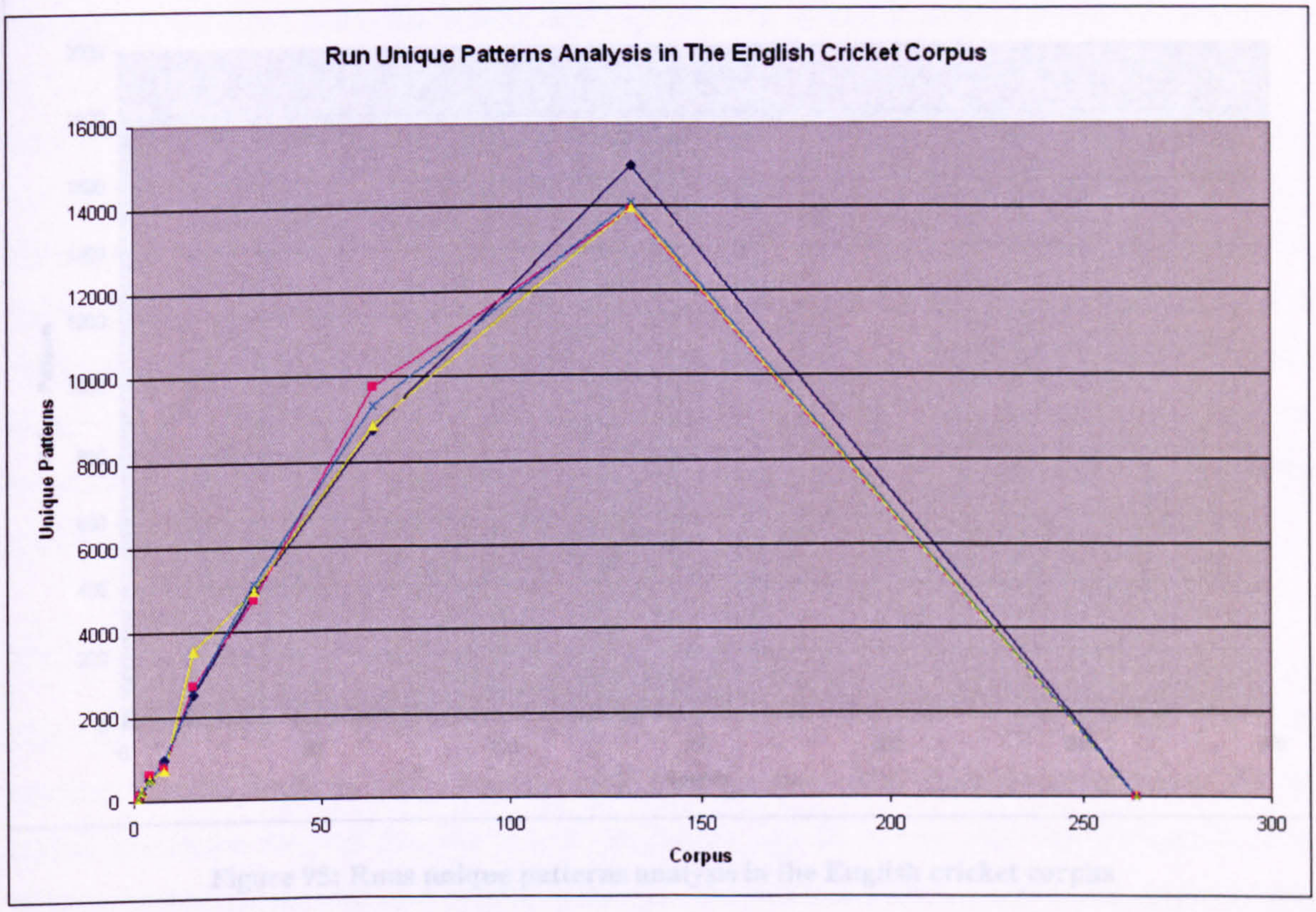


Figure 94: Run unique patterns analysis in the English cricket corpus

Table 87: Kruskal Wallis Test result for Run unique patterns analysis in the English cricket corpus

Variable	Value
Kruskal Wallis Test	5
p-Value	0.916

Table 87 shows the p-Value = 0.916 which means there is insignificant difference between the *kick* unique patterns detection in the 4-groups. This indicates that there is solidity in the *kick* unique patterns appearance.

Figure 95 below shows Runs unique patterns analysis in the English cricket corpus.

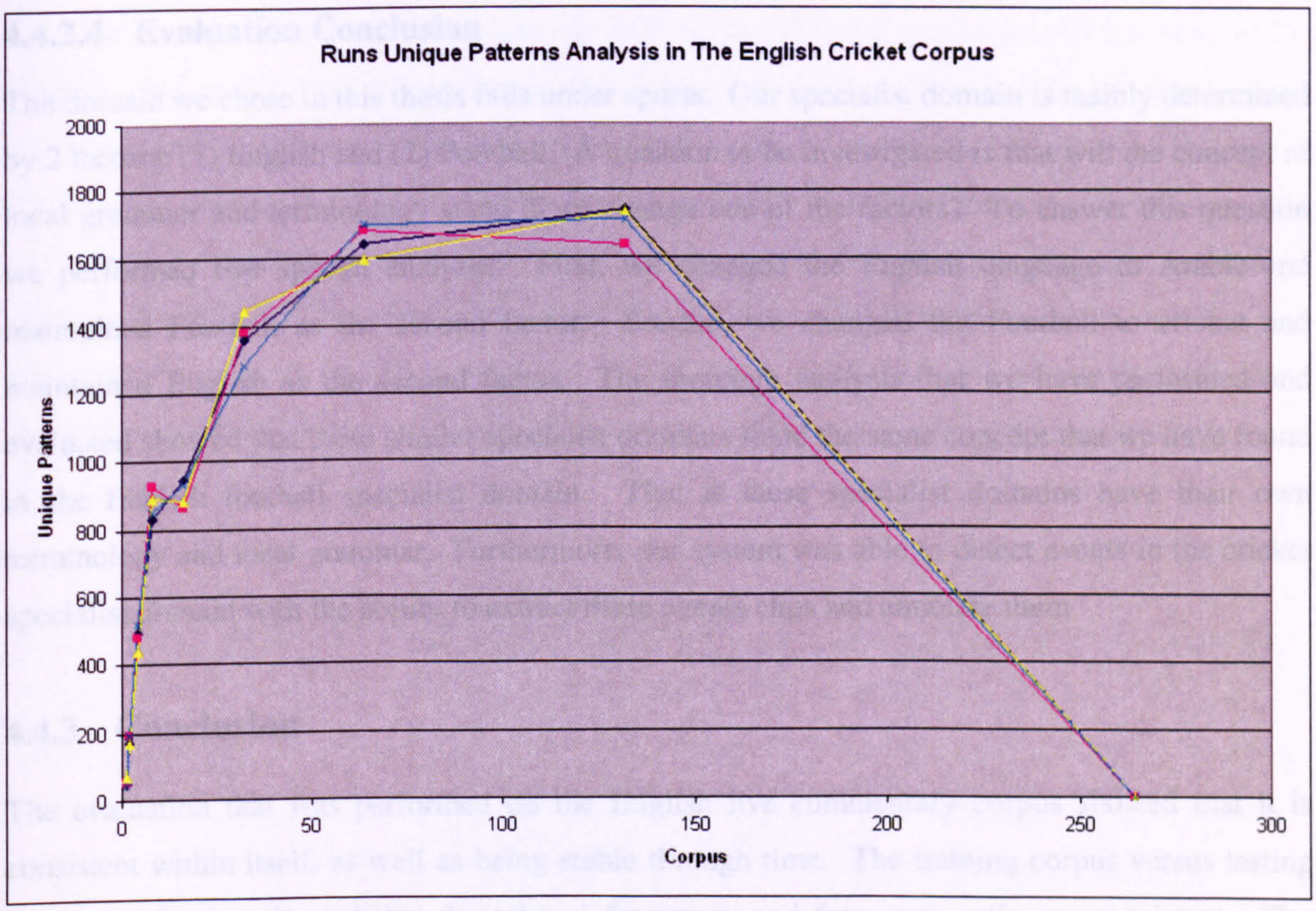


Figure 95: Runs unique patterns analysis in the English cricket corpus

Table 88: Kruskal Wallis Test result for Runs unique patterns analysis in the English cricket corpus

Variable	Value
Kruskal Wallis Test	5
p-Value	0.927

Table 88 shows the p-Value = 0.927 which means there is insignificant difference between the *kick* unique patterns detection in the 4-groups. This indicates that there is solidity in the *kick* unique patterns appearance.

The analysis we have done for the unique patterns in the Arabic football corpus and the English cricket corpus shows that both corpora are solid and unique patterns exist until a certain number of commentary texts collection has been achieved. This confirms the analysis result we obtained from the unique patterns analysis that was undertaken for the English football commentary.

4.4.2.4 Evaluation Conclusion

The domain we chose in this thesis falls under sports. Our specialist domain is mainly determined by 2 factors: (1) English and (2) Football. A question to be investigated is that will the concept of local grammar and terminology stand if we change one of the factors? To answer this question we performed two special analyses. First, we changed the English language to Arabic and maintained Football as the second factor. Second, we changed the Football to cricket and maintained English as the second factor. The thorough analysis that we have performed and evaluated showed that these similar specialist domains share the same concept that we have found in the English football specialist domain. That is these specialist domains have their own terminology and local grammar. Furthermore, our system was able to detect events in the cricket specialist domain with the ability to extract these events clips and annotate them

4.4.3 Conclusion

The evaluation that was performed on the English live commentary corpus showed that it is consistent within itself, as well as being stable through time. The training corpus versus testing corpus evaluation showed that the tokens' frequency and frequency ratio are consistent. The training corpus versus testing-2 corpus evaluation showed that after two years the corpus is still stable as the CLAWS and System Quirk analysis showed consistent results. The text commentary versus the match video evaluation showed that the live commentary reports a high percentage of the actual events which allows the text substitution of video. The Precision and Recall analysis showed that the system has a high event detection percentage and a high retrieval percentage. The system strength evaluation showed that the system presented in this thesis would catch a high percentage of the actual events. The correlation analysis answered the question regarding the corpus size and the smallest optimal size. Applying the thesis method to a different sport (cricket) still using the English language and then to a different language (Arabic) but with the same sport (football) showed that although some primary parameters are missing, the system would still detect events just by analysing their corpora and finding their tokens' frequency and frequency ratio. The evaluation results in English football showed high correlations and were very promising and the results from English cricket and Arabic football test applications showed promising conclusions that supported the thesis and its method of automation. The patterns analysis that was done in English football, Arabic football and English cricket showed consistency in these specialist domain linguistic structures.

In conclusion, the domain we chose in this thesis falls under sports. Our specialist domain is mainly determined by 2 factors: (1) English and (2) Football. A question to be investigated is that

will the concept of local grammar and terminology stand if we change one of the factors? To answer this question we performed two special analyses. First, we changed the English language to Arabic and maintained Football as the second factor. Second, we changed the Football to cricket and maintained English as the second factor. The thorough analysis that we have performed and evaluated showed that these similar specialist domains share the same concept that we have found in the English football specialist domain. That is these specialist domains have their own terminology and local grammar. Furthermore, our system was able to detect events in the cricket specialist domain with the ability to extract these events clips and annotate them

Chapter 5

5 Conclusion and Future Work

5.1 Conclusion

This is a text-based study which investigated how video annotations could be automatically generated from a corpus of text in a visual domain. It has shown that, starting with frequency count and collocation analysis, one can identify and extract repeated patterns of use in the domain to indicate how events take place. The corpus was analysed statistically and keywords were chosen without human interference. This approach has been discussed by Ahmad and Gillam (2005) in their work on how to choose words based on the computation of the z-score of the frequency and of weirdness. Keywords collocation and their phrases collocation were detected throughout the corpus automatically. That led to the introduction of local grammar for these patterns. The local grammar with its patterns covers 66% of the events and the annotation system that was introduced automated the generation of the events' clips fully annotated. Human involvement was minimized to two sections: the accept-reject training when new patterns are detected and to help with text-video synchronization. If the time-stamp in the commentary text files was close enough to the actual event time, the human part would be unnecessary and could be eliminated. Notice that the football video files never got visually analysed, such as colour histogram analysis or motion detection and yet the system managed to cut and annotate clips successfully. The evaluation has shown great consistency throughout the corpus and stability through time. Precision and Recall analysis showed that the method and the system were robust in detecting and retrieving events. The system's advanced search features in the annotating clips section provided many features ranging from the ability to search for general events (for example: free kick) to narrowing the search for a very specific event including the player name and the team name (for example: free kick taken left-footed by Mark Arsenal). In addition, the system showed flexibility in accepting the search keywords in no specific order. Similarly, the system's retrieval section provided the user with the option to search annotated clips using general keywords or specific keywords. The system's method was applied to cricket, and even though its commentary text has fewer events and big gaps between the time-stamps, statistical keywords

were chosen and patterns were detected. Furthermore, the method was applied to the more challenging Arabic language. However, manual analysis showed that using statistically chosen keywords can reveal existing patterns.

This system can significantly reduce the human role in text/video synchronization when working with a visual or audio analysis system that can indicate the start and the end of a match; it could even eliminate it altogether. Furthermore, our system can estimate the event time and then flag the new time-stamp.

It is hoped that this thesis has shown that through using a text corpus, massive automated clips' annotation can be achieved, on a considerable scale.

5.2 Future Work

One of the primary goals in automated annotation is to eliminate human intervention. The system introduced and discussed in this thesis needs to be hooked to a video analysis system that can detect the beginning and the end of the football match in order to align the live commentary text time stamp with the video. This will help when dealing with a football match that has extra time as it varies from one game to another including some games going into two additional overtime halves and possibly a penalty shoot-out. An issue that requires further consideration is that the Arabic sporting articles corpus needs to be bigger to allow additional corpus analysis. Also, a code needs to be written to perform a collocation analysis on the Arabic corpus.

5.3 Summary and Overall Conclusion

When we started our research, our interests lay in establishing whether an automated video indexing and annotating system could be accurate and effective when dealing with a live commentary text corpus. The sports domain has attracted many researchers, particularly the sub-domain of football. Its worldwide popularity and the increased numbers of matches being broadcast made researchers increasingly interested in analysing its videos. Furthermore, football

videos and sports videos in general are considered to pose automation problems of medium level difficulty (Snoek et al, 2006:101 – Table 2) which made the task seem more achievable. Researchers have aimed to automate their systems as much as possible with respect to faster processing time and higher accuracy. Throughout the years, different methods have been introduced. Some like Gross (1993) and Andrade et al (2003) have used object recognition methods. Others including Quenot et al (2002) and Sattar (2002) used events recognition methods. Some researchers took it further to develop methods for video indexing (Nam and Tewfik, 1999; Ronard and Thuong, 2003; Snoek and Worring, 2005, and Zhong and Chang 2000). The techniques that are used involved moving objects and the moving region methods (Lema et al, 2000), colour histogram comparison (Tahaghoghi et al, 2005), layered video data modelling (Petkovic et al, 2001), shot boundary detection (Smeaton and Over, 2002), structure parsing (Zhong and Chang 2000), and image segmentation (Petkovic et al, 2001). Other researchers focused on the speech that accompanies the video, and speech recognition and indexing was their primary goal (exemplified by Adam et al, 2002; Dolbear and Brady, 2003; Snoek and Worring, 2005, and Wolf et al, 2002). Other researchers paid attention to the text commentary that comes with football videos, such as the MUMIS project (Declerck et al 2001 and Wang et al 2005).

The point to mention about these previous systems is that events and keywords are pre-defined or pre-selected. Manually selected or enhanced keywords and automated keyword selection is what differs our system from the others. For example, within the video analysis section of the MUMIS project, the researcher, in order to train his system, must choose the event(s) that his system to recognise and analyse. Also, if *keywords* are used then these *keywords* are manually selected. Researchers who used the text commentary have either pre-selected the *keywords* to be used or selected keywords or events to be ignored. When it comes to video indexing and annotating, these systems also use manually chosen keywords by the researcher. This leads to their systems being static instead of dynamic. In order to build a system that automatically indexes and annotates videos without human interference (the goal of this thesis), human pre-selection of keywords for the system to include/avoid has to be eliminated.

The system that is presented in this thesis has no prior knowledge of the keywords that will be used and no prior knowledge of the events types. It started with collecting a corpus of the live commentary text (Figure 5). The system is then trained based on the 5-step modified algorithm developed by Ahmad and Gillam (2005) (Figure 9). Events patterns are detected (Table 6) with the use of Smadja's outline method (1994). The keywords are then chosen based on their

weirdness value (Table 10). *Globalization* is then used which leads to the local grammar of the corpus (Figure 26). A system shown in Figure 28 was then proposed based on the work of other fellow researchers and the early analysis of this thesis.

To eliminate human intervention as much as possible and work towards a fully automated system, the corpus is analysed using several applications (Figure 30 and Figure 31) to find unique keywords and then to allocate their patterns by using the collocation method (Figure 33). Notice that at this point the corpus is analysed and patterns are detected without any human intervention. Now that patterns are found and local grammar has been introduced, the text file and its video are ready to join together to provide the video indexing and events annotation. It is noted that the live commentary text time-stamp does not really synchronize with the event time in the source video. For evaluation purposes, a human intervention exists at this point to correct the live commentary time-stamp. A sample of 10 games is used and estimated correction is presented (Table 17 and Table 18). The system then synchronizes the text with its video.

The system GUI is provided in this thesis (Figure 44 and Figure 46). Thus this is the GUI that the researcher would use; and much of the work is being done behind the scene. Note that in Figure 46 a human intervention is needed to evaluate new detected patterns as a valid sentence only, but not as a valid event. A different GUI is presented to show most of the work. The GUI is split into 2 sections: 1) File setting and searching, and 2) Video annotating. The first GUI (Figure 48) shows the accepted text and video files. Figure 49 shows the search section where simple and complex search is provided. The second GUI (Figure 57) shows the automated video indexing and annotating; it also shows the ability to search archived clips for specified keyword(s).

An exclusive evaluation has been performed. The testing corpus and testing-2 corpus are drawn together and the results are compared. Testing corpus and training corpus are compared in: Part-of-Speech analysis (Table 20); Token frequency analysis (Table 22); Collocation analysis (Table 24). The testing-2 corpus and training corpus are compared in: Part-of-Speech analysis (Table 26); Token frequency analysis (Table 28); Tokens frequency ratio (Table 28); and Collocation (Table 31). All these analyses showed consistency and accuracy.

The live commentary text corpus is then evaluated in relation to its video. The number of events detected and events missed has been analysed (Table 53 and Table 54). The results showed that the live text commentary is catching over 95% of the events. This is a very favourable result.

The system is then applied to other domains: English cricket and Arabic football. The system was able to detect events in both domains without change to any of the system parameters. The Arabic language has raised more challenges that need to be addressed and dealt with in order to obtain even better results.

Appendix A

- **Goal kick taken pattern**

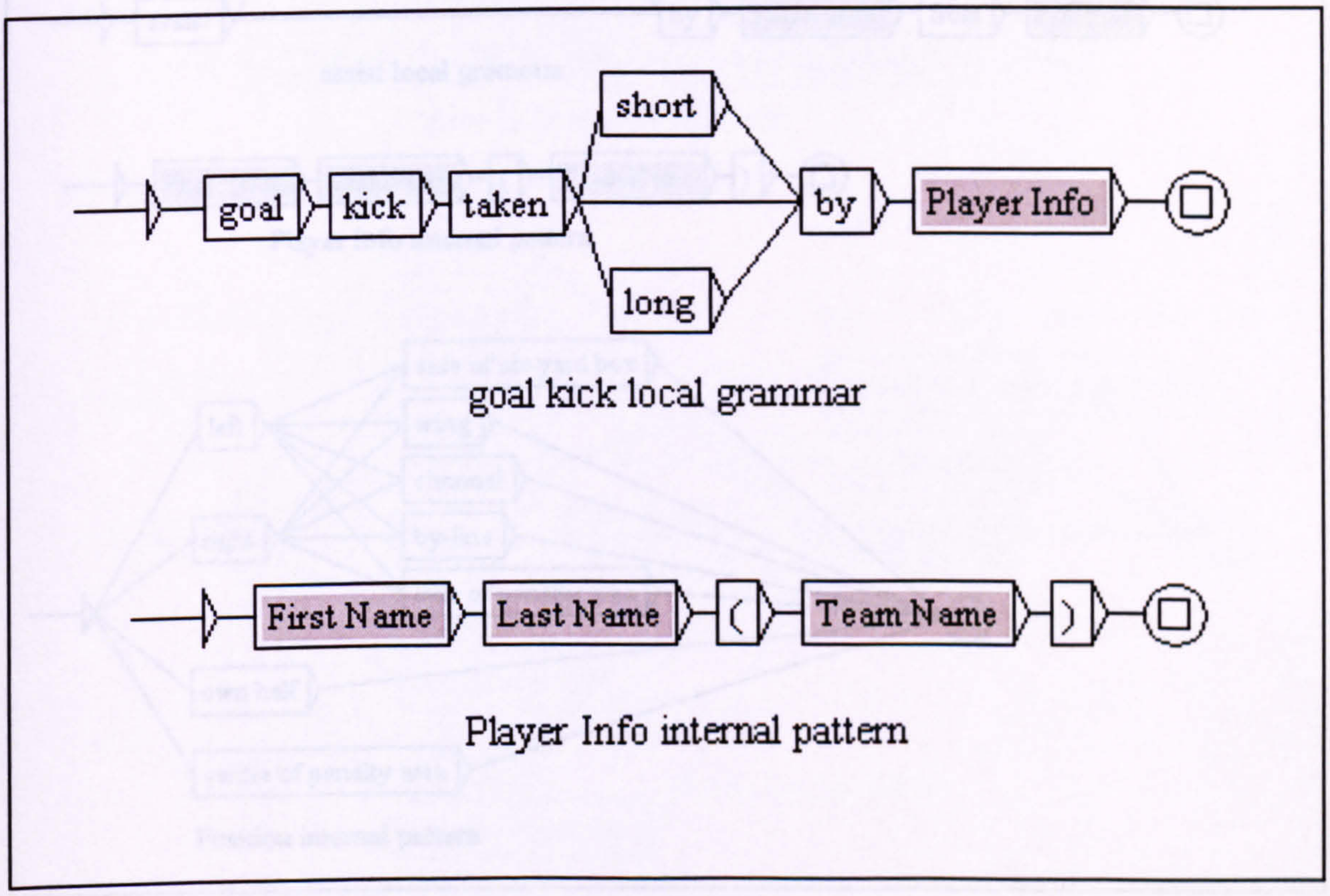


Figure 96: *Goal kick* Pattern

- **Assist pattern**

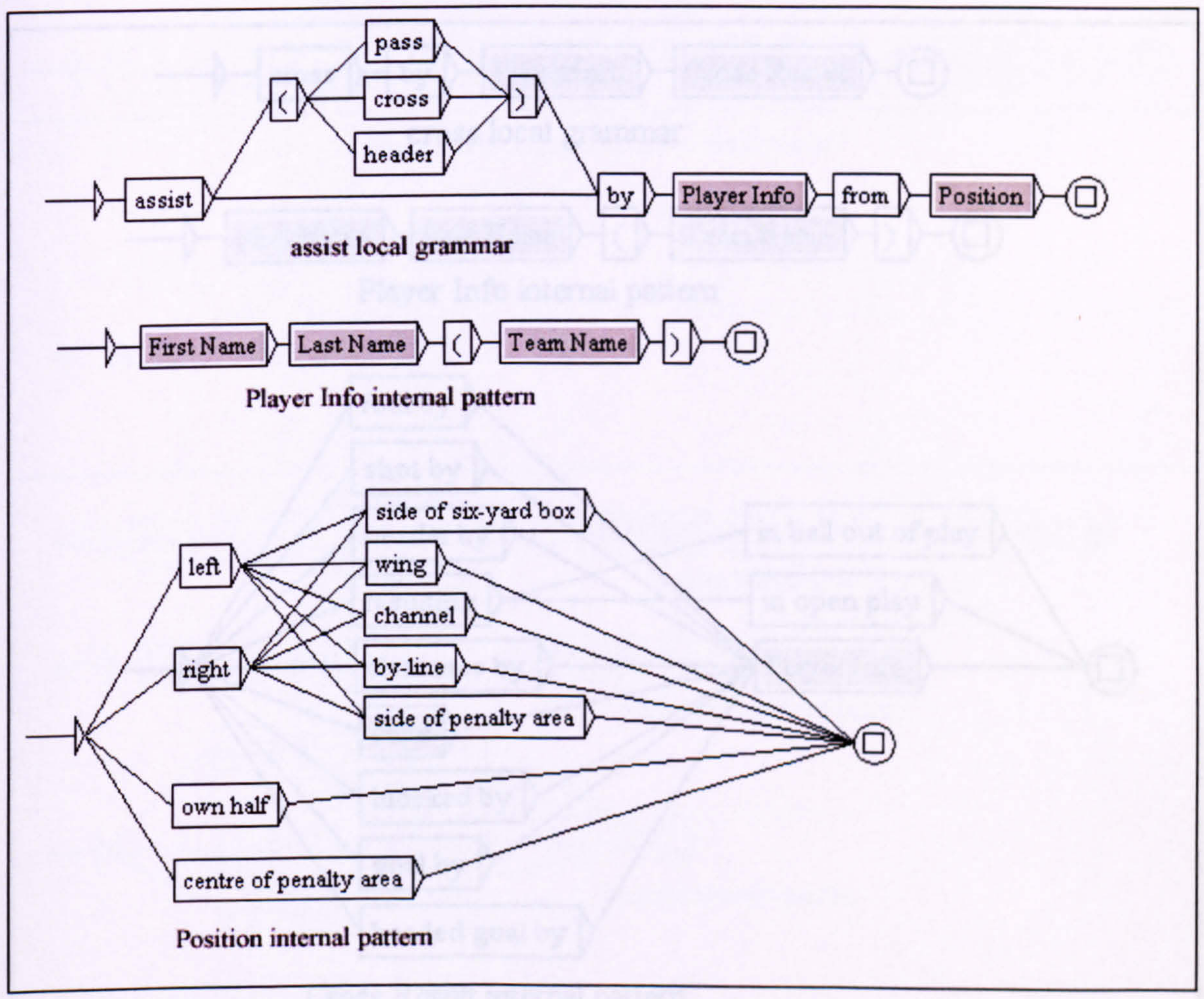


Figure 97: *Assist* Pattern

- **Attacking throw-in pattern**

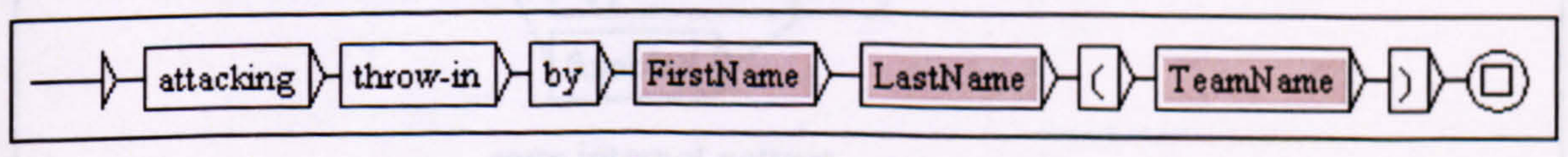


Figure 98: *Attacking throw-in* Pattern

- **Cross by pattern**

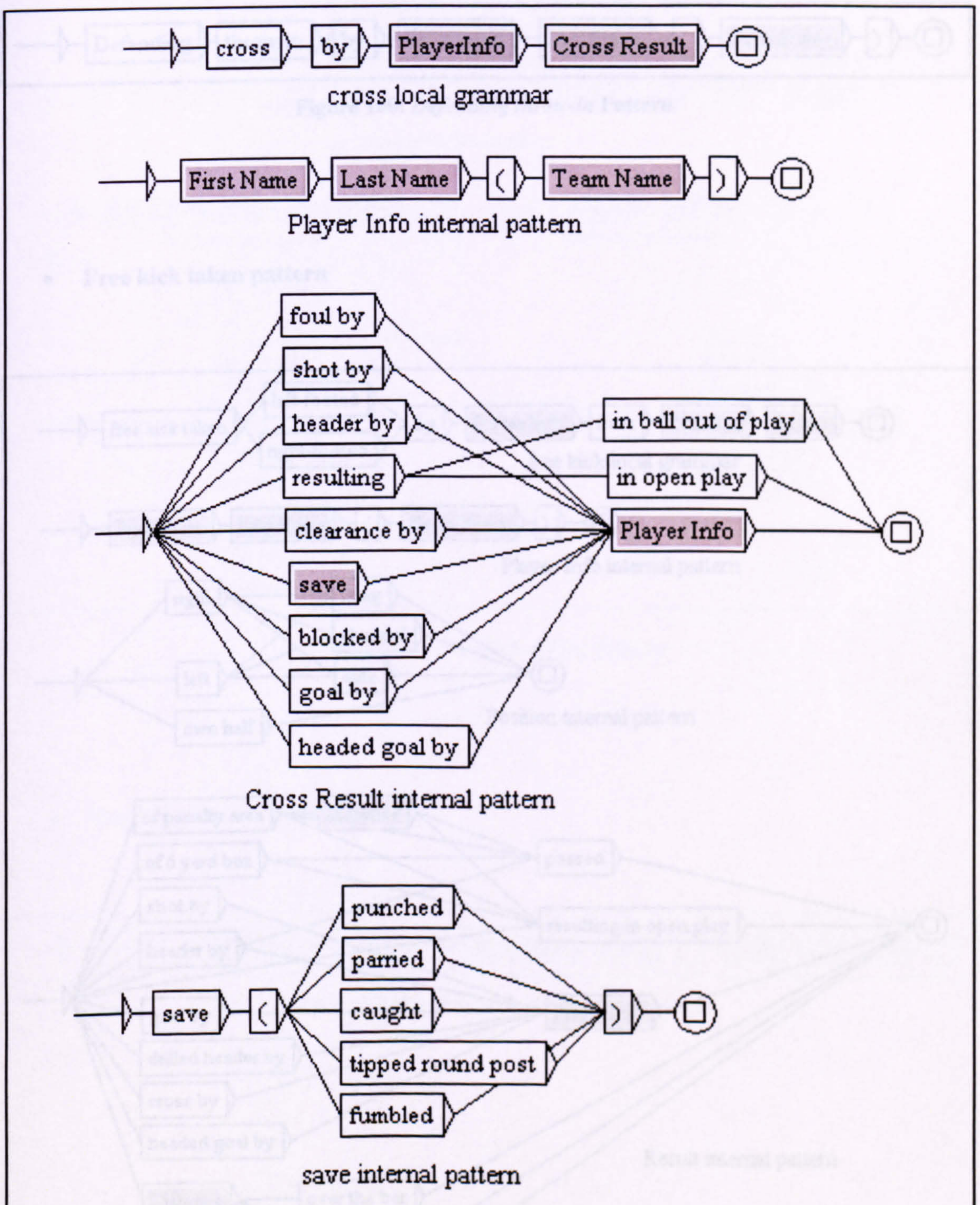


Figure 99: *Cross by* Pattern

- **Defending throw-in pattern**

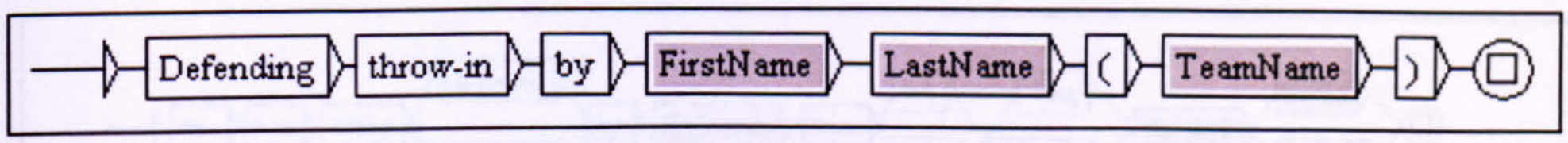


Figure 100: *Defending throw-in* Pattern

- **Free kick taken pattern**

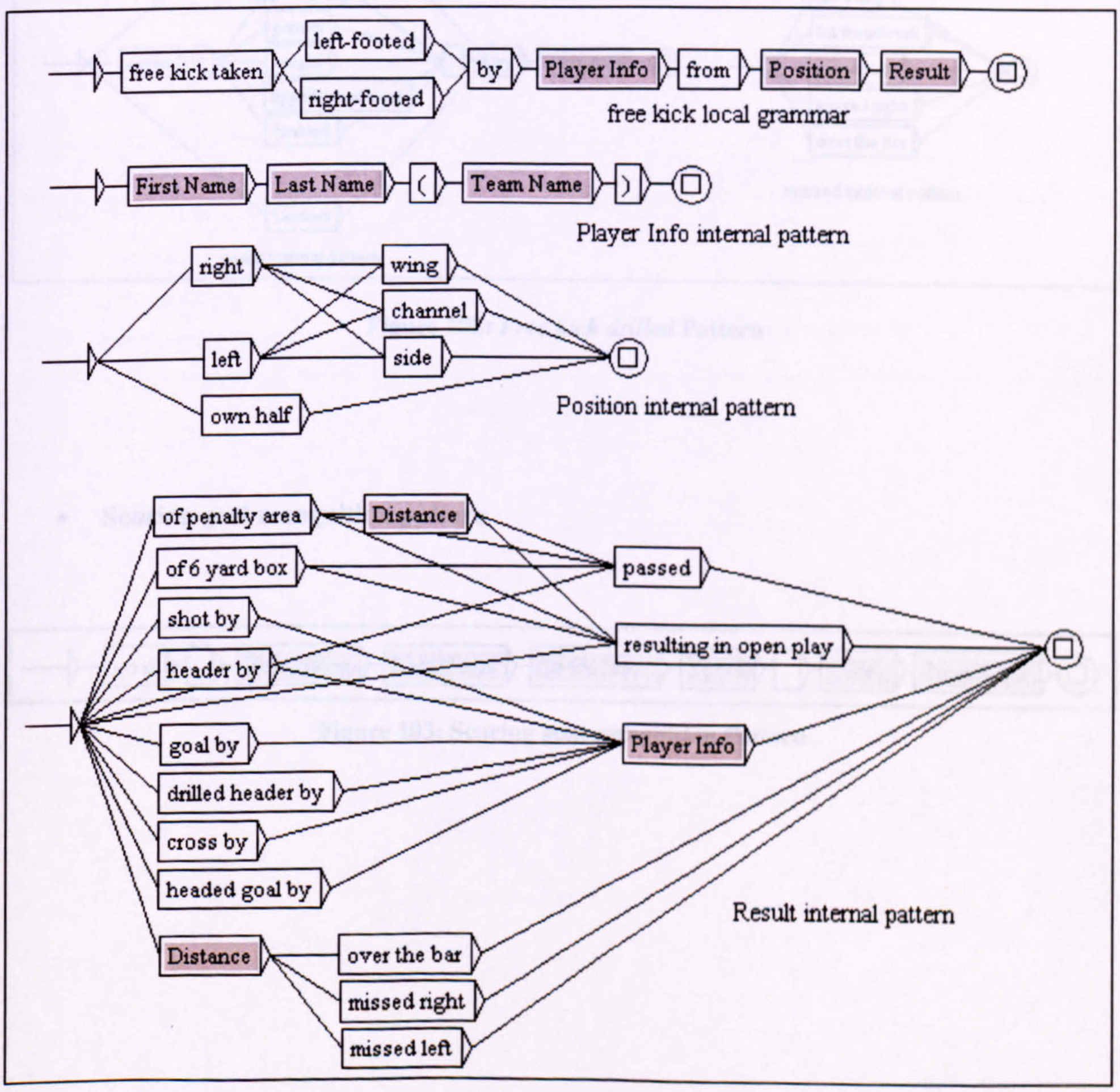


Figure 101: *Free kick taken* Pattern

- **Free kick drilled pattern.**

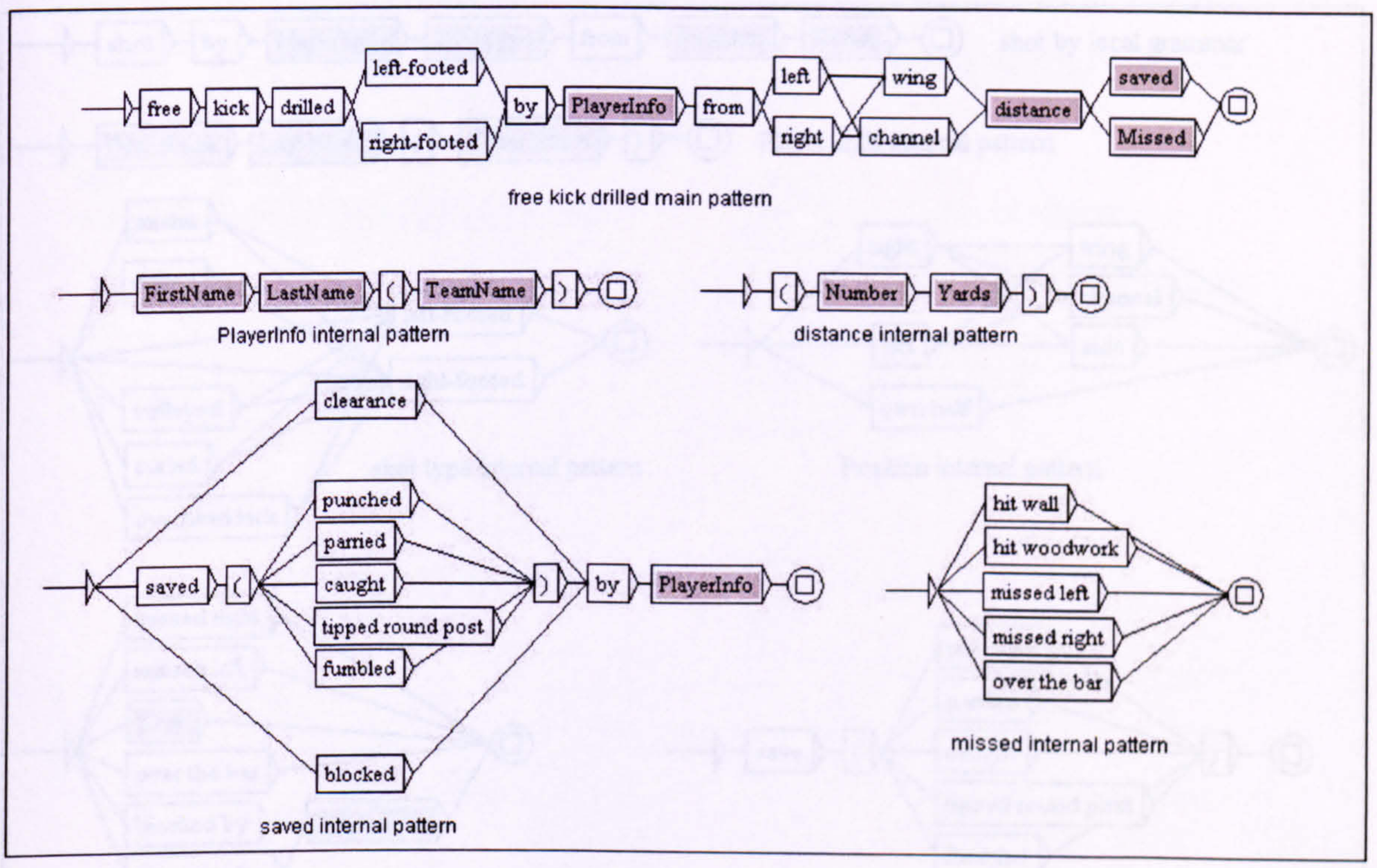


Figure 102: *Free kick drilled* Pattern

- **Scoring goal recognition pattern**

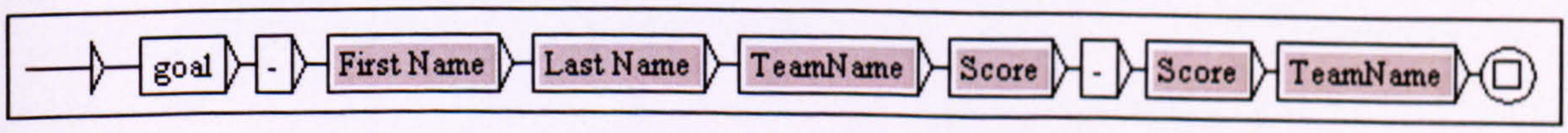


Figure 103: Scoring goal recognition Pattern

- **Shot by pattern**

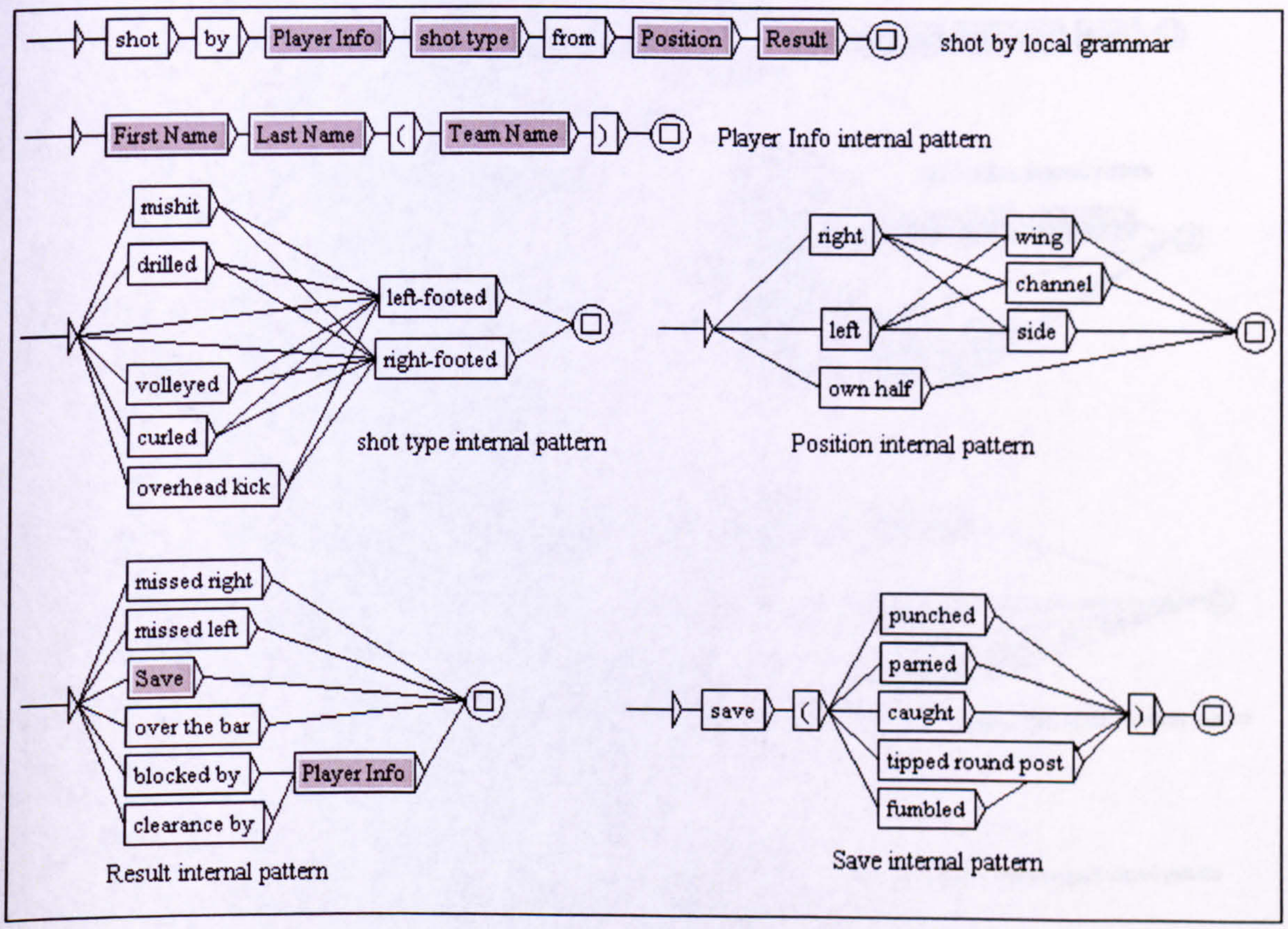


Figure 104: *Shot by* Pattern

- **Corner pattern**

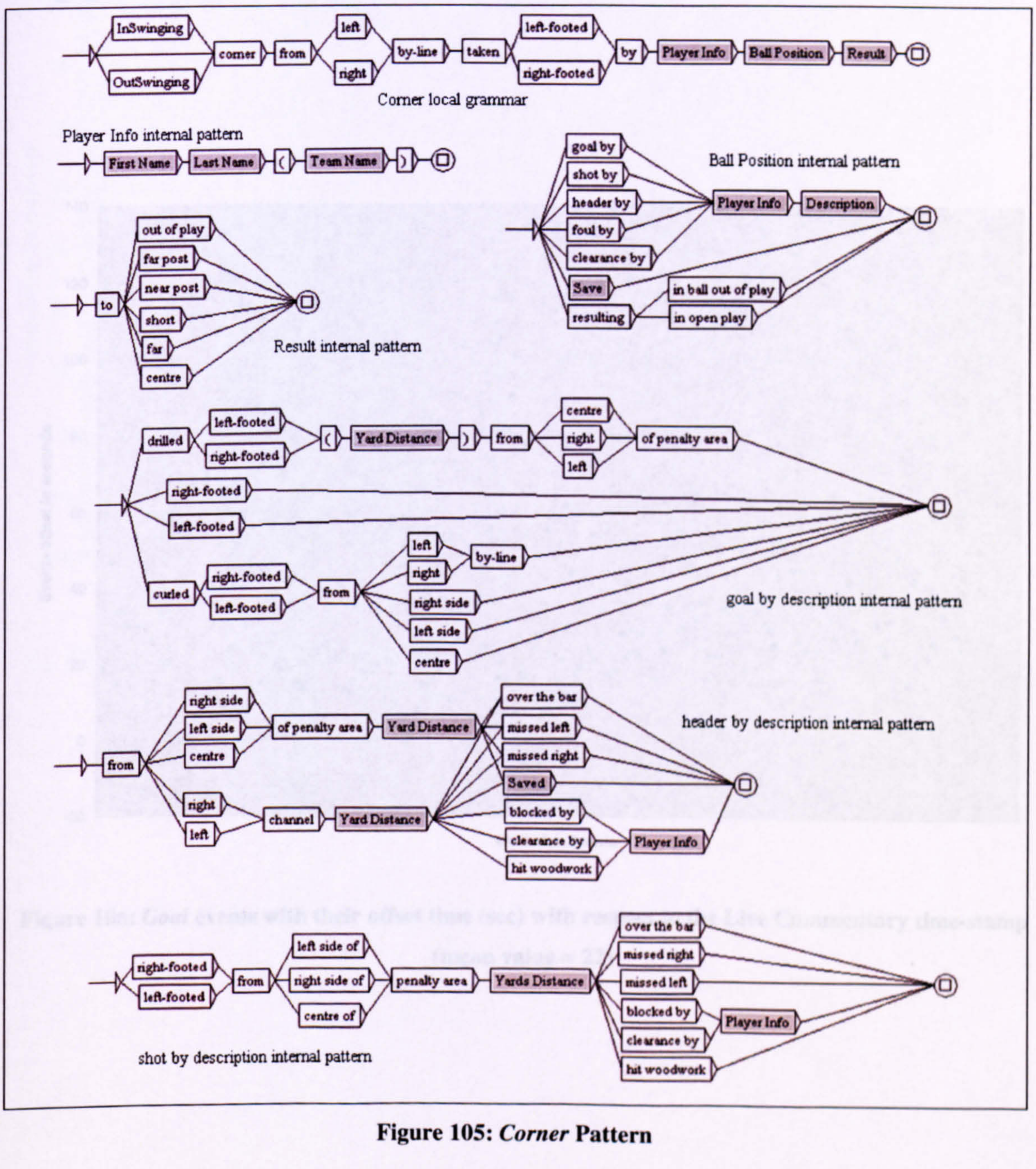


Figure 105: *Corner* Pattern

Appendix B

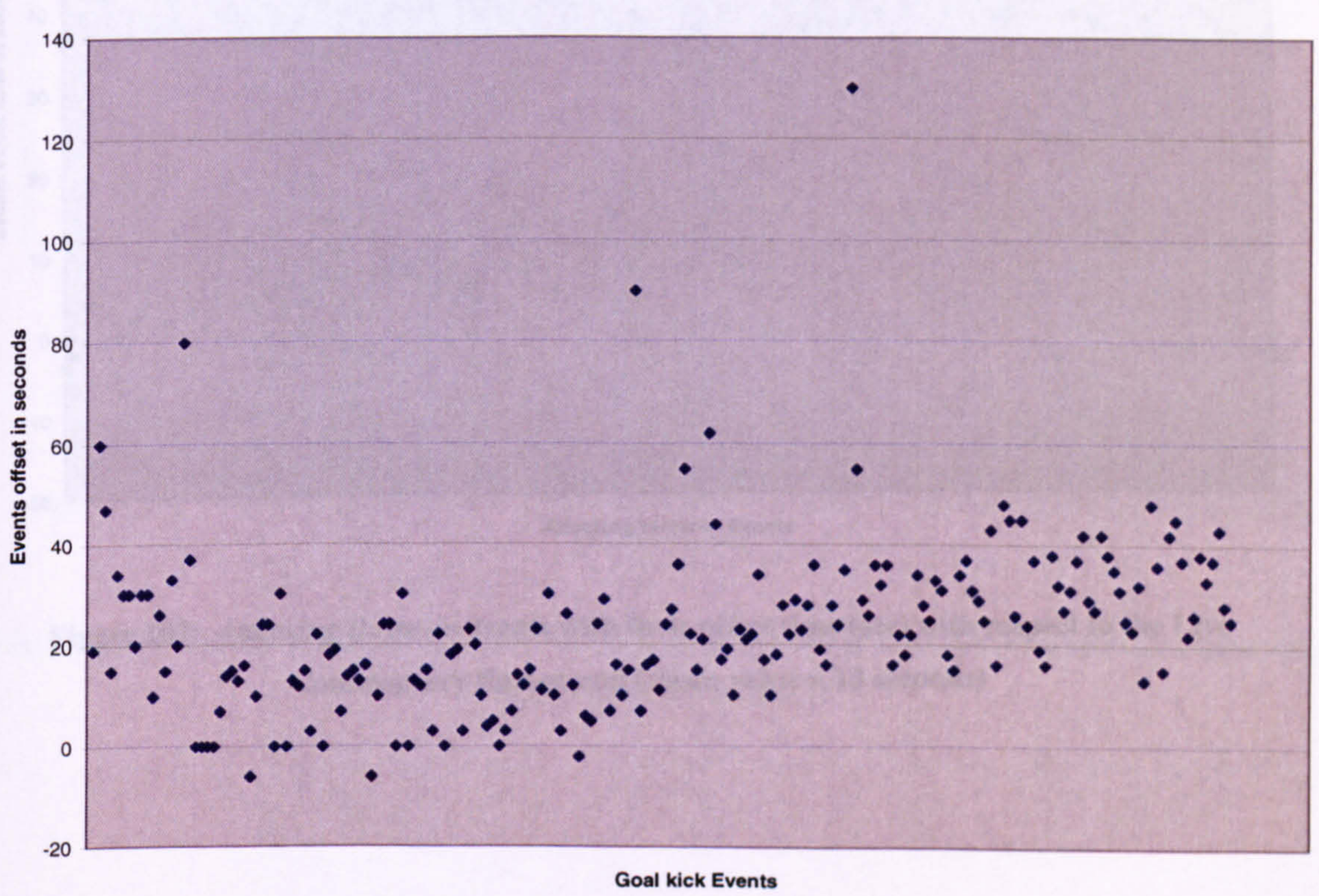


Figure 106: *Goal* events with their offset time (sec) with respect to the Live Commentary time-stamp (mean value = 22)

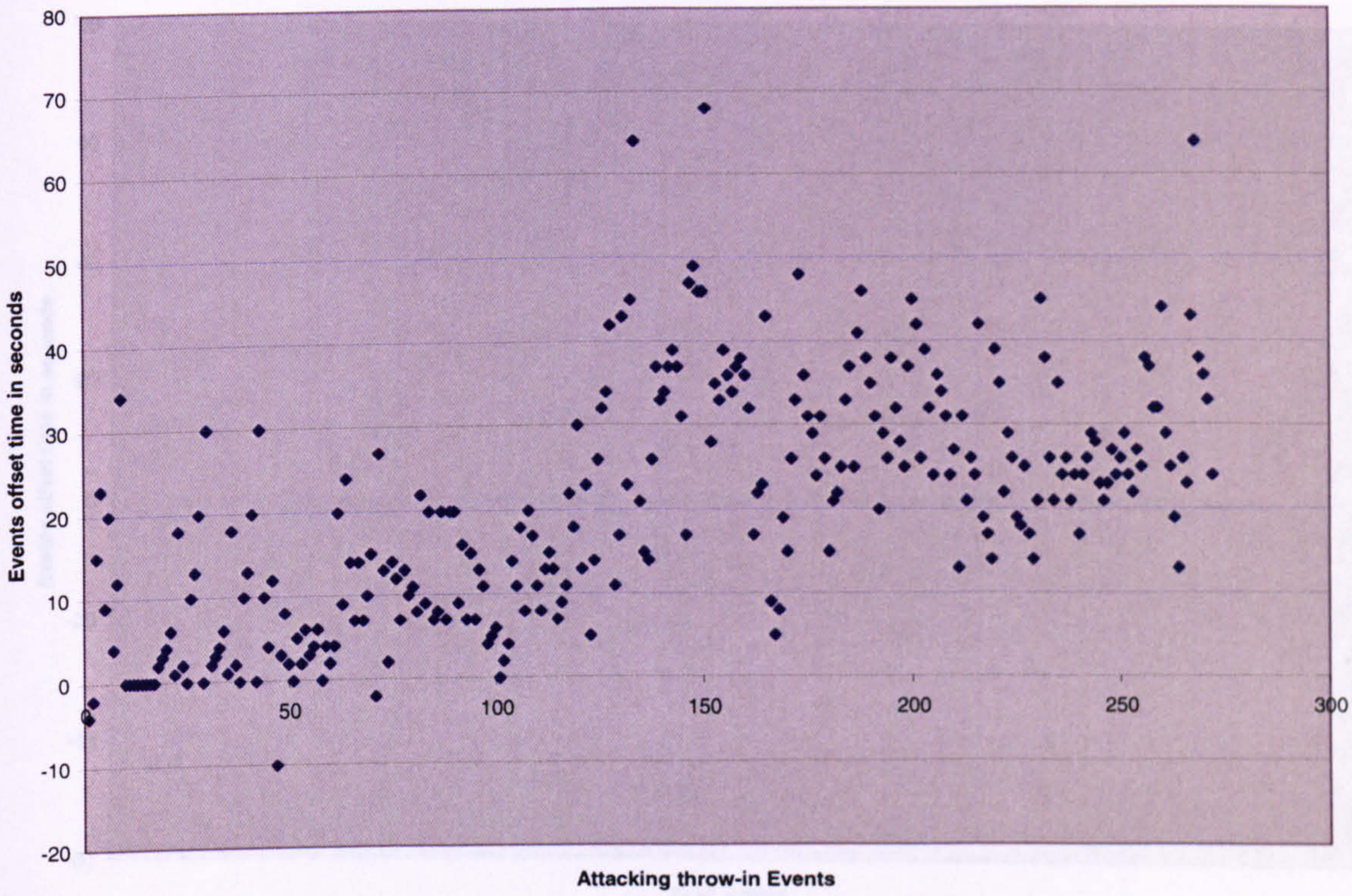


Figure 107: *Attacking throw-in* events with their offset time (sec) with respect to the Live Commentary time-stamp (mean value = 13 seconds)

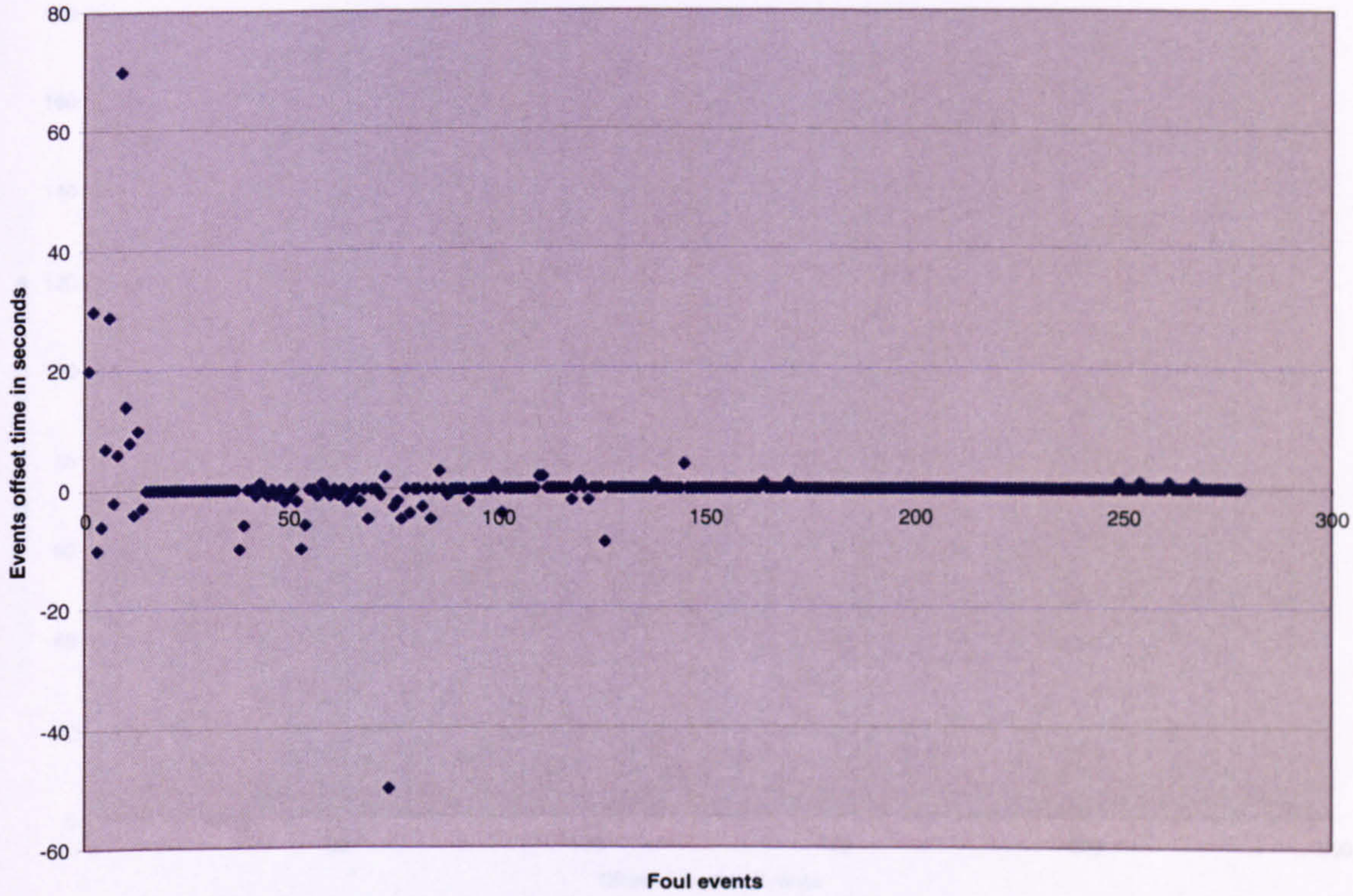


Figure 108: *Foul* events with their offset time (sec) with respect to the Live Commentary time-stamp (mean value = 0 seconds)

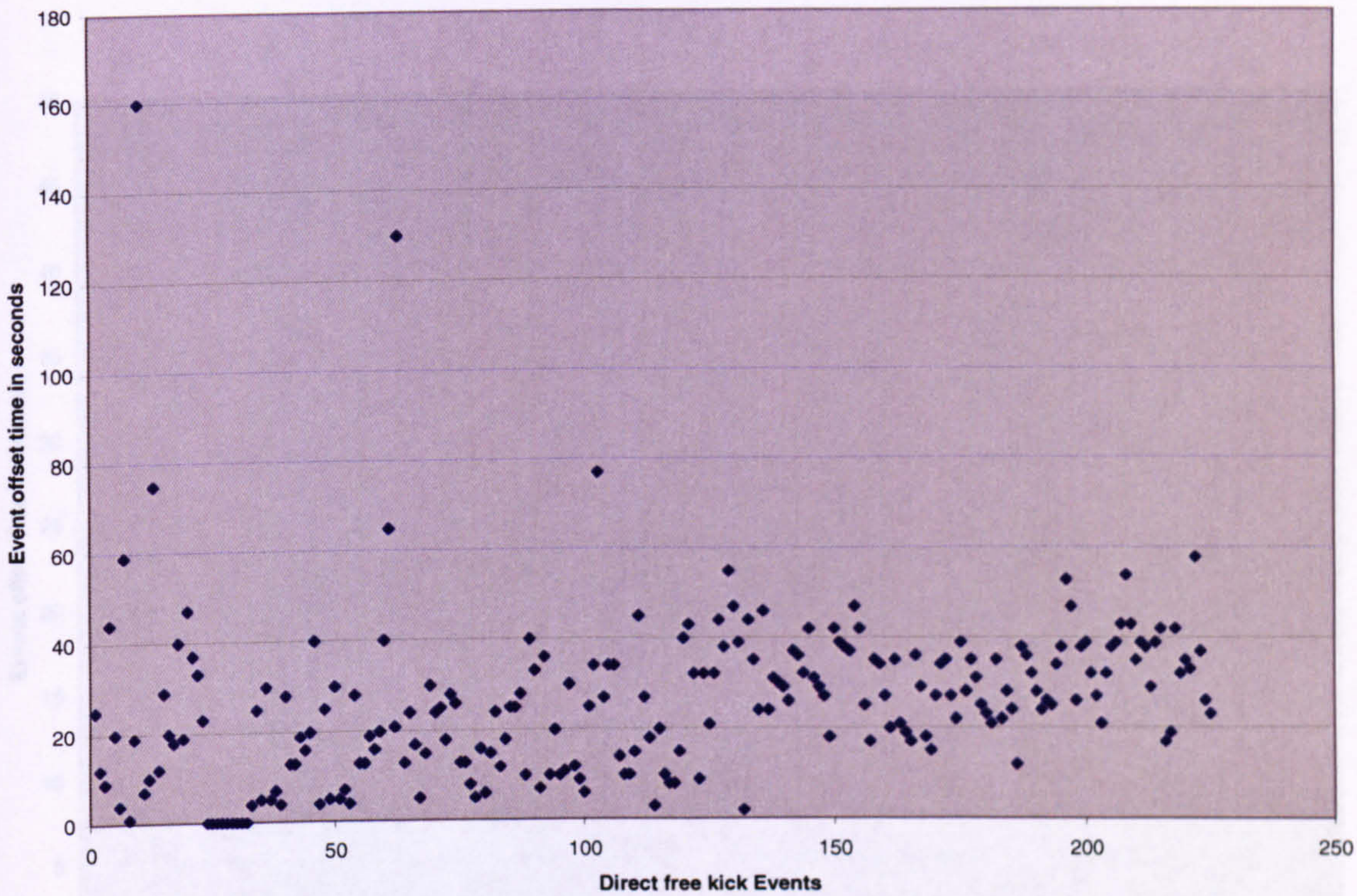


Figure 109: ***Direct free kick*** **events with their offset time (sec) with respect to the Live Commentary time-stamp (mean value = 24)**

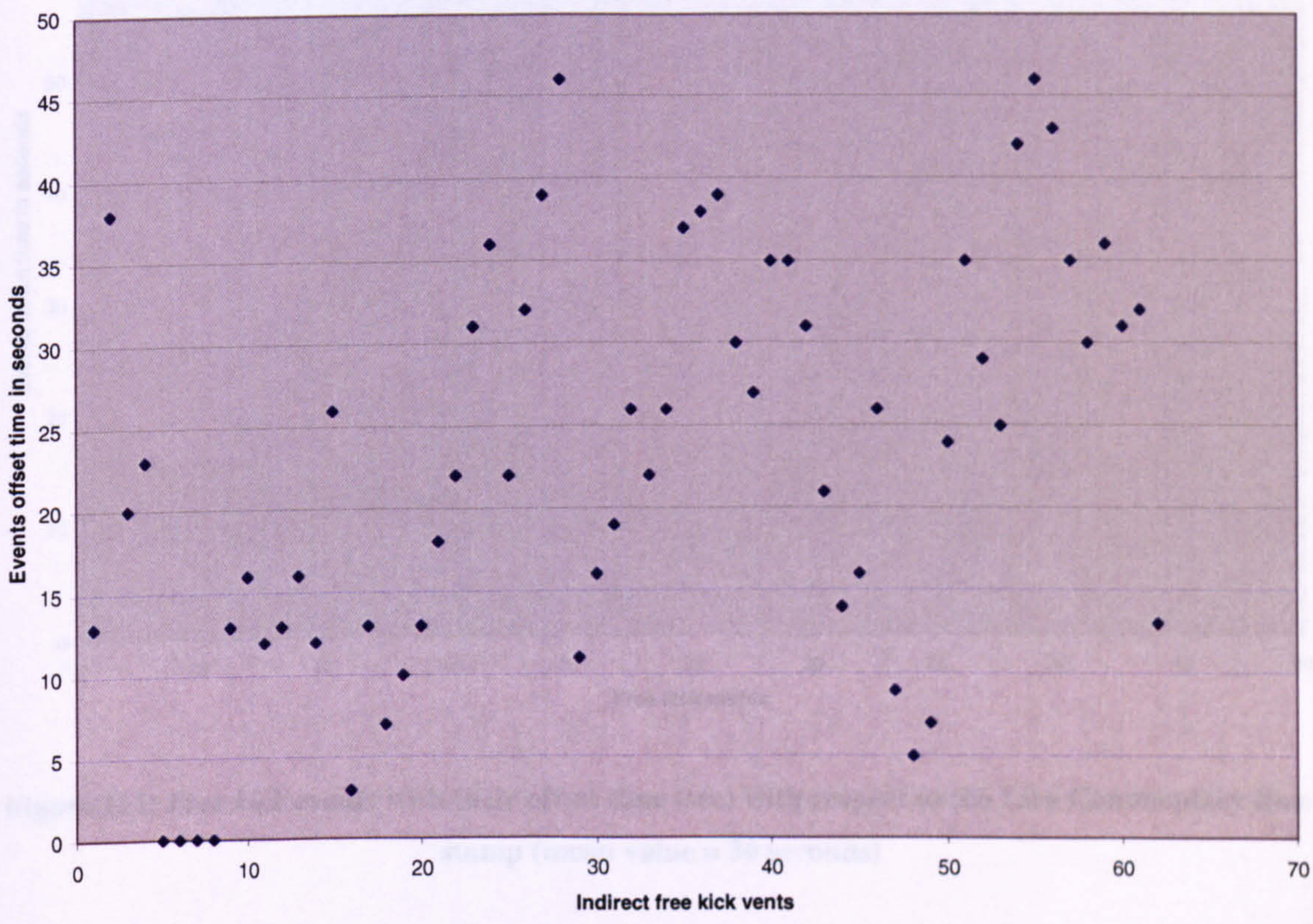


Figure 110: ***Indirect free kick*** **events with their offset time (sec) with respect to the Live Commentary time-stamp (mean value = 22.5 seconds)**

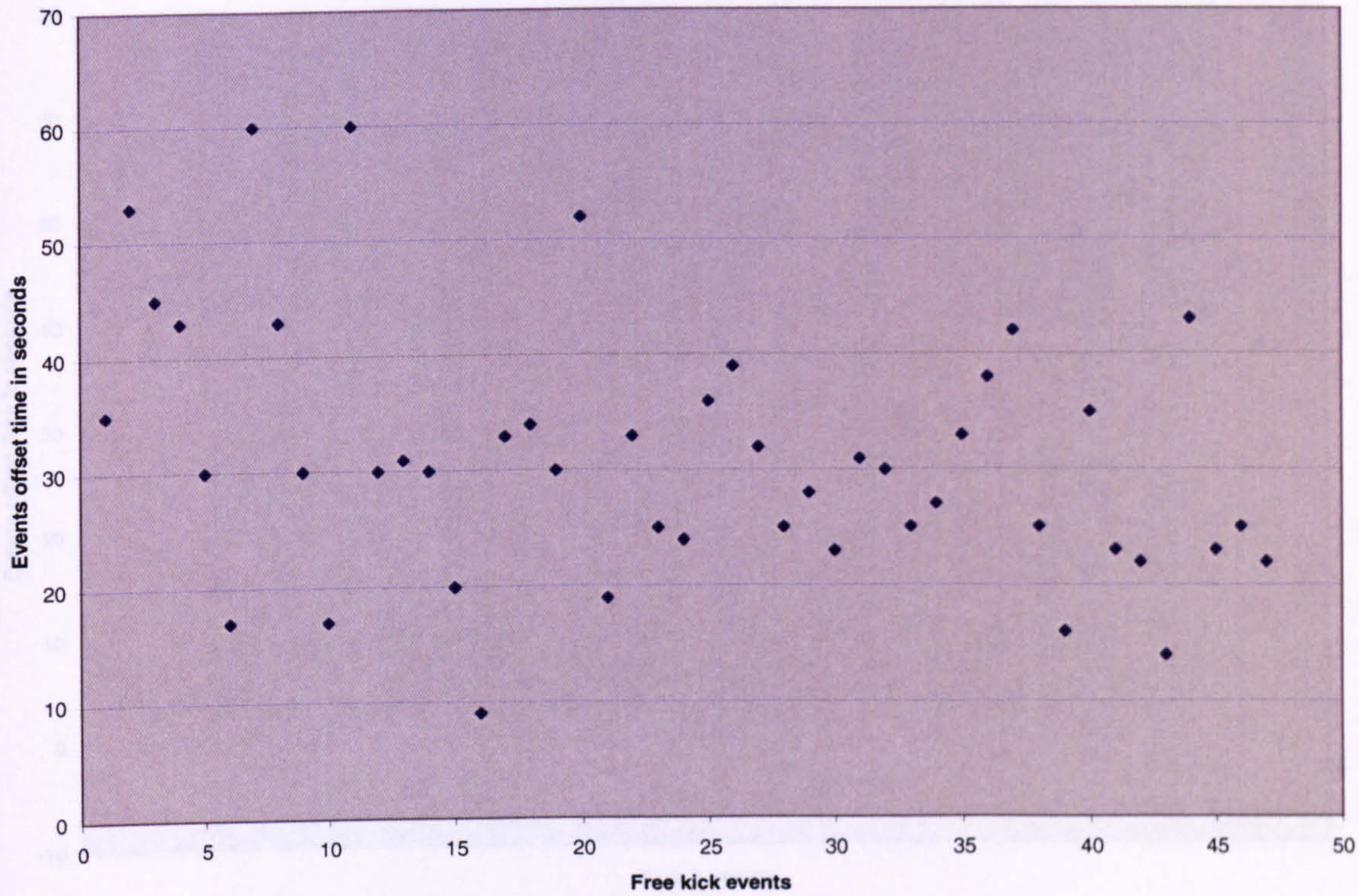


Figure 111: *Free kick* events with their offset time (sec) with respect to the Live Commentary time-stamp (mean value = 30 seconds)

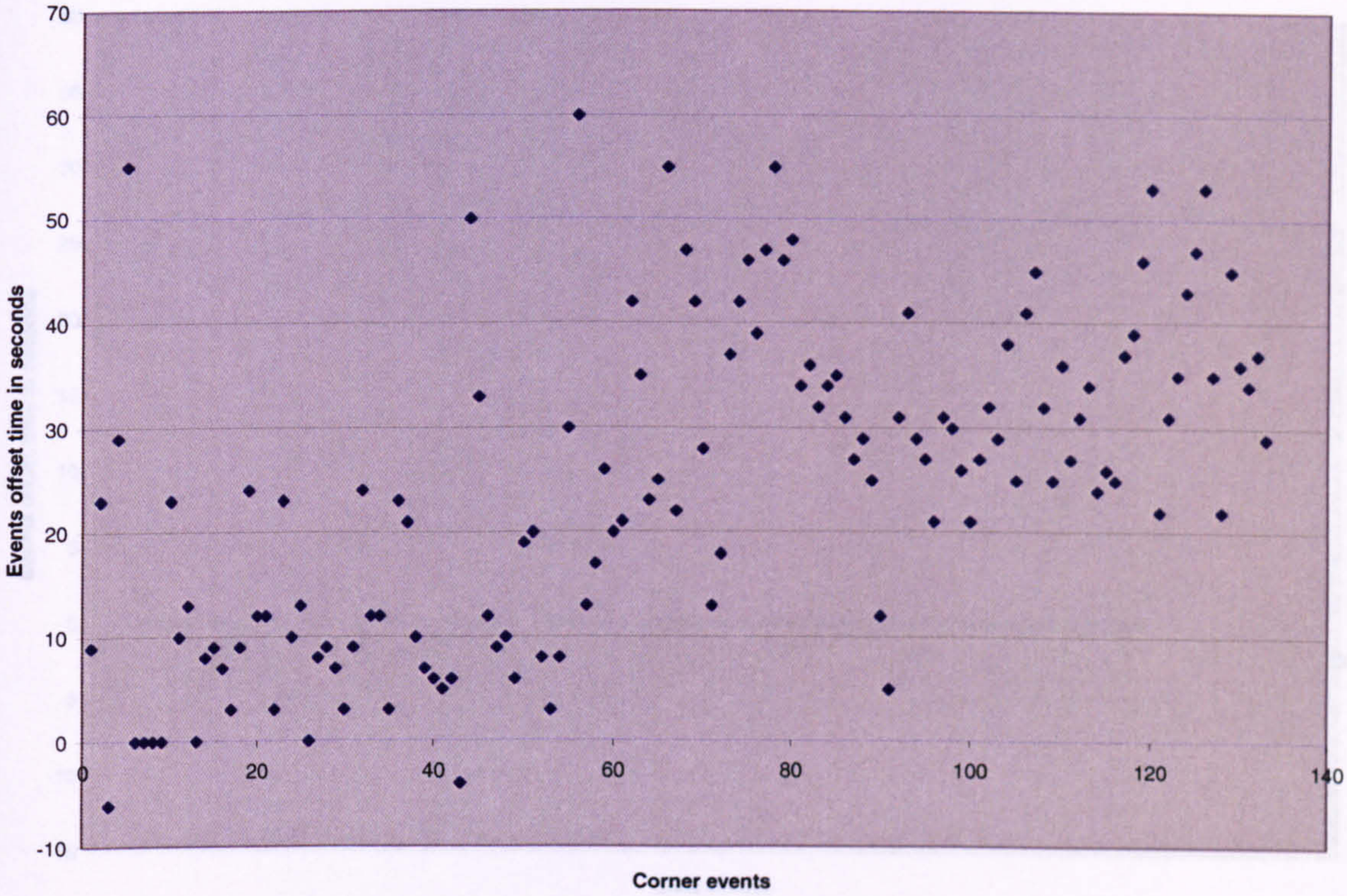


Figure 112: *Corner* events with their offset time (sec) with respect to the Live Commentary time-stamp (mean value = 25 seconds)

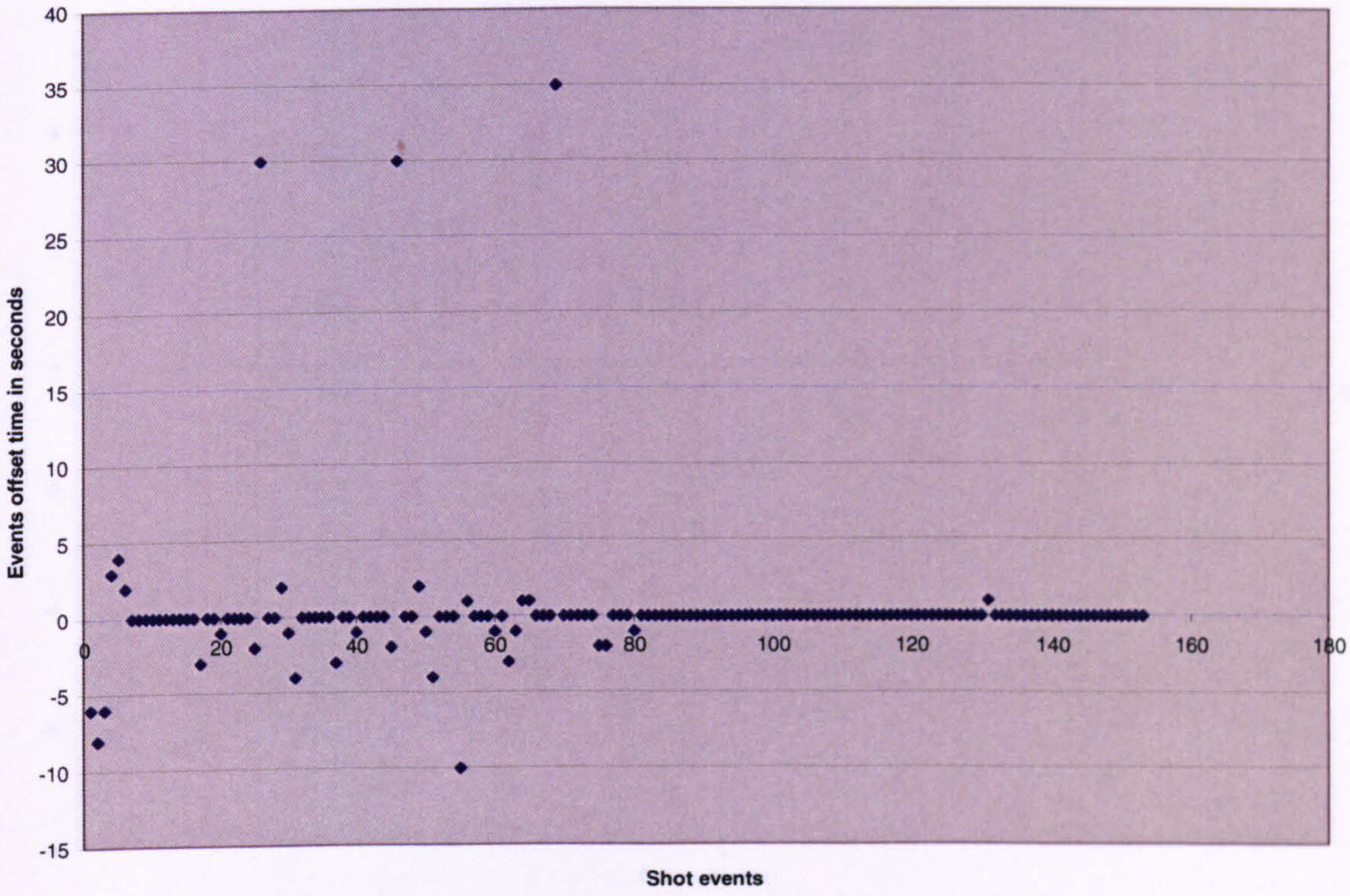


Figure 113: *Shot* events with their offset time (sec) with respect to the Live Commentary time-stamp (mean value = 0)

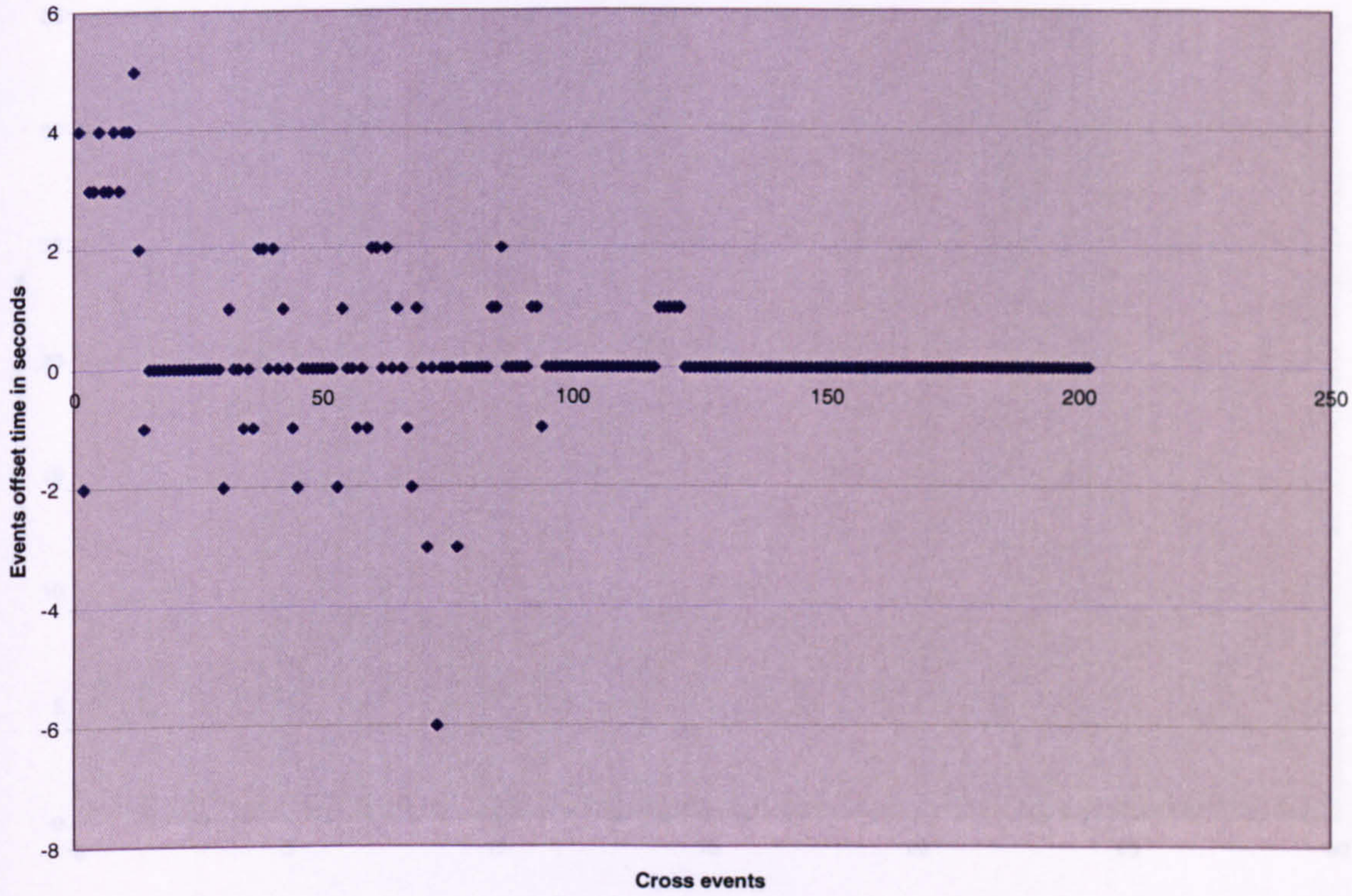


Figure 114: *Cross* events with their offset time (sec) with respect to the Live Commentary time-stamp (mean value = 0)

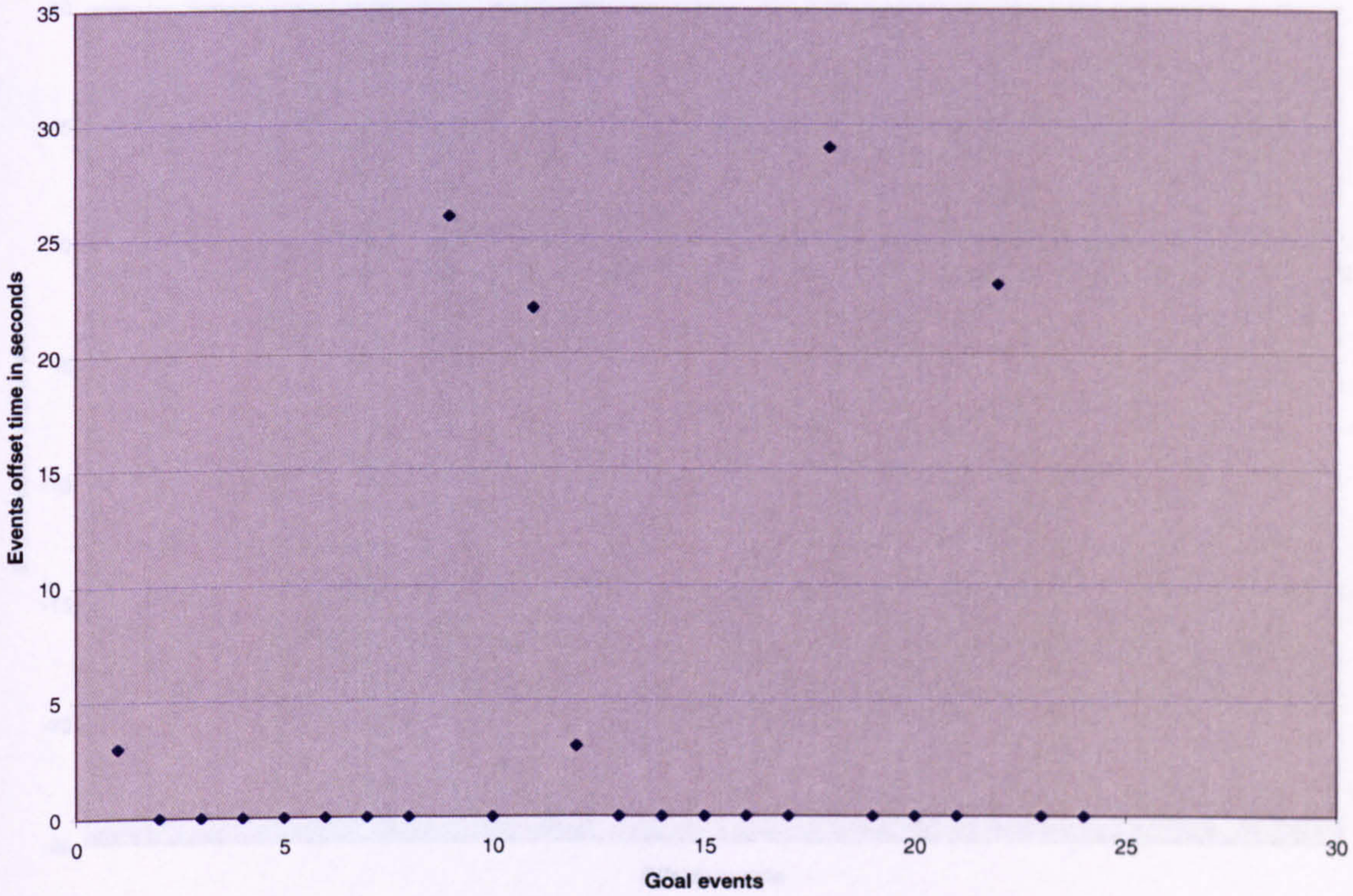


Figure 115: *Goal* (scoring) events with their offset time (sec) with respect to the Live Commentary time-stamp (mean value = 0)

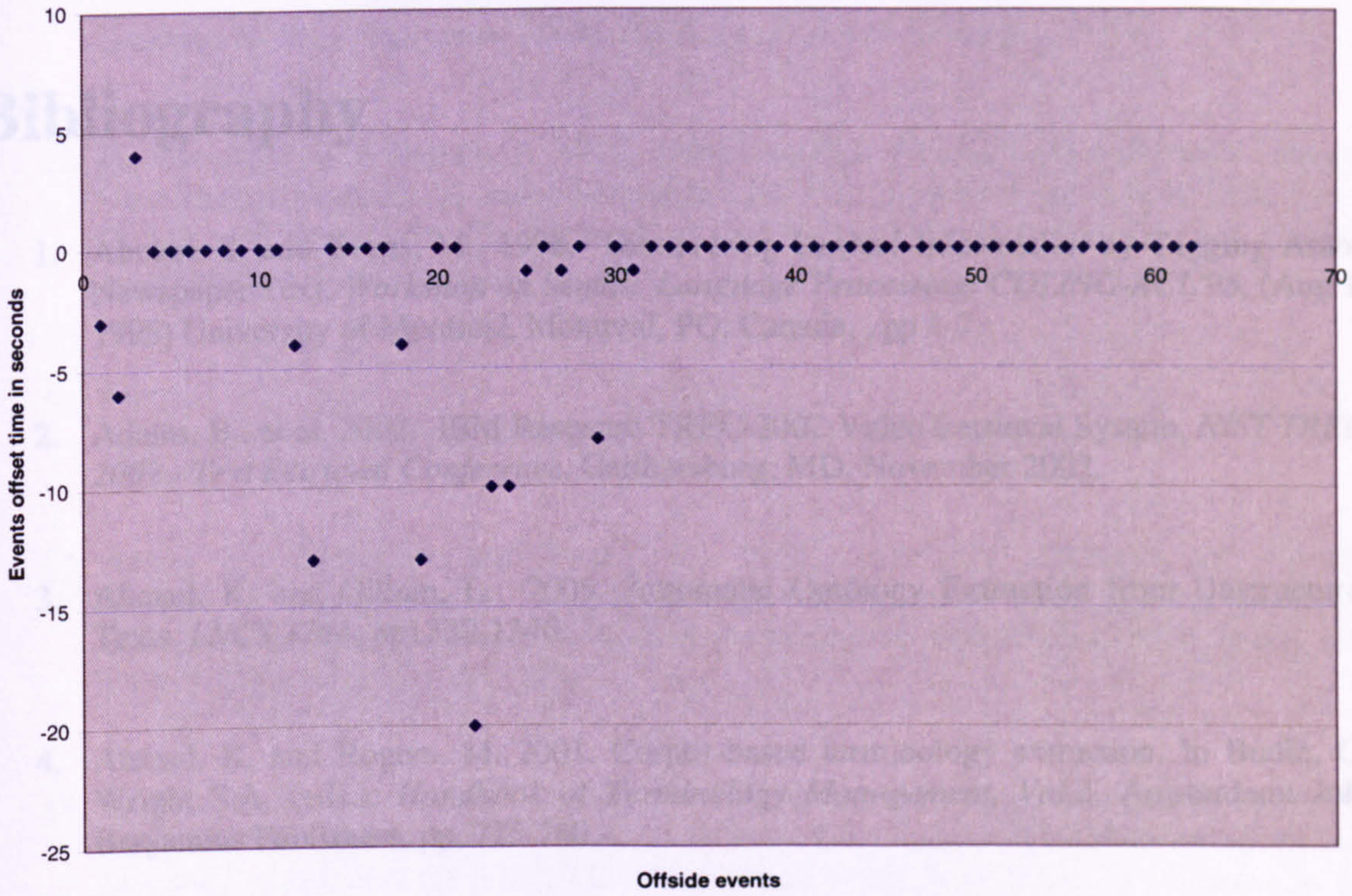


Figure 116: *Offside* events with their offset time (sec) with respect to the Live Commentary time-stamp (mean value = 0)

Bibliography

1. Abuleil, S. and Evens, M., 1998. Discovering Lexical Information by Tagging Arabic Newspaper Text, *Workshop on Semitic Language Processing. COLING-ACL'98*, (Aug 16 1998) University of Montreal, Montreal, PQ, Canada, , pp 1-7.

2. Adams, B., et al. 2002. IBM Research TREC-2002 Video Retrieval System, *NIST TREC-2002 - Text Retrieval Conference*, Gaithersburg, MD, November 2002.

3. Ahmad, K. and Gillam, L. 2005. Automatic Ontology Extraction from Unstructured Texts. *LNCS 3761*, pp1330-1346.

4. Ahmad, K. and Rogers, M. 2001. Corpus-based terminology extraction. In Budin, G., Wright S.A. (eds.): *Handbook of Terminology Management, Vol.2*. Amsterdam: John Benjamins Publishers, pp. 725-760.

5. Ahmad, K., Temizel, T., and Ahmad, S. 2004. Summarizing Time Series: Learning Patterns in 'Volatile' Series. In (Eds.) Z.R. Yang, R. Everson, and H. Yin. *Proc. of 5th Int. Conf. on Intelligent Data Engineering and Automated Learning (Exeter, UK, 25-27 August 2004)*, LNCS Vol. 3177, D-Side publication, pp. 523-532.

6. Alittihad Newspaper. 2007. [online]. Available from: http://www.alittihad.ae. [Accessed 2007].

7. Alshuth, P., Hermes, T., Klauck, C., Kreyß, J., and Röper, M. 1996. IRIS - a system for image and video retrieval. In (eds.) M. Bauer, K. Bennet, M. Gentleman, H. Johnson, K. Lyons, and J. Slonim. *Proc. of the 1996 Conference of the Centre for Advanced Studies on Collaborative Research* (Toronto, 12 – 14 November 1996), IBM Press, pp. 2.

8. Amirijoo, M., TeanoviC, A., and Nadjm-Tehrani, S. 2004. Raising motivation in real-time laboratories: the soccer scenario. In *Proc. of the 35th SIGCSE Technical Symposium on Computer Science Education*, (Norfolk, 03 - 07 March 2004), New York: ACM Press, pp. 265-269.

9. Ando, R., Shinoda, K., Furui, S., and Mochizuki, T. 2006. Robust scene recognition using language models for scene contexts. In *Proc. of the 8th ACM International Workshop on Multimedia Information Retrieval*, (Santa Barbara, 26 – 27 October 2006), New York: ACM Press, pp. 99-106.

10. Andrade, E., Khan, E., Woods, J., and Ghanbari, M. 2003. Player classification in interactive sport scenes using prior information region space analysis and number

recognition. In *Proc. of International Conference on Image Processing 2003 – Volume 3*, (Barcelona, 14-17 September 2003). IEEE, pp. III - 129-32 vol.2.

11. Barnbrook, G., and Sinclair, J. 1993. The Automatic Analysis of Dictionaries - Parsing Cobuild Explanations. EC research contract ET-10/51, progress report 2. In (eds) Baker, Francis and Tognini-Bonelli, Elena, *Text and Technology: in honour of John Sinclair.* Amsterdam: John Benjamins

12. BBC Sport. [online]. Available from: http://news.bbc.co.uk/sport2/hi/football/default.stm. [Accessed 2003-2007].

13. Bertini, M., Bimbo, A., and Torniai, C. 2006. Automatic annotation and semantic retrieval of video sequences using multimedia ontologies. In *Proc. of the 14th Annual ACM International Conference on Multimedia* (Santa Barbara, 23-27 October 2006), New York: ACM Press, pp. 679 – 682.

14. Bertini, M., Bimbo, A., and Torniai, C. 2005. Enhanced ontologies for video annotation and retrieval. In *Proc. of the 7th ACM SIGMM International Workshop on Multimedia Information Retrieval* (Hilton, 10 – 11 November 2005), New York: ACM Press, pp. 89-96.

15. Bertini, M., Bimbo, A., Torniai, C., Cucchiara, R., and Grana, C. 2006. MOM: multimedia ontology manager. A framework for automatic annotation and semantic retrieval of video sequences. In *Proc. of the 14th Annual ACM International Conference on Multimedia* (Santa Barbara, 23 – 27 October 2006), New York: ACM Press, pp. 787-788.

16. Bertini, M., Bimbo, A., and Nunziati, W. 2006. Automatic detection of player's identity in soccer videos using faces and text cues. In *Proc. of the 14th Annual ACM International Conference on Multimedia*, (Santa Barbara, 23 – 27 October 2006), New York: ACM Press, pp. 663-666.

17. Bezerra, F., and Lima, E. 2006. Low cost soccer video summaries based on visual rhythm. In *Proc. of the 8th ACM International Workshop on Multimedia Information Retrieval,* (Santa Barbara, 26 – 27 October 2006), New York: ACM Press, pp. 71-80.

18. Blank, G. 1989. A Finite and Real-Time Processor for Natural Language. *Communications of the ACM, Vol. 32, Issue 10* (October 1989), New York: ACM Press, pp. 1174-1189.

19. Boyland, J. 2005. Remote attribute grammars. *Journal of the ACM (JACM),* Volume 52, Issue 4 (July 2005), New York: ACM Press, pp. 627-687.

20. Carbonell, J. and Hayes, J. 1983. Recovery Strategies for Parsing Extragrammatical Language. *American Journal of Computational Linguistics*, Vol. 9, No. 3-4, 1983, pp. 123-146.

21. Cheng, D. 2007. *Corpus Analysis and Market Sentiment.* Unpublished Ph.D. thesis. University of Surrey.

22. Cleenewerck, T., and D'Hondt, T. 2005. Disentangling the implementation of local-to-global transformations in a rewrite rule transformation system. In *Proc. of the 2005 ACM Symposium on Applied Computing*, (Santa Fe, 13 – 17 March 2005), New York: ACM Press, pp. 1398-1403.

23. Cunningham, C., Maynard, D., Bontcheva, K., and Tablan, V. 2002. GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications. In *Proc. of the 40th Anniversary Meeting of the Association for Computational Linguistics (ACL'02)*. Philadelphia, July 2002.

24. Daimi, K. 2002. Using Modified Semantic Grammar to Generate Natural Language Interfaces. In *Proc. of Applied Informatics,* (Innsbruck, February 18 – 21), ACTA Press.

25. Daniels M., and Meurers, W. 2004. A grammar formalism and parser for linearization-based HPSG. In *Proc. of the 20th International Conference on Computational Linguistics,* (Geneva, 2004), Morristown: Association for Computational Linguistics, Article No. 169.

26. Declerck, T., Wittenburg, P., and Cunningham, H. 2001. The automatic generation of formal annotations in a multimedia indexing and searching environment. *In Proc. of the Workshop on Human Language Technology and Knowledge Management - Volume 2001*, (Toulouse, 06 - 07 July 2001), Morristown: Association for Computational Linguistics, pp. 1-8.

27. Tjondronegoro, D., Chen, Y., and Pham, B. 2006. Extensible detection and indexing of highlight events in broadcasted sports video. In *Proc. of the 29th Australasian Computer Science Conference - Volume 48* (Hobart, 16 – 19 January 2006), Darlinghurst: Australian Computer Society, Inc., pp. 237-246.

28. Dikovsky, A. 2001. Grammars for local and long dependencies. In *Proc. of the 39th Annual Meeting on Association for Computational Linguistics* (Toulouse, 06 – 11 July 2001), Morristown: Association for Computational Linguistics, pp. 156-163.

29. Dolbear, C., and Brady, M. 2003 Soccer Highlights Generation using a priori Semantic Knowledge. *International Conference on Visual Information Engineering, 2003. VIE 2003,* (Guildford, 7-9 July 2003), London: Institution of Electrical Engineers, vol. 495, pp. 202-205.

30. Dubai Sports TV. 2007. [online]. Available from: http://www.dubaisports.ae/home.asp. [Accessed 2007].

31. Gabsdil, M., and Lemon, O. 2004. Combining Acoustic and Pragmatic Features to Predict Recognition Performance in Spoken Dialogue Systems. In *Proc. of the 42nd Annual Meeting of the Association for Computational Linguistics*, (Barcelona, July 21 – 26, 2004), New York: ACM Press. Article No. 343.

32. Garside, R. 1987. The CLAWS Word-tagging System. In (eds.) R. Garside, G. Leech and G. Sampson (eds), *The Computational Analysis of English: A Corpus-based Approach.* London: Longman.

33. Ghoshal, A., Ircing, P., and Khudanpur, S. 2005. Hidden Markov models for automatic annotation and content-based retrieval of images and video. In *Proc. of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (Salvador, 15 – 19 August 2005), New York: ACM Press, pp. 544-551.

34. Goularte, R., Cattelan, R., Camacho-Guerrero, J., Inácio, V., and Pimentel, M. 2004. Interactive multimedia annotations: enriching and extending content. In *Proc. of the 2004 ACM Symposium on Document Engineering*, (Milwaukee, 28 - 30 October 2004), New York: ACM Press, pp. 84-86.

35. Gross, M. 1993. Local Grammars and their Representation by Finite Automata. In (ed.) Hoey, M. *Data, Description, Discourse: Papers on the English Language in Honour of John McH Sinclair.* London: HarperCollins. pp 26-38.

36. Hirst, G. 2002. Patterns of text: in honour of Michael Hoey. *Computational Linguistics Volume 28 , Issue 4 (December 2002).* Cambridge: MIT Press, pp. 560-564.

37. Huayong, L. 2004. Content-Based TV Sports Video Retrieval Based on Audio-Visual Features and Text Information. In *Proc. of the 2004 IEEE/WIC/ACM International Conference on Web Intelligence*, (20 – 24 September 2004), Washington: IEEE Computer Society, pp. 481-484.

38. Hunston, S., and Francis, G. 2000. Pattern Grammar. A corpus-driven approach to the lexical grammar of English. *Studies in Corpus Linguistics 4.* Amsterdam: John Benjamins Publishing Company.

39. King Abdulaziz City for Science and Technology (KACST). [2007]. CD-Rom Database, (+9661) 4883555. Riyadh, Saudi Arabia.

40. Kojima, A., Izumi, M., Tamura, T., and Fukunaga, K. 2000. Generating Natural Language Description of Human Behavior from Video Images. In *Procs. of International Conference on Pattern Recognition (ICPR'00) - Volume 4*, Barcelona, Spain 2000, pp. 728-731.

41. Kuramoto, M., Masaki, T., Kitamura, Y., and Kishino, F. 2002. Video Database Retrieval Based on Gestures and Its Application. *IEEE Transaction on Multimedia*, Vol. 4, No. 4. Dec 2002, pp. 500-508.

42. Lao, S., Smeaton, A., Jones, G., and Lee, H. 2004. A query description model based on basic semantic unit composite petri-nets for soccer video analysis. In *Proc. of the 6th ACM SIGMM International Workshop on Multimedia Information Retrieval*, (NewYork, 15 - 16 October 2004), New York: ACM Press, pp. 143-150.

43. Lee, Y., Papineni, K., Roukos, S., Emam, O. and Hassan, H. 2003. Language Model Based Arabic Word Segmentation. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics,* (Sapporo, 07 – 12 July), New York: ACM Press, pp. 399 - 406.

44. Lema, J., et al. 2001. Algorithms for Moving Objects Database. *The Computer Journal, Volume 46(6)*. pp. 680-712(33), Oxford: Oxford University Press.

45. Lienhart, R. 1997. Automatic text recognition for video indexing. In *Proc. of the fourth ACM International Conference on Multimedia*, (Boston, 18 - 22 November 1996), New York: ACM Press, pp. 11-20.

46. Liu, Song., Xu, Min., Yi, Haoran., Chia, Liang-Tien., and Rajan, Deepu. 2006. 'Multimodal Semantic Analysis and Annotation for Basketball Video', *EURASIP Journal on Applied Signal Processing*, Volume 2006, Article ID 32135, pp 1–13.

47. Madhwacharyula, C., Davis, M., Mulhem, P., and Kankanhalli, M. 2006. Metadata handling: A video perspective. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP), Vol. 2, Issue 4*, New York: ACM Press, pp. 358-388.

48. Madnani, N. 2007. *Getting Started on Natural Language Processing with Python.* The ACM Student Journal: Crossroads, issue 13.4, pp. 10 – 15.

49. Matthews, B. 2006. Grammar, meaning and movement-based interaction. In *Proc. of the 20th Conference of the Computer-Human Interaction Special Interest Group (CHISIG) of Australia on Computer-Human Interaction: Design: Activities, Artefacts and Environments* (Sydney, 20 – 24 November 2006), New York: ACM Press, pp. 405-408.

50. McKeown, K., Barzilay, R., Chen, J., Elson, D., Evans, D., Klavans, J., Nenkova, A., Schiffman, B., and Sigelman, S. 2003. In *Proc. of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology: Demonstrations - Volume 4.* (Edmonton, May 27 - June 01, 2003), Morristown: Association for Computational Linguistics, pp. 15-16.

51. Naidoo, W., and Tapamo, J. 2006. Soccer video analysis by ball, player and referee tracking. In (eds.) B. Judith and K. Derrick. *Proc. of the 2006 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists on IT research in developing countries.* (Somerset West, 09 – 11 October 2006), Republic of South Africa: South African Institute for Computer Scientists and Information Technologists, pp. 51-60.

52. Nakanishi, H., Miyao, Y., and Tsujii, J. 2005. Probabilistic Models for Disambiguation of an HPSG-Based Chart Generator. In *Proc. of the Ninth International Workshop on Parsing Technologies (IWPT),* Vancouver. pp. 93–102.

53. Nam, J., and Tewfik, A. 1999. Dynamic Video Summarization and Visualization. In *Proc. of the seventh ACM International Conference on Multimedia (Part 2),* (Orlando, October 30-November 5, 1999), New York: ACM Press, pp. 53-56.

54. Navalpakkam, V., and Itti, L. 2003. Sharing Resources: Buy Attention, Get Recognition. In *Procs. International Workshop on Attention and Performance in Computer Vision (WAPCV'03),* Graz, Austria, Jul 2003.

55. Negnevitsky, M. 2002. *Artificial Intelligence: A Guide to Intelligent Systems.* Harlow, UK: Pearson Education Ltd.

56. Neyret, F., and Cani, M. 1999. Pattern-based texturing revisited. In *Proc. of the 26th Annual Conference on Computer Graphics and Interactive Techniques.* New York: ACM Press/Addison-Wesley Publishing Co. pp. 235-242.

57. Nijholt, A., Akker, O., and Jong, F. 2003. *Language Interpretation and Generation in Football Commentary.* University of Twente, Centre for Telematics and Information Technology, TKI-Parlevink Research Group.

58. Oliver, M. 2004. Automatic Processing of Local Grammar Patterns. In *Proc. of the 7th Annual Colloquium for the UK Special Interest Group for Computational Linguistics,* University of Birmingham, (6-7 January 2004), p.166-171.

59. Petkovic, M., Jonker, W., and Zivkovic, Z. 2001. Recognizing strokes in tennis videos using hidden Markov Models. In *Proc. of Intl. Conf. on Visualization, Imaging and Image Processing,* Marbella, Spain, 2001.

60. Petkovic, M., and Jonker, W. 2000. A Framework for Video Modelling, *18th IASTED Conference on Applied Informatics,* Innsbruck, Austria, 2000.

61. Protégé. 2006. [online]. Available from: http://protege.stanford.edu. [Accessed 2006].

62. Quenot, G., Moraru, D., Besacier, L., and Mulhem, P. 2002. CLIPS as TREC-11 experiments in video retrieval. *SP 500-251 The Eleventh Text Retrieval Conference (TREC 2002),* Gaithersburg, Maryland, November 19-22, 2002.

63. Quirk, R., Greenbaum, S., Leech, G., and Svartvik., J. 1985. *A Comprehensive Grammar of the English Language*, London, Longman 1985.

64. Rittscher, J., Blake, A., Hoogs, A., and Stein, G. 2003. Mathematical Modelling of Animate and Intentional Motion. *Phil. Trans. R. Soc. Lond. B* (2003) 358, 475–490.

65. Ronard, R., and Thuong, T. 2003. A Framework for Aligning and Indexing Movies with their Script. *IEEE International Conference on Multimedia and Expo*, pp 21-24. July 2003. Baltimore, USA.

66. Saragiotis, P., Vrusias, B. and Ahmad, K. 2005. Learning to Classify a Collection of Images and Texts. In *Proc. Of European Symposium on Artificial Neural Network.* Bruges, (27-29 April 2005), p. 551-556.

67. Sattar, H., 2002. *Fundamentals of Classical Arabic: Volume I: Conjugating Regular Verbs and Derived Nouns.* Chicago: Faqir Publications.

68. Smadja, F. 1994. Retrieving Collocations from Text: Xtract. In (ed.) S. Armstrong*Using Large Corpora*. London: MIT Press.

69. Smeaton, A., and Over, P. 2002. The TREC-2002 Video Track Report, page 69 *NIST Special Publication 500-251: The Eleventh Text Retrieval Conference* (TREC 2002), Gaithersburg, Maryland, November 19-22, 2002.

70. Snoek, C., and Worring, M. 2003. Time Interval-based Modelling and Classification of Events in Soccer Video. In *Proc. of the 9th Annual Conference of the Advanced School for Computing and Imaging (ASCI),* Heijen, The Netherlands, June 2003.

71. Snoek, C., and Worring, M. 2005. Multimodal video indexing: A review of the state-of-the-art. *Multimedia Tools and Applications.* Volume 25, Number 1, pp. 5-35.

72. Snoek, C., and Worring, M. 2005. Multimedia Event-based Video Indexing using Time Intervals. *IEEE Transactions on Multimedia*, Volume 7, Issue 4, pp. 638-647, 2005.

73. Snoek, C., Worring, M. and Hauptmann, A. 2006. Learning rich semantics from news video archives by style analysis. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP)*, Volume 2, Issue 2. New York: ACM Press, pp. 91-108.

74. Srikanth, M., Varner, J., Bowden, M., and Moldovan, D. 2005. Exploiting ontologies for automatic image annotation. In *Proc. of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (Salvador, 15 – 19 August 2005), New York: ACM Press, pp. 552-558.

75. Su, C., Liao, H., and Fan, K. 2005. A motion-flow-based fast video retrieval system. In *Proc. of the 7th ACM SIGMM International Workshop on Multimedia Information Retrieval*, (Hilton, 10 - 11 November 2005), New York: ACM Press, pp. 105-112.

76. System Quirk: Language Engineering Workbench. University of Surrey. Available from: http://www.computing.surrey.ac.uk/SystemQ. [Accessed 2004-2007].

77. Tahaghoghi, S., Williams, H., Thom, J., and Volkmer, T. 2005. Video Cut Detection using Frame Windows. In *Proc. of the Twenty-eighth Australasian Conference on Computer Science - Volume 38* (Newcastle, 2005), Darlinghurst: Australian Computer Society, Inc. pp. 193-199.

78. Tanaka-Ishii, K., Hasida, K., and Noda, J. 1998. Reactive content selection in the generation of real-time soccer commentary. In *Proc. of the 17th International Conference on Computational Linguistics - Volume 2*, (Montreal, 10 - 14 August 1998), Morristown: Association for Computational Linguistics, pp. 1282-1288.

79. Tanaka, K., Nakashima, H., Noda, I., Hasida, K., Frank, I., and Matsubara, H. 1998. MIKE: an automatic commentary system for soccer. In *Proc. IEEE International Conference on Multi Agent Systems,* (Paris, 3-7 July 1998), pp. 285-292.

80. Text Analysis International (TextAI). [online]. Available from: http://www.textanalysis.com. [Accessed 2006]

81. Tgondronegoro, D., and Chen, Y. 2004. Integration Highlights for More Complete Sports Video Summarization. *IEEE MultiMedia Magazine*, Vol 11, No 4, pp. 22-37.

82. Tjondronegoro, D. 2005. *Content-based Video Indexing for Sports Applications using Integrated Multi-Modal Approach.* Unpublished Ph.D. thesis, Deakin University, Australia.

83. Tjondronegoro, D., Chen, Y., and Pham, B. 2006. Extensible detection and indexing of highlight events in broadcasted sports video. In *Proc. of the 29th Australasian Computer Science Conference - Volume 48*, (Hobart, 16 - 19 January 2006), Darlinghurst: Australian Computer Society, Inc. pp. 237-246.

84. Tong, X., Liu, Q., Duan, L., Lu, H., Xu, C., and Tian, Q. 2005. A unified framework for semantic shot representation of sports video. In *Proc. of the 7th ACM SIGMM International Workshop on Multimedia Information Retrieval*, (Hilton, 10 - 11 November 2005), New York: ACM Press, pp. 127-134.

85. Volkmer, T., Smith, J., and Natsev, A. 2005. A Web-based System for Collaborative Annotation of Large Image and Video Collections. In *Proc. of the 13th Annual ACM International Conference on Multimedia* (Hilton, 06 - 11 November 2005), New York: ACM Press, pp. 892-901.

86. Wang, J., and Parameswaran, N. 2004. Survey of sports video analysis: research issues and applications. In *Proc. of the Pan-Sydney Area Workshop on Visual Information Processing*, Darlinghurst: Australian Computer Society, Inc. pp. 87.90.

87. Wang, J., Xu. C., Chng, E., Duan, L., Wan, K., and Tian, Q. 2005. Automatic generation of personalized music sports video. In *Proc. of the 13th Annual ACM International Conference on Multimedia* (Hilton, 06 - 11 November 2005), New York: ACM Press, pp. 679-682.

88. Ward, W., and Pellom, B. 1999. The CU Communicator System. In *Proc.of the IEEE Automatic Speech Recognition and Understanding Workshop (ASRU 1999).* Keystone, Colorado, USA, pp. 341-344.

89. Wolf, C., Doermann, D., and Rautiainen, M. 2002. Video Indexing and Retrieval at UMD. In: NIST Special Publication: SP 500-251 *The Eleventh Text Retrieval Conference (TREC 2002)*, Gaithersburg, Maryland, November 19-22, 2002.

90. Xu, C., Wang, J., Wan, K., Li, Y., and Duan, L. 2006. Live sports event detection based on broadcast video and web-casting text. In *Proc. of the 14th Annual ACM International Conference on Multimedia,* (Santa Barbara, 23 – 27 October 2006), New York: ACM Press, pp. 221-230.

91. Yankova, M., and Boytcheva, S. 2003. Focusing on Scenario Recognition in Information Extraction. In *Proc. of the tenth Conference on European Chapter of the Association for Computational Linguistics - Volume 2*, (Budapest, 12 – 17 April 2003), Morristown: Association for Computational Linguistics, pp. 41-48.

92. Yu, X., Xu, C., Leong, H., Tian, Q., Tang, Q., and Wan, K. 2003. Content analysis: Trajectory-based ball detection and tracking with applications to semantic analysis of broadcast soccer video. In *Proc. of the eleventh ACM International Conference on Multimedia* (Berkeley, 02 - 08 November 2003), New York: ACM Press, pp. 11-20.

93. Zelle, J., and Mooney, R. 1993. Learning Semantic Grammars with Constructive Inductive Logic Programming. *In Proc. of the Eleventh National Conference on Artificial Intelligence.* Washington, DC: AAAI Press / MIT Press, pp 817-822.

94. Zhong, D., and Chang, S. 2000. *Structure Parsing and Event Detection for Sports Video using Domain Models.* Columbia University ADVENT technical report #091, Dec 2000.

95. Zhang, R., Sarukkai, R., Chow, J., Dai, W., and Zhang Z. 2006. Joint categorization of queries and clips for web-based video search. In *Proc. of the 8th ACM International Workshop on Multimedia Information Retrieval,* (Santa Barbara, 26 – 27 October 2006), New York: ACM Press, pp. 193-202.

96. Zhou, X., Zhou, X., and Shen, H. 2007. Efficient similarity search by summarization in large video database. In (eds.) J. Bailey and A. Fekete. *Proc. of the eighteenth Conference on Australasian Database Volume 63* (Ballarat, January 30 - February 02, 2007), Darlinghurst: Australian Computer Society, Inc., pp. 161-167.