UNIVERSITY OF BAHRAIN

College of Arts

Department of English
Language and Literature

Analysis of Lexical Bundles in Football Commentary by Professionals and Fans: A Systemic Functional Approach

A Thesis Submitted in Partial Fulfillment of the Requirements for the Master's Degree in Applied English Studies

Submitted by
Mohamed Ebrahim A Nabi Ebrahim
ID Number: 20122284

Supervised by
Dr. Haleema Al-Aali
Assistant Professor
University of Bahrain

Kingdom of Bahrain
May 2022

UNIVERSITY OF BAHRAIN

College of Arts

Department of English Language and Literature

Analysis of Lexical Bundles in Football Commentary by Professionals and Fans: A Systemic Functional Approach

A Thesis Submitted in Partial Fulfillment of the Requirements for the Master's Degree in Applied English Studies

Submitted by

Mohamed Ebrahim A Nabi Ebrahim

ID Number: 20122284

Supervised by

Dr. Haleema Al-Aali

Assistant Professor

University of Bahrain

Kingdom of Bahrain

May 2022

University of Bahrain

College of Arts

Department of English Language and Literature

Defense Committee's Approval Page

M.A. program: Applied English Studies

Student **Mohamed Ebrahim Hasan** ID number **20122284** defended his thesis by which he fulfilled the requirements for the Master's degree in Applied English Language Studies entitled (Analysis of Lexical Bundles in Football Commentary by Professionals and Fans: A Systemic Functional Approach).

Thus, the committee recommends awarding of M.A. Degree with the grade of: S

Defense committee

Name	Signature
Dr. Haleema Kadhem Al-Aali as Supervisor	
Dr. Nuri Ageli as Internal examiner	
Dr. Ali Al-Jasim as External examiner	

Abstract

This study adopts the notion that football is like any other scientific and academic fields in the way the proficiency of its texts is assessed. Hence, with the emergence of social media (namely, Twitter) the fans became able to describe and analyze football matches as they are being played, making the description and analysis of football matches not limited to football commentators and pundits. Therefore, we investigate and compare the linguistic aspects of the professionals and the fans using SFL Bundles, a framework that combine two theories which share the concept of formulaicity (Systemic Functional Linguistics and Lexical bundles). The aim of this investigation and comparison is to measure the difference in linguistic proficiency between professionals and fans in the discourse of football. To achieve this purpose, two corpora of 5000 words are compiled, and the clauses within these corpora are extracted in excel sheets according to the fitting SFL bundle. The results show that the SFL Material bundle is the predominant bundle in both corpora (professionals and fans). Nevertheless, the quantitative similarity is superficial as the qualitative analysis reveals that some linguistic variations are actually due to the high linguistic competency of the professionals, but other variations are due to other factors such the difference in the mood of delivery (written and spoken). The study ends with recommending further research to be done on a larger corpus, in addition to exploring the subsidiary SFL Bundles which are beyond the scope of this study.

Dedication

To my family and friends,

Without your support, I would not be able to make it this far.

Acknowledgments

I would like to express my gratitude for the supervisor of this thesis Dr. Haleema Alaali, who provided her invaluable support and feedback throughout the stages of this study. I am also deeply indebted to Dr. Dana Ebrahim who was more than generous with her expertise and precious time. Thanks are due to Dr Diana Al-Jahromi who sat me on the right path as I was preparing my proposal.

Table of Contents

Defense committee's approval page ..... B

Abstract C

Dedication ..... D

Acknowledgments ..... E

List of Tables ..... H

List of Figures I

1. Chapter one: Introduction 1

1.1. Overview 1

1.2. Research questions 4

1.3. Research Purpose & Research Gap 4

1.4. Structure of the study 6

2. Chapter Two: Literature review 7

2.1. Overview 7

2.2. Halliday's Transitivity Framework 8

2.2.1. Material clauses (process): The process of doing and happening 10

2.2.2. Mental clauses (process): The process of sensing 11

2.2.3. Relational clauses: The process of being 12

2.2.4. Subsidiary processes: behavioral, verbal, and existential 14

2.3. Lexical bundles and Formulaic Language: An SFL perspective. 17

2.3.1. Formulaic Language 18

2.3.2. Lexical bundles 19

2.3.3. Previous studies of Lexical bundles 22

2.4. Discourse of football 23

2.4.1. Simplification 24

2.4.2. Inversion 25

2.4.3. Heavy Modifiers 25

2.4.4. Tense usage 26

2.4.5. Routines 26

3. Chapter Three: Methodology 28

3.1. Theoretical Framework: SFL bundles 28

3.2. Methods of Data collection 32

3.2.1. Corpus 1: Professional corpus 32

3.2.2. Corpus 2: Fans corpus 33
3.3. Methods of data analysis 35
3.3.1. SFL Material bundle 36
3.3.2. SFL Relational bundle 38
3.3.3. SFL Mental bundle 39
3.3.4. Approaching the analysis 40
3.4. Research Ethics 41
4. Chapter Four: Findings 43
4.1. Professional corpus 43
4.1.1. SFL Material bundle 44
4.1.2. SFL Relational bundle 49
4.1.3. SFL Mental bundle 52
4.1.4. Summary of the professional corpus 53
4.2. Fans corpus 54
4.2.1. SFL Material bundle 55
4.2.2. SFL Relational bundle 58
4.2.3. SFL Mental bundle 60
4.2.4. Summary of the fans' corpus 61
5. Chapter Five: Data analysis 62
5.1. Overall analysis of SFL Material bundle 64
5.2. Overall analysis of SFL Relational bundle 69
5.3. Overall analysis of SFL mental bundle 72
5.4. Answering the research questions 73
6. Chapter Six: Conclusion 79
6.1. Summary 79
6.2. Research Recommendations 81
7. References 82

List of Tables

Table 1: components in Material processes 11
Table 2:Principal types and functions of Relational Processes 14
Table 3: Types of verbs in Behavioral clauses 16
Table 4: Commentators' teams 32
Table 5: The Twitter accounts 34
Table 6: Types under SFL Material bundle 36
Table 7: Structures of the SFL Mental bundle 39
Table 8: Frequency of SFL bundles - professional 44
Table 9: Frequency of participants in the professional corpus 45
Table 10: Frequency of types - SFL Material bundle 45
Table 11: Types and counts of SFL Relational bundle - Professional 49
Table 12: Frequency of SFL bundles- Fans' corpus 55
Table 13: Participants in the SFL Material bundle -Fans' corpus 55
Table 14: Frequency of types in SFL Material bundle 56
Table 15: Types and counts of SFL Relational bundle 59
Table 16: Frequency of SFL bundles- comparison between professional and fans 62
Table 17: Types of SFL Material bundles- count and percentage 64
Table 18: comparison of Lexical density- process verbs 68
Table 19: SFL Relational bundle- Types, counts, and percentages 69

List of Figures

Figure 1: terminology of the study 31
Figure 2: Type A of SFL Material bundle 37
Figure 3: Type B of SFL Material bundle 37
Figure 4: Type C of SFL Material bundle 38
Figure 5: Type D of SFL Material bundle 38
Figure 6: Type A of the SFL Relational bundle 39
Figure 7: Type B of the SFL Relational bundle 39
Figure 8: Excel sheet of Type A of SFL Mental bundle 40
Figure 9: Excel sheet of Type B of the SFL Mental bundle 40
Figure 10: Refusal of publishing the name of the account 42
Figure 11: Percentage of SFL bundles in professional corpus 54
Figure 12: Percentages of SFL bundles in the fans' corpus 61

1. Chapter one: Introduction

1.1. Overview

The use of language varies based on the social contexts that are created by social variables such as class, age, or gender. The study of these social variables was developed further as sociolinguists began to identify specific linguistic aspects within each type of speech context or discourse, hence, we see distinguishable linguistic aspects in the discourse of lawyers, doctors, and so on. Moreover, use of language varies based on register. Several sociolinguistic studies were conducted on the variations of using language in term of formality. An early study is of Irvine (1978) which discussed the social variables that determine the distinction between formal and informal speech. Another relatively new study is Larrson's (2019) which discussed the ability of non-native writers to produce formal linguistic patterns that fit the register and academic norms of their writing. These "stylistic variations" among discourses are sometimes attributed to social factors for which the speaker chooses to adopt a certain level of formality (style), and sometimes to the context where the speaker shows adherence to the academic norms. In studying these variations, Wardhaugh (2006) argued that *register* is an essential aspect of determining key vocabulary associated with a certain group. He defined *register* as "sets of language items associated with discrete occupational or social groups" (p. 52). Wardhaugh's definition restricts the concept of register to vocabulary, mirroring the formal traditions which viewed grammar and vocabulary as separate entities while disregarding the relationship between

contextual meaning and language. On the other hand, a broader description of register is provided by Ferguson (1994) who states that "people participating in recurrent communication situations tend to develop similar vocabularies, similar features of intonation, and characteristic bits of syntax and phonology that they use in these situations" (1994, p. 20). Ferguson's description explicitly signals the importance of formulaic sequences in building fast mutual communication among social and occupational groups. In other words, Ferguson investigated the uniqueness of "lexicon, lexical collocations, sentence structure, and intersentential linking devices" in each of the registers he studied, rather than limiting the study of register to vocabulary (lexicon) only. Ferguson's approach could be aligned with Halliday's Systemic Functional linguistics (Hereafter SFL) and lexicogrammar theory in which he viewed lexis (vocabulary) and grammar as a combined structure rather than discrete. The application of lexicogrammar was vividly presented in his analysis of transitivity within SFL. Halliday's views differed from the formal traditions such as Kats and Fodor (1963) or Lyons (1963); these traditions viewed transitivity as a matter of language form and measured it by the number of objects the verb in each clause can take. Halliday's tradition, on the other hand, viewed the relation between language form (grammar) and language meaning as "organic" and not arbitrary. Therefore, Halliday (1994) proposed six types of processes (Material, Relational, Mental, Behavioral, Verbal, and existential), which provide semanticized analysis of transitivity based on the organic relation between language form and meaning, discarding the idea of opposition between syntax and semantics that formal traditions had proposed (Refer to section 2.2 for a detailed description of the processes).

This research study aims to examine the language of football commentary; for this purpose, it adopts the transitivity system of Halliday's SFL as a mean to identify the relation between language structure and formality (proficiency) within the discourse of football. It intends to examine the structures of the SFL processes which professional commentators predominantly use to describe and analyze the events of football matches. The study further compares between the structures professionals use and the ones fans use for the same purpose. The study also combines the concept of lexical bundles within SFL processes to examine the linguistic patterns that signals proficiency in the discourse of football. For the purpose of this research, the two concepts (SFL processes and Lexical bundles) are combined, coining the term *SFL bundle*; this is explained further in section (3.1).

Moreover, it is important to note that this study adopts the term **"*discourse of football*"** rather than "register of football" or "language of football", as the latter might imply singular, unified, homogenous language of football. From a SFL perspective, "*discourse of sport*" indicates a constitution of different registers and genres that are realized through lexicons and grammar. Football -similar to any other scientific field- is subject to linguistic variations due to different dialects and speech styles. Hence, the term "discourse of football" was selected to account for all these variations. In this study, discourse of football is examined through analyzing two types of corpora: (1) The professional corpus consisting of the language of the professional commentators transcribed from YouTube videos and recorded matches and (2) The fans corpus consisting of the language of the football fans posted as comments and tweets on social media applications (namely, twitter).

In the next section, the research questions are presented, which revolve around the coined concept "SFL bundles".

1.2. Research questions

This study aims to answer three research questions which are:

1- What are the Material, Relational, and Mental SFL bundles that are predominantly used in the corpus of professional commentators in the context of football matches?

2- What are the Material, Relational, and Mental SFL bundles that are predominantly used in the fans corpus in the context of football matches?

3- What are the similarities and differences in the frequency of structural types, and functional meaning between the SFL bundles used/found in the corpus of professionals and fans?

These questions are formed to guide us through the process of collecting and analyzing the data. Answering these questions leads us to achieve the purpose of this study, which is discussed in the next section.

1.3. Research Purpose & Research Gap

The significance of this investigation primarily lies in the scarcity of research in this area. Several studies examined different linguistic aspects of the football discourse whether as one register or as a comparison between radio commentary and TV commentary. Examples of comparative studies between TV and Radio commentary were carried out by Kern (2010) and Trouvain (2011), among others. While some studies viewed the language of TV and Radio commentary as types of the same register (register of football) and overlooked the differences between them, such as in the case of Ferguson (1983). Almost all of the previous studies discussed the linguistic aspects that professional

commentators/pundits use on TV and Radio, however, the emergence of social media - especially Twitter- allowed the fans to commentate on the events of football matches as they happen in specific hashtags, making the language that is used in these hashtags a part of the "*Discourse of football*" (See section 1.1). No studies -as far as the knowledge of the researcher goes- discussed the language that fans use for the purpose of describing and analyzing football matches. Therefore, this study aims to shed lights on the fans' language as a part of the "*discourse of football*" through comparing it to the language used by professionals.

This study also emphasizes that football is like any other scientific/academic field in the way its texts are assessed. Therefore, it applies the notion of Altenberg (1998) which suggests that studying the frequency and functions of lexical bundles in the corpus of a certain discourse proves quantitively whether the text in question signals competence in the field or not. Therefore, we are looking for these formulaic aspects of the discourse of football. The only difference is that this study adopts its own operational definition of lexical bundles (SFL bundle) to identify the linguistic similarities and differences between the professionals and fans within the same discourse (see Section 3.1).

To collect the needed data for recognizing the SFL bundles in the discourse of football, the commentary of the football matches is transcribed and compiled in excel sheets in the case of corpus of professional commentary (See section 3.2.). The fans corpus is also compiled in excel sheets from the social media website (twitter) from the hashtags created by the Premier Leauge social media team.

1.4. Structure of the study

The study is organized as follow: The second chapter provides a review of the related literature that sets the ground for the theoretical framework for this study. The theoretical framework is based on two theories, *Halliday's SFL processes* which is discussed in the first section of the chapter, and *Lexical bundles* which is discussed in the second section of the chapter. Moreover, aspects of formulaic language that binds the two concepts together is presented. The third section of the literature review provides a summary of the characteristics of the "*Discourse of football*", which is the discourse we are targeting in this study.

The third chapter presents the theoretical framework of the study in which the SFL bundles are presented. Then the study moves to the methodology which consists of two sections, methods of data collection and methods of data analysis. Then, we end the section with the research ethics and the various challenges encountered in recording, transcribing, and analyzing such a large corpus.

The fourth chapter reports the data that were collected from the two corpora. The first section reports the data that were found in the professional commentators' corpus, and the second section reports the data that were found in the fans' corpus. Each section is divided into three parts according to the three types of SFL bundles. Finally, we end the chapter with a summary of findings of both corpora.

The fifth chapter analyzes the data and compares between both corpora for the purpose of identifying general linguistic characteristics and specify which of them signal proficiency in the discourse. The analysis has three sections according to the three SFL bundles that

are analyzed in the chapter. The analysis is focused on the contextual use of each SFL bundle.

Last but not least, the conclusion section summarizes the notions and results of this study by directly answering the research questions in section (1.2) and providing recommendations for further research.

2. Chapter Two: Literature review

2.1. Overview

For the purpose of this research, it is essential to review the overarching theories that discuss the relation between language structure and meaning, which results in the formulaicity of discourses. This study presents two theories that adopt this notion which are SFL (Section 2.2) and formulaic language (Section 2.3.1). In SFL (section 2.2), we discuss the transitivity system of representation that provides a practical application of the organic relation between grammar and semantics through categorizing the structure of the sentences into types of processes. In section (2.3), we begin by providing a general overview of the notion of formulaicity in discourses, and how it is essentially adopted in the development of SFL. Then, we discuss the theory of formulaic language in section (2.3.1), which shows evidence of the similar linguistic textuality in the texts of the same discourse. Finally, we specifically discuss lexical bundles (section 2.3.2) -which is a form of formulaic language- as signals of proficiency. Moreover, we view the alignment in core concepts between formulaic language and SFL as both theories suggest that structure and meaning are interrelated. After presenting these theories, we dedicate section (2.4) for

identifying the characteristics of the football discourse, which is the discourse we are analyzing in our corpora.

2.2. Halliday's Transitivity Framework

Halliday (1994) built the transitivity framework based on the notion that language form (grammar) and language meaning (semantics) are combined rather than discrete as viewed by the formal traditions such as Katz and Fodor (1963) and Lyons (1963). The debate between the formal traditions and the Hallidayan tradition (functional grammar) is beyond the scope of this study, but it is essential for our framework to point out that functional grammar, unlike formal traditions, viewed semantics and grammar as interdependent in one level in the coding process from meaning to expression. Formal traditions on the other hand, viewed language as "a system of forms, to which meanings are then attached" (Halliday, 1994, p. F40).

The study of language in formal traditions can be summarized in three sequential stages: 1- the study of the words forms (morphology). 2- the study of forms of sentences (Syntax) 3- And once the previous two steps are established, the question of meaning is then posed (Halliday M. , An Introdction to Functional Grammar , 1994). On the other hand, Halliday interpreted language in his tradition as "system of meanings, accompanied by forms, through which meanings can be realized" (Halliday, 1994, p. F40). Based on that, Halliday presented the *systemic* aspect of language where grammar is seen as "a network of interrelated meaningful choices". To prove this notion, Halliday adopted authentic texts as the tool for linguistic analysis instead of the virtual examples that formal traditions' grammarians used to match the described type of grammar. Halliday argued that all

descriptions of grammar can be found in these authentic spoken and written texts, and with the growing use of computerized linguistics at the time, it finally became possible for the field of linguistics to acquire reliable data that proved what had been theorized. Based on that, in his book "An Introduction to Functional Grammar", Halliday provided quantitative evidence of syntagmatic and paradigmatic relations of lexical and grammatical kind, these quantitively proven relations supported the argument that the relation between form and meaning is not arbitrary as perceived in the formal traditions. The four relations are explained in detail in Halliday and Matthiessen (2004, pp. 37-43) and can be summarized as follow:

1- Syntagmatic lexical relations: the degree to which the probability of a word (lexical item) increases given the presence of a certain other word (the node) within a specified range (the span).
2- Syntagmatic grammatical structures: repeated sequences of word classes.
3- Paradigmatic lexical relations: lexical items function in sets having shared semantic features and common patterns of collocation.
4- Paradigmatic grammatical system: grammatical categories are organized in systems.

One important aspect that was used to highlight the features of experiential meaning in SFL is the analysis of transitivity in clauses using semantic-based approach of analyzing structures, relying on the notion that language form and language meaning are organically related (Halliday, 1994; Martin et al., 1997). To analyze transitivity within SFL in clausal sentences, Halliday (1994) proposed the following processes, which are described and

explained in this study based on the explanations of Martin et al. (1997) and Halliday and Matthiessen (2004).

2.2.1. Material clauses (process): The process of doing and happening

Halliday (1994) defined material clauses as "the clauses of doing and happening" (p. 179). In other words, these clauses represent physical actions in the material world that can be observed or perceived. Therefore, sentences like *"she melted the cheese; she drove to Manama"* are concreate examples of changes in the material world, thus, they are material clauses. Moreover, Martin et al (1997) suggested that material process also includes changes in abstract phenomena as well, "e.g., *prices fell throughout the season; the committee was dissolved*" are sentences that represent an abstract change. Therefore, it was concluded that "Material processes" include two types, concrete and abstract.

Moving on to the elements that construct the material clauses, they obligatorily have an *Actor*- which is the one doing the deed- and A *Process* -which is the verb that represents the doing or happening-. Therefore, sentences like "*the lake froze, he tripped*" are material clauses. Moreover, there are elements which are not necessarily found in each material clause, such as, *Goal* -the one impacted by the doing (done to/with), the goal was either created by the process of doing (e.g., *built a hut, composed a song, and so on*), or there was a change in quality or location for instance that was the result of the process "e.g., *cleaned the house; scrubbed the floor*". *Range* also is another element; it differs from goal as the range is not directly affected by the process (e.g., *climbed the mountain*).

Beneficiary is also another participant marked by prepositions to indicate it as a recipient beneficiary or client beneficiary. The difference between the recipient and the client is that

"the recipient appears when there is a transfer of existing goods represented by the preposition *to*", for example "*She gave the house to the kids*". The client is involved when there is a provision of service including creating new goods represented by the preposition *for* (e.g., *she built the house for the kids*). The table below provided by Martin et al (1997) summarized the roles in Material processes.

Table 1: components in Material processes

Type	**Actor**	**Process**	**Goal**	**Beneficiary**	**Range**
Action	She	Built	the house	(for the kids)	-
	She	Gave	the house	(to the kids)	
	She	Moved	the chair		
Event	The chair	Moved	-	-	-
	She	Climbed	-	-	the mountain

2.2.2. Mental clauses (process): The process of sensing

As discussed in (section 2.2.1), the material process represents concreate process in the physical words (e.g., *building, driving*) and abstract ones such as (e.g., *dissolving, prices falling*). However, in instances such as "*Ali liked the show*", applying the concepts of actor and goal does not make much sense as Ali did not do a physical action that affected the goal (the show), because the process of "liking" has a different functional interpretation. Therefore, Halliday (1994) grouped the clauses of feeling, thinking, perceiving under the mental process.

In mental processes, the obligatory elements are *the Senser* and the *mental process verb*. The senser is the participant that has the rational ability to do the sensing process, or as described by Halliday (1994, p. 114) "endowed with consciousness". Mental process verbs

are the ones that convey the feelings, thinking, or perceiving (e.g., *liked, knew, felt*). Moreover, mental processes could involve *the phenomenon* that is being sensed by the senser. Keeping in mind that the phenomenon is not restricted to the "things" as in our example earlier but could be extended to "facts" and "acts" as well. For example (*She liked that he took responsibility of his actions*). Therefore, the phenomenon is not always represented as a single participant within the clause, "It may also be represented by a separate clause. For example: (*David thought -» the moon was a balloon)*. The second clause here is a *projected clause* which represents "an idea brought into existence by the mental processing" (Martin et al., 1997, p.106).

2.2.3. Relational clauses: The process of being

This section talks about the process of being which is arguably the second most frequent process in the English language (Halliday, 1994). There are two main types of meaning in the relational process which are, the *'attributive'* and the *'identifying'*. Both types are detected by the copula verbs. In the *attributive* type, there is "an entity to which a quality is ascribed or attributed" (Halliday, 1994, p. 120). The entity ascribed is identified structurally as *the carrier*, and the quality is identified structurally as *the attribute*. An example of an attributive relational clause is (*Jane*[carrier] *is* [attributive relational process] *wise* [attribute]) where Jane is *the carrier* of the wisdom quality, and wise is *the attribute* ascribed to Jane. In the *identifying* type, "something has an identity assigned to it. What this means is that one entity is being used to identify another" (Halliday, 1994, p. 122).

To clarify the difference in structural function between the two types, let us analyze this example: *(Trump [token/identified] is [identifying relational process] the president [value/identifier])*. In this example, we have a token (*Trump*) that was identified by the value

(the presidency), but the distinction between them might be tricky as the sentences are reversible in the identifying type. For example, (*Trump is the president*) can be reversed to *(The president is Trump)*. Moreover, copula verbs do not have passive forms even though they have passive function as in (*The president is Trump*), the passive function could be represented clearly when we rhetorically ask the questions (*which is he?*) and (*which position does he occupy?),* the answers would be (*he is the president*), (*the president is him*), with *which* being the identifier value (*the presidency*), and *he* being the token identified by the value (*Trump*).

To clarify the example systemically in term of active and passive voice, we can refer to Halliday's rule "the active voice is the one which the subject is also the token" (Halliday, 1994, p.124). The rule directly leads us to conclude that (*Trump is the president*) is the active voice. We can also provide the following answers to further distinguish between the active and the passive: (*Trump occupies the president seat*) and (*the president seat is occupied by Trump*). In other instances, the functionality of token and value might be reversed as in this example, (*Holmes is the tall one*): (Holmes *[identified/value]* is *[identifying relational process]* the tall one *[identifier/token]*). In the previous example, *Holmes* was identified by assigning a token to him as an identifier, which was the opposite of the functionality of token and value in the example (*Trump is the president*). The rule of Halliday to determine the active and passive still stands, "*the active voice is the one which the subject is also the token*", therefore, we can deduce that our example was in the passive voice because the value (*Holmes*) was the subject in it. To clarify the active and voice we applied the same methodology of asking the following questions: (which one is him?), with

which being the identifier (*the tall one*), and him being the identified (*Holmes*), the answer in active voice would be (*the tall one is Holmes*).

To sum up, reversibility is the key distinctive feature between the identifying type and the attributive type. In identification type, the key to determine the identifier and the identified is to analyze the structural functions of tokens and values. In the types of meaning (identification and attribution), there are three functions of relations that operate within them which are (intensive, circumstantial, and possessive). The table below by Halliday (1994) represents the types and the functions of relation that operate within them. The table also articulates the feature of reversibility in the identifying type that was explained earlier.

Table 2: Principal types and functions of Relational Processes

Type / Function	**Attributive**	**Identifying**
Intensive	Sarah is wise.	Tom is the leader; the leader is Tom
Circumstantial	The fair is on Tuesday	Tomorrow is the 10th; the 10th is tomorrow.
Possessive	Peter has a piano	The piano is peter's; Peter's is the piano.

2.2.4. Subsidiary processes: behavioral, verbal, and existential

The three processes discussed in the previous section are the principal processes of English clauses. The principal processes (Material, Mental, and Relational) represent three different structural patterns that cannot be confused, the subsidiary processes on the other

hand are combinations of two of the principal processes. This section briefly discusses the subsidiary processes in relation to the principal processes. According to Halliday (1994), the three subsidiary processes share the characteristics of the three principal processes as follow:

1- Behavioral: combination of material and mental characteristics

2- Verbal: combination of mental and relational characteristics

3- Existential: combination of relational and material characteristics

2.2.4.1. Behavioral clauses: physiological and psychological processes

Behavioral processes do not have distinctive characteristics of their own, instead they have shared characteristics of mental and material processes. In other words, the main participant (*behaver*) who is doing the process of behaving is a "conscious being" like the senser in the mental process, but the clause is grammatically structured as "Material processes" as they adopt the present continuous to describe present time. Behavioral processes are described as "(typically human) physiological and psychological behavior, like breathing, coughing, smiling, dreaming, and staring" (Halliday, 1994, p. 139). Many behavioral clauses can be determined to be near another process; table (3) below by Halliday and Matthiessen (2004) clarifies the resemblance of behavioral verbs to other processes, in addition to the typical kinds.

Table 3: Types of verbs in Behavioral clauses

Near mental	Conscious processes represented as behaviors	Look, watch, stare, listen, think, worry.
Near verbal	Verbal processes as forms of behavior	Chatter, grumble, talk, gossip, argue
-	Conscious physiological processes	Cry, laugh, smile, frown, sigh, sob, whine, nod
-	Other physiological processes	Breath, sneeze, cough, burp, faint, yawn
Near material	Bodily postures and pastimes	Sing, dance, lie (down), sit (up, down)

2.2.4.2. Verbal clauses: processes of saying

Halliday (1994) describes verbal processes as "clauses of saying, as in What did you say? — I said it's noisy in here" (p. 140). Such clauses are important in discourses as they enable the reporting of what has been said. The reporting in such process is not restricted to the traditional forms of "saying" as in (*I said it's noisy in here*), it could also be extended to include symbolic exchanges that are not necessarily verbal, as in (*the notice is indicating that they start at 7*).

To detect verbal processes, it is important to recognize important elements other than *the Sayer*. First, *the receiver*, which is the one addressed by the interaction, as in (*They asked me to leave the place).* Second, the content can be represented in a separate clause as in the example (*the notice is indicating that they start at 7*), such clauses can be referred to as projected clauses or reported clauses. It is important to notice that verbal processes in the previous examples are consisted of two clauses, the primary clause is the only clause that

is a verbal process, while the other clause could be of any process. Moreover, the content can also be represented as "verbiage" of a verbal clause as in (*they told me a story*).

2.2.4.3. Existential clauses

In existential processes, "things are simply stated to exist" (Eggins, 2004, p. 237). Existential processes and relational processes are both encoding the state of being, however, the difference between the two is that relational processes are describing the state of existence in relation to other things (attributes or identification) as stated in the section (relational clauses: the process of being). Existential processes on the other hand are only concerned with introducing *the existent* using "there". *there* signals the process type but does not function as a Location, circumstance; nor does it represent a participant" (Halliday & Matthiessen, 2004, p. 110). Hence, the existence of "there" in the process is simply because all English clauses need a subject. The following examples represent existential processes in discourse: (*There was a cat in your yard; There will be tea at the conference room; There is a car in the garage*).

In the following section, we discuss two concepts that share the same core notion of SFL that suggest the organic relation between language structure and meaning. These two concepts are key in enabling us to assess the "*Discourse of football*" using the theoretical framework that we developed.

2.3. Lexical bundles and Formulaic Language: An SFL perspective.

SFL and Formulaic language theory share the core notion of combining meaning to structures. In SFL, Halliday concluded that the relation between the language elements (form and meaning) is interdependent and not arbitrary by analyzing clauses in authentic

spoken and written texts. On the other hand, Formulaic language is concerned with how the human mind processes language and how language is mentally stored and retrieved. Formulaic language theory suggests that language is stored as "word strings" that are retrieved as single units (Wray & Perkins, 2000). This core concept could be aligned with Halliday's syntagmatic and paradigmatic lexical relations and grammatical structures (Halliday & Matthiessen, 2004) (See section 2.2). One important relation that Halliday presented is "the degree to which the probability of a word (lexical item) increases given the presence of a certain other word (the node) within a specified range (the span)" (Halliday & Matthiessen, 2004, p. 38); such relation is called syntagmatic lexical relation. This alignment between formulaic language and SFL in analyzing the relation between form and meaning provides solid ground to the methodology used in this study to analyze the discourse of football. Therefore, in the following sections, we briefly define formulaic language as the main theory from which the concept of lexical bundles was derived (2.3.1). Then we move to the concept of lexical bundles (2.3.2), as it is the concept to which the aspects of SFL are added to coin the term SFL bundles, which is used in this study (See section 3.1).

2.3.1. Formulaic Language

Wray (2002) defined Formulaic Language as "a sequence, continuous or discontinuous, or words of other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar" (p. 9). In other words, structures are stored as empty slots and associated with meanings and depending on the context, these structures are retrieved, and the empty slots of the structures are filled with words. Formulaic Language

theory aims to analyze how the human mind acquires and processes language; the studies conducted by Ellis (1996) and Cowie (1996) signify the importance of acquiring formulaic sequences to be able to produce writing or speech in first and second language acquisition. However, it is difficult to draw the borderline of formulaicity in language. Jespersen (1924) claimed that there are things in languages that are of "formula character", and there is nothing to change about them. According to Jespersen, "A phrase like 'How do you do?' is entirely different from such a phrase as " I gave the boy a lump of sugar." In the former everything is fixed: you cannot even change the stress, saying 'How do you do?'" (1924, p. 18). His study was solely focused on such fixed expressions that might carry a different meaning than its components. Another study of Formulaic Language is by Firth (1951) who examined the collocations and colligations, the co-occurrence of words in relation to other words, and the occurrence of grammatical categories in certain syntactic structures. All these studies and theories are aligned with the core concept of Formulaic Language, however, for the sake of this study we needed to draw the borderline of the types of formulaic language we want to investigate for the purpose of identifying the linguistic formula that distinguishes between the professional discourse and non-professional discourse. Therefore, lexical bundles analysis was the adopted theory in combination with Halliday's processes to precisely identify the signals of proficiency in the discourse of football. The following section briefly explains the concept of lexical bundles and how it is a formulaic tool to identify proficiency in discourses (Altenberg, 1998).

2.3.2. Lexical bundles

Lexical bundles, multi-words expressions, fixed expressions, clusters, all these concepts refer to "words which follow each other more frequently than expected by chance, helping

to shape text meanings and contributing to our sense of distinctiveness in a register" (Hyland, 2008, p. 7). In other words, these strings of words are stored in the mind and retrieved as single "big words". Sinclair (1991) described these strings as "single choices, even though they might appear to be analyzable into segments" (p. 110). Therefore, we can see through the study of collocations such as *"Nice to meet you"* and *"always a pleasure seeing you here"* that they do not go through the traditional process of grammatical construction. Hence, through the study of collocational patterns of language especially in the works of Jespersen (1924) and Firth (1951), it has become evident how language is formulaic in nature. According to Altenberg (1998), "80% of the language can be studied as patterns and collocations, as lexical items appearing together with such frequency are greater than random probability" (pp. 107-109). Hence, linguists such as Sinclar and Hoey developed unconventional theories of language. They suggested that instead of regarding the lexical choices as filling the slots that the grammar makes available, we should consider the lexis as systematically structured through the repeated patterns of use (Sinclair, 1991, pp. 108-110; Hoey, 1991; Hoey, 2005).

It is important to point that 'fixed expressions' as identified by Jespersen (1924) convey contextual meanings that are not necessarily derived from its parts but rather from the context. Lexical bundles are not exactly idiomatic as they do not convey meanings that are different from their parts, but they function as important blocks of discourse. The functionality of lexical bundles is analyzed mainly by their frequency across texts. They are described as "recurrent sequences of three or more words" (Biber & Barbieri, 2007). A more descriptive definition of lexical bundles was provided by Hyland as "words which follow each other more frequently than expected by chance, helping to shape text

meanings, and contributing to our sense of distinctiveness in a register" (2008, p. 7). In other words, lexical bundles were mostly studied in term of words frequency rather than their structure, as some of these "recurrent sequences" might not necessarily represent a complete structure. Such bundles were studied thoroughly in the corpora of English for Academic Purposes (EAP) to determine formulaic sequences that adhere to the norm and could be considered indicators of competence in a field. Examples of these studies are by Haswell (1991) and Hoey (2005). Such notion of formulaic sequences being indicators of language proficiency is explained in detail by Yorio (1989) where he identified the role of "conventionalized language and idiomaticity" as indicators of language proficiency in the language production of second language learners. More focus on the importance of controlling the use of lexical bundles as an essential component of proficiency of language production was the subject of a study by Hyland (2008 b) where he emphazied the importance of lexical bundles as central to signal the academic discourse and differentiate between writing geners. According to Hyland "[lexical bundles are] means of differentiating genres, with implications for more evidence-based instructional practices in advanced writing contexts". (2008b, p. 41).

Therefore, we can conclude from the previous studies the importance of formualic sequences in distinguishing professional writers from non professional ones. Different studies suggested different definitions and terminologies to identify these formualic sequences, but all these studies take aspects from the Formualic Language theory. Therefore, this study embraces a broader perspective of the phenomena of lexical bundles by adding concepts of Systemic Functional Linguistics to study the discourse of football (See section 3.1).

The following section aims to present some key linguistic characteristics of the "Discourse of football", which is the discourse this study targets.

2.3.3. Previous studies of Lexical bundles

In this section, we briefly review three studies that applies the concept of Lexical bundles in analyzing the academic writing. These three studies compare the differences between the professionals and non- professionals, natives and non-natives in the use of lexical bundles. Shin (2019) compares between two groups. The first group consists of essays written by native speakers, and the second group consists of essays written by EFL learners. Both groups' members are from the same age group, and strictly matched for writing prompts and time constraints. The results did not show significant difference in the general percentage of lexical bundles. However, there are noticeable difference within the types of lexical bundles as the results for instance, showed more usage of noun phrases with embedded of-phrase fragments such as "*the end of the*", and the learners used more noun phrases with relative clauses such as "*the person who*". Therefore, this study confirms the insufficiency of depending on statistical analysis only, and the need for in-depth qualitative analysis for the types of lexical bundles, which is what is applied in this study. Another study of Chen & Baker (2010) examines two different types of texts. The first text (corpus) is of published academic papers, and the other is of undergraduate students' essays. Both texts included papers that are written by native and non-native speakers. The results showed a significant difference between L1 speakers and L2 speakers in the types of lexical bundles used. For example, the natives generally exhibited more tendency to use NP-based lexical bundles in a wide range of their written texts. Moreover, there are certain lexical bundles that are overly used in the L2 texts, but are rarely used by natives (*e.g., all over*

the world). The study also concludes that the use lexical bundles grows with the advancement of writing proficiency. This conclusion is important as it confirms the use of lexical bundles as an indicator to proficiency. The study of Wright (2019) is different in its approach as it investigates the existence of a core set of bundles for written academic prose, through comparing stand-alone literature reviews to other sections of studies (introduction, methodology, results, and discussion). The study aimed to identify these core set of lexical bundles through compiling a corpus of 3.4 million words from 417 studies in different academic fields such as psychology, medicine, and education. The results that lexical bundles in stand-alone literature reviews were identical to those found in other sections of the studies. Such conclusions validate the assumptions that each academic field has its unique lexical bundles that are essential to the textuality of its text, which is the notion that is adopted in this study.

2.4. Discourse of football

Since this study aims to mainly analyze the discourse of football commentators, it is essential to understand that English commentary is done by two commentators, the first is the descriptive commentator and the other is the analytical commentator. Ferguson defined commentary -or sportscasting- as "the oral reporting of an ongoing activity [descriptive], combined with provision of background information and interpretation [analytical or color commentator]" (1983, p. 155). An iconic duo that is well-known in the world of football commentary is Martin Taylor who usually played the role of descriptive commentator, and Alan Smith who usually played the role of analytical commentator. To convey the definition of the discourse of sports, it is important to identify its borderlines. As stated in section (1.1), this study adopted the term "discourse of football" because it

embraces the whole register with its dialectical and formality variations as long as they serve the same social purpose. Ferguson (1983) suggested that the discourse of sports has the distinctive function of reporting the events of the game as they happen while also briefing about other aspects of the game such as: information about the players, previous results and so on. To convey the purposes of the discourse, five syntactic characteristics are suggested by Ferguson (1983) as markers of the football discourse: simplification, inversion, results expressions, heavy modifiers, tense usage, and routines[1]. These will be briefly explained below:

2.4.1. Simplification

As the events of the game accelerate in speed, commentators' sentences tend to lack some expected elements. Ferguson identified the elements that are most commonly lacking: a- sentence initial noun phrase or noun phrase plus copula, and b- post-nominal copula. The phenomenon of lacking the initial material -sometimes called prosiopesis- was first recognized by Jespersen (1922) in variety of registers. And even though studies such as of Bowman (1966) suggests that such phenomenon indicates informality, Ferguson argued that "prosiopesis" in commentators' speech is influenced by the conditions of the game. "It may also simply represent the erosion of less stressed elements of recurrent formulas, as Jespersen originally pointed out". In other words, not vocalizing the initial sounds as in *good morning!* becoming *morning!*. Some examples of simplification in commentators' discourse are:

1- [it's] *a great cross into the penalty area.*

[1] Six characteristics was stated by Ferguson (1983) as markers of sports discourse in general, five of them are applicable to the discourse of football.

2- [it's] *a wonderful tackle by Puyol.*

There are also some limited instances of copula verbs deletion, usually comes after names of players such as in the following examples:

1- Vidic [is] *fighting to win the ball back.*

2- Messi [is] *asking for the ball on the wing.*

2.4.2. Inversion

A very common characteristic in the discourse of football is inversion, it is defined as "structures in which the predicate precedes the subject" (Ferguson, 1983, p. 160). Inversion gives commentators time to ascertain the identity of the one whose action is being reported, however, in some instances the identity of the player is very well known to the commentators, but they still prefer to use the inversion, which, according to Ferguson, indicates that it is a marker of the football discourse, regardless of the plausible explanation provided earlier. Some examples of inversion are:

1- Holding the possession of the ball is Xavi.

2- Getting ready to take the freekick is Beckham.

2.4.3. Heavy Modifiers

The use of heavy modifiers is similar to the use of inversion in that it could be used as an irregular conversational style, but it is very common in the discourse of football and sports in general. They creates opportunities to provide information about the players - mainly- without deriving from describing the events of the game. According to Ferguson, "[heavy modifiers are] brief, incidental identification along with the name of a player, or

less often an umpire, referee, or someone else referred to in the broadcast. The identification is typically a characteristic of his playing (e.g., left-hander) or his previous performance". (1983, p. 163). Being able to use referential functions (inversion and heavy modifiers) is an indicator of competence in the discourse of football in general and in commentating specifically. Some examples of heavy modifiers in the discourse of football are as follows:

1- *Neymar, the 222-million-euro man, is in the starting lineup tonight.*

2- *Iniesta, the scorer of the winning goal in the world cup final of 2010, is getting ready to take the corner.*

2.4.4. Tense usage

In linguistic studies and analyses of the discourse of sports, linguists focused mainly on the tense usage within the discourse to identify its characteristics. Ferguson (1983) claims that the studies had unanimous findings which indicated that the use of simple present is used in reporting events of short duration nature, such as: *Pique passes the ball to Jordi Alba*. Moreover, the use of the progressive is used to describe actions that takes longer duration, such as: *they are making excellent progress with streak of neat passes*, or for summing up the performance, such as: *Messi is tormenting Real Madrid's defense*. While the past simple is used to describe actions of a quick nature, such as: *Ronaldo takes the shot... and he placed it wide*.

2.4.5. Routines

Many routines' aspects are aligned with the core concepts of lexical bundles and formulaic language. Therefore, to avoid redundancy of explaining formulaic sequences and

idiomaticity, the focus of this part would be on the negative outlook perceived from such extensive use of routines. Commentators being under the pressure of describing and analyzing recurrent events tend to use recurrent syntactic patterns. Such routines were addressed negatively in studies like Wilton (2017) in which he criticized the lack of content in post-match interviews because of the huge amount of formulaic language. Even commentators have had fair share of criticism as another study of Pfeiffer (2014) analyzed the discourse of sports commentators in English and German, claimed that the discourse is filled with pointless set of phrases and meaningless metaphors. However, while the claim of the lack of content in the post-match interviews is true, neither of the studies denied the benefits of formulaic sequences in helping commentators cope with the fast nature of the game. Moreover, the studies that were discussed earlier in the lexical bundles section proved that these formulaic sequences when used appropriately are indicators of competence in the field.

The theories that are discussed in sections (2.2) and (2.3) should justify the direction we took in developing the theoretical framework of this study (see section 3.1). Moreover, the aspects of the "Discourse of football" that we present in section (2.4) clarify the type of discourse that is targeted in this study. The next chapter addresses the methodology of this study, and further explains the theoretical framework, including the concept of SFL bundles coined for the sake of this study.

3. Chapter Three: Methodology

As stated in section (1.3), the main purpose of this study is to examine the fans' discourse as a new type of the "Discourse of football" alongside the discourse of professional commentators. Previous studies were fixated on the professional commentators on TV and Radio, as the fans did not have a channel to commentate and analyze the events of the matches as they happen prior to the emergence of Twitter. Therefore, we aim to compare between the professional discourse of football commentary and the fans' discourse in term of formality, proficiency, and linguistic characteristics using the theoretical framework (SFL bundles) which is explained in the first section of this chapter (3.1). In the second section (3.2), we present the methods of data collection as we describe the two types of corpora that were analyzed. In the last section (3.3), we present the methods of analysis and the framework followed in categorizing and analyzing the clauses according to their correct SFL bundle.

3.1. Theoretical Framework: SFL bundles

As stated in section (2.2), SFL signifies the role of clauses in linguistic analysis, whereas lexical bundles as defined in the previous studies are not concerned with clauses as much as with words that occur frequently together (See section 2.3.2). Therefore, in this study, the analysis of football discourse from an SFL perspective is based on clause analysis. Hence, we do not adopt the previous theoretical definitions that were fixated on words frequently occurring together, instead we have built our operational definition of lexical bundles by combining aspects of SFL to coin the term *SFL bundles* which aims to:

1) expand the concept of "recurrent sequences" to include SFL word functions instead of just the probability that a certain word occurs in the range of another. In other words, the elements that constitute the clausal processes (carrier, actor, senser, etc.) and their recurrent sequences (structures) are what build the SFL bundles we intend to analyze in this study.

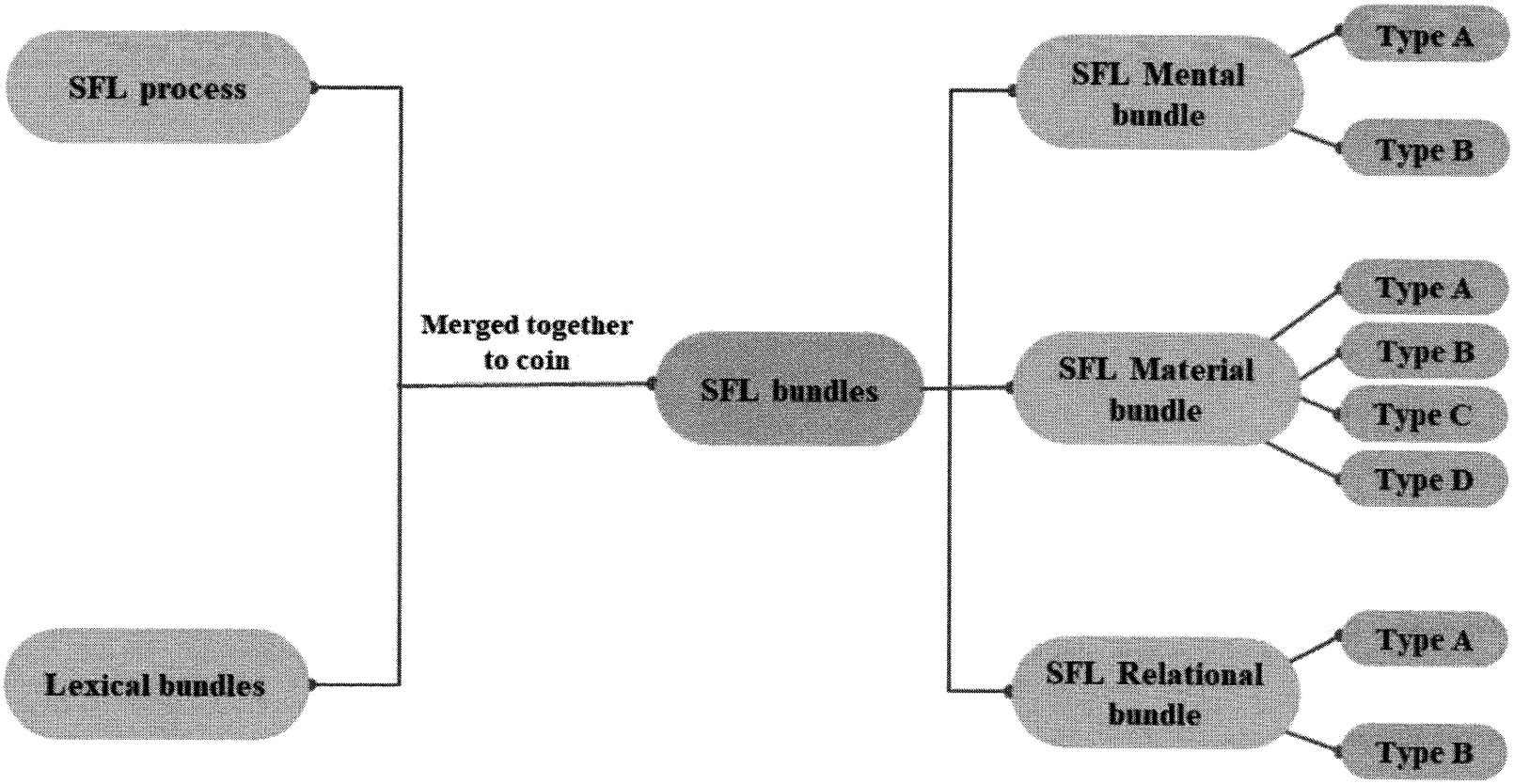
2) exclude non-clausal utterances, as SFL analyzes the functionality of discourses in term of their clausal sentences. Therefore, SFL bundles are bundles that are limited to clauses, and hence abides the rules of SFL theory.

Therefore, what we intend to do in our analysis of the discourse of football is merge aspects of the two concepts, SFL and Lexical bundles, to create a framework based on the notion of formulaicity that both these concepts essentially share (See section 2.3 & Figure 1). This merger requires us to provide the clausal structures of SFL bundles that we need to look for in our corpora.

To sum up, this framework adopts clausal sentences as means of identifying the SFL bundles which is a different direction than the previous studies of lexical bundles such as Altenberg (1998) and Hyland (2008b), which were concerned with words occurances in discourse rather than the structures. However, this study still gives the same significance to the formulaic nature of language in signaling professional discourse that was stated by Yorio (1989), Haswell (1991), Hoey (2005), and Hyland (2008b), among others. This framework also signifies the importance of frequency in generating the textuality of the discourse but in a more holistic manner, as it generalizes the concept of lexical bundles from looking for the probability of a word appearing with another, to include the functional types in the processes such as: carrier, attribute, actor, and so on instead of being limited

to a word by itself. In other words, adding the functional aspects of SFL enhances the notion which combines the language structure to the semantics and context of the language, which is a notion that is related to the formulaicity from which lexical bundles were studied (See section 2.3). Figure (1) summarizes the merger that is applied in this study for coining the term *SFL bundles.* It also summarizes the types under each SFL bundle which are further explained in section 3.3.

Figure 1: terminology of the study



3.2. Methods of Data collection

For the data collection process, we used a mixture of quantitative and qualitative methods to investigate two sets of corpora within the discourse of football: the corpus of professional commentators as they describe and analyze the ongoing matches, and the corpus of football fans on Twitter on the live hashtags created by Premier League social media team for fans who want to comment on the matches as they go on. Qualitatively, we collected the data through identifying the clauses in the professional and fans' corpora and assigning them into the correct SFL bundle based on the participants within each clause. Quantitively, we recorded and compared the number of clauses under each of the principal SFL bundle in both corpora.

3.2.1. Corpus 1: Professional corpus

To collect the professional corpus, we started by transcribing football commentary of previously recorded matches followed by counting and recording the number of clauses under each SFL bundle, and finally categorizing the clauses based on the types of processes explained in section (2.2). As explained in section (2.4), Professional English football commentary is normally performed by two commentators, one specialized in describing the ongoing activities of the game, and the other with providing background information in addition to analytical interpretations of the events. The professional corpus is composed of approximately 5000 words from four different commentator teams (approximately 1250 words for each team) (See table 4).

Table 4: Commentators' teams

Teams	Descriptive commentator	Analytical commentator
1	Alan Green	Ian Brown
2	Martin Taylor	Andy Gray

3	Ian Drake	Darren Fletcher
4	Peter Dury	Rob Green

The nature of the commentary shows repetition of the same clauses inside each bundle, therefore, inside each bundle we provided details of frequency in the corpus in general, and details about range of use among the commentators' team. The reason for collecting data from four different teams is to ensure the reliability and consistency of the linguistic phenomena in the data, as opposed to depending on a single commentary team. Moreover, we chose to collect the data in separate timelines to ensure that the SFL bundles are built into the discourse. To collect the corpus of football commentary, we first tried to rely on programs of auto captioning to provide the needed data. However, after our revision of the texts that were extracted by the program, we realized that the Auto-captioning is only 60-80% accurate. Moreover, in occasions, these auto captioning programs disregard punctuation, and provide inappropriate line breaks. Hence, we decided to transcribe the commentary of the videos by listening, as these programs were not providing us with the needed accuracy. During the transcription, clauses in the data are extracted manually in excel sheets to identify the participants of each clause according to Halliday's processes (See section 2.2 & 3.1).

3.2.2. Corpus 2: Fans corpus

The fans corpus is collected from Twitter (a social media application) through compiling commentary tweets from the live hashtags of the Premier League. The corpus primarily consisted of tweets from four different accounts, these tweets describe and analyze football matches. This type of corpus would have been difficult to collect before the emergence and widespread use of social media applications. However, now the Premier League social media team creates hashtags

on Twitter for each match in which fans can commentate on the events of the match simultaneously as it being played. Therefore, these hashtags allow the fans to describe, commentate, and share their thoughts on the events of the matches. This makes the fans' discourse an integral part of the "*Discourse of football*". In these hashtags, fans write tweets about the match as if they are impersonating the role of commentators. There are tweets of descriptive nature, and others of analytical nature.

Similar to the analysis of the professional corpus, the tweets were studied to identify the clauses in the corpus, and the participants of the clauses were extracted into an excel sheet and categorized according to the correct SFL bundle. The extracted clauses were then used to compare the use of SFL bundles (structures), their frequency, and their functions with the results of the professional SFL bundles. 5000 words were collected from these tweets. Within the SFL bundle, we identified the elements of each clause as can be seen in section (3.3).

The data were taken from four different twitter accounts which are always active in the special hashtags that are created by the English Premier League social media team. These accounts are popular on twitter because of their minute-by-minute description and analysis of football matches. We could not acquire consents from the owners of the accounts to post the name of their accounts in our research (See section 3.4). Therefore, we identify them as follow in table 5:

Table 5: The Twitter accounts

Account	Description
Account A:	Describes the events of matches in PL, interested in providing statistical data when analyzing the events.

Account B:	Minute by minute description of PL games, especially Man City.
Account C:	Detailed analytical comments in addition to description of the important events in the PL matches.
Account D:	A Liverpool fan who is interested in providing analytical comments on the important events of Liverpool matches and other important PL matches.

3.3. Methods of data analysis

In this section, we present the process of organizing and categorizing the clauses into the excel sheets and identifying each type of SFL bundle based on the participants. Moreover, we also provided the frequency and the number of times this clause has been used in corpus, in addition to the range which describes the number of different commentators' teams [2] (out of four) which used the clause. Section (3.3.1) presents the excel sheet in which the clauses under the SFL Material bundle were analyzed. Section (3.3.2) presents the excel sheet in which the clauses under the SFL Relational bundle were analyzed. Section (3.3.3) presents the clauses in which the SFL Mental bundles were analyzed. Notice that the same templates of excel sheets were used to extract the clauses from the professional corpus and from the fans' corpus, as the aim of the analysis is to

[2] The different twitter accounts in case of the fans' corpus.

examine the types of structures within the SFL bundles, the frequency of these structures, and their function within the SFL bundle framework.

To achieve this objective, we divided the SFL bundles into three main bundles based on the three main processes that were discussed in Halliday (1994) and Martin et al (1997), which are (Material, Relational, and Mental). The categorization of these processes is explained in the next sections:

3.3.1. SFL Material bundle

For this study, we categorized the types of clauses under SFL Material bundle into two categories which are "goal-directed" and "non-goal directed". Goal-directed are the types that have the goal participant in their clauses, and non-goal are the types that do not have the goal participant in their clauses. Such categorization should make our types more focused, and our analysis more concreate, especially the functional analysis. The fifth structure of Martin et al (1997) was excluded as there was only one example that was found in the professional corpus, and none was found in the fans' corpus. Therefore, it was alternatively replaced with Type D, which has an actor, a process verb, and a beneficiary (See table 6).

Table 6: Types under SFL Material bundle

Types under the SFL Material bundle	
Goal-directed clauses	**A:(Actor + process verb+ goal + Beneficiary)**
	B:(Actor + process verb+ goal)
Non-goal directed clauses	**C:(Actor + process verb)**
	D: (Actor+ process verb+ Beneficiary)

3.3.1.1. SFL Material bundle: Excel sheet

As stated in section (2.2.1), the participants of the material processes are actors, process verbs, goal, and beneficiary. However, not all of these participants are necessarily found in all material clauses, that is why the four types (A, B, C, and D) of SFL Material bundle were created in our framework (See section 3.1). Hence, to be able to identify each clause into the correct type, we used an excel sheet, which allows us to exclude or include any participant in order to get the frequency of each type. In the clauses in which one of the participants is not present, we use the word NILL to indicate that it is not existent in the clause. Figures (2,3,4 & 5) shows how the excel sheet organized the clauses according to the types. Figure (2) shows an example of type A of the SFL Material bundle in which we can see all the participants are in the clause. Figure (3) shows an example of Type B in which there is no beneficiary. Figure (4) shows Type C in which there is only an actor and a process verb. Finally, figure (4) shows an example of Type D in which there is an actor, process verb, and a beneficiary (refer to appendix xx for further examples).

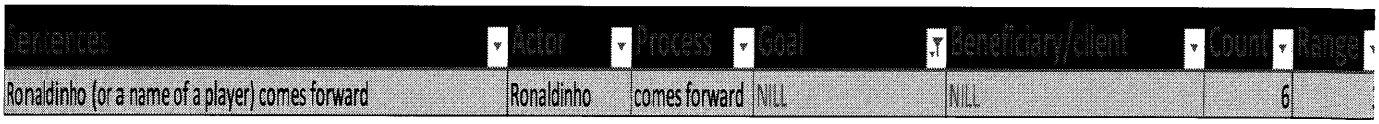
Figure 2: Type A of SFL Material bundle

Sentences	Actor	Process	Goal	Beneficiary/client	Count	Range
Puyol (or name of a player) wins the ball back for Barcelona	Puyol	wins back	the ball	for Barcelona	5	

Figure 3: Type B of SFL Material bundle

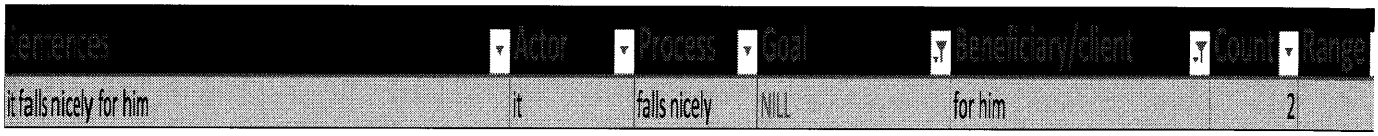
Sentences	Actor	Process	Goal	Beneficiary/client	Count	Range
Iniesta (or name of a player) hits the post	Iniesta	hits	the post	NILL	5	
He turned around the defender	He	turned around	the defender	NILL	2	

Figure 4: Type C of SFL Material bundle



Sentences	Actor	Process	Goal	Beneficiary/client	Count	Range
Ronaldinho (or a name of a player) comes forward	Ronaldinho	comes forward	NILL	NILL	6	

Figure 5: Type D of SFL Material bundle



Sentences	Actor	Process	Goal	Beneficiary/client	Count	Range
it falls nicely for him	it	falls nicely	NILL	for him	2	

3.3.2. SFL Relational bundle

Halliday categorized the relational processes into two types (Attributive and identifying) and three functions (intensive, possessive, and circumstantial) under each type (See section 2.2.3). For this study, we use the types to categorize the relational clauses in the findings section, but we leave the distinction of functions to the analysis chapter to avoid the confusion of having too many categories. Therefore, the attributive clauses under the SFL Relational bundle is called Type A, and the identifying clauses is called Type B. And as explained in section (2.2.3), the rule of reversibility is what distinguishes between the two types.

3.3.2.1. SFL Relational bundle: Excel sheets

As stated in section (2.2.3), there are two types of relational process which are the attributive type and the identifying type. In our framework (see section 3.1), we adopt the clauses of attributive type as Type A of the SFL Relational bundle. In Type A, clauses need to have carriers and attributes in addition to the relational process verb to form the clause. Therefore, the excel sheet was structured to apprehend these three participants. In Type B, different participants are involved and hence we had to create a different excel sheet to apprehend its participants which are tokens,

values, and the relational process verbs. Figure (6) shows an example of Type A in the SFL Relational bundle, while Figure (7) shows an example of Type B.

Figure 6: Type A of the SFL Relational bundle

Sentence	Carrier	relational-attribut	Attribute	Count	Range (out of 4)
that looked close	that	looked	close	5	4
it is a little dinked freekick into the penalty area	it	is	a little dinked freekick	4	4

Figure 7: Type B of the SFL Relational bundle

Sentence	token	identifying -relational proce	Value	count	Range (out of 4
Amazing was the word that Frank Rijkard used to describe Xavi's first leg	Amazing	was	the word that Frank Rijkard used to describe Xavi's first leg	1	1
Chelsea's Stamford Bridge is the place of the second encounter	Chelsea's Stamford bridge	is	the place of the second encounter	1	1

3.3.3. SFL Mental bundle

As the framework of SFL is concerned mainly with the clausal analysis, the types of mental processes that were explained in section 2.2 were simplified into two types in the SFL Mental bundle according to the phenomenon in the clause. Type A contains the clauses in which the phenomenon is a participant within the clause, while Type B contains the clauses in which the phenomenon is in a separate clause. Table (7) shows our categorization.

Table 7: Structures of the SFL Mental bundle

Senser	**Mental process verb**	**The phenomenon (Participant within the clause)**
You	can hear	the cries of Liverpool fans
Senser	**Mental process verb**	**The phenomenon (Separate clause)**
They	Thought	it was a clear handball not given

3.3.3.1. SFL Mental bundle: Excel sheet

In our SFL bundle framework (see section 3.1), we decided to divide the mental clauses into types based on the phenomenon. Therefore, figure (8) shows a sample from the excel sheet that was used to identify the clauses under Type A (phenomenon as participant within the clause), while figure (9) shows the excel sheet that was used to identify the clauses under Type B (phenomenon as a separate clause).

Figure 8: Excel sheet of Type A of SFL Mental bundle

Sentence	Senser	Mental process	phenomenon	Count	Range
You heard the cries of the fans	You	heard	the cries of the fans	2	2

Figure 9: Excel sheet of Type B of the SFL Mental bundle

Sentence	Senser	Mental process	phenomenon	Count	Range
They clearly thought it was a handball	They	thought	it was a handball	2	2

3.3.4. Approaching the analysis

It was previously stated in section (2.2) that SFL is inherently related to social context (Halliday, 1994). Therefore, part of the SFL bundles analysis is the functional analysis of structures, which will be organized according to the nature of the SFL bundle. For instance, in the SFL Material bundle, the analysis was based on the similarities and differences in the contextual use of the goal directed clauses and the non-goal directed clauses. Moreover, we adopted a part of Halliday's lexical density hypothesis (which is the weight of the lexical item in the density scale) to investigate which corpus is denser and hence more proficient. In the SFL relational bundle, we adopted the functions of relational processes in Halliday (1994), to identify the contexts and

aspects of the game in which the SFL relational bundle is used. In the SFL mental bundle, we looked at functionality of the two types of phenomena as discussed by Halliday and Matthiessen (2004), as they looked at the functionality of the phenomenon as a participant within the mental clause, and the phenomenon that constitute a separate clause.

3.4. Research Ethics

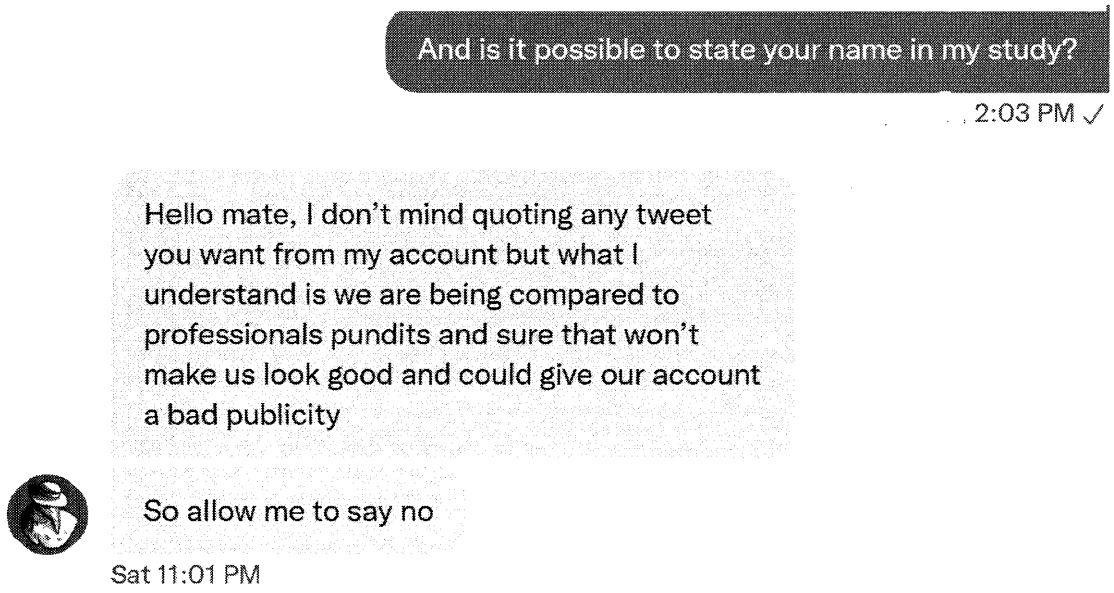
A number of ethical considerations were taken into account in the process of collecting and analyzing the data. Some are related to the data collection, and others related to the use of the actual names of the accounts.

Two types of corpora were used in this study. The first corpus is the corpus of football commentary which is a corpus that was transcribed from recorded football matches on sports channel like BBC, BT sports, BEIN sports, and SKY among others. The researcher made sure that the commentary is transcribed accurately by listening, as the results from the auto-captioning program were only 60-80% accurate after our revision. Moreover, the researcher made sure to collect these data from different commentators' teams to ensure the representativeness of the transcribed text to the discourse of football, and to avoid what might be idiosyncrasies of a single commentator. This is in addition to collecting the data in separate timelines to ensure that the SFL bundles are built into the discourse. Moreover, the data were presented in several numerical figures to ensure that it is not misrepresented. Finally, some clauses can be confusing to categorize under the processes, therefore, the researcher was asking for the opinion of two colleagues to decide the category of these clauses, and make the final decision based on the majority.

The second corpus is the fans' corpus, which is a corpus that was compiled from four twitter accounts that are dedicated to describing and analyzing the matches of the English Premier League.

The researcher collected a corpus of the same size as the first corpus (approximately 5000 words) to ensure the validity of quantitative comparison between the two. Moreover, informed consents are required for stating the names of the twitter accounts in any media or public studies (according to Twitter terms and conditions), however, informed consents are not applied to the tweets as they are available for public use. Hence, direct messages were sent to the owners of these four accounts on twitter. Three of them did not respond to the message until the time of submitting this study, and one account accepted the use of the account's content but refused to state the name of the account in the study (see figure 10). Therefore, as can be seen in table 5, we settled for a description of the accounts as the names of the accounts do not affect the course of this study.

Figure 10: Refusal of publishing the name of the account



4. Chapter Four: Findings

The collected data can be categorized into two parts: the corpus of professional commentary and the corpus of the fans. As stated in section (3.2), the corpus of professional commentary contained an approximate of 5000 words, these are divided into 4 commentary teams. Clauses were extracted from the data and categorized into the three principle SFL processes (Material, relational, and Mental). In the Professional corpus we found a total of 334 clauses, 309 out of 334 were clauses which fall under these three principle SFL bundles (Material, relational, and mental), while 25 clauses are considered "subsidiary SFL bundles". The fans corpus, on the other hand, contained a total of 313 clauses, 287 clauses fell under the principle SFL bundles. The scope of this study is concerned with the principal SFL bundles; therefore, within the findings section, an independent sub-section is dedicated to each SFL bundle (Material, Relational, and Mental respectively). Moreover, each clause was reported in term of frequency of occurrence, and the range of occurrence among the commentators' teams (out of 4). The significance of these occurrences is discussed in chapter 5.

In this section, the findings of the professional corpus and fans corpus are presented respectively; within each corpus, we start by the SFL Material bundle and the four types that are under it. Next, we move to the SFL Relational bundle and its two types. Then, we report the SFL Mental bundle, and we finish the section of the professional corpus with a summary of the findings.

4.1. Professional corpus

The data set is composed of approximately 5000 words from four different commentator teams (approximately 1250 words for each team). The four commentators teams are Alan Green

and Ian Brown, Martin Taylor and Andy Gray, Ian Darke and Darren Fletcher and Peter Dury and Rob Green (refer to table 4 for further details).

Based on the analysis of all the commentators, The corpus is dominated by the SFL Material bundle which were 160 clauses. The SFL relational bundle came in second with a total of 131. The SFL Mental bundle was significantly lower in frequency than the first two as only 18 clauses were found in the data.

Table 8: Frequency of SFL bundles - professional

SFL Material bundle	SFL Relational bundle	SFL Mental bundle
160 clauses	131 clauses	18 clauses

4.1.1. SFL Material bundle

The structures of the targeted bundle are the ones that contain the following: actor, process verb, goal, and beneficiary. Each component was identified in the examples for the purpose of categorizing each clause under the four types in the SFL Material bundles. Therefore, we found that some of these examples contained all the four components, while others missed either the beneficiary, the goal, or both. The actor and the process are the essential components that build the clause, therefore, all the examples contained at least these two. The total of the clauses in this SFL bundle was 160, only 14 clauses (9%) out of the 160 did not contain a goal. On the other hand, only 28 clauses (18%) contained a Beneficiary, while only 9 clauses (6%) contained neither of them. The data can be summed up in the tables 9 and 10 below, table 9 shows the count of each component in the data, table 10 shows the structural types of the SFL Material and the count of the clauses for each structural type. Based on these tables, we can see the dominance of the 'goal directed' clauses in the material SFL bundle in the professional corpus.

Table 9: Frequency of participants in the professional corpus

Actor	**Material process verb**	**Goal**	**Beneficiary**
160	160	146	24

Table 10: Frequency of types - SFL Material bundle

	Types under the SFL Material bundle	Number of clauses
Goal-directed clauses	**Type A:(Actor + process verb+ goal + Beneficiary)**	19
	Type B:(Actor + process verb+ goal)	127
Non-goal directed clauses	**Type C:(Actor + process verb)**	9
	Type D: (Actor+ process verb+ Beneficiary)	5

4.1.1.1. SFL Material bundle (Type A)

Within the SFL material bundle, Type A contains all the components of the SFL bundle. We found 19 clauses that fall under this structural type, which is considered relatively infrequent in the data as it represents roughly 12% of the clauses of Material SFL bundle. An example of this type of SFL Material bundle is [1- *Puyol wins the ball back for Barcelona*]. The previous example contains all the components of the SFL Material bundle where *Puyol* is the actor, *wins back* is the material process verb, *the ball* is the goal, and *for Barcelona* is the Beneficiary. To make the categorization clearer, we can mark them in the example as follow:

Example 1: *Puyol wins the ball back for Barcelona*

Actor	**Material process verb**	**Goal**	**Beneficiary**
Puyol	wins…back	the ball	for Barcelona

In example (1), we notice that the beneficiary is marked with the preposition *for*, in other examples we can find beneficiaries/clients who are not marked by preposition such as [*2- Messi gives Barcelona a golden opportunity*]. In this example we see that the beneficiary is placed before the goal, and hence not marked by a preposition. The example can be categorized as follow:

Example 2: *Messi gives Barcelona a golden opportunity*

Actor	**Material process verb**	**Beneficiary**	**Goal**
Messi	gives	Barcelona	a golden opportunity

4.1.1.2. SFL Material bundle (Type B)

Within the SFL Material bundle, Type B is the most frequent in the data. We found 127 clauses under this type, which represent 79% of the SFL Material bundle. The clauses under this type vary in the goal component as some of the goals are nouns (pronouns included), and others are noun phrases. An example of a clause which has a noun as the goal is [3- *the defender blocks the cross*] in which *the defender* is the actor, *blocks* is the material verb process, and *the cross* is the goal. An example of a clause which has noun phrase as the goal is [4- *Clichy pounces on Xavi's lost ball*] in which *Clichy* is the actor, *pounces on* is the material process verb, and *Xavi's lost ball* is the noun phrase that represents the goal in the clause. We also found clauses in which the noun

phrases are the actor, such as [5- *The through ball opened up the defense*]. In the previous example, *The through ball* is the actor, *opened up* is the material process verb, and *the defense* is the goal component. The following examples (3,4, and 5) have been broken down to show the components of the SFL material bundles:

Example 3: *the defender blocks the cross*

Actor	**Material process verb**	**Goal**
The defender	blocks	the cross

Example 4: *Clichy pounces on Xavi's lost ball*

Actor	**Material process verb**	**Goal**
Clichy	Pounces on	Xavi's lost ball

Example 5: *The through ball opened up the defense*

Actor	**Material process verb**	**Goal**
The through ball	Opened up	The defense

4.1.1.3. SFL Material bundle (Type C)

Even though Type C is significantly less frequent than the previous two types (9 clauses), it is still a part of the SFL material bundle. This type includes neither goal nor beneficiary, as the Material process verb stands with the actor to make the clause. The Material process verbs in this type are intransitive in nature, therefore, the effect is on the actors themselves. We see that this type describes the locations of the players in the field, for example [6- *Drogba cuts inside*] or [7-

Ronaldinho comes forward]. As clear in examples (6) and (7), the actors are *Drogba* and *Ronaldinho*, the Material process verbs are *cuts inside* and *comes forward* which are descriptive of the locations of the players with the ball on the field.

Example 6: *Drogba cuts inside*

Actor	**Material process verb**
Drogba	Cuts inside

Example 7: *Ronaldinho comes forward*

Actor	**Material process verb**
Ronaldinho	Comes forward

4.1.1.4. SFL Material bundle (Type D)

Within the SFL Material bundle, Type D is similar to the previous type in the intransitivity of the verb, but it differs in the notion of specifying a beneficiary from the action. For example, the clause [8- *it falls nicely for Anelka*] indicates that there is someone (Anelka) who is directly benefited from the intransitive action of the fall. We found only 5 clauses that can be categorized under this type, and the following table explains how the participants of this type are categorized.

Example 8: *it falls nicely for Anelka*

Actor	**Material process verb**	**Beneficiary**
It (the ball)	Falls nicely	For Anelka

4.1.2. SFL Relational bundle

Following the typology of Martin et al (1997), there are two types of relational processes that can be found in the English language which are attributive relational clauses (Type A) and identifying relational clauses (Type B). In the first type, one value is equated with another, therefore, the participants of the identifying relational type are the token and the value, in addition to the relational process verb. In the latter, a quality is ascribed to an entity, hence, the participants in the attributive relational type are the carrier and the attribute, in addition to the relational process verb. In the professional corpus, the total of the clauses under the SFL relational bundle is 131. we found 36 clauses that fall under the identifying relational type, and 95 clauses that fall under the attributive relational type. Distinguishing between the two modes of the relational processes could be problematic, therefore, we follow Halliday's (1994) method of distinguishing between the two types. Halliday suggested that identifying relational clauses are reversible, while the attributive relational clauses are not. Therefore, to make sure that our categorization is correct, we are going to apply this rule on the examples in our findings.

Table 11: Types and counts of SFL Relational bundle - Professional

Type	Count
Relational Attributive (Type A)	95
Relational identifying (Type B)	36

4.1.2.1. Attributive Relational clauses (Type A)

This type answers the questions of what and how. Therefore, we can see the highly frequent use of demonstratives and pronouns as the carriers in the clauses of this type, for example [9- *that was an optimistic effort*], [*that was a very good interception*], and [10- *He was spot on with that clearance*]. There were also other clauses which contained the nouns as the carriers, such as [11-

Ronaldinho has the ability to turn the game on its head] and [12- *the fans are delirious with joy*]. The total of the attributive relational clauses is 95, and the participants of examples (9, 10, 11, and 12) can be categorized as follow:

Example 9: *that was an optimistic effort.*

Carrier	**Relational attributive process**	**Attribute**
that	Was	An optimistic effort

Example 10: *He was spot on with that clearance.*

Carrier	**Relational attributive process**	**Attribute**
He	Was	Spot on

Example 11: *Ronaldinho has the ability to turn the game on its head.*

Carrier	**Relational attributive process**	**Attribute**
Ronaldinho	has	The ability to ....

Example 12: *the fans are delirious with joy*

Carrier	**Relational attributive process**	**Attribute**
The fans	are	Delirious with joy

4.1.2.2. Relational identifying clauses (Type B)

As stated earlier in section (2.2.3), this type answers the questions of who and which, but this could still be confusing to identify in text. Therefore, we applied the reversibility rule to the examples. For example, [13- *Chelsea's Stamford Bridge is the place of the second encounter*] is a clear identifying relational type, and the rule of reversibility could easily be applied to it when saying [*the place of the second encounter is Chelsea's Stamford Bridge*]. In this example, the

token/identifier is *Chelsea's Stamford Bridge*, and the value/identified is *the place of the second encounter*.

Another example which is common in football discourse but might not be as common in other discourses is [14- *Amazing was the word that Frank Rijkard used to describe Xavi's first leg*]. Here, we can see that the adjective *amazing* which serve the role of token/identifier is placed at the beginning of the clause, *the word that Frank Rijkard used to describe Xavi's first leg* is the value/identified]. The rule of reversibility could also be applied to the previous example, so it becomes [*the word that Frank Rijkard used to describe Xavi's first leg was amazing*]. Another type of clause that is common in the use of identifying relational process is [*On the edge of the box is Etoo*], where Etoo plays the role of the token/identifier, and on the edge of the box is the value/identified. The exact same example was used but the token and value places in the clause were reversed to become [*Etoo is on the edge of the wall*]. A clearer categorization of the participants of examples (13,14, and 15) is in the following:

Example 13: *Chelsea's Stamford Bridge is the place of the second encounter*

Token/identifier	**Identifying relational process verb**	**Value/identified**
Chelsea's Stamford Bridge	is	the place of the second encounter

Example 14: *Amazing was the word that Frank Rijkard used to describe Xavi's first leg*

Token/identifier	**Identifying relational process verb**	**Value/identified**
Amazing	was	the word that Frank Rijkard used to describe Xavi's first leg

Example 15: *On the edge of the box is Etoo.*

Token/identifier	**Identifying relational process verb**	**Value/identified**
Etoo	is	On the edge of the box/wall

4.1.3. SFL Mental bundle

This type of SFL bundle has the least occurrences in the data with 18 clauses only. SFL Mental bundle being the least frequent is not surprising as such clauses are usually used to express the commentators' perceptions outside the "in-play" phases of the game. In other words, they are not commonly used to describe the ongoing events of the game, but to fill in the gap when there are no significant events. This is discussed further in the analysis section (5.3).

The mental processes verbs are transitive in nature, hence, each clause under the mental SFL bundle must have (1) a senser, (2) mental process verb, and (3) a phenomenon. The phenomenon in the mental SFL bundle could be either a participant within the main clause (Type A) or a separate clause (Type B). We identified ten clauses under Type A and eight clauses under Type B. An example of Type A is [16- *You hear the cries of the fans*], here the phenomenon is a participant within the clause, which means the phenomenon *the cries of the fans* is not a clause that stands by itself. On the other hand, [17- *they thought it was a handball*] is an example of Type B. We can see that the phenomenon in example (17) is a clause that can stand by itself [*it was a handball*]. The mental process of thinking [*thought*] is done by the senser *they*. The following shows the categorization of the clauses within the SFL Mental bundle.

Example 16: *You hear the cries of the fans.*

Senser	**Mental process verb**	**Phenomenon**
You	hear	The cries of the fans.

Example 17: *they thought it was a handball.*

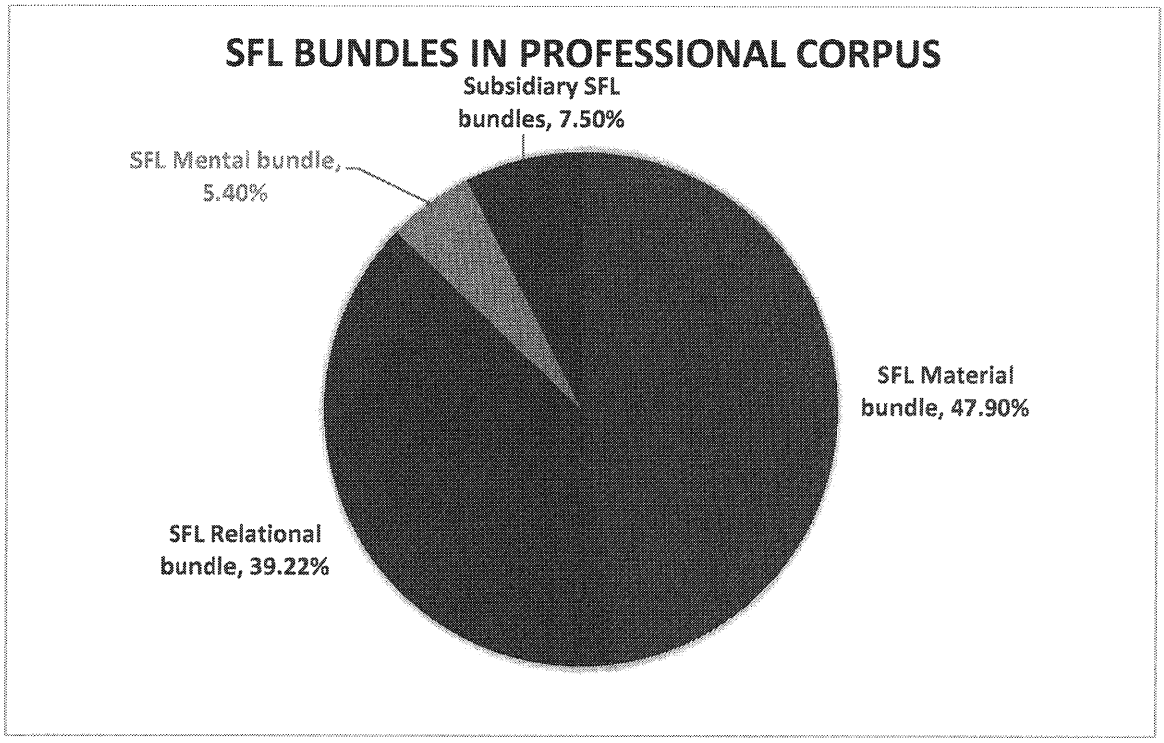
Senser	**Mental process verb**	**Phenomenon (as separate clause)**
they	thought	It was a handball.

4.1.4. Summary of the professional corpus

We can sum up the principle three SFL bundles in the professional corpus by the percentage of each of the SFL bundles as shown in the following graph. We can see that the Material SFL bundle dominates the corpus with an approximate of 48% of all clauses in the data. The relational SFL bundle is also highly frequent in the data as it has an approximate percentage of 42%. The

mental SFL bundle was the least frequent of the principal bundles with 5%, and 7% were subsidiary SFL bundles.

Figure 11: Percentage of SFL bundles in professional corpus



Since the purpose of this research is to distinguish between the professional commentary discourse and the fans' discourse, the results of the fans corpus are presented in the next section (4.2).

4.2. Fans corpus

The data of the fans corpus was extracted from four different twitter accounts which are always active in the special hashtags that are created by the Premier League social media teams. It is also important to note that the data is relatively new (2019 to 2021) compared to the professional corpus that allowed us to compile data from different time frames (2005-2015). We compiled a corpus of 5000 words from these accounts (approximately 1250 words from each account) and

found a total of 313 clauses, 287 of them were clauses that fall under the three principal SFL bundles. The corpus was also predominated by the Material SFL bundle (157 clauses) and the relational SFL bundle (112 clauses). The mental SFL bundle was considerably less frequent than the two as we found only (18 clauses) under it. In addition to 26 clauses which were subsidiary SFL bundles.

Table 12: Frequency of SFL bundles- Fans' corpus

SFL Material bundle	SFL Relational bundle	SFL Mental bundle
157 clauses	112 clauses	18 clauses

4.2.1. SFL Material bundle

In categorizing the clauses under the SFL Material bundle in the fans corpus, we followed the same approach we used in the professional corpus data (See section 5.1.1). There are four structural types under the Material SFL bundle, all the types must include at least two participants which are the actor and the Material process verb. Therefore, in the 157 clauses, we identified 157 actors and material process verbs. The goal participant was also highly frequent as it was found in 142 of the clauses. The beneficiary participant, however, was the least frequent as it was found in 54 of the clauses. Table (13) summarizes the count of the participants in the SFL Material bundle, and Table (14) shows the distribution of clauses under the four types of the SFL Material bundle.

Table 13: Participants in the SFL Material bundle -Fans' corpus

Actor	**Material process verb**	**Goal**	**Beneficiary**
157	157	142	54

Table 14: Frequency of types in SFL Material bundle

	Types under the SFL Material bundle	Number of clauses
Goal directed clauses	**Type A: (Actor + process verb+ goal + Beneficiary)**	45
	Type B: (Actor + process verb+ goal)	97
Non-goal directed clauses	**Type C: (Actor + process verb)**	6
	Type D: (Actor+ process verb+ Beneficiary)	9

4.2.1.1. SFL Material bundle (Type A)

A total of 45 clauses were found to fall under this structure type which contains all the four components of the Material SFL bundle. This structure type represents an approximate of 29% of all the clauses under the SFL Material bundle. An example of this type is [18- *the referee gives a dangerous freekick for Newcastle*] and [19- *De Bruyne passes an intelligent through ball to Sterling*]. The following table categorize the components of the previous clauses.

Example 18: *the referee gives a dangerous freekick for Newcastle*

Actor	**Material process verb**	**Goal**	**Beneficiary**
The referee	Gives	A dangerous freekick	For Newcastle

Example 19: *De Bruyne gives an intelligent through ball to Sterling*].

Actor	**Material process verb**	**Goal**	**Beneficiary**
De Bruyne	passes	An intelligent through ball	To Sterling

4.2.1.2. SFL Material bundle (Type B)

As in the professional corpus, this type is also dominant with 97 clauses, which represent an approximate of 68% of the clauses under the SFL material bundle in the fans corpus. [20- *Silva makes an important interception*] and [21- *Van Dijk intercepts Foden's through ball*] are examples of this type in the fans corpus. We can also find examples in which the whole team is the actor in the clause, as in [22- *Barcelona are dominating the possession of the ball*].

Example 20: *Silva makes an important interception*

Actor	**Material process verb**	**Goal**
Silva	Makes	An important interception

Example 21: *Van Dijk anticipates Foden's through ball*

Actor	**Material process verb**	**Goal**
Van Dijk	Intercepts	Foden's through ball

Example 22: *Man City are dominating the possession of the ball*

Actor	**Material process verb**	**Goal**
Man City	Are dominating	The possession of the ball

4.2.1.3. SFL Material bundle (Type C)

This type is the least frequent in the Fans corpus. We found only Intransitive verbs are commonly used in the examples of this type, such as [*Sterling falls down*], but we also found a clause with a verb that can be transitive or intransitive -in the case of this type it is considered intransitive-, which is [*Dias goes down*]. The following table shows the two components of this type (Actor and Material process verb).

Example 23: *Sterling falls down*

Actor	**Material process verb**
Sterling	Falls down

Example 24: *Dias goes down*

Actor	**Material process verb**
Dias	Goes down

4.2.1.4. SFL Material bundle (Type D)

In this type, intransitive verbs are also used but the difference is in stating a beneficiary or a client that is affected by the action. For example, in the clause [*The ball comes to Mahrez on a golden plate*], we can see that there is no object in this clause, but Mahrez is the receiver of the ball, and therefore can be considered the beneficiary.

Example 25: *The ball comes to Mahrez on a golden plate*

Actor	**Material process verb**	**Beneficiary**
The ball	Comes	To Mahrez

4.2.2. SFL Relational bundle

As in the professional corpus, we categorized the clauses of the Relational SFL bundle into the two main modes, which are attributive and identifying. The same grammatical feature of reversibility was used to distinguish and categorize the clauses into the two modes. In the fans corpus we found a total of 112 clauses, 90 clauses are attributive relational clauses, and 22 identifying relational clauses. As stated in section (2.2.3), the participants of the identifying

relational clauses are tokens and values, while in the attributive relational clauses the participants are carriers and attributes.

Table 15: Types and counts of SFL Relational bundle

Type	Count
Relational Attributive (Type A)	90
Relational identifying (Type B)	22

4.2.2.1. Attributive Relational clauses (Type A)

90 clauses were found to fall under this category. Some of the carriers are nouns, such as [*Stones was aware of his intentions*]. Others are demonstratives and pronouns, such as [*that was a big miss*] and [*They look vibrant today*]. Examples (26, 27, and 28) show the categorization of the participants in this type.

Example 26: *Stones was aware of his intentions*

Carrier	**Relational attributive verb**	**Attribute**
Stones	Was	Aware of his intentions

Example 27: *That was a big miss.*

Carrier	**Relational attributive verb**	**Attribute**
That	Was	A big miss

Example 28: *They look vibrant today*

Carrier	**Relational attributive verb**	**Attribute**
They	Look	Vibrant

4.2.2.2. Relational identifying clauses (Type B)

We found 22 clauses under this type. To ensure that these clauses are of the identifying relational nature, we applied the rule of reversibility. For example, [*Sterling's shot was the highlight of the first half]* and [*this corner might be the last chance for an equalizer*]. Both sentences can be reversed into [*The highlight of the first half is Sterling's shot*] and [*the last chance for an equalizer might be this corner*].

Example 29: *Sterling's shot was the highlight of the first half*

Token/identifier	**relational Identifying verb**	**Value/identified**
Sterling's shot	Was	the highlight of the first half

Example 30: *this corner might be the last chance for an equalizer*

Token/identifier	**relational Identifying verb**	**Value/identified**
This corner	Might be	the last chance for an equalizer

4.2.3. SFL Mental bundle

Based on our analysis, the mental SFL bundle is the least frequent among the three principal types. We found 11 clauses under this SFL bundle. As in the professional corpus, we found the two types of phenomena (within the same clause, and separate clause). Eight clauses had the phenomenon within the clause (Type A), and 3 clauses had the phenomenon as a separate clause. An example of the phenomenon within the clause is [*Spurs' manager senses the weakness on the right flank*]. An example of the phenomenon in a separate clause (projected clause) [*We feel that the equalizer is a matter of time*].

Example 31: *Spurs' manager senses the weakness on the right flank*

Senser	**Mental process verb**	**Phenomenon (within the clause)**
Spurs' manager	Senses	the weakness on the right flank

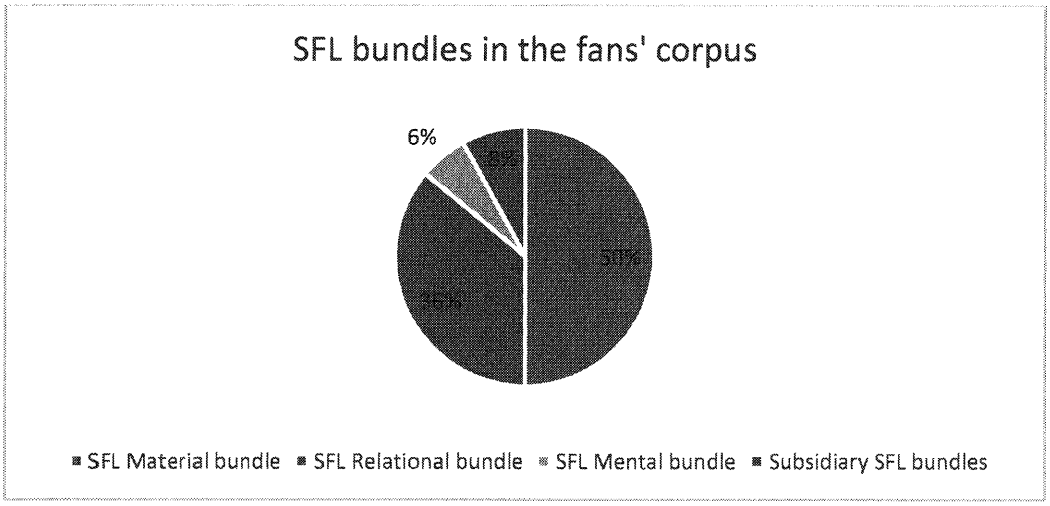
Example 32: *We feel that the equalizer is a matter of time*

Senser	**Mental process verb**	**Phenomenon (a separate clause)**
We	Feel	the equalizer is a matter of time

4.2.4. Summary of the fans' corpus

We can summarize the percentage of the clauses in the fans corpus in the following graph. The Material SFL bundle dominates the fans corpus with an approximate of 50%, the Relational SFL bundle comes next with 36%, the Mental SFL bundle is the least of frequent with an approximate percentage of 6%, and the subsidiary SFL bundles were 8% of the data. Figure (12) visually represents the percentages of the SFL bundles.

Figure 12: Percentages of SFL bundles in the fans' corpus



Now that the data of both corpora are clearly represented, we have the information we need to proceed to the next section (Data analysis), which is the key section to answer the research questions (see section 1.3).

5. Chapter Five: Data analysis

In this section, we present an overall analysis of both corpora (professional and fans) in which we compare between the uses of SFL bundles in the discourse of football commentators and the fans. We then discuss the three SFL bundles in term of frequency, structure, and function. And at the end of the section, we answer the research questions (section 1.3).

As apparent in the findings section, formulaically, we can see that in both data sets the results of formulaic language revealed similar transitivity profiles in the use of process types. This was represented clearly as both corpora were predominated by SFL Material bundle and SFL relational bundle, which are terms coined to organize the structure types of Halliday's processes in our framework (See section 3.1). On the other hand, the SFL mental bundle is considerably less frequent in both corpora. Table (16) below shows the similarity in SFL bundles.

Table 16: Frequency of SFL bundles- comparison between professional and fans

Corpus / **SFL bundle**	**Professional**	**Fans**

Material	47.9%	50%
Relational	39.22%	36%
Mental	5.4%	6%
Subsidiary	7.5%	8%

Nevertheless, this similarity is apparent in general term only as the use of structure types within each SFL bundle was markedly different in the two corpora, such difference in the functional use of the SFL bundle between professional and fans can be found throughout both corpora. We see in this section that some of these differences are deemed to the variation in the proficiency level in the discourse, these variations engender the professional commentary discourse as more sophisticated and technical than the fans' discourse. However, some of these differences can also be attributed to the different mode of delivery of language in each corpus. In the fans' corpus, language is used as the only tool to communicate the idea of their description or analysis without any visual assistance. Professional corpus on the other hand, the language is only an aid that describes the actual live coverage of the game. Therefore, the following sections carefully highlight the contextual differences and their effects on creating the similarities and variances between the two corpora. Therefore, this means that the analysis of the SFL bundles in the following sections differentiate between (1) the different contextual variances in language use due to the varying proficiency level, and (2) the differences that result from the nature of the corpus. Section (5.1) below presents the analysis of the SFL Material bundle, section (5.2) presents the analysis of the SFL Relational bundle, and section (5.3) presents the analysis of the SFL Mental bundle.

5.1. Overall analysis of SFL Material bundle

In this section, we analyze the four different types under the SFL material bundle. The first type is [Actor + Material process verb + goal + Beneficiary] (hereafter Type A), the second is [Actor + Material process verb + goal] (hereafter Type B), the third is [Actor + material process verb] (hereafter Type C), and the fourth is [Actor + Material process verb + Beneficiary] (hereafter Type D). The table below shows a comparison between professional and fans corpora in frequency of each type next to the percentage within the SFL Material bundle.

Table 17: Types of SFL Material bundles- count and percentage

Types under the SFL Material bundle	**Professional**		**Fans**	
	Count	Percentage	Count	Percentage
A:(Actor + process verb+ goal + Beneficiary)	19	12%	45	29%
B:(Actor + process verb+ goal)	127	79%	97	61%
C:(Actor + process verb)	9	6%	6	4%
D:(Actor+ process verb+ Beneficiary)	5	3%	9	6%

In addition to being the prevalent SFL bundle in the professional commentary corpus and fans' corpus, we can see that the 'goal-directed' types of SFL Material bundle are found in similar proportions in the two corpora. There are two types under the SFL Material bundle that has the goal participant which are **Type A** and **Type B.** Both types represent 91% of the professional corpus and 90% of the fans' corpus. This is not surprising as football matches are basically physical actions and bodily movements that have consequences, therefore, these three components (Actor-process verb- goal) are essential to describe important events of the game like scoring goals, passing the ball, defending and so on. There are instances where the same clause was used to

describe similar events in both corpora such as *(Actor which is always a player) wins the ball back for (Beneficiary which is always a team)* and *(Actor which is either a team or a player) gives the ball away*. The fact that such structures are repetitive among different commentators and different fans support the claim of Halliday (1994) that language is inherently related to social context in which it is used. The frequency of these structures prove that professional commentators and fans have shared fixed structured to describe certain events and all they do is fill in the empty slots with the words they see fit, which is essentially the notion adopted in SFL. Moreover, such phenomena prove the centrality of these structures to the proximity between the producer of the language and the receiver of the language that make the descriptions of events easily communicated through these structures.

Moreover, the frequency difference between the Fans corpus and the professional corpus in these 'goal directed' clauses is in the presence of the beneficiary component in the clauses. It is of no surprise to find that the beneficiary is significantly less frequent than the other 'optional' component which is the goal. The goal is the component in the clause that is affected by the process. The nature of actions and physical movements in football usually requires an impacted participant, whether something that is brought into existence because of the process such as in the example "33- *Barcelona created enough chances*" or a thing that is manipulated as a result of the action "34- *Ronaldinho tore the defense apart*". Therefore, we can see that the goal component is found in 91% of the clauses in professional corpus, and 90% of the clauses in the fans' corpus, which shows almost the same proportion. However, we can see a clear distinction between the two corpora in the frequency of beneficiary participant in the 'goal directed' clauses. In the professional corpus, the frequency of the beneficiary participant is 12% only, while in the fans' corpus we find a frequency of 29%. One apparent cause of this distinction is the nature of the corpus; in the

professional corpus the commentary is assisted by the visuals on TV, therefore, we believe that commentators sometimes do not feel the need to state the obvious beneficiary of some actions. However, we want to look at this distinction with an increased delicacy through the examples, hence, let us look at two examples that describe similar attacking action (crossing a ball), one from the professional corpus and the other is from the fans' corpus respectively as follows: (35- *He swings a wonderful cross, but Anelka could not convert*) and (36- *Cantwell plays a great cross for Pukki, who could not hit it right*). The obvious distinction is in stating the beneficiary in the fans' corpus example (35) as we can see a preference of stating the name of the actor who crosses the ball. The example of professional corpus (34), on the other hand, uses the pronoun *He* to describe the actor, disregards the beneficiary from the process, and moves directly to declare the missed opportunity which is the result of the action. Therefore, we can clearly realize that the simultaneousness of the professional commentary with the visual live coverage is key in choosing the wording to be used to describe the action. In fact, language in the professional commentary seems to be an ancillary to the televised coverage, while in the Fans corpus, language is the only medium to describe the action for their twitter followers. Hence, this distinction is not a matter of language competence as it is a matter of the nature of the corpus. Therefore, we can see that the quantitative analysis alone does not provide us with enough information regarding the proficiency. This is why we decided to look from a different direction and apply a mixture of quantitative and qualitative analysis to examine the differences between the two corpora in the SFL Material bundle.

Another distinction between the two corpora, which is less related to the nature of corpus and more related to the referential linguistic competence of the speaker/writer, is the choice of content words -or in this case, the participants in the clause-. Choosing the appropriate technical words to describe

the actions signals competence in the field as they increase the 'lexical density' in the text. Measuring the lexical density based on a Hallidayan framework would not be a simple comparison between content words which consist of nouns, verbs, and so on, against grammatical items such as determiners and preposition. According to Halliday, "For a systematic and formal investigation of lexical density in texts, we should adopt some weighting system whereby lexical items of lower frequency 'score' more highly than common ones" (1985, p. 65). And since calculating the lexical density of the corpora is beyond the scope of this study, and we only intend to examine whether the professional commentary uses denser, and more proficient terms, we decided to conduct a simple search in COCA corpus to compare the frequency of three Material process verbs used in similar situations. The results of this search were aligned with our assumption, the material process verbs that were used in the fans' corpus were more common than the ones used in the professional corpus, which makes the latter a denser discourse. The examples of crossing the ball *(35)* and *(36)* show a choice of the verb swing in the professional corpus, and the verb play in the fans' corpus. Following Halliday's hypothesis which can be summarized as 'the lesser the frequency, the denser the lexical item feels', we can deduce that the clauses of the professional corpus are denser, and thus more proficient in the jargon. This becomes clearer in several other events in our data. In describing a missed chance, the professional commentator chose to say (37- *Etoo squandered the opportunity*), while in similar event the fan chose to tweet (38- *Sterling wasted a big chance to score*). The table below (18) shows the comparison in the choice of material process verbs, and their frequency in COCA corpus. We also added these examples of tackling to the comparison which are (*39- Abidal dangerously lunges at him*) from the professional corpus, and (*40- Ruben Dias slides on the loose ball*) from the fans corpus.

Table 18: comparison of Lexical density- process verbs

Verbs used and their frequency according to COCA				
Event	**Fans**	**Frequency**	**Professional**	**Frequency**
Crossing a ball	Play	286583	swing	27301
Tackling	Slide	19471	lunge	1190
Missed chance to score	Waste	50556	squander	694

According to Halliday's assumption stated earlier, the significant difference in frequency among the verbs used in the fans' corpus examples (36), (38), and (40) in comparison to the verbs used in the professional corpus examples (35), (37), and (39) indicate that the professional corpus is generally denser, and hence more proficient than the fans corpus in the SFL Material bundle. The non-goal directed clauses are significantly less frequent because of the nature of discourse that we are targeting. In such clauses, the verbs are intransitive with the actor being the only one affected by the verb.

Finally, we can find Two types of clauses that are non-goal directed which are **Type C** and **Type D.** In these clauses, the frequency is low in both corpora, therefore, we decided to shed the light on them from a qualitative perspective, as the quantitative analysis alone is not enough to provide us with sufficient information about their usage. An important function we noticed in the clauses under the type **(Actor+ process verb)** is that they always serve a locational function in both corpora. For example, in the professional corpus we find (7- *Ronaldinho comes forward*) and in the fans' corpus (41- *Mo Salah cuts inside*); these are both clauses that indicate an actor moving toward a position on the pitch. Such clauses are usually used when the ball is in places that are the most newsworthy on the football pitch and being in those places would create a potential goal-

scoring opportunity. Interestingly, example (41) from the fans' corpus is identical with example (6) from the professional corpus (the only difference is the name of the player), which is a clear indication of the similar functionality of types **C & D** in both corpora, in addition to the notion of SFL that aligns language use with the context and type of discourse in which it was used.

5.2. Overall analysis of SFL Relational bundle

In the SFL Relational bundle, we have examined two types of structure which are: the relational attributive (Type A) and the relational identifying (Type B). To analyze the variations between the two types, we are going to analyze the functionality of the clauses according to the functional categories provided by Halliday and Matthiessen (2004) and Martin et al (1997) who classified them into intensive, possessive, and circumstantial. This is in addition to analyzing the context of the clauses in relation to the aspects of the game. Table (19) shows the distribution of the clauses under the SFL Relational bundle to two types in the two corpora.

Table 19: SFL Relational bundle- Types, counts, and percentages

Types under the SFL Relational bundle	**Professional (131)**		**Fans (112)**	
	Count	Percentage	Count	Percentage
Type A: Relational attributive	95	73%	90	80%
Type B: Relational identifying	36	27%	22	20%

The majority of clauses under the SFL Relational bundle in both corpora are attributive clauses in which there are attributes being assigned to carriers, or as Halliday (1994) described them as "qualities that are ascribed to entities". The dominance of the attributive relational type is expected as the main function of such clauses is the description of entities, which is arguably considered the main function of football commentary. However, this similarity on the surface is

met with some distinctions within the function of the clauses. An interesting observation emerged from the analysis is that within the SFL Relational bundle, professional commentary corpus included more evaluative judgments and personal opinions on the events than the fans' corpus. This can be attributed to the visual assistance to the professional commentators; comparatively, the fans corpus is constrained by more obligation to describe the events as they operate under the assumption that their audience do not necessarily have access to the televised coverage. This becomes evident as we see separate clauses in the professional commentary in which both the analytical and the descriptive commentators are more likely to express their opinions in separate clauses, while the opinions of the fans are included within the same descriptive clause or in non-clausal utterances. For example, in the professional corpus, we can see examples like (42- *The through ball by Ronaldinho was ably anticipated by Terry, that was a superb reading of the game*) or (43- *Chelsea's Stamford Bridge is the place of the second encounter, it is one of the most prestigious stadiums in the world of football*). Here we can see a description of the event, -which the anticipation of a through ball in the first clause-, then after it directly a clause which represents the admirative opinion of the descriptive commentator. The fans might not always get the luxury of conducting a clause to express their opinion. In a similar context, we can see in the fans' corpus an example like (44- *Salah passes it through, but Stones was there to intercept it intelligently*) or clauses that include phrases such as (45-*It was a one-on-one wasted opportunity for Sterling, What a miss!*). Such distinction does not imply that the professional commentary is easier to practice, as the professional commentary has its own complications too. For instance, In the televised coverage, there are the action-replays which are replays of important actions that happened in the game. These replayed important actions incline the professional commentators toward a greater frequency of evaluation; therefore, clauses of evaluative judgements and expressions of opinions

are required for appropriating the commentary with what is on screen. On the other hand, "reply talks" as Marriott (1996) describes them, are not required from the fans, which explains the lack of them in the fans' corpus.

Moving from the distinctions to the similarities, we stated earlier that in both corpora, the attributive relational type is the dominant over the identifying relational one. Both types have three functional categories which are intensive, possessive, and circumstantial (Halliday, 1994). The similarity in dominance is not in the type only but runs further in term of the functional categories and the context of use. The clauses that serve the intensive function are the most dominant in both corpora. In the professional corpus, we see that 54 of the 95 attributive relational clauses, and 19 out of 36 identifying relational clauses are serving an intensive function. A clear intensive function within the SFL relational bundle can be found in example (13) of the professional corpus, "13- *Chelsea's Stamford Bridge is the place of the second encounter*", which is of the identifying relational type. The intensive function in example (13) specifies a timely phase of identifying to which X becomes A, or as Halliday & Matthiessen (2004) described it "X identifies A". In example (13), (*Chelsea's Stamford Bridge*) is the token (A), which is construed to be identified by (*the place of the second encounter*) which is the value (X). In the previous example, the token identifies the value because of an event unfolding in a certain time frame. This becomes clear when asking the question "Where is the place of the second encounter?" and we use Chelsea's Stamford Bridge to identify it. The exact same functional analysis can also be applied to example (29) of the fans' corpus (29- *Sterling's shot was the highlight of the first half*) when we ask, "What is the highlight of the first half?".

The intensive function is inversely conveyed in the (Type A) as in example (9) of the professional corpus, (*9- that was an optimistic effort*) and example (27) of the fans' corpus (*27-*

that was a big miss). In these examples, the attributes are being assigned to carriers, but the attribute in these examples does not identify the carriers as equals to the attributes like values do with tokens in Type B. The possessive function is clearly demonstrated in example (14) of the professional corpus, (11- *Ronaldinho has the ability to turn the game on its head*) and this example from the fans' corpus, (46- *Salah has got a wonderful dribbling ability*). We can see the resemblance in function between the two corpora in examples (11) and (46) as they describe the players' ownership of certain abilities using attributive relational clauses. We also find examples that show resemblance in describing teams as in this example of professional corpus, (47- *Barcelona definitely have more shots on target*) or in this example of the fans' corpus, (48- *Spurs have a chance to equalize from this corner*).

The last function is the circumstantial function which in Type B mainly describes relationships between the two participants in the clause (time, place, manner, and so on). Such function is found in this example of professional corpus, (48- *The best tactic for United was to fall back and wait for the opportunity to counterattack*). In (48) we can see a relationship that is based on manner between the token which is (*the best tactic for United),* and the value, which is the prepositional phrase (*to fall back....to counterattack*). In the fans' corpus, we see the use of the circumstantial function operates the relationship of place in the identifying relational clauses in this example (49- *Matip's positioning was far away from the near post*).

5.3. Overall analysis of SFL mental bundle

The SFL mental bundle is not a significant part of both corpora, as it was the least frequent of the principal bundles (which also include Material and Relational). This is not surprising as actions of personal perceptions (hearing, seeing, and so on), personal cognition (understanding,

thinking), or even personal desiderative (hoping, wanting) are not related to the purpose of reporting the events of football matches. Therefore, with the little need to represent these actions, very few clauses were found under the SFL mental bundle. We have observed that these instances fall within certain themes and/or functions. To start with, we have noticed that the clauses under the SFL mental bundle are used by professional commentators to fill in the gap when there are no significant actions on the field. For instance, example (16) of the professional corpus, (16- *You hear the cries of the fans*) is used while the game was stopped due to an injury. The descriptive commentator here used the mental clause to strike a conversation with the analytical commentator instead of keeping silent. The fans, on the other hand, use such clauses to communicate their prediction based on the flow of the game, as in example (32), (32- *We feel that the equalizer is a matter of time*). One function that both corpora have in common is communicating the feeling of the players or managers. We see example (17) of the professional corpus (17- *they thought it was a handball*) is communicating the feeling of the players after a controversial verdict from the referee of not giving a penalty. We also see example (31) of the fans' corpus, (31- *Spurs' manager senses the weakness on the right flank*) after the manager decides to substitute the right back after a shaky performance.

Now we have all the needed information to answer our three research questions.

5.4. Answering the research questions

In this section, we directly answer the research questions. These answers are built up throughout the previous chapters and compiled in this section.

Question 1: What are the Material, Relational, and Mental SFL bundles that are predominantly used in the corpus of professional commentators in the context of football matches?

To answer this question, it is important to revise and summarize the distribution of SFL bundles within the professional corpus as reported in the findings section (See section 4.1). First, The SFL Material bundle is the predominant SFL bundle. We conclude that this dominance is natural as SFL Material bundle is based on the material processes which are descriptive of events and actions. Describing the events and actions in the football pitch is crucial in the football context. Therefore, we can see that within a corpus of 5000 words that contained 334 clauses, the clauses of the SFL Material bundle represents a 48% of the total clauses (See section 4.1.1). The type of clauses that predominates the corpus are the ones that have actor, material process verb, and goal (Type B). Moreover, Type A which contained all the components of Type B in addition to the beneficiary component is the second in frequency. Hence, we deduce that in the context of football commentary, the dominant clauses in SFL Material bundle are the ones that are "goal-directed". The remaining two types "non-goal directed" are also present in corpus but with significantly less frequency.

SFL Relational bundle, which has the two types, Attributive (Type A) and Identifying (Type B), is the second most frequent SFL bundle after the SFL Material bundle as it represents 39% of the total clauses. Both types and their functions are essential in describing and identifying the elements in the football context (see section 5.2 for detailed description of the functions). The predominant type within the SFL Relational bundle is Type A which takes within its umbrella the clauses that ascribe qualities to entities.

The clauses within the SFL Mental bundle were the least frequent as they represented only 5% of the total clauses. Mental clauses, being descriptive of sensical actions (See section 2.2.2) are least likely to be used by commentators during the in-play actions of the football match. The lack of clauses in the corpus makes difficult to pinpoint the linguistic aspects of the SFL Mental bundle, however, we can conclude from the contextual analysis that such clauses are used to describe some events and aspects of the game other than the action on the pitch.

Question 2: What are the Material, Relational, and Mental SFL bundles that are predominantly used in the Fans corpus?

The answer for this question lies in section 4.2 in which the results of the fans corpus are reported. The total clauses in the fans corpus are 313 in which the SFL Material bundle represents 50% of the total clauses. Within the SFL Material bundle, Type B is the predominant type with 95 clauses as "goal-directed" types of SFL Material bundle dominates the fans corpus as well. Type A, which is also a goal directed type has a 47 clause. The "non-goal directed" types were present but with much less frequency.

The second bundle in frequency is the SFL Relational bundle which represents 36% of the clauses in corpus. Type A (Attributive relational) is also the predominant within the SFL Relational bundle in the fans corpus with 90 clauses. Type B (Identifying relational) is also present in corpus with 22 clauses.

The SFL Mental bundle has the least frequency as it is only 3% of the clauses in corpus with a total of 11 clauses. Type A has 8 clauses and Type B contained 3 clauses.

Question 3: What are the similarities and differences in frequency of structural types, and in functional usage between the SFL bundles used/found in the corpus of professionals and fans?

As the aim of this study is to compare the professional discourse with the fans discourse, an essential part of the research is to identify the similarities and differences. The answer for this question is scattered throughout the sections and can be summarized in the following points.

1. SFL Material bundle is the Most frequent SFL bundle in both corpora

. The similarity in frequency of structural types can clearly be realized through the answers for question 1 and question 2 as both corpora are predominated by the SFL Material bundle.

2. Goal directed types of SFL Material bundle: predominant in both corpora.

Within SFL Material bundle, the two types which are identified as "goal directed"[3] are most frequent in both corpora as they represent 44% of the total clauses in the professional corpus and 45% of the clauses in the fans corpus.

3. Beneficiary is considerably more frequent in fans corpus.

Regardless of the apparent similarity in the frequency of structure, breaking down the use of Type A and Type B within each corpus revealed that the beneficiary component is used considerably more often in the fans corpus than in the professional corpus. This distinction was regarded to the difference in nature between the professional corpus -which is accompanied by visuals- and fans corpus -which the action is described in written form-.

[3] Type A and Type B

4. Clear difference of competence as aspects of lexical density are applied to the clauses of SFL Material bundle.

An aspect of lexical density theory is used to identify differences in competence, which is the weight of content words in clauses. By a simple comparison between similar events that are described in both corpora, it becomes evident that the professionals use vocabulary (content words) that are denser and more proficient than the fans. Such difference in the choice of words in similar functions is regarded to the gap in the referential linguistic competence between one and the other (See section 5.1).

5. Type A of the SFL Relational is prevalent in both corpora, but the difference lies in the functions.

Within the SFL Relational bundle, Type A (Relational attributive) is also the prevalent in both corpora. However, Type A clauses in the professional corpus contained more evaluative judgments and personal opinion than the fans corpus. This is regarded to the visual assistance to the professional commentators, comparatively, the fans corpus is constrained by more obligation to describe the events as they assume their audience do not necessarily have access to the televised coverage. Hence, we see that the fans prioritize the description of events over expressing their personal opinions. On the other hand, professional commentators are obliged to provide more evaluative analysis as the televised coverage replays the important events of the game and the commentators must appropriate what they say with what is on screen.

6. Functional categories of Halliday, similar frequency in both corpora, similar functions.

Moving to the analysis based on Halliday's functional categories, it was evident that the majority of the clauses in both corpora are of an intensive function, which is not a surprise as one of the

main aims of commentary is to ascribe qualities to the carriers within a certain even in the match, which the purpose of the intensive function (See section 5.2 for the examples). The second function that is frequent in both corpora is the possessive function, in which the clause is used -in the context of football- to describe the players' ownership of some attributes. The least frequent function is the circumstantial in which the clause conveys a relationship of a certain manner between carriers and attributes in case of Type A and the token and value in case of Type B.

7. SFL Mental bundle: the least frequent in both corpora.

18 clauses in each corpus are the clauses within the SFL Mental bundle. It is not a surprise as the SFL Mental bundle contains clauses that describe actions of personal perceptions and personal cognition. And even though such actions are present in the context of football, yet they are not the main aim of it. Therefore, such clauses are used in professional commentary to fill in the gap when there are no significant actions on the pitch. The fans, on the other hand, use such clauses to communicate their prediction based on the flow of the game. Moreover, the clauses of SFL Mental bundle is also used to communicate the feelings of the players and managers on the pitch, which is a function that is found in both corpora.

6. Chapter Six: Conclusion

6.1. Summary

The purpose of this research is to compare the linguistic aspects of the Fans discourse to the Professional discourse within the discourse of football. The Fans discourse being a new addition to the discourse of football -with the emergence of social media- was not examined before. Therefore, to achieve this purpose, the theoretical framework of this study combines two theories that share the concept of formulaicity (SFL and Lexical bundles). The combination of these two theories in one framework has led us to coin the term SFL bundles, which is a concept coined to include the aspects of both SFL and Lexical bundles. All sections of this study revolve around the three research questions in section (1.2), therefore, the answers for these questions are built up throughout the methodology (chapter 3), findings (chapter 4), and the analysis (chapter 5) sections, until we were able to directly answer them in the last section of chapter 5.

Based on the discussion of the three principal SFL bundles, we see that the data partially proves the notion adopted in this study which is "Football is like any other academic or professional field in the way its texts are assessed" as some of the differences that are found between the two corpora are regarded to the difference in the mood of delivery rather the proficiency in the discourse of football. Now, we summarize the most important findings of this study based on our theoretical framework (SFL Bundles).

In the SFL Material bundle, we investigated aspects of Halliday's lexical density (1985) to compare the verbs used in both corpora (See section 6.1 & Table 18). The results showed that professional commentators used denser, more technical verbs, therefore; we can -based on these results- prove the assumption that professional commentators are more proficient than fans' in

describing the events of the football matches. This was also a clear indication of the appropriateness of the researcher's approach that treated the discourse of football like the other academic, scientific, and professional discourses whose proficiency can be assessed through the formulaicity of language. This, however, does not indicate a lack of ability in describing the events of the football matches in the fans' corpus, because there were more linguistic similarities than differences when it came to the quantitative analysis of data. In addition to the clauses within the SFL Material bundle being the most frequent in both corpora, Type B and Type A were also the most dominant in the SFL Material bundle. It was our assumption that the quantitative data were not enough to provide us with the information needed to distinguish the proficiency level between the two corpora, therefore, we incorporated qualitative analysis of functions to our discussion.

In the SFL Relational bundle, we shed lights on the way the evaluative judgments were used. We noticed that professional commentators were frequently using separate clauses to convey these evaluative comments or judgments, while the fans were including them within the same clause or using phrases to express but not complete clauses. The quantitative analysis was also in favor of equalizing the proficiency of the fans with the professional commentators as no significant difference were evident in the usage of types. Moreover, the functional analysis showed more similarities than difference in the contextual use of the SFL Relational bundle.

Last but not least, the SFL Mental bundle was of less importance in the discourse of football as football is a game of physical movements and actions, while the mental clauses are descriptive of cognitive actions. Nevertheless, both corpora shared the same function of filling the gap when there were no physical actions on the pitch, or when the play stops.

The next paragraph includes recommendations for further research.

6.2. Research Recommendations

This study considers only the principal processes of Halliday which are material, relational, and mental as the types that formulate the contextual texture of the discourse of football [4]. And even though these principal processes are the dominant in the context of football, not tackling the subsidiary processes leaves a linguistic area unexplored. Therefore, we call for further research which takes into account exploring all types of Halliday's processes. Moreover, this study compiled a relatively small corpus (5000 words for each corpus), which might not be enough to explore all the formulaic aspects within the SFL bundles. Therefore, the second recommendation is to apply the concept of SFL bundles on larger corpora.

[4] Hence, the term SFL bundle is coined based on them.

7. References

Altenberg, B. (1998). On the Phraseology of Spoken English: The Evidence of Recurrent Word-Combinations. In A. Cowie, *Phraseology: Theory, Analysis, and Applications* (pp. 101-122). Oxford: OXFORD University Press.

Austin, J. L. (1962). *How to Do Things with Words* . OXFORD.

Bache, C. (2008). *English Tense and Aspect in Halliday's Systemic Functional Grammar: A Critical Appraisal and an Alternative (Discussions in Functional Approaches to Language).* Sheffield: Equinox Publishing Limited.

Bauer, L., & Renouf, A. (2011). A corpus based study of compounding in English. *Journal of linguistics.*

Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes 26*, 263-286.

Chen, Y.-H., & Baker, P. (2010). Lexical Bundles in L1 and L2 in Academic Writing . *Language Learning & Technology*, 30-49.

Cowie, A. P. (1996). Multiword Lexical Units and Communicative Language Teaching. In P. J. Arnaud, & H. Béjoint, *Vocabulary and Applied Linguistics* (pp. 1-12). London: Palgrave Macmillan.

Duţă, O. A. (2010). A Brief History of Theories on Idiomatic Expressions. *BULETINUL*, 37-42.

Eggins, S. (2004). *An Introduction to Systemic Functional Linguistics* . London: Continuum international Publishing Group.

Ellis, N. (1996). Sequencing in SLA: Phonological memory, chunking, and points of. *Studies in Second Language Acquisition, 24*, 143-188.

Ferguson, C. (1994). Dialect, Register, and Genre: Working Assumptions about Conventionalization. In D. Biber, & E. Finegan, *Sociolinguistic Perspectives on Register* (pp. 15-30). Oxford University Press.

Ferguson, C. A. (1983). Sports announcer talk: Syntactic aspects of register variation . *Language in society* , 153-172.

Firth, J. R. (1951). Modes of meaning. Essays and Studies. *The English Association*, 118-149.

Halliday, M. (1985). *Spoken and written language.* Oxford: Oxford University Press.

Halliday, M. (1994). *An Introdction to Functional Grammar* . London: Edward Arnold.

Halliday, M. (1994). *Introduction to Functional Grammar.* London: Arnold Publishing.

Halliday, M., & Matthiessen, C. (2004). *An Introduction to Functional Grammar.* London: Hodder Arnold.

Haswell, R. (1991). *Gaining ground in college writing: Tales of development and interpretation.* Dallas: Methodist University Press.

Hilpert, M. (2019). *Construction Grammar and its Application to English.* Edinburgh: Edinburgh University Press.

Hoey, M. (1991). *Patterns of lexis in text.* Oxford: Oxford University Press .

Hoey, M. (2005). *Lexical priming: A new theory of words and language.* London: Routledge.

Hoey, M. (2005). *Lexical priming: A new theory of words and language.* London: Routledge.

Hoey, M. (2005). *Patterns of lexis in text.* Oxford: Oxford University Press.

Hyland, K. (2008). Academic clusters: text patterning in published and postgraduate writing. *International Journal of Applied Linguistics*, 41-62.

Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *Science Direct*, 4-21.

Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *Science Direct*, 4-21.

Irvine, J. T. (1978). Formality and Informality in Speech Events. *Working Papers in Sociolinquistics, No. 52*, 1-24.

Jespersen, O. (1922). *Language, its nature, development, and origin.* New York: Henry Holt.

Jespersen, O. (1924). *The philosophy of grammar.* London: Allen & Unwin.

Jespersen, O. (1924). *The Philosophy of Grammar.* Great Britian : Unwin Brothers LTD.

Johansson, V. (2008). Lexical density and Lexical diversity in speech and writing: developmental perspective . *Lund University*, 61-79.

Katz, J., & Fodor, J. (1963). The Structure of A Semantic Theory. *Language 39*, 170-210.

Labov, W. (1966). The Social Stratification of (r) in New York City Department Stores. *Washington Center for Applied Linguistics.*

Larsson, T. (2019). Grammatical stance marking across registers: Revisiting the formal-informal. *Register Studies 1 (2)*, 243-268.

Lyon, J. (1963). *Structural Semantics: An Analysis of Part of the Vocabulary of Plato.* London: Blackwell .

Martin, J. R., Matthiessen, C., & Painter, C. (1997). *Working with Functional Grammar.* London: Arnold, Hodder Headline Group.

Pfeiffer, C. (2014). .Phraseologie in der Fußballberichterstattung der Printmedien. In V. J. Dobrovol'skij, *Phraseologie und Kultur* (pp. 491-515). Maribour: Univerza v Mariboru.

Shin, Y. K. (2019). Do native writers always have a head start over nonnative writers? The use of lexical bundles in college students' essays . *Journal of English for Academic Purposes*, 1-14.

Sinclair, J. (1991). *Corpus, concordance, collocation.* Oxford: OUP.

Ure, J. (1971). Lexical density and register differentiation. *Applications of linguistics: selected papers of the second International Congress of Applied Linguistics, Cambridge*, 442-452.

Wardhaugh, R. (2006). *An Introduction to Sociolinguistics.* Blackwell.

Wilton, A. (2017). The interactional construction of evaluation in post-match football interviews. In D. Caldwell, J. Walsh, E. W. Vine, & J. Jureidini, *The discourse of sport: analyses from social linguistics* (pp. 92-112). London: Routledge.

Wray, A. (2002). *Formulaic Language and the Lexicon* . Cambridge: CAMBRIDGE UNIVERSITY PRESS.

Wray, A., & Perkins, M. (2000). The functions of formulaic language: An integrated. *Language 20 (1)*, 1-28.

Wright, H. R. (2019). Lexical bundles in stand-alone literature reviews: Sections,frequencies, and functions. *English for Specific Purposes*, 1-14.

Yorio, C. A. (1989). Idiomaticity as an indicator of second language proficiency. *Bilingualism across the Lifespan: Aspects of Acquisition, Maturity and Loss. Cambridge University Press*, 55-72.

الملخص

هذه الدراسة تتبنى فكرة أن الخطاب المستخدم في التعليق وتحليل كرة القدم هو كأي خطاب أكاديمي أو مهني آخر من ناحية تقييم الجودة اللغوية، فبطبيعة الحال يكون المتخصص في المجال قادرًا على إنتاج خطاب أكثر جودة من غير المتخصص؛ لذلك مع ظهور برامج التواصل الاجتماعي (تويتر تحديدًا)، أصبح للمشجعين دور في خطاب التعليق وفي خطاب تحليل كرة القدم، مما جعل ذلك الخطاب غير محصور على المحللين والمعلقين الرياضيين. في هذه الدراسة، نهدف إلى تحري الخطاب المستعمل من قبل المحللين والمعلقين في القنوات الرياضية ومقارنته مع الخطاب المستعمل من قبل المشجعين على برنامج التواصل الاجتماعي (تويتر) باستعمال مفهوم (الحزم النظامية الوظيفية).

والحزم النظامية الوظيفية مصطلح صِيغَ خصيصاً لهذه الدراسة من خلال دمج مفهومي (الحزم المعجمية) و (اللغويات النظامية الوظيفية) اللذَين يشتركان في فكرة أن التركيب اللغوي للنصوص يميل للتشابه في المفردات المستعملة في كل نوع من أنواع الخطاب، وهو ما يطلق عليه بالـ .Formulaicity هذه الدراسة تتبنى هذه المفاهيم في محاولة لإثبات أن خطاب المعلقين والمحللين الرياضيين أكثر جودة لغوية من المشجعين. ولتحقيق هدف هذه الدراسة قام الباحث بتجميع متن لغوي من خمسة آلاف كلمة للمحللين والمعلقين الرياضيين، وتجميع متن لغوي آخر للمشجعين من خمسة آلاف كلمة. وقد قام الباحث بتوزيع كل جملة في هذه المتون في برنامج الاكسل وفقًا لانتماء الجملة للحزمة النظامية الوظيفية المناسبة. أظهرت النتائج في كلا المتنين غلبة كمية واضحة لـ) حزمة الأحداث النظامية الوظيفية (SFL Material bundle -وهي التي تندرج تحتها الجمل المتعلقة بالأفعال البدنية حسب تصنيف هاليدي. غير أن هذا التشابه على المستوى الكمي كان سطحيًا، فقد رُصدَتْ اختلافات جوهرية عند تحليل الأنواع تحليلًا نوعيًا داخل كل حزمة من الحزم النظامية الوظيفية. فعلى سبيل المثال، لم تكن الاختلافات اللغوية فقط بسبب القدرة اللغوية والتخصصية في المجال؛ بل إن هنالك أسبابًا أخرى، منها أسباب متعلقة بطبيعة المتن اللغوي المأخوذ (مكتوبًا أو متحدَّثًا به). اختتمت الورقة البحثية بتوصية أن يكون البحث على متون لغوية أكبر من خمسة آلاف كلمة كي تتاح الفرصة لدراسة جميع الحزم النظامية الوظيفية سواء كانت أساسية أو فرعية

جامعة البحرين

كلية الآداب

قسم اللغة الإنجليزية وآدابها

تحليل الحزم المعجمية في خطاب رياضة كرة القدم بين المعلقين المحترفين والمشجعين: نهج نظامي وظيفي

أطروحة مقدمة كجزء من متطلبات الحصول على الماجستير في
دراسات اللغة الإنجليزية التطبيقية

إعـــداد

محمد إبراهيم عبدالنبي إبراهيم

الرقم الجامعي: ٢٠١٢٢٢٨٤

إشـــراف

د. حليمة كاظم العالي

أستاذ مساعد

جامعة البحرين

مـملكة البـــحرين

مايو ٢٠٢٢م