

Vilnius University
Faculty of Philology
Department of English Philology

Živilė Kulytė

A Contrastive Analysis of Lexical Bundles in Automatised and Human-
Written Football Live Text Commentaries

Thesis submitted in partial fulfilment of requirements

for the degree of BA in English Philology

Supervisor: Assoc. Prof. Dr Rita Juknevičienė

Abstract	3
1.Introduction	3
2.Data and methods.....	7
3.Results and Discussion	11
3.1.Length and overlap of lexical bundles in Corpus_{Guardian} and Corpus_{Automated}	11
3.2.Structural features of lexical bundles in Corpus_{Guardian} and Corpus_{Automated}	14
3.3.Functional features of lexical bundles in Corpus_{Guardian} and Corpus_{Automated}	21
3.4.General discussion of features of football commentary language	27
4.Conclusions	28
Summary in Lithuanian	31
References	32
Appendix	40

Abstract

The present paper focuses on the comparison of two types of football live text commentaries, those written solely by reporters and those written with the help of automation systems. The aim of the paper is to find whether the two language varieties exhibit any differences or similarities in terms of formulaic language. The method chosen to fulfil this aim is the lexical bundle approach (Biber et al. 2004). The results of the analysis indicate that human-written and automated live text commentaries differ greatly with respect to lexical bundles. The two varieties are completely different structurally: human-written commentaries largely consist of prepositional and noun phrase fragments, while automated ones are predominately composed of independent clauses. Functionally, the varieties do exhibit a leaning towards the referential function but prioritise to indicate different aspects of a football match. Automated live text commentaries contain many stance expressions, while human-written commentaries rarely express this function. The principal conclusion of this study asserts that the language varieties under investigation are inherently different in terms of recurrent lexical bundles, their structure and functions.

Key words: football discourse, lexical bundles, automated language, human-written language, live text commentaries

1. Introduction

Being the sport that captivates billions of people across the world (Pifer et al. 2018), football garners interest not only from the fans of the game but also from linguists. The language of football, just as any other language, has undergone considerable changes with time. Nowadays, when every facet of life is being altered by new technologies, it is no surprise that language production is also being affected and football discourse is no exception. Since it is important to keep up with the innovations and critically assess them, the paper aims to compare human-produced football language and language that has been generated with the help of automated tools.

Football discourse has attracted the attention of numerous linguists. All types of football discourse have been analysed in a myriad of different ways. Football match reports which are usually published on news sites and contain the account of the most important events of the already concluded match, are very popular texts for investigation. Match reports have been investigated cross-linguistically (Ebeling 2021), in terms of the presence of theologising language (Kaur 2019)

and by comparing computer- and human-generated language (Juknevičienė and Viluckas 2019). Chants or songs produced by fans within a stadium at a given game are another type of football discourse. The analysis of chants has shown how they can reflect the sociocultural environment (Huddleston 2022) and how alterations in chants coincide with people's changing values (Tamir 2019). Football match conferences and interviews have also been used for linguistic analysis, researchers have looked at the role of the interpreter in such settings (Sandrelli 2015) and the design that interviewers employ for their question turn taking (Wilton 2021; Rhys 2016). Online football forums have been analysed in terms of female experiences of the medium (Hynes and Cook 2013) and racism (Cleland 2013). Spoken football commentary has been investigated with a variety of methods, by looking into the multimodality of commentating (Gerhardt 2008), the differences between radio and tv commentating (Trouvain 2011) and comparing the emotional content of commentating in different languages (Lavric and Wolfgang 2008). These are just a few examples of different papers written on the various types of football language. The sheer number of linguistic papers dealing with different aspects of football discourse shows the huge interest that this sport generates among linguists.

The type of football discourse that the current research is concerned with is live text commentary. This type of commentary is a fairly new genre which emerged together with the new technological developments of the Internet (Jucker 2010: 58). Live text commentaries are described as 'written accounts of sports events that are produced and published incrementally on the Internet while the event is unfolding' (*ibid.*). Meier-Vieracker adds a list of characteristics that are indicative of the genre, namely, 'a series of entries listed in reverse chronological order' (2021: 277) the entries consist of a time stamp and a description of events; the entries can be either human- or computer-written or a combination of the two; the commentaries often include certain multimodal aspects; they can be read both live or after the event has finished (*ibid.*). With the spread of the Internet in the 1990s, such commentaries emerged because the written word could be published and read within seconds. News websites have been thought of as the originators of the genre, with *The Guardian* being presumably one of the first to start making such accounts of football matches in English in the late 90s (McEnnis 2016 as cited in Meier-Vieracker 2021: 277). Such commentaries of football matches are referred to in a few different names which include 'play-by-play', 'minute-by-minute' commentary, 'event tracker' and many more. Even though each name has its specific nuances the umbrella term 'live text commentary' (henceforth – LTC), seems to encompass all of the specificities of the genre.

Live text commentaries have been subjected to a variety of analyses. Chovanec has looked at the competitive interactions that occur within these commentaries between the commentator and the readers' opinions which are interpolated in some of the texts (2006). Certain LTCs include an interactional element in which the commentator encourages readers to send in their thoughts and then picks out certain comments to include and answer in the public commentary (Chovanec 2006: 25). Chovanec, after analysing four different live text commentaries from the news site *The Guardian*, has concluded that with the use of humour, irony, wordplay and other types of verbal wit, a rivalry is formed between the commentator and the readers that is good-natured and forms a sense of community among football enthusiasts (2006: 34). Another linguist has investigated seven live text commentaries of football matches posted on *The Guardian* website with regards to the conceptual metaphor theory (Bergh 2011). The aim of the research was to check how pervasive was the already established notion of the FOOTBALL IS WAR metaphor in the medium of live text commentaries (*ibid.*). In his case study, Bergh found that the language that commentators use in this medium is filled with words and expressions that are taken from the conceptual domain of war and are applied to football, intensifying the dramatic elements of the game (2011: 91). Another paper analysed how the genre of live text commentaries, while being a written medium, incorporates elements of spoken language (Chovanec 2009). Graphological and lexical features are cited among others as being related to speech, for instance the written representations of sounds ('Peep!'), lengthening and capitalisation of words to showcase emotionality ('GOALLLLLLLLLLL!'), using slang, shortenings and expletives (Chovanec 2009: 117). In LTC, features that connect the written text with the oral medium create a sense of immediacy which is vital to the genre since its format relies on the notion of relaying events happening in real time (Chovanec 2009: 125). From competitive interactions, conceptual metaphors to spoken language imitation, live text commentaries encapsulate many fascinating language features worth analysing by linguists.

Football commentary is known to be filled with formulaic expressions (Chovanec 2020: 1) and it is no surprise that these have been analysed in the live text commentary genre. Meier-Vieracker has compared German and English LTCs and match reports in terms of their formulaicity (2019). He found that in both languages and in both types of discourse schematic patterns of expression can be found, templates describing actions on the pitch are filled with synonymous verbs and adjectives thus appearing unique even though their structure is clearly formulaic (2019: 15). An

example of a formulaic template is a combination of a description of an assist, an effort and an interception frequently occurring together but in various different words (Meier-Vieracker 2019: 16). Levin checked formulaic constructions in football match reports and LTCs with regards to the words *net*, *minute(s)* and *whistle* (2008: 143). He concluded that football language is ‘built up from semi-fixed chunks of language’ (2008: 153). According to Levin, such phrases which include *[Verb] the ball into his own net, in the nth minute, the final whistle* are so closely associated with the football reporting register that they can be considered as its markers (*ibid.*). Findings from published research suggest that football discourse and, more specifically, live text commentaries encompass features of formulaic language.

Formulaic nature of football discourse facilitates the application of automatised text production. Nowadays, ‘robot-journalism’ is starting to gain traction in news agencies and most often algorithms are used to write sports related content due to it being a very straightforward, data driven medium (Fanta 2017: 10). Meier-Vieracker, who has outlined the evolution of live text commentaries in his study, has stated that messages describing football actions on the pitch ‘could be algorithmically generated’ (2021: 296). According to Meier-Vieracker this automatised production of LTCs ‘would be the most consistent continuation’ of the evolution of the genre (*ibid.*). Knowing that there are numerous internet sites which are dedicated to the publication of sports information live, it is possible to deduce that the implementation of algorithmic commentary already exists. For instance, sites such as flashscore.com or livescore.in offer live text commentary for all major football competitions around the world. They do not indicate the authors of the commentary and offer it in a myriad of languages. These aspects make it possible to hypothesise that the sites utilise some type of algorithmic language production, even though this is not disclosed directly.

Taking into consideration the formulaic nature of the live text commentary genre and the possibilities of it being written not only by people but also by algorithms, a question naturally arises of whether there are any differences in terms of formulaicity between the two types of texts. The present paper is concerned with answering this exact question. Since there are many approaches to defining formulaicity, the present research relies on the lexical bundle approach proposed by Biber et al. as the definition of formulaic language (1999). Lexical bundles here are understood as a type of formulaic sequence falling under the broad definition by Wray (2002). The term *lexical bundles*, first described by Biber et al. is defined as ‘<...> sequences of word forms that commonly go

together in natural discourse' (1999: 991). The first studies on these recurrent sequences conducted by Biber et al. were focused on the distribution of lexical bundles across different registers of English, namely, conversation and academic prose. However, as the papers which followed this seminal work have established, lexical bundles prevail in many other discourses as well, for instance, employee work emails (Wright 2014), spoken business language (Allan 2015), diverse legal genres (Breeze 2013), Wikipedia articles (Hiltunen 2018), press releases (Granger and De Cock 2021) etc. Knowing that people use 'building blocks' and prefabricated sequences to produce coherent language (Hyland 2008: 44, Wray and Perkins 2000: 1), it is logical to suppose that the football commentary genre will also contain recurrent word sequences, termed in linguistic literature as lexical bundles. Different structural and functional types of lexical bundles (henceforth: LBs) have been described by Biber et al. and the widespread and effective use of these classifications among linguists have further proven their validity (2004). Studies which employ the lexical bundle approach include cross linguistic analyses (Kostromitina 2022, Granger 2014), research of different varieties of English, namely native and non-native English (Chen and Baker 2010, Gungor and Uysal 2016, Adel and Erman 2012) and different academic disciplines (Jalali and Moini 2014, Lee and Lee 2018) to name a few. The present paper will make use of Biber et al.'s (2004) methodology in order to analyse the LTC genre and answer the following research questions:

1. How different or similar are human-produced and automated live text commentaries in terms of lexical bundles:
 - 1.1 What are their structural features?
 - 1.2 What are their functional features?
2. Are live text commentaries more similar to spoken or written language?

2. Data and methods

The present section will provide a detailed account of the data processing involved in the study. The study was set up as a corpus-driven analysis of recurrent sequences of words, termed here lexical bundles (Biber et al. 1999). The analysis was carried out in four stages to fulfil the aims of the research.

The first stage involved the compilation of two corpora. Both corpora of LTCs have the same target audience of avid football fans, who might not be able to tune in to the live broadcast but nevertheless want to follow the action. The first corpus representing the language of automated texts was compiled using data taken from the [livescore.in](https://www.livescore.in) site. The site is one of numerous internet web pages which offer real time results of various sports competitions across the world. The site was chosen based on its large coverage of sports events and a high number of live text commentaries. The data taken from the site was the LTCs of football matches of the English Premier League. The most recent full season was chosen, which lasted from August of 2021 until May of 2022. The season had 380 matches which were all accompanied by live text commentaries. All commentaries were retrieved from the site and carefully revised by leaving out the pre-match information or welcome messages and deleting the occasionally inserted pictures. This type of data retrieval produced 380 texts (600,716 words), henceforth referred to as Corpus_{Automated}.

The second corpus representing commentaries written by football reporters was compiled using data from the site [The Guardian](https://www.theguardian.com), a prominent English news site. The site was chosen due to its long-standing tradition of covering football matches. To ensure comparability of the data the same exact football league and season were chosen for the retrieval of texts. Some manual text revision had to be done for this data as well with the deletion of pre-match build-up sections and post-match evaluations, leaving solely the texts written for specific minutes of the match. Pictures and their descriptions as well as posting time stamps were also erased. The resulting corpus, titled Corpus_{Guardian}, was comprised of 173 LTCs (352,522 words) written by 15 different journalists. For both corpora the texts retrieved were saved as txt. files.

The second stage of data processing was the extraction of data relevant to the study. For this purpose, the language processing software Lanksbox (version 6.0, Brezina et al. 2020) and WordSmithTools (Scott 2020) were employed. The two programs were used to generate lists of lexical bundles from the corpora using the n-grams function. The WordSmithTools software was chosen to generate lists which include LBs of different lengths, i.e. 3- to 5-word bundles. Yet this program failed to process Corpus_{Automated} successfully as the corpus contained a number of very long recurrent sequences, therefore, to process data from Corpus_{Automated} Lanksbox software was used. The automatically generated frequency lists of LBs were then exported to Excel for further analysis.

During the third stage, a careful manual revision of lists of lexical bundles in Excel spreadsheets was undertaken. This task proved to be challenging since the two extracted lists differed greatly. The alterations made to the lists were also quite different. The list of LBs for Corpus_{Guardian} included mostly 3-word bundles with very few 4- and 5-word bundles. Since certain bundles overlapped with each other, some had to be deleted or merged to obtain a better sample for analysis, for instance the 4-word LB *down the left and* (141 repetitions) was included while the 3-word *the left and* (119 repetitions) which is a less informative, overlapping bundle was deleted. Frequency and dispersion data of each LB was considered in order to establish which of the items should be kept in the final study sample. A longer LB was preferred to the shorter one as its analysis proved to be more meaningful. Corpus_{Automated} yielded a list which contained very long (9-word) LBs. This list was also checked and manually revised to merge certain items into longer ones or split into shorter ones based on their frequency and dispersion data. Additionally, in Corpus_{Automated} most frequent lexical bundles of 4- and 5-words were checked manually so as not to leave out shorter bundles which were not part of longer sequences. After the pilot analysis, it was decided for the purposes of this study to choose the first 200 bundles from each corpus for the study. The normalised frequency cut-off point, applied in both corpora is 5 occurrences per 100,000 words. This corresponds to 50 occurrences per million words and is a rather stringent cut-off point in comparison to other studies, for instance Biber et al. chose to analyse lexical bundles which occurred 40 times per million words (2004: 376). Even though the two corpora of the present study do not have comparable scope the samples included in both have high currency in their respective corpora.

For the fourth stage of data processing each lexical bundle was categorised in terms of its structure and function. For this purpose, Biber et al.'s (2004) methodology and grammatical categories provided in *Longman Student Grammar* (Biber et al. 2002: 248) were employed.

Structural categories presented in Biber et al. (2004), although fit Corpus_{Guardian} could not be efficiently used for Corpus_{Automated}. LBs retrieved from Corpus_{Automated} included not merely fragments of phrases but rather full independent clauses, so the structural categories of different types of independent clauses were used (Biber et al. 2002: 248) for their analysis. The types are Simple, Complex or Coordinated independent clauses, for instance (1), (2) and (3) respectively:

(1) *the ball is easily cleared out*

(2) *the defence is alert as they clear a dangerous cross from*

(3) *the ball is off the pitch and it's a goal kick*

LBs classified as Simple clauses (1) are those that have no fragments of dependent or coordinating clauses. LBs categorised as Complex clauses (2), consist of more than one subject and predicate, and fragments of dependent clauses, Coordinated clauses (3) have a coordinating conjunction.

The categories of functions provided by Biber et al. (2004) had to be interpreted more broadly to accommodate data in the study sample. Functional categories for Reference that fit the purposes of this paper were Identification, Place and Specification of Attributes. Identification was understood as any phrase identifying the actions on the pitch or the factual information about the game, for example (4), see below. LBs tagged for Place had to denote the positions of players, balls, shots, etc. as in (5). Finally, the Specification of Attributes dealt with the objective denotation of a type of shot, ball, action, etc. that took place, as in (6):

(4) *picks up a knock and the physio has to come on* (CorpusAutomated)

(5) *on the touchline* (CorpusGuardian)

(6) *a low shot* (CorpusGuardian)

For CorpusAutomated in which the LBs were long and more often than not included multiple categories, the category by which such LB was classified was decided based on its role in the sequence. If two categories had equal weight in the phrase they were classed as multifunctional, for example Place+Identification :

(7) *the ball is off the pitch and it's a goal kick* (CorpusAutomated)

The Functional category of LBs denoting Stance was also identified in the study corpora. Stance was treated as any indication of a subjective, opinionated remark, for example:

(8) *is comfortably intercepted by one of the defending players* (CorpusAutomated)

Remarks such as *in attempt to*, *attempts to*, *tries to* were also treated as subjective as it is a speculation of the intentions of players. If any long sequence had a word or a phrase indicating

subjectivity it was marked for Stance, because even if indications of different types of references were present, the most salient part of the sequence was the expression of opinion.

Stance and Reference (Specification of Attributes) differed in this subjective or objective denotation. Words such as *low*, *long*, *loose* were seen as objective attributes rather than subjective remarks because the words refer to types of situations which do not imply subjective judgments but rather describe objectively a common strategy or ball position.

Following the completion of classification of the lexical bundles mentioned above the analysis of the findings was carried out.

3. Results and Discussion

This section will provide the quantitative and qualitative results of the data analysis with the aim of answering the research questions. As mentioned in the data and methods section, all quantitative results have been obtained from the 200 most frequent bundles in both corpora and their frequencies have been normalised to 100,000 words. The key findings will be presented in tables and charts in the form of percentages. Although the differing nature of the two corpora made it difficult to compare them quantitatively, the quantitative tendencies established in both corpora still provide important insights into the repetitive patterns of both varieties of football commentaries.

3.1. Length and overlap of lexical bundles in Corpus_{Guardian} and Corpus_{Automated}

Before delving into the structural and functional features of the two language varieties under investigation, it is important to overview some general tendencies in the distribution of lexical bundles found in both. Considering the length of lexical bundles across both corpora, the quantitative comparison is presented in Table 1 below. The lexical bundle length across corpora was determined based on different approaches, as was mentioned in the data and methods section, in Corpus_{Guardian} 3-word bundles were the most frequent, while for Corpus_{Automated} the full length of bundles was prioritised over frequency.

Table 1. Distribution of lexical bundles based on length

Type of bundle	Corpus _{Guardian}	Corpus _{Automated}
3-word	79,5 %	6 %
4-word	16 %	2,5 %
5-word	2,5 %	4,5 %
6-word	2 %	6 %
7-25-word	-	81 %

Table 1 reflects the disparity between the corpora, showing how inherently different are chunks of repetitive language in both varieties in terms of length.

Corpus_{Guardian} mainly consists of 3-word bundles, with a few 4-6-word bundles also appearing frequently enough to be part of the sample. Although many previous studies have chosen to focus solely on 4-word lexical bundles by justifying that they are the most meaningful (Juknevičienė 2017: 53), the present paper found that such bundles did not account for a large proportion of repetitive LBs in the corpus. It was apparent that 3-word bundles which were not parts of longer ones were the most pervasive type of bundles. This in itself provides important information on the writing of football journalists. It seems that journalists do not overuse the same pre-fabricated sequences and only employ very short phrases repetitively. The frequencies of the bundles in Corpus_{Guardian} ranged from 115 repetitions of the same bundle to 9. The short lexical bundles of Corpus_{Guardian} contrast greatly with Corpus_{Automated}.

Corpus_{Automated} produced much longer lexical bundles. Since it became apparent early in the bundle extraction process that short 3- or 4-word bundles were more often than not a part of longer sequences, it was important to reconstruct the complete bundles. 9-word lexical bundles seemed to present the clearest picture of bundle length. After some piloting and trials, it was found that the length of LBs varied from 3 to 25 words. The most frequent bundles were 9- (15%), 10- (12%) 11- (8,5%) and 13-word (6,5%). Although bundles of other lengths were not as common in Corpus_{Automated}, they nonetheless had high frequencies of usage. In comparison to Corpus_{Guardian}, 3-6 word bundles had much lower frequencies in Corpus_{Automated} (see Table 1). The frequencies of the bundles found in Corpus_{Automated} range from 64 instances of the same bundle to 5. The finding that Corpus_{Automated} not only has bundles which are of staggering length but that they are repeated a considerable number of times, can be highlighted as a significant feature of the LTC register. The

length of repetitive lexis in Corpus_{Automated} makes this variety markedly different from natural human-produced language.

Other studies pertaining to formulaic language have not indicated any findings of lexical bundles of the length found in this paper. Biber et al. while analysing the registers of conversation and academic prose went up to only 6-word bundles, explaining that these were very uncommon and in order to include them he needed to significantly lower the cut-off point of the study (1999: 992). A paper analysing legal texts found that genres of legislation and documents consisted of a significant number of bundles up to 8-word length (Breeze 2013: 233). Additionally, the sheer number of studies which focus solely on 4-word lexical bundle analysis might also suggest that longer bundles did not appear during the process of research. Therefore, the bundle length found in Corpus_{Automated} of this paper, contrasts greatly with the findings from human-written texts. This shows that language produced with the help of automatisisation utilises lengthy repetitive phrases which is simply not at the disposal of human writers.

Further comparison of the two corpora revealed that a significant proportion of Corpus_{Guardian} bundles (38%) can be found in Corpus_{Automated}. This suggests that the language employed in the two corpora has considerable overlap. Additionally, it was also found that from the other 62% of Corpus_{Guardian} bundles 30% were also identified in the language of Corpus_{Automated} but not frequently enough to be included in the study sample, i.e., they appeared in the list below the cut-off point. However, there were 32% of Corpus_{Guardian} bundles which did not occur in Corpus_{Automated} at all. Thus, even though there is a significant overlap of the language across corpora, over one third of lexical bundles used by journalists are completely non-existent in Corpus_{Automated}. This finding reveals that even within the same type of discourse, covering identical matches, there is a certain portion of journalist language that is distinctly different from language produced with the help of automatisisation.

For further and more in-depth comparisons of the repetitive language across corpora, a structural and functional analysis of lexical bundles was carried out.

3.2. Structural features of lexical bundles in Corpus_{Guardian} and Corpus_{Automated}

The structural analysis presented below draws on the aforementioned classifications by Biber et al. (2002: 248 and 2004: 381), which were adapted to the present data.

The principal structural difference between Corpus_{Automated} and Corpus_{Guardian} was that while the speech of journalists consisted of repetitive phrase fragments, namely prepositional, noun or verb phrase fragments (10), (11) and (12) respectively, discourse composed with automation largely consisted of clausal material such as independent clauses (13):

(10) *out for a goal kick* (Corpus_{Guardian})

(11) *man of the match* (Corpus_{Guardian})

(12) *has a shot* (Corpus_{Guardian})

(13) *the first half of today's match has just finished* (Corpus_{Automated})

Naturally such a difference was expected considering the markedly different length of LBs in both corpora. As can be seen in Figure 1, The short fragments of Corpus_{Guardian} were most often prepositional or fragments of noun phrases. The prepositional fragments were just slightly more common than noun phrases.

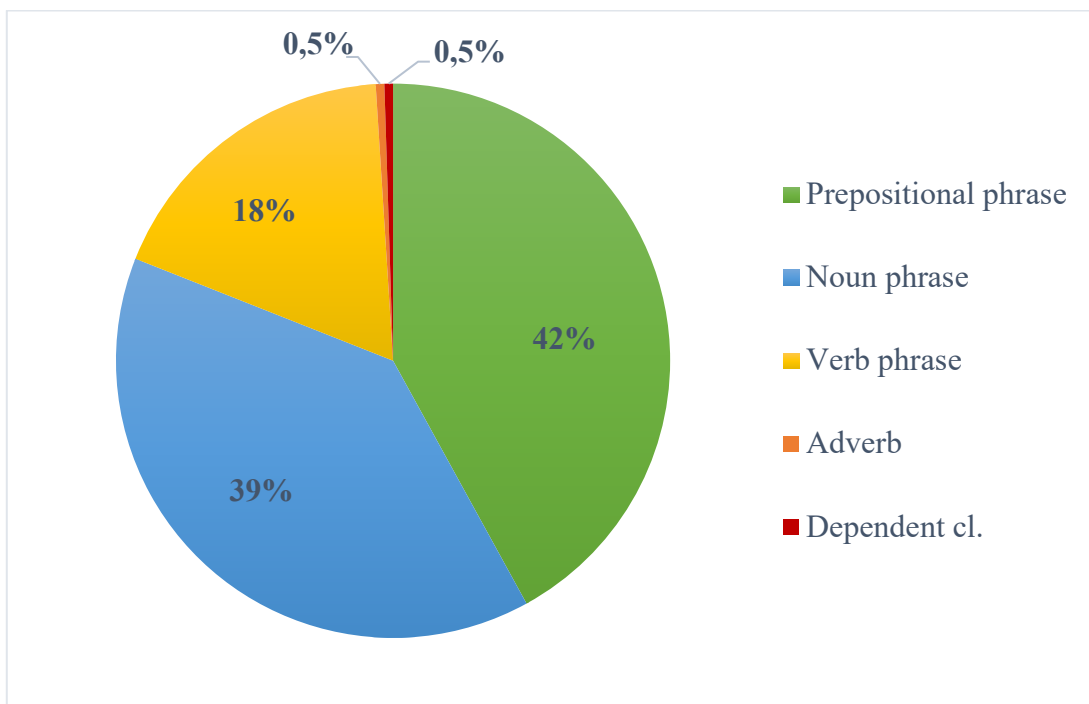


Figure 1. *Distribution of structural types of LBs in Corpus_{Guardian}*

The most common prepositions used were *on, out, in, down, to, from* which shows that these fragments primarily dealt with locations. Prepositional expressions most often indirectly denoted movement or placement of either players or the ball e. g. *from the edge of the box, over the bar* (82%). Many of these expressions were used interchangeably for both players and the ball, for instance *a cross **in from the right**, Redmond cuts **in from the left***. The rest of the prepositional expressions were miscellaneous and could not be put into more specific categories, for instance *for a foul on, under pressure from*. Therefore, it is clear that in order to express actions on the pitch, journalists utilise a myriad of prepositional fragments, specifying the placement of the ball and positions of players.

Noun phrase fragments, which were almost as frequent as prepositional expressions, were classified into three different types. The most common being simple noun phrase expressions (50%), e. g. *the penalty area*, followed by noun phrases with post-modifiers (37%), e. g. *a shot from*, and finally noun phrases with of-phrase fragments (13%), e. g. *the outside of*. The majority of simple noun phrases had the definitive article *the* e. g. *the centre circle, the free kick, the home fans*. As illustrated from the latter examples, a large portion of these noun phrase expressions were locations on the pitch, others dealt with common actions and general concepts surrounding football. Nouns with post-modifiers were most often modified by prepositional phrases which denote placement e. g. *the ball into the, pass down the*. Half of the small number of nouns with of-phrases also was found to have been related to locations e. g. *side of the, edge of the d, corner of the*. Thus, lexical bundles composed of noun phrase fragments most often denote a certain location on the pitch.

Considering that the nature of football discourse is inherently action-oriented, there was a relatively small number of verb phrase fragments in Corpus_{Guardian} (see Figure 1). Out of the verb phrase fragments, three types of actions emerged. The most frequent type was actions which were either directly or indirectly caused by a player or team e. g. *he tries to, is replaced by* (53%). The rest of the actions were almost equally divided between ones caused by the ball e. g. *the ball breaks to, shot is blocked* (25%) and general or non-specific actions e. g. *the flag goes up, there will be* (22%). Thus, it seems that repetitive verb phrase expressions, however scarce, are centred on the actions of players. However, since lexical bundles of this type were not common, the writing of journalists is possibly more varied when it comes to descriptions of the action.

Lexical bundles in Corpus_{Guardian} are most often used to convey the spatial location of the action through prepositional phrase fragments. Even in the category of noun phrases, a considerable number pertain to spatial parameters. A claim could possibly be made that journalists most often repeat phrases that involve locations, either because the sphere of locations is fixed and inherently does not involve a lot of variability or because LTCs are meant to describe the action as understandably to the reader as possible which requires lots of mentions of where the action is taking place. Most likely the combination of both reasons adds up to the high number of spatial mentions in both prepositional and noun phrase lexical bundles.

Research on football match reports corroborate the finding of a large number of lexical bundles involving spatial features. For instance, Ebeling analysed match reports written by staff journalists of different Premier League teams and the lexical bundle list that he obtained had a large number of bundles concerning a certain location of the pitch, Ebeling concluded that English football match reports ‘rely heavily on where on the pitch the action takes place’ (2019: 45). Therefore, this feature of focusing on location seems to be a common characteristic between football LTCs and match reports written by journalists. Another feature found in the present paper concerning the small number of verb phrase fragments has been noted in the analysis done on match reports produced both by human writers and computers. Verb phrase fragments were found to be less prevalent than noun and prepositional fragments in both types of texts being examined (Juknevičienė and Viluckas 2019: 75). However, the latter paper found noun phrases to be the most pervasive by a large margin, whereas in the present study prepositional and noun phrases appear almost equally often. Perhaps this disparity is due to the different discourses with LTCs possibly being more favourable towards prepositional phrases than match reports are.

The structural material of Corpus_{Automated} was classified into four different types, as can be seen in Figure 2, coordinated, simple, complex, and dependent clauses. Since the lexical bundles were of such considerable length, it is no surprise that the most frequent type of clauses was coordinated.

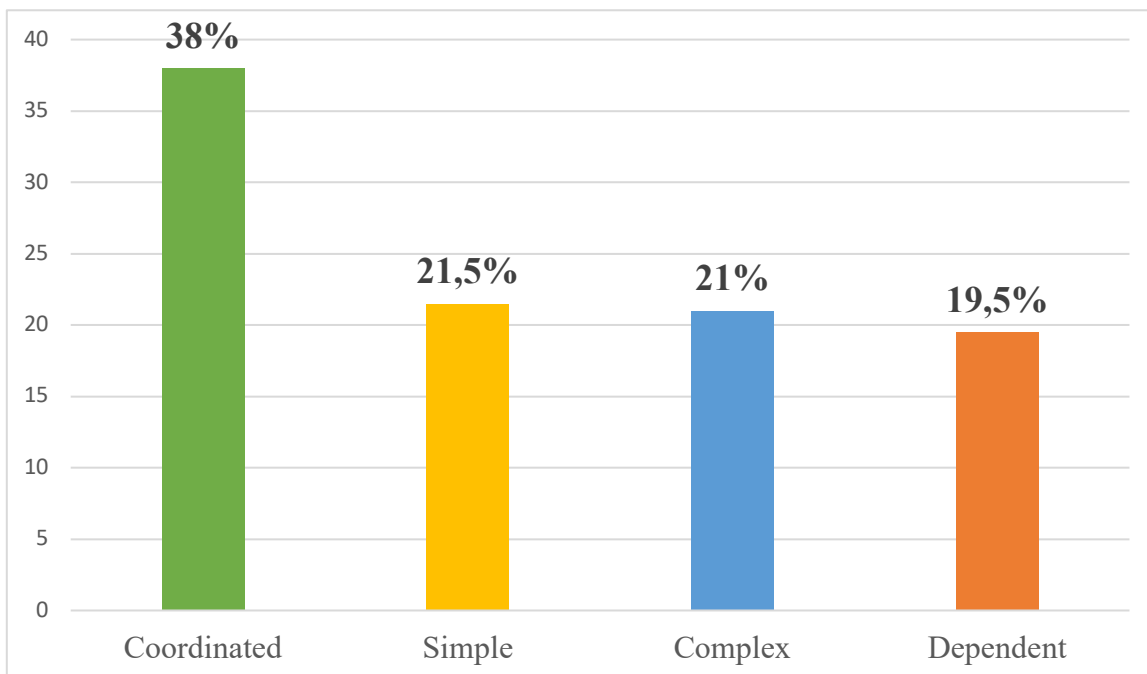


Figure 2. *Distribution of lexical bundle clause types in CorpusAutomated*

Clauses were categorised as coordinated if at any point in the bundle a coordinating conjunction was used. These clauses in themselves incorporate other types and, thus, can also be seen as multi-structural. The fact that there was such a significant number of coordinated clauses in *CorpusAutomated* underlines once more the rigid nature of the language and shows how limited it is in terms of range and variation. Multiple clauses are merged together into longer stretches of text using identical means of cohesion, as in examples below:

(14) *won't tolerate such behaviour on the pitch **and** blows for a foul*

(15) *slaloms his way past challenges **but** an opposing player does well to get the ball away*

Such a characteristic has not been observed in human-written texts, as has been noted in the section analysing the length of lexical bundles.

The coordinated clauses, as illustrated in the examples above, most often contained the conjunction *but* closely followed by *and*. Although journalists may not be able to write exact phrases multiple times, other studies have pointed out that human writers of LTCs utilise fixed syntactic patterns, one of which includes a highly frequent adversative conjunction *but* (Meier-Vieracker 2019: 8).

Meier-Vieracker, who compared English and German language LTCs has concluded that *but* is utilised as a basic way to present a very common scene in football matches which is a failed attempt at a shot, that is either a simple miss or an intervention (*ibid.*). Therefore, the finding of the present paper of the high number of coordinated lexical bundles with *but* could be seen as an exaggerated version of a common writing strategy used by human writers. While journalists employ varying lexis before and after the conjunction *but*, in automated texts examined in the present study, complete repetitive constructions with conjunction *but* are present.

Simple and Complex clauses were used an almost identical number of times. This finding suggests that within LTCs generated with automation there is no preferability between the two structural types. The Simple clauses generally denote an action on the pitch with no elaboration followed, such as (16), while Complex clauses incorporate both information on an unfolding action and an explanation or speculation of intent, as in (17):

(16) *the first half of today's match has just finished*

(17) *wastes a good opportunity as his pass into the box is blocked by the defence*

Coordinated clauses, in comparison, consist of a bit of everything with both simple and complex fragments merged into one lexical bundle by use of a conjunction, as can be seen in (18):

(18) *goes on a solo run, but he fails to create a chance as an opposition player blocks him*

An important structural pattern that emerged from the analysis of the clausal material of Corpus_{Automated} was that a very large proportion (53%) of all types of clauses lacked a subject at the beginning of the phrase. This finding was not surprising, the lexical bundles are repeated numerous times, hence they are used in a variety of matches. The subject slot is filled by a name of a player or team, agents of the action, each time. This finding reveals that lexical bundles are used as templates to describe a common sequence of events that a player or team initiated. For example sentence (19) has various players as subjects, such as *Christiano Ronaldo, Bukayo Saka, Heung-Min Son* and sentence (20) has a variety of teams in the subject position such as *Chelsea, Manchester City, Tottenham*:

(19) *[SUBJECT] goes down in the box after a challenge from the defender*

(20) [SUBJECT] quickly take a corner kick with a short pass

It is important to point out that names of players are always written in full as in the examples above and are followed by their team name in brackets, as in *Christiano Ronaldo (Manchester United)*. For a medium that is highly reliant on timing and time constraints it would seem time consuming to include such long titles for a subject, therefore, it seems possible that players and teams are also part of the building blocks of LTCs of CorpusAutomated.

The remaining 47% of bundles in CorpusAutomated do start with a subject and these subjects are limited to being either a pronoun (*he, it, we, they*), the referee, the ball or a generic denotation (*the players, one of the defenders, the defence, an opposition defender*) without mentioning the name of a player or team.

Out of the lexical bundles without a subject, there were a few (8%) which also lacked a verb at the beginning of the phrase. These lexical bundles, although few, show that the variability is possible not only in the subject but also in the verb slot. Although CorpusAutomated does display a lot of rigidity, there are a few instances where there seems to be a variation of verbs used for certain bundles as in example (21).

(21) [SUBJECT] [VERB] the corner but fails to find any of his teammates

A closer look into the corpus data reveals that the latter bundle has a varying player as the subject and two varying verb options, namely, *takes* and *swings in*. Other lexical bundles without subjects and verbs display a similar tendency of having from one to three options of different verbs attached to the same clause. Although this is a sign of a certain degree of variability, the limited number of verb options only reinforces the stiff nature of CorpusAutomated, where even the choice words come from small, predetermined sets.

Overall, the part of the structural analysis presented above contrasts greatly from the structural characteristics found in CorpusGuardian. The long-winding clauses found in the lexical bundles of CorpusAutomated present a type of writing that is completely unfathomable for human writers and that seems to reflect the limitless capacity of computer software. There are no traces of such lexical

bundles in Corpus_{Guardian}, however, Corpus_{Automated} does have the features of shorter and fragmented lexical bundles of Corpus_{Guardian}.

Fragments of dependent clauses, similar to those in Corpus_{Guardian}, were present in Corpus_{Automated}, however, they did not account for a very considerable proportion, even though in Figure 2. they appear to be as frequent as Simple and Complex clauses. Knowing that Coordinated clauses are mostly composed of the Simple and Complex types, dependent clause distribution is, in fact, much lower than the other types. Considering the sub-categories of dependent clauses found in Corpus_{Automated}, prepositional fragments are the most common (see Table 2). They are followed by noun phrase fragments with verb phrases being the least frequent.

Table 2. Distribution of dependent clause lexical bundles from Corpus_{Automated}

Structural type	Percentage
Prepositional phrase fragment	54 %
Noun phrase fragment	36 %
Verb phrase fragment	7 %
To-clause fragment	3 %

This pattern is identical to the one found in Corpus_{Guardian} just with slightly different percentages. It is possible to suggest that the discourse of LTCs garners more prepositional and noun phrase fragments than it does verb phrases. This finding possibly points out that verb-based sequences in LTCs are more varied and do not become fixed as easily as noun/preposition-based expressions do.

Another important structural finding pertains to the tenses of the lexical bundles. Analysis into the tenses of lexical bundles in Corpus_{Automated} revealed that 74,5% of all bundles were of the present tense. Most of the lexical bundles retrieved from Corpus_{Guardian} were not marked morphologically for tense (81,5%) because only a very small portion of them were verb-based fragments. However, looking solely at verb-based bundles 86% contained a verb in the Present tense. It is not possible to make any generalisations regarding tense for the Corpus_{Guardian} because the sample is too small. However, from the findings of Corpus_{Automated}, it seems that the LTC discourse, at least one written with the help of automation, is abundant with Present tense expressions. This would fall in line with other studies which have deemed Present tense as a characteristic feature of LTCs. For

instance, Lewandowski has found that LTCs (which he calls OSCs) utilise Present tense both very frequently and even more often than spoken TV commentary (2012: 70). Another researcher has found Present tenses to be the most prevalent in LTCs with regards to descriptions of major incidents (Vandenhoeck 2015: 57). Thus, both in this study and in previous analyses Present tense plays a major role. One explanation could be that the goal of LTC is to transmit information which is happening in real-time and even though there is a certain delay, the action is still written out as happening at the moment of writing. This choice of tense helps immerse the reader into the action and the spectator can feel as though they are in the middle of the action rather than catching up.

3.3. Functional features of lexical bundles in Corpus_{Guardian} and Corpus_{Automated}

Below is the account of the functional analysis of lexical bundles following the methodology described by Biber et al. (2004: 381). Out of the three types of functions proposed by Biber, only two, Stance and Referential were found in the study sample.

The distribution of the two types of functions can be seen in Figure 3 below. As is evident, Corpus_{Automated} shows a more balanced distribution between the two functions, while in Corpus_{Guardian} the referential function is clearly the dominant one, with an extremely small portion of LBs expressing stance. Generally, across both corpora the referential function is favoured which is not surprising considering the nature of LTCs, whose aim is to inform of what is taking place during the match.

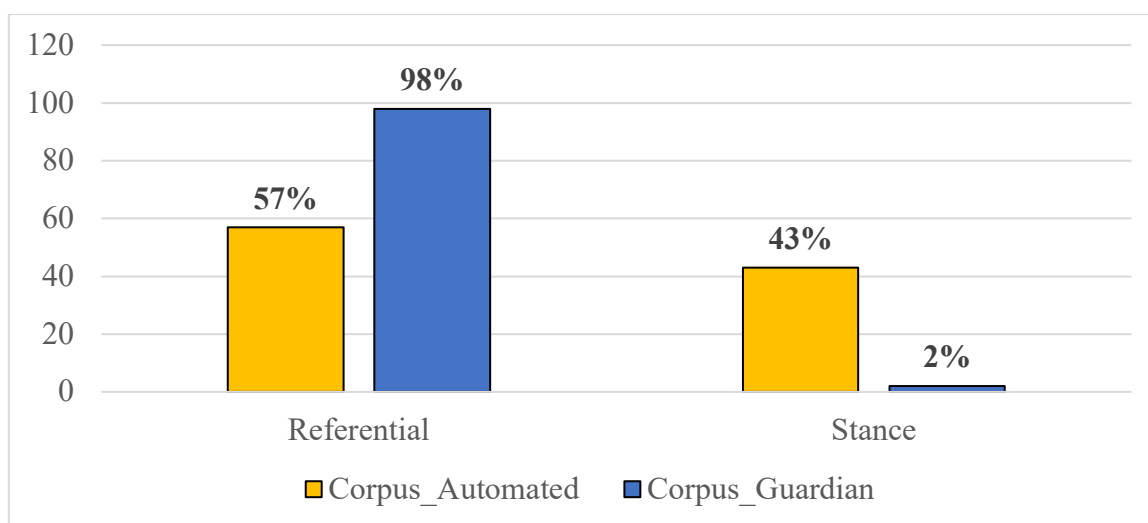


Figure 3. *The distribution of function types across the two corpora*

However, it is quite unexpected to find such disparity regarding the stance function, both in that automated language has such a considerable percentage and that human reporters' texts have such an insignificant share. To further illustrate the disparity a closer look into the two functions was required. First, more in-depth results of the referential function will be presented, followed by the detailed findings in the category of stance.

The referential function was subdivided into categories of identification, place, time, specification of attributes and, additionally, a multifunctional reference emerged, combining both identification and place. The distribution of all sub-categories can be seen in Figure 4 below.

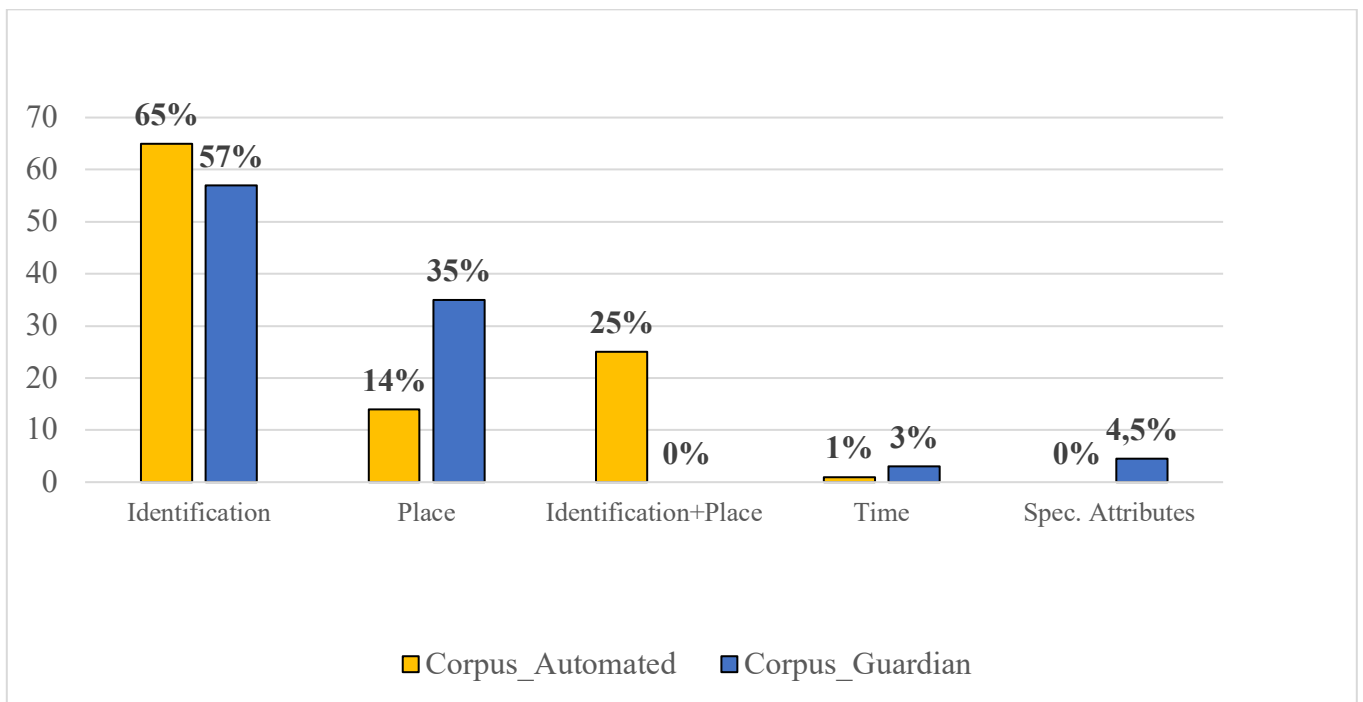


Figure 4. *The distribution of the reference sub-categories across the two corpora*

The referential sub-category of identification was found to be the most common function in both Corpus_{Guardian} and Corpus_{Automated}. Across both corpora, five main types of identification based on meaning were outlined:

- 1) Actions of players/team, e. g. *picks up a knock and the physio has to come on, he tries to*
- 2) Movement of the ball/shot/pass/cross, e. g. *shot towards the, the ball is, ball into the*

- 3) Combination of the player and the ball, e. g. *steps up to take the corner but the ball just sails in among the defenders*
- 4) Actions of the referee, e. g. *the referee signals that it is time for a substitution and*
- 5) General situation, e. g. *the set piece, a free kick is awarded, man of the match*

Within Corpus_{Automated} actions of players were the preferred identification (30%), whereas in Corpus_{Guardian} general situations were identified most often (57%). This possibly stems from the fact that the long lexical bundles of Corpus_{Automated} allow for a whole movement or action to be described, while the short Corpus_{Guardian} sequences can only convey limited details on the surrounding context. However, general situations were often the object of identification in Corpus_{Automated} as well (23%), which possibly points out that much of the language of LTCs references common occurrences that do not pertain to actions or movements.

Another important finding within the sub-category of identification is that while the movement of the ball was the second-most frequent reference in Corpus_{Guardian} (25%), it was the least common in Corpus_{Automated} (8%). It seems that while journalists repeat similar linguistic phrases describing shots of the ball but not players, within automated language it is vice versa. It is possible to suppose that players evoke more varied descriptive language in Corpus_{Guardian} than they do in Corpus_{Automated}. It is also important to note that Corpus_{Automated} had quite a few identification LBs which were a combination of references to the actions of players and movements of the ball (19%), these were very scarce in Corpus_{Guardian} (2%). While it is understandable that the language of the latter corpus is not compatible with combinations of different identifications due to the length of bundles, it is quite surprising that these references were not used more often in Corpus_{Automated}. While the game of football is inherently focused on the player's actions with the ball, having only 19% of identification references be a combination of the two seems quite low.

An unexpected number of references to the actions of the referee was found in Corpus_{Automated} (20%), while it was barely present in Corpus_{Guardian} (1%). Within automated language mentioning the referee was a common occurrence, non-specific phrases of the referee signalling, pointing, whistling as to enact the rules of the game were made of use a myriad of times. In Corpus_{Guardian} referees were rarely mentioned, which might signal that they, as do players, evoke more colourful language or are in general not referred to as much. There were instances in Corpus_{Guardian} in which an action which could have involved a referee was made to be a general comment of events on the

pitch, e. g. *the flag goes up, is booked for, minutes of added time*, excluding the mention of the referee who was the sole instigator of the event. Journalists possibly prefer to shorten the information involving refereeing, focusing more on the enactment of a rule. Considering the time constraints in LTC writing, this would be a more effective way to present information. Since automated language possibly involves the use of templates, there is no need to shorten certain expressions and many clauses contain the full description of actions of referees.

Another reference typical to the language of LTCs is reference to the place of action. The place reference was more prevalent in Corpus_{Guardian} (see Figure 4). That is partly because many place references of Corpus_{Automated} fell into the multifunctional category of identification+place. Taking into account the place references put into a different category in Corpus_{Automated}, between the two corpora such references are similarly distributed. In both corpora the repetitive place references are used to denote the placement of players, e. g. *down the left and* and the ball, e. g. *into the net*, often the same place reference can be used for both player and ball, e. g. *edge of the D*. Interestingly, the place references in both corpora almost always present the area around or within the box. This shows that the emphasis in LTCs falls on the more exciting action rather than players preparing for attacks and passing in the middle of the pitch.

As for Corpus_{Automated}, the multifunctional reference of identification+place most often manifests itself as a template to refer to a player kicking the ball in various ways into/from different parts of the pitch, for instance:

(22) *he launches the ball into the penalty area*

(23) *the ball is whipped in from the corner by*

This type of pattern appears in the majority of lexical bundles categorised as having the multi-function of identification+place. Meier-Vieracker while investigating the formulaicity of human-written LTCs, found ‘text routines’ such as *Assist_Effort_Intervention* and others as evidence of the ‘schematic nature’ of LTCs (2019: 9). The finding of the present study within the identification+place function could be categorised in a similar text routine, as *Effort_Location*. While the templates of journalists emerge from varying language of multiple sentences, the pattern from Corpus_{Automated} reveals itself solely from lexical bundles. The present study confirms the existence of these ready-made templates in LTCs written with the aid of automation.

The two remaining categories within the reference function are time and specification of attributes which were very scarce in both corpora. Time reference appeared more often in Corpus_{Guardian} with lexical bundles such as *at the moment, for a second, just in time*. The general lack of time indications might have to do with the nature of LTCs which present a minute indication before every entry, therefore, the timing of the action is already referred to. Studies on football match reports have pointed out that time references appear extremely frequently (Levin 2008: 150, Ebeling 2019: 45). Levin has indicated that the phrase *'in the nth minute* is probably the most common way of specifying time in football' (2008: 150), while that may be true for match reports, from the findings of the present paper it seems that LTCs have no use for such time indications.

The specification of attributes sub-category was only present in Corpus_{Guardian}, where most of these references specified the type of ball that was being passed, e. g. *a low shot, a long ball, the loose ball*. Although the type of ball would in some cases be indicated in Corpus_{Automated}, it was never the central reference of the lexical bundle.

All in all, the function of reference revealed the numerous ways in which the two corpora differ. The identification references in Corpus_{Automated} mostly denoted actions of players, whereas in Corpus_{Guardian} they dealt with general situations. These findings point to the fact that the writing of journalists is repetitive within already restrictive domains, general situations, movement of the ball, common places are referred to with high frequency. The lexical bundles for such references may be used due to the time constraints. On the other hand, in Corpus_{Automated} most frequent phrases denote actions of players, referees, combinations of two references. This shows that the time constraints pose no issue and the long-winding, elaborate descriptions are ones which are repeated the most.

Considering the function of stance, LBs expressing stance were almost as frequent as those serving the referential function in Corpus_{Automated}, whereas in Corpus_{Guardian} stance was very rare (see Figure 3). While it is not possible to determine generally whether the human-written texts lack expressions of personal opinion, it can be supposed that, at least within repetitive language, more specifically within lexical bundles, stance is almost never relayed. The opposite is true for Corpus_{Automated} which displays an extremely high proportion of opinionated expressions. The latter finding is quite unexpected since language produced with automation is usually viewed as rigid and technical.

Although, another paper has indicated that match reports produced by computers do exhibit a tendency towards very subjective evaluative expressions (Juknevičienė and Viluckas 2019: 79). There seems to be a certain fixedness to the stance expressions of Corpus_{Automated} with a number of evaluative words appearing continually throughout the corpus (see Table 3). Overall, there are 33 different evaluative words for 86 stance lexical bundles.

Table 3. Ten evaluative lexemes which appear in Corpus_{Automated} more than once.

Fail (x9)	Have a chance/opportunity (x2)
Lovely (x4)	Looks to/like (x2)
Promising (x4)	Dangerous (x2)
Comfortably (x3)	Bad (x2)
Far too much/close (x2)	Quality (x2)

These numbers possibly show that while stance expressions are extremely prevalent, they are chosen from quite a small set of possibilities. The evaluative vocabulary of Corpus_{Automated} has a range of adjectives, adverbs and verbs. The most frequently used words indicated in Table 3 are employed in lexical bundles such as:

(24) *fails to pick a body out in the box with an over-hit*

(25) *delivers a lovely cross into the penalty area*

(26) *finds himself in a promising position after receiving a free kick*

Such stance expressions may be employed to attract readers who wish not to simply receive dry factual information but to feel involved and engaged in the action of the match. The fact that similar evaluations are used in the language of automated texts is most likely not apparent when reading one specific LTC.

Almost all stance expressions of Corpus_{Automated} fall into the sub-category of attitudinal stance (Biber et al. 2004), so much so that subdivision into various categories proved to be insignificant. Attitudinal stance as the name suggests denotes the writer's attitude or evaluation. Taking into consideration that the writer as such does not figure in Corpus_{Automated} where texts are written with a certain degree of automation, it is especially interesting that attitude plays such a significant role in these LTCs.

The fact that so many opinionated expressions are part of lexical bundles in CorpusAutomated showcases that even colourful language can be put into a fixed and repetitive phrase. This takes away from originality and uniqueness that each football match possesses. On the other hand, in CorpusGuardian where stance expressions were few, journalists possibly evaluate actions on the pitch in language that does not become a recurrent phrase. Examples of stance expressions that were found in CorpusGuardian include *it looks like*, *does well to*, *would have been*. None are overt statements of opinion.

3.4. General discussion of features of football commentary language

Much of the scientific research involving LTCs has pointed out that the genre stands in between the oral and written mediums of spoken commentary and written match reports. There is a recurrent theme in various studies of LTC to check its leaning to one or other medium (Perez-Sabater et al. 2008, Chovanec 2009, Lewandowski 2012). The present paper set out to add to this debate by analysing whether the lexical bundles found in automated and human-written LTCs display features of spoken or written language.

From a structural point of view, the lexical bundles retrieved from CorpusAutomated do not fit with the 3-4-word bundles that were found to be predominant in conversation and academic prose (Conrad and Biber 2005: 60). Thus, CorpusAutomated does not resemble either discourse structurally. CorpusGuardian, on the other hand, can be compared with the two discourses. The major structural finding of the present study highlights that prepositional and noun phrase lexical bundles are prevailing in CorpusGuardian, while verb phrase fragments are relatively infrequent. Thus, LTCs written by journalists, in terms of structural features, are closer to written language, more specifically, academic prose, which also utilises more noun and prepositional expressions rather than verb phrases (Conrad and Biber 2005: 64). LTCs drift quite far from conversation which typically exhibits verb phrase lexical bundles as prevailing (*ibid.*).

The functional aspects of both CorpusGuardian and CorpusAutomated, show a similar trend. In both corpora the referential function is more prevalent, which is the case in academic prose as well (Conrad and Biber 2005: 66). However, the type of references found in this paper differ greatly

from ones found in academic prose, seeing as LTCs largely showcase identification and academic writing centres on the specification of attributes (*ibid.*). It is also important to point out that Corpus_{Automated} had a considerable number of stance expressions, which is the predominant function of conversation, however, the stance of conversation usually is epistemic while in Corpus_{Automated} it is attitudinal. Thus, while Corpus_{Automated} does have more resemblance to conversation than Corpus_{Guardian}, both types of LTCs lean more towards written language.

Both structural and functional aspects of lexical bundles found in LTCs analysed in the current study show a proclivity towards the written medium. However, this presents a clash with previous findings, which have outlined the many ways in which LTCs incorporate spoken language. For instance, Perez-Sabater et al. have concluded that ‘elements of orality’ are introduced into LTCs of various languages in a number of ways, such as through ‘the commentator’s personality, interaction with readers’, the frequent use of slang and taboo words (2008: 30 ff). Chovanec has found that LTCs incorporate spoken features in ‘phonological, lexical, syntactic, pragmatic’ levels in order to simulate an immediacy of events (2009: 125). Finally, Lewandowski has outlined that from their linguistic features, LTCs are a hybrid of both spoken commentary and written match reports (2012: 74). Therefore, the finding of the present paper that lexical bundles found in LTCs lean towards the written medium does not present an accurate picture of the overall language of this discourse. The previous findings help highlight that the proclivity towards the written medium is only found within the lexical bundles of LTC.

4. Conclusions

Although the language of football has been extensively examined through a variety of approaches, the present paper set out to fill a niche in this research area by comparing the language of human-written and automatised live text commentaries using the lexical bundle approach (Biber et al. 2004). After an in-depth comparison of the lexical bundles extracted from corpora representing the two varieties of language numerous significant findings were uncovered.

The foremost finding of the paper was the contrasting nature of LBs in both corpora. While human-written LTCs consisted of mostly short 3-word bundles, the automated texts were composed of bundles of staggering length, most frequently 9- and 10-word, reaching even 25-word LBs. This finding shows the extent to which human and automated repetitive language differs, highlighting

the limited use of LBs by human reporters and the extensive use within automated production. Thus, it is obvious that the amount of repetitive language is considerably larger in automated language production.

The structural features of LBs in the two varieties of football language hinge on the contrasting length. While the limited LBs of reporters fit into categories of phrase fragments, the lengthy automated LBs require to be categorised into various independent clauses. Corpus_{Guardian} was found to contain mostly prepositional phrase and noun phrase fragments, while Corpus_{Automated} consisted of predominantly coordinated clauses. Based on the length and structure of LBs in both language varieties, automated language was found to be much more rigid.

The functional features revealed that both language varieties favour the referential function, which was largely used to identify actions of players and the referee, the movement of the ball and general situations. The distribution within these categories revealed, that within automated language the most frequent identifications were of the more elaborate categories of the actions of players and the referee, a combination of the player and the ball, whereas within human-written LTCs most frequent identifications referred to general situations. Considering the function of stance, it was observed that Corpus_{Automated} displayed much more stance expressions than Corpus_{Guardian}. While human reporters did not express overt opinions within LBs, the automated language was filled with evaluative expressions. However, it became apparent that the stance LBs of Corpus_{Automated} nonetheless displayed a fixed nature, due to the evaluative words being chosen from a relatively limited selection.

Both structural and functional findings are consistent in showing how greatly automated LTCs differ from those written by reporters. It seems that the notion that LTCs generally display economised language due to the time constraints does not exist within automated texts and is a feature of human writing. While reporters seem to utilise LBs to efficiently denote general situations, movements of the ball, common placement of various actions through prepositional and noun phrase fragments, automated language conveys full sequences of events on the pitch, together with evaluations of them, most often using coordinated clauses.

Another important finding of the study involves the answer of whether automated and human-written LTCs are more similar to spoken or written language. The present research concludes that

considering solely the lexical bundles found within both varieties of LTCs, the proclivity is towards the written medium. However, this should not be seen as representative of the whole medium of LTC, as it has been established by other researchers that this discourse contains many spoken language features.

The results of the current study are subject to certain limitations. Due to the limited study sample, both in terms of corpora size and the number of lexical bundles analysed, the generalisations of the findings might not be fully representative. Considering the corpora, both Corpus_{Automated} and Corpus_{Guardian} rely on only one source each, livescore.in and guardian.com sites respectively, additionally, Corpus_{Automated} relies on texts taken from a site which does not directly state to what extent the LTCs are automated. To obtain more definitive results corpora of the two varieties of language, human-written and automated, could be compiled from numerous different sources. This would account for a wider variety of human written language and different approaches to automated production. For automated texts, sites which directly state the nature of LTC production should be used for further study. Moreover, certain limitations should also be noted involving the lexical bundle analysis. Due to the nature of the lexical bundles produced by Corpus_{Automated}, the process of assigning the structural and functional categories involved a lot of subjective interpretation. Even though the categorisation was checked multiple times and discussed with the supervisor of the study, for more definitive tagging multiple researchers should be involved in the process.

LTC is an ever-evolving genre that is being shaped by the rapid progress of automated language production. Further research will be necessary to keep up with the newest developments and analyse the language used within LTC from various angles. The analysis into formulaic language of LTC specifically is scarce and requires more research. There are numerous other approaches to formulaicity that could be utilised in examining LTC language, for instance, research of collocations of most frequent terms.

Even though the current study has its limitations, there are a few potential applications of the findings. Supposing that the objective of LTCs written with automation is to mimic human-produced language, the findings of the differences between the two discourses could direct towards a less rigid and more varied language production in automated LTCs. Additionally, readers of LTCs

would benefit from knowing to what extent does formulaic language differ in the two varieties. This would help make a more informed decision on which type of texts to choose in the future.

Summary in Lithuanian

Viena populiariausių pasaulyje sporto šakų – futbolas – sulaukia ne tik sporto entuziastų, bet ir lingvistų dėmesio. Futbolo kalba yra dažna lingvistinių darbų tema, skirtų įvairiems šio diskurso aspektams išnagrinėti. Rungtynių santraukos, žaidėjų interviu, žiūrovų skanduočių – tai futbolo diskursas, kuris buvo tirtas lyginant skirtingas kalbas, interviu klausimų dizainą, vartojamas metaforas ir kt.

Futbolo diskurso atmaina, nagrinėjama šiame tyrime, tai kartu su interneto atsiradimu išpopuliarėję realiu laiku publikuojami rašytiniai komentarai (ang. live text commentary). Tokie komentarai yra apibūdinami kaip „sporto varžybų aprašymai, kuriami ir viešinami internete dalimis, realiu laiku vykstant varžyboms“ (Jucker 2010: 58). Meier-Vieracker šią futbolo diskurso atmainą apibūdina keliais esminiais bruožais - komentarus sudaro įrašų eilė publikuojama atvirkštine chronologine tvarka, įrašas viešinamas su laiko žyma ir veiksmo aprašymu, įrašai gali būti sukurti žmogaus, kompiuterio ar derinant abu autorystės tipus, komentarai dažnai turi multimodalinį aspektą, jie gali būti skaitomi tiek gyvai, tiek varžyboms pasibaigus (2021: 277). Šie komentarai publikuojami tiek naujienų puslapiuose, tiek specialiose naujausios sporto informacijos svetainėse. Komentarai skirti futbolo mėgėjams, kurie dėl įvairių priežasčių neturi galimybės žiūrėti rungtynių gyvai, tačiau nori sekti rungtynių eigą realiu laiku.

Šio darbo tikslas yra palyginti dvi komentarų rūšis - komentarus rašomus sporto apžvalgininkų ir komentarus rašomus pasitelkiant automatizuotus teksto generavimo įrankius. Žinant, jog lingvistiniai tyrimai skirti futbolo diskursui, rodo, kad ši kalbos atmaina turi tam tikrą formuliškumą (Chovanec 2020: 1), gretinamoji analizė buvo atlikta susitelkiant į pasikartojančių leksinių žodžių sekų kaip formuliškumo tipo tyrimą, pasitelkiant tekstynų lingvistikoje plačiai taikomą leksinių samplaikų metodą (Biber ir kt. 1999, 2004).

Analizės rezultatai rodo, jog rašytiniai komentarai, kurti sporto apžvalgininkų, leksinių samplaikų atžvilgiu labai skiriasi nuo kurtų pasitelkiant automatizavimą. Dviejų komentavimo rūšių

struktūriniai požymiai yra visiškai skirtingi – „Guardian“ apžvalgininkų kalbą reprezentuojančiame tekстыne nustatytos leksinės samplaikos dažniausiai yra daiktavardiniai ar prielinksniniai frazių fragmentai, tuo tarpu automatizuotuose tekstuose randama daugiausiai sudėtinių sudurtinių struktūros sekų. Funkciniu atžvilgiu, abiejuose komentarų tipuose dažniausiai kartojamos referencinės samplaikos, tačiau referuojama yra į skirtingus varžybų aspektus. Autoriaus poziciją išreiškiančios samplaikos labai dažnos automatizuotus komentarus reprezentuojančiame tekстыne, tačiau itin retos apžvalgininkų komentaruose. Pagrindinė šio tyrimo išvada – dvi futbolo komentarų rūšys, tirtos šiame darbe, yra skirtingos, tiek pagal struktūrinius, tiek pagal funkcinis leksinių samplaikų požymius.

Raktiniai žodžiai: futbolo diskursas, leksinės samplaikos, sporto apžvalgininkų kalba, automatizuoto teksto generavimo kalba, realiu laiku publikuojami komentarai.

References

Software:

Brezina, V., Weill-Tessier, P., and McEnery, A. 2020. #LancsBox v. 5.x. [software]. Available at: <http://corpora.lancs.ac.uk/lancsbox>.

Scott, M., 2020, WordSmith Tools version 8, [software] Stroud: Lexical Analysis Software

Primary sources:

Guardian News & Media Limited. 2023. Available at:

<https://www.theguardian.com/football/live/2021/nov/20/liverpool-v-arsenal-premier-league-live>

Livescore.in. 2023. Available at: <https://www.livescore.in/football/england/premier-league-2021-2022/#/6kJqdMr2/table/overall>

Secondary sources:

- Adel, A. and B. Erman. 2012. Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for Specific Purposes*, 31/2: 81-92. Accessed on May 19th, 2023. Available at: https://www.researchgate.net/publication/251600597_Recurrent_word_combinations_in_academic_writing_by_native_and_non-native_speakers_of_English_A_lexical_bundles_approach
- Allan, R. 2015. Lexical Bundles in ELF Business Meetings. *The Linguistics Journal Volume*, 10/1: 141-163. Accessed on May 19th, 2023. Available at: https://www.researchgate.net/publication/308918045_Lexical_Bundles_in_ELF_Business_Meetings
- Bergh, G. 2011. Football is war: A case study of minute-by-minute football commentary. *PPG LINGUISTICA/UFJF*, 2: 83-93. Accessed on May 19th, 2023. Available at: <https://periodicos.ufjf.br/index.php/veredas/article/view/25079>
- Biber, D., S. Conrad, and V. Cortes, 2004. 'If you look at ...': Lexical Bundles in University Teaching and Textbooks. *Applied Linguistics*, 25/3: 371–405. Accessed on May 19th, 2023. Available at: https://jan.ucc.nau.edu/biber/Biber/Biber_Conrad_Cortes_2004.pdf
- Biber, D., S. Johansson, G. Leech, S. Conrad and E. Finegan. 1999. *The Longman Grammar of Spoken and Written English*. Harlow: Longman.
- Biber, D., S. Conrad, and G. Leech, 2002. *Longman Student Grammar of Spoken and Written English*. London: Pearson Education Limited.
- Breeze, R. 2013. Lexical bundles across four legal genres. *International Journal of Corpus Linguistics*, 18/2: 229-253. Accessed on May 19th, 2023. Available at: https://www.researchgate.net/publication/259484470_Lexical_bundles_across_four_legal_genres

- Chen, Y.-H. and P. Baker. Lexical bundles in L1 and L2 academic writing. *Language Learning and Technology*, 14/2: 30–49. Accessed on May 19th, 2023. Available at: <https://scholarspace.manoa.hawaii.edu/server/api/core/bitstreams/f8e51275-cc1c-4161-b20f-78df49b6d909/content>
- Chovanec, J. 2006 Competitive Verbal Interaction in Online Minute-by-Minute Match Reports. *Brno Studies in English*, 32: 23-35. Accessed on May 19th, 2023. Available at: https://digilib.phil.muni.cz/_flysystem/fedora/pdf/104090.pdf
- Chovanec, J. 2009. Simulation of Spoken Interaction in Written Online Media Texts. *Brno Studies in English*, 35/2: 109-128. Accessed on May 19th, 2023. Available at: https://www.researchgate.net/publication/284420631_Simulation_of_spoken_interaction_in_written_online_media_texts
- Chovanec, J. 2020. "The battle fever's on": Phrasal Routines and Multilingual Play in Sports Commentary in Lavric, E. and M. Meulleman (eds.) *Le corps et la langue du footballeur : match interdisciplinaire*. Innsbruck: Innsbruck University Press. Accessed on May 19th, 2023. Available at: https://www.researchgate.net/publication/339128618_The_battle_fever's_on_Phrasal_routines_and_multilingual_play_in_sports_commentary
- Cleland, J. 2013. Racism, Football Fans, and Online Message Boards: How Social Media Has Added a New Dimension to Racist Discourse in English Football. *Journal of Sport & Social Issues*, 38/5: 415-431. Accessed on May 19th, 2023. Available at: https://www.researchgate.net/publication/270601130_Racism_Football_Fans_and_Online_Message_Boards_How_Social_Media_Has_Added_a_New_Dimension_to_Racist_Discourse_in_English_Football
- Conrad, S. and D. Biber. 2005. The Frequency and Use of Lexical Bundles in Conversation and Academic Prose. *Lexicographica*, 20: 56-71. Accessed on May 19th, 2023. Available at: https://pdxscholar.library.pdx.edu/cgi/viewcontent.cgi?article=1002&context=ling_fac

- Ebeling, S. 2019. The language of football match reports in a contrastive perspective, in Callies, M. and M. Levin (eds.) *Corpus Approaches to the Language of Sports*. London: Bloomsbury Academic. 37-63.
- Ebeling, S. 2021. Minutes of action! A contrastive analysis of time expressions in English and Norwegian football match reports, in Čermáková et al. (eds.) *Time in Languages, Languages in Time*. Amsterdam: John Benjamins Publishing Company. 229-254.
- Fanta, A. 2017. Putting Europe's Robots on the Map: Automated Journalism in News Agencies. *Reuters Institute Fellowship Paper*, 24: 1-23. Accessed on May 19th, 2023. Available at: <https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2017-09/Fanta,%20Putting%20Europe's%20Robots%20on%20the%20Map.pdf>
- Gerhardt, C. 2008. Turn-by-turn and move-by-move: A multimodal analysis of live TV football commentary in: Lavric E. et al. (eds.), *The Linguistics of Football*, 283-294. Tübingen: Gunter Narr. Accessed on May 19th, 2023. Available at: [https://publikationen.sulb.uni-saarland.de/bitstream/20.500.11880/27131/1/Gerhardt_2008_Turn by turn for repository.pdf](https://publikationen.sulb.uni-saarland.de/bitstream/20.500.11880/27131/1/Gerhardt_2008_Turn%20by%20turn%20for%20repository.pdf)
- Granger, S. 2014. A lexical bundle approach to comparing languages: Stems in English and French. *Languages in Contrast*, 14/1: 58-72. Accessed on May 19th, 2023. Available at: https://www.researchgate.net/publication/259740773_A_lexical_bundle_approach_to_comparing_languages_Stems_in_English_and_French
- Granger, S. and S. De Cock. 2021. Stance in press releases versus business news: a lexical bundle approach. *Text & Talk: An Interdisciplinary Journal of Language, Discourse & Communication Studies*, 41/5-6: 691-713. Accessed on May 19th, 2023. Available at: https://www.researchgate.net/publication/353835770_Stance_in_press_releases_vs_business_news_a_lexical_bundle_approach

- Gungor F. and H. H. Uysal. 2016. A Comparative Analysis of Lexical Bundles Used by Native and Non-native Scholars. *English Language Teaching*, 9/6: 176-188. Accessed on May 19th, 2023. Available at: <https://files.eric.ed.gov/fulltext/EJ1101230.pdf>
- Hiltunen, T. 2018. Lexical bundles in Wikipedia articles and related texts : Exploring disciplinary variation in Tyrkko J. and J. Kopaczyk (eds.) *Applications of Pattern-driven Methods in Corpus Linguistics. Studies in Corpus Linguistics*. Amsterdam: John Benjamins. 189-212. Available at: https://helda.helsinki.fi/bitstream/handle/10138/325605/HIL3_acceptedmanuscript.pdf?sequence=1
- Huddleston, W. 2022. Kicking off: Violence, honour, identity and masculinity in Argentinian football chants. *International Review for the Sociology of Sport*, 57/1: 34–53. Accessed on May 19th, 2023. Available at: <https://journals.sagepub.com/doi/pdf/10.1177/1012690220988651>
- Hyland, K., 2008. Academic clusters: text patterning in published and postgraduate writing. *International Journal of Applied Linguistics*, 18/1: 41-62. Accessed on May 19th, 2023. Available at: <https://cupdf.com/document/hyland-2008-academic-clusterstext-patterning-in-published-and-postgraduate.html?page=1>
- Hynes, D. and A. Cook. 2013. Online belongings – a study of female fans’ experiences in online football forums in Rowe, D. and B. Hutchins (eds.) *Sport, Technology, Power and Culture in the Network Society*. London: Routledge. Accessed on May 19th, 2023. Available at: https://www.researchgate.net/publication/280574588_Hynes_D_and_Cook_A_2013_Online_belongings_a_study_of_female_fans'_experiences_in_online_football_forums
- Jalali, Z. S. and R. Moini. 2014. Structure of Lexical Bundles in Introduction Section of Medical Research Articles. *Procedia – Social and Behavioral Sciences*, 98: 719-726. Accessed on May 19th, 2023. Available at:

https://www.researchgate.net/publication/275542868_Structure_of_Lexical_Bundles_in_Introduction_Section_of_Medical_Research_Articles

Jucker, A. H. (2010) “Audacious, brilliant!! What a strike!”: Live text commentaries on the Internet as real-time narratives. In Christian R. Hoffmann (ed.). *Narrative Revisited. Telling a Story in the Age of New Media*. (Pragmatics & Beyond New Series 199). Amsterdam/Philadelphia: John Benjamins, 57-77. Accessed on May 19th, 2023. Available at:
https://www.academia.edu/26379777/Live_text_commentaries_on_the_Internet_as_real_time_narratives

Juknevičienė, R. 2017. *English Phraseology and Corpora*. Vilnius: Vilniaus Universiteto leidykla.

Juknevičienė, R. and P. Viluckas. 2019. Lexical Features of Football Reports: Computer- vs. Human-mediated Language, in Callies, M. and M. Levin (eds.) *Corpus Approaches to the Language of Sports*. London: Bloomsbury Academic. 63-83.

Kaur, S. 2019. Theologising the Discourse of Representation through ‘Faith’ and Religion in Football News. *The Southeast Asian Journal of English Language Studies*, 25/3:150-162. Accessed on May 19th, 2023. Available at:
https://www.researchgate.net/publication/336143974_Theologising_the_Discourse_of_Representation_through_'Faith'_and_Religion_in_Football_News

Kostromitina, M. 2022. Cross-linguistic transfer in academic journal writing: Preliminary evidence from lexical bundle analysis in Russian and English. *Research in Corpus Linguistics*, 10/2: 70–112. Accessed on May 19th, 2023. Available at:
<https://ricl.aelinco.es/index.php/ricl/article/view/188>

Lavric E. and S. Wolfgang. 2008. "Zidane, Zidane, What Have You Done?" – Emotions on TV in Six Language, in Lavric E. et al. (eds.), *The Linguistics of Football*. Tübingen: Gunter Narr. 283-294. Accessed on May 19th, 2023. Available at:
<https://www.researchgate.net/profile/Eva->

[Lavric/publication/280009603 Zidane Zidane What Have You Done - Emotions on TV in Six Languages/links/55a3887a08ae69432ac53be8/Zidane-Zidane-What-Have-You-Done-Emotions-on-TV-in-Six-Languages.pdf](https://lavric/publication/280009603_Zidane_Zidane_What_Have_You_Done_-_Emotions_on_TV_in_Six_Languages/links/55a3887a08ae69432ac53be8/Zidane-Zidane-What-Have-You-Done-Emotions-on-TV-in-Six-Languages.pdf)

Lee, J.-Y. and H. J. Lee. 2018. Lexical Bundles in Computer Science Research Articles: A Corpus-Based Study. *International Journal of Contents*, 14/4: 70-75. Accessed on May 19th, 2023. Available at:
<https://koreascience.kr/article/JAKO201810760746690.pdf>

Levin, M. 2008. Hitting The Back of The Net Just Before the Final Whistle?: High Frequency Phrases in Football Reporting, in Lavric E. et al. (eds.) *The Linguistics of Football*. Tübingen: Gunter Narr. 143-163.

Lewandowski, M. 2012. The Language of Online Sports Commentary in a Comparative Perspective. *Lingua Posnaniensis*. 54/1: 65-76. Accessed on May 19th, 2023. Available at:
https://www.researchgate.net/publication/259591385_The_Language_of_Online_Sports_Commentary_in_a_Comparative_Perspective

Meier-Vieracker, S. 2019. Formulaic language and text routines in football live text commentaries and match reports. A cross-and corpus-linguistic approach, in Marcus Callies, Magnus Levin (eds.) *Corpus Approaches to the Language of Sports*. London: Bloomsbury. 13-37.

Meier-Vieracker, S. 2021. The Evolution of Football Live Text Commentaries. A Corpus Linguistic Case Study on Genre Change. *AILA Review*, 3/2: 274-299. Accessed on May 19th, 2023. Available at:
<https://www.ingentaconnect.com/content/jbp/aila/2021/00000034/00000002/art00006?crawler=true&mimetype=application/pdf>

Pérez-Sabater, C., G. Peña-Martinez, E. Turney, B. Montero-Fleta, 2008. A spoken genre gets written: Online football commentaries in English, French, and Spanish. *Written Communication*. 25/2: 235-261. Accessed on May 19th, 2023. Available at:

https://www.researchgate.net/publication/235735501_A_Spoken_Genre_Gets_Written_Online_Football_Commentaries_in_English_French_and_Spanish

- Pifer, N., G. Scremin, B. Pitts and Y. Wang. 2018. Contemporary global football industry: An introduction, in James Zhang, Brenda Pitts (eds.) *The Global Football Industry - Marketing Perspectives*. London: Routledge. 3-35.
- Rhys, C. 2016. Grammar and Epistemic Positioning: When Assessment Rules. *Research on Language and Social Interaction*, 49/3: 183-200.
- Sandrelli, A. 2015. “And maybe you can translate also what I say”: interpreters in football press conferences. *The Interpreter’s Newsletter*, 20: 87-105. Accessed on May 19th, 2023. Available at: <https://www.openstarts.units.it/bitstream/10077/11854/5/Sandrelli.pdf>
- Tamir, I. 2019. ‘I am grateful that god hates the reds’: persistent values and changing trends in Israel football chants. *Sport in Society*, 24/1: 1-13. Accessed on May 19th, 2023. Available at: https://www.researchgate.net/publication/333715757_'I_am_grateful_that_god_hates_the_reds'_persistent_values_and_changing_trends_in_Israel_football_chants
- Trouvain, J. 2011. Between Excitement and Triumph - Love Football Commentaries in Radio vs. TV. Conference: 17th International Congress of Phonetic Sciences. At: Hong Kong. Accessed on May 19th, 2023. Available at: <https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2011/OnlineProceedings/RegularSession/Trouvain/Trouvain.pdf>
- Vandenhoeck, T. 2015. Minute-by-Minute Football Commentary: A Corpus-based Discourse Analysis of Tense and Community within Online MBM. *CELE Journal*, 50-65. Accessed on May 19th, 2023. Available at: <https://core.ac.uk/download/72791585.pdf>
- Wilton, A. 2021. Epistemics in post-match interviews: A focus on questioning turn design. *International Journal of Applied Linguistics*, 31/1: 1-19. Accessed on May 19th, 2023.

Available at: https://www.researchgate.net/publication/348512982_Epistemics_in_post-match_interviews_A_focus_on_questioning_turn_design

Wray, A. 2002. *Formulaic Language and the Lexicon*. Cambridge & New York: Cambridge University Press

Wray, A. and M. Perkins. 2000. The functions of formulaic language: An integrated model. *Language & Communication*, 20/1: 1-28. Accessed on May 19th, 2023. Available at: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.558.6248&rep=rep1&type=pdf>

Wright, D. 2014. *Stylistics versus Statistics: A corpus linguistic approach to combining techniques in forensic authorship analysis using Enron emails*. [PhD thesis] Accessed on May 19th, 2023. Available at: <https://etheses.whiterose.ac.uk/8278/1/David%20Wright%20%28200327943%29%20Final%20Thesis%20-%20Stylistics%20versus%20Statistics%20%284.3.15%29.pdf>

Appendix

CorpusAutomated data sheet:

N-grams	Raw frequency	Normalised fr.	Structure	Function
the ball is off of the pitch and it's a goal kick for	387	64,42312174	coordinated	Referential (identification+place)
side of the goal	334	55,60031696	dependent (3a - nominal)	Referential (place)
failed to take advantage of the corner as the opposition's defence was alert and averted the threat	264	43,94755592	noS+coordinated	Stance
is comfortably intercepted by one of the defending players	233	38,78704746	simple_noS	Stance

win a corner	232	38,62057944	dependent (1d - verb phrase)	Referential (identification)
have a chance to score from a corner kick but the ball is cleared by a defender	226	37,62177135	noS+coordinated	Stance
take the corner, but their hopes of scoring a goal end with a nice clearance by the defence	219	36,45649525	noS+coordinated	Stance
from close range	196	32,62773091	dependent (3d - prepositional)	Referential (place)
shot towards the	194	32,29479488	dependent (3b - nominal)	Referential (identification)
goes on a solo run, but he fails to create a chance as an opposition player blocks him	154	25,63607428	noS+coordinated	Stance
will have an opportunity to threaten the opposition's goal	152	25,30313826	complex_noS	Stance
into the bottom left corner	152	25,30313826	dependent (3d - prepositional)	Referential (place)
the referee stops play so that a substitution can be made	151	25,13667024	complex	Referential (identification)
sends a cross into the box but the opposition's defence	144	23,97139414	noS_coordinated	Referential (identification)
looks to break free but an opposing player clears the ball away	142	23,63845811	noS+coordinated	Stance
races towards goal but the defender gets back well to make a challenge	142	23,63845811	noS+coordinated	Stance
into the bottom right corner	134	22,30671399	dependent (3d - prepositional)	Referential (place)

towards the bottom left corner	122	20,30909781	dependent (3d - prepositional)	Referential (place)
signals for the physio to come onto the pitch	112	18,64441766	complex_noS	Referential (identification+place)
the corner but fails to find any of his teammates	111	18,47794965	noS_noV+coordinated	Stance
gets on the end of a	106	17,64560957	dependent (1d - verb phrase)	Referential (identification)
a first time shot	106	17,64560957	dependent (3c - nominal)	Referential (identification)
slaloms his way past challenges but an opposing player does well to get the ball away	105	17,47914156	noS+coordinated	Stance
into the back of the net	101	16,8132695	dependent (3d - prepositional)	Referential (place)
at the moment	101	16,8132695	dependent (3d - prepositional)	Referential (time)
picked up a minor injury but it looks like he will be able to continue	98	16,31386545	coordinated_noS	Stance
towards the bottom right corner	97	16,14739744	dependent (3d - prepositional)	Referential (place)
picks up a knock and the physio has to come on	94	15,64799339	coordinated_noS	Referential (identification)
swings a cross into the box but it's far too close to	92	15,31505736	noS+coordinated	Stance
sends a teasing cross into the area but	91	15,14858935	noS+coordinated	Stance
the referee signals that it is time for a substitution and	90	14,98212134	complex	Referential (identification)
a goal-scoring opportunity from a corner	90	14,98212134	dependent (3b - nominal)	Stance
the ball is out of play	90	14,98212134	simple	Referential (identification+place)

force a corner	89	14,81565332	dependent (1d - verb phrase)	Referential (identification)
their opponents will face another attacking threat	89	14,81565332	simple	Stance
a lovely ball into the penalty area but the	86	14,31624928	dependent (3b - nominal)	Stance
takes a shot from the edge of the box	86	14,31624928	simple_noS	Referential (identification+place)
is down injured and the referee stops play so that he can receive medical treatment	83	13,81684523	noS+coordinated	Referential (identification)
the ball goes behind for a corner	82	13,65037722	simple	Referential (identification)
fails to send a pass into the box as his effort is cut out	79	13,15097317	complex_noS	Stance
over the bar	76	12,65156913	dependent (3d - prepositional)	Referential (place)
is clearly asking for some medical attention with his painful gestures	74	12,3186331	complex_noS	Stance
has a clear sight of it and blows his whistle	74	12,3186331	noS_coordinated	Stance
the extent of his injury is yet to be discovered	74	12,3186331	simple	Stance
the ball is out of play and the linesman points to the corner flag	73	12,15216508	coordinated	Referential (identification+place)
slides a pass forward but one of the defenders cuts it out	73	12,15216508	coordinated_noS	Referential (identification)
the referee points to the corner flag	72	11,98569707	simple	Referential (identification)

it's a corner to	72	11,98569707	simple	Referential (identification)
a cross from a mid-range free kick taken by	71	11,81922905	dependent (3b - nominal)	Referential (identification+place)
and goes into the book for a previous late challenge	71	11,81922905	starts w/coordinated	Referential (identification)
into the penalty area but the opposition's defence is ready	70	11,65276104	noS_noV+coordinated	Referential (identification+place)
the referee and his assistant both point at the corner flag	70	11,65276104	simple	Referential (identification)
the ball goes out for a corner	70	11,65276104	simple	Referential (identification)
requires medical treatment so the referee stops play and signals for the physio to come onto the pitch	68	11,31982501	noS+coordinated	Referential (identification)
the left post	68	11,31982501	dependent (3c - nominal)	Referential (identification)
to find the head of one of his teammates	68	11,31982501	dependent (infinitive)	Referential (identification)
fails to find any of his teammates inside the box as his pass is blocked	67	11,15335699	complex_noS	Stance
sends a pass into the box but his attempt is thwarted and cleared	66	10,98688898	noS+coordinatedx2	Referential (identification+place)
instead of sending the ball into the penalty area	66	10,98688898	dependent (3d - prepositional)	Referential (identification+place)
they send men into the box	66	10,98688898	simple	Referential (identification+place)

in order to find one of his teammates but	64	10,65395295	dependent (3d - prepositional)	Referential (identification)
the referee and one of his assistants signal for a corner kick to	64	10,65395295	simple	Referential (identification)
hasn't suffered any serious injury which would see him leave the pitch	63	10,48748493	complex_noS	Referential (identification)
a shot that	63	10,48748493	dependent (3b - nominal)	Referential (identification)
he's back in the game now	63	10,48748493	simple	Referential (identification)
tries to slide the ball through to a teammate but it's well blocked by a defender	61	10,1545489	noS+coordinated	Stance
has a clear sight and sees a foul from	61	10,1545489	coordinated_noS	Stance
attempts to find his teammate with a pass from outside the box but the opposition's defence is quick to react and thwarts the attack	60	9,98808089	noS+coordinated	Stance
the referee and one of his assistants both point at the corner flag	60	9,98808089	simple	Referential (identification)
have been awarded a corner kick	60	9,98808089	simple_noS	Referential (identification)
a throw in for	59	9,821612875	dependent (3b - nominal)	Referential (identification)
and the referee blows his whistle for a foul	58	9,65514486	starts w/coordinated	Referential (identification)

should know this behaviour is beyond the rules	57	9,488676846	complex_noS	Stance
goes over to take the corner kick and it and it looks like he will send the ball into the penalty box	57	9,488676846	noS+coordinated	Stance
he was too careless with his challenge	57	9,488676846	simple	Stance
from the edge of the area	56	9,322208831	dependent (3d - prepositional)	Referential (place)
whips a promising cross into the box	56	9,322208831	simple_noS	Stance
produces a lovely ball into the penalty area but the defender manages to intercept and comfortably averts the danger	55	9,155740816	noS+coordinated	Stance
has ended the first half by blowing the whistle	55	9,155740816	simple_noS	Referential (identification)
dinks a cross into the box in an attempt to find one of his teammates	54	8,989272801	complex_noS	Stance
unfortunately for him the opponent's defence deals with his cross comfortably and intercepts it	54	8,989272801	coordinated	Stance
sends a lofted cross into the penalty area	54	8,989272801	simple_noS	Referential (identification+place)
sends a beautiful cross into the penalty area	54	8,989272801	simple_noS	Stance

but the opposition's defence quickly intercepts the ball	54	8,989272801	starts w/coordinated	Stance
but the opposition's defence manages to clear the ball	54	8,989272801	starts w/coordinated	Referential (identification)
the referee blows his whistle to end the first half	53	8,822804786	complex	Referential (identification)
the defence is alert as they clear a dangerous cross from	53	8,822804786	complex	Stance
seems to be alright and is okay to carry on	53	8,822804786	noS+coordinated	Stance
the players are now heading to their respective dressing rooms	53	8,822804786	simple	Referential (identification+place)
was too fierce in his attempt to get the ball from his opponent	52	8,656336771	complex_noS	Stance
another attempt to send the ball beyond the defence by	52	8,656336771	dependent (3b - nominal)	Referential (identification+place)
wastes a good opportunity as his pass into the box is blocked by the defence	51	8,489868757	complex_noS	Stance
ends up in no man's land and the attacking effort comes to an end	51	8,489868757	noS+coordinated	Referential (identification+place)

that is it for the first half after the referee blows his whistle and the players head to their respective dressing rooms	51	8,489868757	coordinated	Referential (identification+place)
wide of the right post	51	8,489868757	dependent (3d - prepositional)	Referential (place)
one of the defenders reacts at the last minute to brilliantly block a pass from	50	8,323400742	complex	Referential (identification)
whips the ball into the penalty area but one of the defenders is alert and spansk it	50	8,323400742	noS+coordinated	Referential (identification)
serves up a nice cross but the defence clears the danger	50	8,323400742	coordinated_noS	Stance
but his shot is blocked	50	8,323400742	starts w/coordinated	Referential (identification)
a precise pass on the edge of the box and unleashes a	49	8,156932727	noS_noV+coordinated	Stance
swings in a cross but the ball is cleared to safety by one of the defending players	49	8,156932727	noS+coordinated	Referential (identification)
robs an opponent of the ball and explodes in anger	49	8,156932727	noS+coordinated	Stance
puts far too much on the pass and the chance is gone	49	8,156932727	noS+coordinated	Stance
in the middle of the	49	8,156932727	dependent (3d - prepositional)	Referential (place)

was too aggressive in the battle for the ball	48	7,990464712	simple_noS	Stance
the opposition's defence clears the ball away to eliminate the danger	47	7,823996697	complex	Referential (identification)
is unsuccessful as the opponent's defence quickly clears the ball out of the penalty area	47	7,823996697	complex_noS	Stance
produces a dangerous lofted cross into the box which is intercepted	47	7,823996697	complex_noS	Stance
delivers a promising cross into the box but it's cleared	47	7,823996697	coordinated_noS	Stance
attempts to find the head of one of his teammates	47	7,823996697	simple_noS	Stance
makes a reckless foul in order to win the ball from his opponent	46	7,657528682	complex_noS	Stance
seem to be finding their feet as they enjoy some possession	46	7,657528682	complex_noS	Stance
attempts to send over a cross in order to find one of his teammates but an opposition defender averts the danger by intercepting the effort	46	7,657528682	coordinated_noS	Stance

decides to strike the ball towards goal but his effort fails to threaten the goalkeeper	46	7,657528682	coordinated_noS	Stance
just outside the box but his attempt is blocked by	46	7,657528682	coordinated_noS_noV	Referential (identification+place)
with a promising cross into the box but a defender outjumps the attackers and averts the threat	46	7,657528682	coordinated_noS_noV	Stance
after being well blocked by one of the defenders	46	7,657528682	dependent (3d - prepositional)	Stance
we are about to find out how serious his injury is	45	7,491060668	complex	Referential (identification)
is obviously not following the golden rule treat others as you want to be treated	45	7,491060668	complex_noS	Stance
the players are back out on the pitch following the half-time break and the second half is about to start	45	7,491060668	coordinated	Referential (identification)
he makes a bad challenge and leaves the referee with no other option than to blow for a foul	45	7,491060668	coordinated	Stance
is having a really tough time right now	45	7,491060668	simple_noS	Stance
take a short corner instead of sending the ball into the penalty area	44	7,324592653	complex_noS	Referential (identification+place)

strike but the shot is blocked by the defender	44	7,324592653	coordinated_noS_noV	Referential (identification)
is writhing in pain and can now receive medical treatment after the referee signals for the physio to come onto the pitch	44	7,324592653	coordinated_noS	Stance
the referee blows his whistle for half-time bringing to a close the first 45 minutes of the match	43	7,158124638	complex	Referential (identification)
the referee blows his whistle to signal the start of the second half	43	7,158124638	complex	Referential (identification)
goes down in the box after a challenge from the defender	43	7,158124638	complex_noS	Stance
possibly unsighted, the referee just waves play on	43	7,158124638	complex_noS_noV	Stance
nevertheless an opposition defender is alert and averts the danger	43	7,158124638	coordinated	Referential (identification)
delivers a fine lofted cross into the penalty area but it's cleared by an opposing player	43	7,158124638	coordinated_noS	Stance
on the ball	43	7,158124638	dependent (3d - prepositional)	Referential (identification)
gives away a foul for a clumsy challenge	43	7,158124638	simple_noS	Stance

delivers a lovely cross into the penalty area	43	7,158124638	simple_noS	Stance
cross into the box from the free kick	43	7,158124638	simple_noS_noV	Referential (identification)
sends a lofted cross into the box which is headed clear	42	6,991656623	complex_noS	Referential (identification)
is struggling after picking up a knock hopefully he can run it off	42	6,991656623	complex_noS	Stance
fails to pick a body out in the box with an over-hit	42	6,991656623	complex_noS	Stance
produces a quality cross into the box but the opposition's defence intercepts the ball and spans it away	42	6,991656623	noS+coordinated	Stance
delivers a long ball but it lacks enough power to reach his teammates and is intercepted by one of the defenders	42	6,991656623	coordinated_noS	Stance
pulls the jersey of one of the opposition players	42	6,991656623	simple_noS	Referential (identification)
floats the ball in from the corner but it's intercepted	40	6,658720593	noS+coordinated	Referential (identification)
is lying on the ground after being tackled and the physio comes onto the pitch to give him medical treatment	40	6,658720593	noS+coordinated	Referential (identification+place)

takes the corner kick and sends the ball among a huddle of players inside the penalty area	40	6,658720593	noS+coordinated	Referential (identification+place)
whips the ball in but it fails to reach any of his teammates	40	6,658720593	noS+coordinated	Stance
won't tolerate such behaviour on the pitch and blows for a foul	40	6,658720593	noS+coordinated	Stance
he doesn't connect as he wanted and it's cleared by the well-organized defence	40	6,658720593	coordinated	Stance
the fourth official shows 4 min of added time	40	6,658720593	simple	Referential (identification)
crosses into the box from near the side line	40	6,658720593	simple_noS	Referential (identification+place)
everything is set for the start of the second half	39	6,492252579	simple	Referential (identification)
brings one of his opponents down with a strong tackle	38	6,325784564	complex_noS	Stance
waiting to open the opposition's defence and hit them swiftly on the break	38	6,325784564	noS+coordinated	Stance
into the net	38	6,325784564	dependent (3d - prepositional)	Referential (place)
the first half of today's match has just finished	38	6,325784564	simple	Referential (identification)
a free kick is awarded	38	6,325784564	simple	Referential (identification)
the ball is whipped in from the corner by	38	6,325784564	simple	Referential (identification+place)

launches a powerful cross from the corner	38	6,325784564	simple_noS	Stance
is asking for medical treatment after picking up an injury	37	6,159316549	complex_noS	Referential (identification)
fails to reach the box as the defence clears the ball away to safety	37	6,159316549	complex_noS	Stance
steps up to take the corner but the ball just sails in among the defenders	37	6,159316549	noS+coordinated	Referential (identification)
is penalised for an offensive foul	37	6,159316549	simple_noS	Referential (identification)
quickly take a corner kick with a short pass	36	5,992848534	complex_noS	Stance
finds himself in a promising position after receiving a free kick	36	5,992848534	complex_noS	Stance
takes the corner but it's intercepted by the defender	36	5,992848534	noS+coordinated	Referential (identification)
takes the corner and sends it into the penalty area but the opposition's defence is ready and makes a good clearance	36	5,992848534	noS+coordinated	Stance
the set piece	36	5,992848534	dependent (3c - nominal)	Referential (identification)
was trying to get to the ball but clattered into the legs of the opponent	35	5,826380519	noS+coordinated	Stance
the half-time break is over and the second half is about to start	35	5,826380519	coordinated	Referential (identification)

ball on the edge of the box	35	5,826380519	dependent (3b - nominal)	Referential (place)
by exchanging a combination of one or two touch passes	35	5,826380519	dependent (3d - prepositional)	Referential (identification)
the ball is easily cleared out of danger	35	5,826380519	simple	Stance
was in the right place at the right time to	35	5,826380519	simple_noS	Stance
hold on there's a var review to see whether there was	34	5,659912504	complex	Referential (identification)
steps up to take the corner	34	5,659912504	complex_noS	Referential (identification)
departs the field of play to be replaced by	34	5,659912504	complex_noS	Referential (identification)
finds the head of an opposing player who clears the danger	34	5,659912504	complex_noS	Referential (identification)
receives a signal from the referee and is now allowed to enter the pitch	34	5,659912504	noS+coordinated	Referential (identification)
sends a long ball in but	34	5,659912504	coordinated_noS	Referential (identification)
commits a foul and is shown a yellow card	34	5,659912504	coordinated_noS	Referential (identification)
a cross from the free kick	34	5,659912504	dependent (3b - nominal)	Referential (identification+place)
in the build up to the goal	34	5,659912504	dependent (3d - prepositional)	Referential (identification)
over the top	34	5,659912504	dependent (3d - prepositional)	Referential (identification)
the fourth official shows 2 min of added time	34	5,659912504	simple	Referential (identification)

he launches the ball into the penalty area	34	5,659912504	simple	Referential (identification+place)
but one of the defenders leaps highest to head the ball away	34	5,659912504	starts w/coordinated	Referential (identification)
is heading towards the penalty spot to take it	33	5,49344449	complex_noS	Referential (identification+place)
makes a bad challenge and the referee blows his whistle	33	5,49344449	noS+coordinated	Stance
looks like he will be the one to try and deliver a quality ball into the penalty box from the corner	33	5,49344449	noS+coordinated	Stance
fails to beat the offside trap and the linesman puts his flag up	33	5,49344449	noS+coordinated	Stance
after the ball breaks to him in the box	33	5,49344449	dependent (3d - prepositional)	Referential (identification)
in front of goal	33	5,49344449	dependent (3d - prepositional)	Referential (place)
there will be a minimum of 3 min of added time	33	5,49344449	simple	Referential (identification)
that's the end of the game for	33	5,49344449	simple	Referential (identification)
but one of the defenders reacts well and rises highest to head the ball away	33	5,49344449	starts w/coordinated	Stance
one of the defenders reacts superbly to get it clear	32	5,326976475	complex	Stance
unleashes a shot which is brilliantly blocked by a defender	32	5,326976475	complex_noS	Stance

puts a dangerous cross into the box but the corner is cleared by the first man	32	5,326976475	noS+coordinated	Stance
pass on the edge of the box and shoots	32	5,326976475	dependent (3b - nominal)	Referential (identification+place)
are showing great team chemistry	32	5,326976475	simple_noS	Stance
but one of the defending players gets there first to head it clear	32	5,326976475	starts w/coordinated	Referential (identification)

Corpus_{Guardian} data sheet:

N-grams	Raw frequency	Normalised fr.	Structure	Function
DOWN THE RIGHT	405	115	3d	referential (place)
DOWN THE LEFT	403	114	3d	referential (place)
ON THE RIGHT	356	101	3d	referential (place)
ON THE LEFT	342	97	3d	referential (place)
A FREE KICK	283	80	3c	referential (identification)
INTO THE BOX	278	79	3d	referential (place)
OVER THE BAR	247	70	3d	referential (place)
IN FROM THE	241	68	3d	referential (place)
THE FAR POST	237	67	3c	referential (identification)
FROM THE RIGHT	235	67	3d	referential (place)
FROM THE LEFT	215	61	3d	referential (place)
THE SECOND HALF	204	58	3c	referential (identification)
A COUPLE OF	182	52	3a	referential (identification)
THE FREE KICK	161	46	3c	referential (identification)
THE SIX YARD BOX	152	43	3c	referential (identification)

DOWN THE MIDDLE	144	41	3d	referential (place)
DOWN THE LEFT AND	141	40	3d	referential (place)
DOWN THE RIGHT AND	139	39	3d	referential (place)
IT WAS A	131	37	1b	referential (identification)
OUT FOR A CORNER	131	37	3d	referential (identification)
IS BOOKED FOR	129	37	1e	referential (identification)
WIDE OF THE	120	34	3d	referential (place)
IN FROM THE RIGHT	120	34	3d	referential (place)
THE BALL AND	117	33	3c	referential (identification)
BALL INTO THE	113	32	3b	referential (identification)
CROSS FROM THE	112	32	3b	referential (identification)
AT THE NEAR POST	112	32	3d	referential (place)
THERE WILL BE	110	31	1b	referential (identification)
YARDS FROM GOAL	109	31	3b	referential (identification)
IN THE MIDDLE	107	30	3d	referential (place)
IN FROM THE LEFT	103	29	3d	referential (place)
COMES ON FOR	100	28	1d	referential (identification)
WINS A CORNER	97	28	1d	referential (identification)
AT THE BACK	97	28	3d	referential (place)
OUT OF PLAY	94	27	3d	referential (place)
A BIT OF	93	26	3a	referential (sp. attributes)
THE BOX AND	93	26	3c	referential (identification)
INTO THE NET	91	26	3d	referential (place)
CORNER OF THE	90	26	3a	referential (identification)
THE BALL TO	90	26	3c	referential (identification)

AT THE FAR POST	89	25	3d	referential (place)
IS REPLACED BY	88	25	1e	referential (identification)
THE BALL AWAY	88	25	3c	referential (identification)
IS BOOKED FOR A	85	24	1e	referential (identification)
THE BALL IN	84	24	3b	referential (identification)
IN FRONT OF	84	24	3d	referential (place)
INTO THE AREA	83	24	3d	referential (place)
DOWN THE INSIDE RIGHT CHANNEL	83	24	3d	referential (place)
OF THE GAME	82	23	3d	referential (identification)
IN THE BOX	82	23	3d	referential (place)
WIN A CORNER	81	23	1d	referential (identification)
THE BALL BACK	81	23	3c	referential (identification)
CROSS INTO THE	80	23	3b	referential (identification)
ON THE BALL	80	23	3d	referential (place)
DOWN THE INSIDE LEFT CHANNEL	80	23	3d	referential (place)
SIDE OF THE	79	22	3a	referential (place)
UP THE OTHER END	78	22	3d	referential (place)
IN THE AREA	78	22	3d	referential (place)
THE PENALTY AREA	75	21	3c	referential (identification)
A LOW CROSS	74	21	3c	referential (sp. attributes)
FROM A TIGHT ANGLE	74	21	3d	referential (place)
NOTHING COMES OF THE	73	21	1b	stance
OF THE BALL	73	21	3d	referential (identification)
THE BALL INTO THE	72	20	3b	referential (identification)
OVER THE TOP	72	20	3d	referential (place)

WOULD HAVE BEEN	71	20	1d	stance (prediction)
THAT WAS A	71	20	3b	referential (identification)
THE SET PIECE	71	20	3c	referential (identification)
THE FIRST HALF	71	20	3c	referential (identification)
OUT FOR A GOAL KICK	71	20	3d	referential (identification)
THIS IS A	70	20	3b	referential (identification)
OUT ON THE	70	20	3d	referential (place)
THE BACK OF	69	20	3a	referential (identification)
ON THE EDGE OF THE BOX	69	20	3d	referential (place)
SPACE DOWN THE	68	19	3b	referential (identification)
A SHOT FROM	68	19	3b	referential (identification)
THE HOME FANS	68	19	3c	referential (identification)
IS BLOCKED BY	67	19	1e	referential (identification)
ON THE EDGE OF THE AREA	67	19	3d	referential (place)
MINUTES OF ADDED TIME	66	19	3a	referential (identification)
IT INTO THE	66	19	3b	referential (place)
SHOT IS BLOCKED	64	18	1b	referential (identification)
IN THE FIRST	64	18	3d	referential (sp. attributes)
OUT OF THE	63	18	3d	referential (place)
FROM 25 YARDS	63	18	3d	referential (place)
FREE KICK IS	62	18	3c	referential (identification)
JUST OUTSIDE THE	62	18	3d	referential (place)
AT THE MOMENT	62	18	3d	referential (time)
CROSSES FROM THE	61	17	1d	referential (identification)

DOES WELL TO	61	17	1d	stance (attitudinal)
THE BACK POST	61	17	3c	referential (identification)
OUT ON THE RIGHT	61	17	3d	referential (place)
JUST TO THE	61	17	3d	referential (place)
THE BALL IS	60	17	1b	referential (identification)
THE RIGHT HAND	60	17	3c	referential (identification)
TO HIS RIGHT	60	17	3d	referential (place)
ON THE TOUCHLINE	60	17	3d	referential (place)
BEHIND FOR A CORNER	60	17	3d	referential (place)
EDGE OF THE D	59	17	3a	referential (place)
BOX FROM THE	59	17	3b	referential (identification)
A LOW SHOT	59	17	3c	referential (sp. attributes)
A LOT OF	58	16	3a	referential (sp. attributes)
THE BALL OFF	58	16	3c	referential (identification)
THE FLAG GOES UP	57	16	1b	referential (identification)
PASS DOWN THE	57	16	3b	referential (identification)
THE BAR FROM	56	16	3b	referential (identification)
THE RESULTING CORNER	55	16	3c	referential (identification)
JUST INSIDE THE	55	16	3d	referential (place)
A FIRST TIME	54	15	3c	referential (identification)
HAS A SHOT	53	15	1d	referential (identification)
MAN OF THE MATCH	53	15	3c	referential (identification)
A SHOT THAT	52	15	3b	referential (identification)

THE PENALTY SPOT	52	15	3c	referential (identification)
FOR A PENALTY	52	15	3d	referential (identification)
FROM SIX YARDS	52	15	3d	referential (place)
FROM THE EDGE OF THE BOX	52	15	3d	referential (place)
SHOT TOWARDS THE	51	14	3b	referential (identification)
INTO THE MIXER	51	14	3d	referential (place)
AND TRIES TO	50	14	1d	referential (identification)
TO THE FAR POST	50	14	3d	referential (place)
WITH THE BALL	49	14	2b	referential (identification)
SHOT INTO THE	49	14	3b	referential (identification)
A LONG PASS	49	14	3c	referential (sp. attributes)
TOWARDS THE TOP RIGHT	49	14	3d	referential (place)
GET THE BALL	48	14	1d	referential (identification)
ONE ON ONE	48	14	3a	referential (identification)
THE PREMIER LEAGUE	48	14	3c	referential (identification)
THE LEFT HAND	48	14	3c	referential (identification)
THE FAR CORNER	48	14	3c	referential (identification)
A YELLOW CARD	48	14	3c	referential (identification)
A LOOSE BALL	48	14	3c	referential (sp. attributes)
TOWARDS THE BOTTOM LEFT	48	14	3d	referential (place)
CROSS IS HEADED	47	13	1b	referential (identification)
IT LOOKS LIKE	47	13	1b	stance (attitudinal)
IS HEADED AWAY	47	13	1d	referential (identification)

SIX YARDS OUT	47	13	3b	referential (identification)
BALL ON THE	47	13	3b	referential (identification)
IN THE LAST	47	13	3d	referential (time)
HIS SHOT IS	46	13	1b	referential (identification)
AND SENDS A	46	13	1d	referential (identification)
THE BALL AROUND	46	13	3b	referential (identification)
A THROW IN	46	13	3c	referential (identification)
TO THE BALL	46	13	3d	referential (identification)
BACK INTO THE	46	13	3d	referential (identification)
TO HIS LEFT	46	13	3d	referential (place)
ACROSS THE FACE OF GOAL	46	13	3d	referential (place)
FOR A SECOND	46	13	3d	referential (time)
HE TRIES TO	45	13	1b	referential (identification)
A CORNER ON THE	45	13	3b	referential (identification)
OUT OF HIS	45	13	3d	referential (identification)
FOR A FOUL ON	45	13	3d	referential (identification)
INTO THE BOTTOM LEFT	45	13	3d	referential (place)
FROM CLOSE RANGE	45	13	3d	referential (place)
FROM THE EDGE OF THE AREA	45	13	3d	referential (place)
THE BALL BREAKS TO	44	12	1b	referential (identification)
GET THE SECOND	44	12	1d	referential (identification)
BALL IN THE	44	12	3b	referential (identification)

HIS LEFT FOOT	44	12	3c	referential (identification)
A LONG BALL	44	12	3c	referential (sp. attributes)
OFF THE BALL	44	12	3d	referential (identification)
TOWARDS THE BOTTOM RIGHT	44	12	3d	referential (place)
TO THE GROUND	44	12	3d	referential (place)
HIGH AND WIDE	44	12	adverbs of manner	referential (identification)
WHO HAS BEEN	43	12	1b	referential (identification)
A CROSS FROM THE	43	12	3b	referential (identification)
START TO THE	43	12	3b	referential (time)
UNDER PRESSURE FROM	43	12	3d	referential (identification)
AND OUT FOR A	43	12	3d	referential (identification)
ON THE BREAK	43	12	3d	referential (identification)
OFF THE LINE	43	12	3d	referential (place)
JUST IN TIME	43	12	3d	referential (time)
CORNER IS HEADED	42	12	1b	referential (identification)
TAKES A TOUCH	42	12	1d	referential (identification)
GOES UP FOR OFFSIDE	42	12	1d	referential (identification)
GOES INTO THE BOOK	42	12	1d	referential (identification)
THE LOOSE BALL	42	12	3c	referential (sp. attributes)
CLEAR DOWN THE	42	12	3d	referential (identification)
THE OUTSIDE OF	41	12	3a	referential (place)
THE LEFT WING	41	12	3c	referential (place)
UP ON THE	41	12	3d	referential (place)
FROM 20 YARDS	41	12	3d	referential (place)

TO THE LEFT OF THE	41	12	3d	referential (place)
FOR THE FIRST TIME	41	12	3d	referential (time)
WIN A FREE KICK	40	11	1d	referential (identification)
THROUGH ON GOAL	40	11	1d	referential (identification)
REACHES THE BYLINE	40	11	1d	referential (identification)
CROSS TO THE	40	11	3b	referential (identification)
A CHANCE FOR	40	11	3b	referential (identification)
BUT THE REFEREE	40	11	3c	referential (identification)
INTO THE BOTTOM RIGHT	40	11	3d	referential (place)
THE CORNER IS	39	11	1b	referential (identification)
THE BALL ON	39	11	3b	referential (identification)
BREAK IN PLAY	39	11	3b	referential (identification)
THE CENTRE CIRCLE	39	11	3c	referential (identification)
OF THE FAR POST	39	11	3d	referential (place)
A CORNER DOWN THE	38	11	3b	referential (identification)
TO THE RIGHT OF THE	33	9	3d	referential (place)
THE RIGHT WING	30	9	3c	referential (identification)
ON THE END OF	30	9	3d	referential (place)
THE HALFWAY LINE	29	8	3c	referential (identification)