Book Review: Corpus Approaches to the Language of Sports (Texts, Media, Modalities)

| Correspondence: | Shuyi Amelia Sun
<shuyi.amelia.sun@uq.net.au> | School of Languages and Cultures, The University of Queensland, Australia |

1. Summary

To date, the world of sports has witnessed fundamental changes with regard to a diversification of sports types and events, commercialization and globalization of major spectator sports, and an ever-increasing public attention and intensive coverage in various media and modalities. The emergence of new sports genres in the age of computer-mediated communication (Herring, 1996) has opened up an innovative way of studying sports discourse, i.e., by means of large online electronic resources, while at the same time, linguistic research has greatly benefitted from corpus-based and corpus-driven investigations of real-world language due to the compilation and accessibility of computer corpora and software tools.

Accordingly, this timely volume *Corpus Approaches to the Language of Sports: Texts, Media, Modalities*, edited by Marcus Callies and Magnus Levin, brings together empirical studies that adopt a usage-based perspective and use corpus data and corpus linguistic methods to examine language occurring in a variety of genres and pragmatic contexts of different types of sports. The editors aim to, on the one hand, extend the scope of applied linguistic research on sports beyond football/soccer, which has been very much at the center of attention, and on the other hand, advance the scope of corpus linguistic research more generally by throwing light on both the potential and the necessity of exploring sports language in association with its accompanying audio-visual modes of communication from a multimodal perspective. Given the above scopes, this timely volume not only bridges the gap by collecting a range of innovative studies representing state-of-the-art applications of corpus linguistic approaches, but also serves as a helpful reference for the researchers who are interested in cutting-edge discursive practices in the field of sports. From this aspect, the volume is expected to be of interest to a broad readership, including those researchers working on (sports) discourse analysis and corpus linguistics, and in the larger fields of applied linguistics and social studies.

Structurally, the volume comprises of an introductory chapter and ten empirical studies, where the introductory chapter (Chapter 1) lays down the theoretical and methodological context required to appreciate the subsequent studies, while the following ten empirical studies (Chapters 2-11) are divided into three parts. Part one consists of 'Texts. Contrastive and comparative aspects of the phraseology of football match reports' (Chapters 2-4) explores the phraseology of football reporting across different text types and languages by adopting a comparative/contrastive linguistic approach. Part two 'Media. Expanding the scope of research to new contexts of use' (Chapters 5-8) extends the existing research to new media and the relatively downplayed sports discourse outside of football. The final part 'Modalities. Multimodal studies' (Chapters 9-11) addresses sports language from a multimodal perspective which has rarely been applied to the language of sports.

2. Review by Section

The volume begins with Chapter 1 'Introduction' (penned by Marcus Callies & Magnus Levin) which first introduces sports-related linguistic research with a view to main research topics and recent trends. It then presents several new research initiatives and useful resources regarding the use of corpora in sports discourse, and finally gives an overview of the remaining chapters in this volume. The opening chapter lays foundation for the following studies and indicates ways they are arranged in the three broader parts (i.e., texts, media, and modalities) of the present volume.

To come to the first part, Chapter 2 'Formulaic language and text routines in football live text commentaries and match reports – A cross and corpus linguistic approach' (Simon Meier) studies authors' strategies to produce online football coverage while meeting the challenge of reconciling linguistic routines and emotional involvement under high time pressure. Meier adopts data-driven corpus linguistic methods to investigate two types of formulaicity, namely recurrent schematic constructions and idioms. A large corpus of over 12 million tokens were built based on German and English data from two online football-related genres, i.e., live text commentaries and match reports. The results suggest that the production of online football coverage oscillates between preconstructed patterns and word-for-word combinations. To be more detailed, syntactic patterns and idioms serve as text routines to present texts in a cross-cultural register-specific way that is tied to the communicative and social needs in the domain of sports coverage, while there are still enough open choices to modify these 'templates' so

as to demonstrate creativity and make narratives more appealing. As is argued by the author, such findings give evidence to what Sinclair (1991) has pointed to as the alternation between 'idiom principle' (word-for-word combinations) and 'open choice principle' (preconstructed multi-word combinations) in the production of texts, indicating that the very interaction of idiomaticity and open-choice is also constitutive for the register of football coverage.

Chapter 3 'The language of football match reports in a contrastive perspective' (Signe Oksefjell Ebeling) explores the English-Norwegian Match Report Corpus (ENMaRC) with the aim of filling the gap that little contrastive research between English and Norwegian has been done on more specific and homogeneous non-fictional text types, such as football match reports in the present study. Relying on ENMaRC, with the Premier League match reports reaching more than 500,000 tokens and the 'Eliteserie' match reports reaching roughly 155,000 tokens, the author applies corpus-driven extraction methods in the forms of word lists, n-gram lists and keyword lists. Results of this study show that post-match reports in the two languages under study are similar to other text types in the use of time and space expressions. Besides, there are cross-linguistic differences when reporting on victories and defeats. Given the pioneering observations based on ENMaRC, this study is impressive in its exploratory nature and avenues for future research. At the same time though, since ENMaRC is still under construction, the uncompleted corpus itself could have prevented the author from offering more in-depth studies of specific linguistic tendencies.

Chapter 4 'Lexical features of football reports: Computer- vs. human-mediated language' (RitaJuknevičienė & PauliusViluckas) presents a cross-mode comparison between human-mediated and computer-mediated football reports in a bid to fill in the niche that lexical (dis)similarities between the two modes remain largely unknown. This study chooses 'Football Manager 2017' reports and BBC online football reports to represent the two modes, and correspondingly two corpora were compiled, each containing 200 texts and spanning around 20,000 tokens. The researchers then use a corpus-driven approach to analyze keywords that distinguish one mode from the other and the relationships regarding lexical bundles. The results reveal a number of prominent differences in terms of both individual lexemes (particularly among function words) and four-word lexical bundles where only 11 out of the 200 most frequent lexical bundles are shared by both corpora. The study also shows a limited use of conjunctions, cohesive and linking devices in computer-generated football reports, which may explain a major cross-mode difference related to textual cohesion. However, findings of this study are not without shortcomings as the analyzed corpora are still relatively small with only 20,000 tokens in each sub-corpus, which may affect the generalizability of the present findings.

Part two begins with Chapter 5 'Such a nice guy who loved racing is bike: Framing in media accounts of fatal crashes involving competitive cyclists' (Turo Hiltunen) which explores framing in news media accounts of cycling crashes. This chapter aims to investigate the structure and functions of such reports, describe the ways different social actors are represented, and investigate what is identified as the cause of the crash and whether the cause is expressed neutrally. To achieve these goals, Hiltunen applies a corpus-based approach to study framing in these reports on the basis of a 79,000-word corpus of 230 English reports collected from the Internet, identifies and discusses the main textual functions and lexico-grammatical patterns from a discourse-analytical perspective. His findings suggest that similar textual strategies are employed for framing crash and representing social actors, although the shortest and most other texts still vary in the specific aspects they stress on. Besides, Hiltunen's findings highlight clear differences in the representation of riders and drivers of motor vehicles involved in the crash, which could be interpreted as evidence for the existence of such media bias against cycling.

Chapter 6 'When did I do dangerous driving then?: Structures and functions of Formula One race radio messages' (Jukka Tyrkkö & Hanna Limatius) studies race radio interactions between drivers and teams (i.e. race engineers) in Formula One, attempting to fill in the gap that no previous linguistic research has touched upon race radio interactions or any other similar context of language use. Based on a newly compiled corpus, which consists of 5432 individual messages (63,183 tokens) from the 2016 and 2017 seasons of Formula One, the authors apply corpus-based quantitative and qualitative methods to investigate the dialogic turns for structure and complexity, and present a breakdown analysis of the stylometric markers of both driver and team broadcasts. Findings suggest that the effects of stress are observed by examining most of the selected linguistic markers though there are significant differences among individual drivers and race engineers during a race. While the sampling method for constructing corpus could be "opportunistic" (p. 117) as is openly acknowledged by the authors, as a pilot study into the language of Formula One during the actual race, the present research takes a preliminary step into this field and provides heuristic implications for future studies.

Chapter 7 'The emotional content of English swearwords in football chat speak: WTF and other pragmatic devices' (Isabel Balteiro) investigates the use of English *f*-expressions (WTF, fucking, and fuck) used by Spanish football followers in spontaneous and synchronic comments while chatting online. The corpus consists of over 390,500 authentic online messages and comments produced by Spanish football followers between 2007 and 2018, manually compiled from the comments sections and messages in chatrooms in the online version of Spanish sports newspaper *Marca*. The actual hits of expletives in the corpus are 144 examples (28 examples of fuck, 16 examples of fucking and 100examples of wtf) by 139 different users. Findings show that the *f*-expressions used by Spanish football followers in online chats have lost their taboo by and large. Instead, they are adopted primarily to organize discursive sequentiality and communicate speakers' attitude and emotions. Besides, the distribution patterns of *f*-expressions indicate that Spanish users imitate the native use of those words, which could be motivated by context and in-group norms. As is claimed by the author, the main interests of this study lie in the functions of these expletives as pragmatic markers and their interactional significance, position, and distribution in football-related discourse. Hence, this study is expected to provide interesting insights into both the language of football followers and cross-linguistic pragmatic role of expletives.

Then Chapter 8 'Fighting for integrity against a corrupting disease: The legal metaphors of sports fraud' (Miguel Ángel Campos-Pardillos) addresses sports-related legal discourse by illustrating the presence of a metaphor in the description of sports fraud (e.g., bribing) and the fight against it. The author manually extracted 203 metaphors and constructed a corpus of 72,809 words associated with sports fraud and fight measures from nine academic studies, including four journal papers and five book chapters, all of which deal with this topic from a legal or law enforcement perspective. Following a qualitative approach, the author analyzes how scholars in the field of law use metaphorical language to justify the fight against sports fraud, and how the metaphorical discourse is influenced by its identification with other criminal activities. His analysis shows that, on the one hand, some metaphors have a clear and objective ontological basis which usually pertains to the world of concrete objects, on the other hand, process and event scenarios are employed to justify measures and actions that seem strict or controversial at first but are eventually accepted as something inevitable within "war against fraud" (p. 176). This chapter concludes with a reminder of the fundamental role of metaphors in creating a discourse, and therefore to increase readers' awareness of the use of metaphors in sports law contexts and in general discourse on sports.

In the final part, Chapter 9 'A multimodal analysis of football live text commentary' (Valentin Werner) explores the multimodal nature of football live text (FLT) and the role of audience participation in the electronic medium. Based on a corpus of 68 FLT reports (around 160,000

tokens in total) from *The Guardian* and *Spiegel Online*, Werner examines the corpus data from a multimodal perspective and provides a detailed analysis of communicative situations, linguistic features and non-textual elements. As "a multimodal artifact" (p. 184) characterized by media convergence, FLT is found to combine both linear and non-linear data, tap into elements from various external sources (e.g., information from a commercial statistics provider and images), and thus enables audience participation. From the methodological perspective, this chapter goes beyond a more traditional textual study towards a multimodal analysis of FLT, which arguably does more justice to this type of communication due to the very multimodal nature of this artefact. Furthermore, the present study leaves room for individual aspects to be elaborated on in future work, for instance, as to how participant roles are realized or as to how emotionality is expressed. Although FLT has become an established form of online reporting, technological advances will continue to shape its general and linguistic form. Therefore, it is likely that future analyses will have to take account of other aspects or elements that are not yet foreseeable at this stage.

Chapter 10 'Fear and disgust' – a corpus study of sentiment towards sporting events as expressed multi modally on 4chan's/sp/board' (Peter Crosthwaite & Joyce Cheung) studies multimodal discourse practices in *4chan*, an anonymous online community where users post images and texts on a wide range of topics. Their corpus consists of 11 full threads (including 35,850 posts and 1,169 images posts) relating to the Ultimate Fighting Championship's (UFC) 2017-2018 New Year's Eve flagship event UFC 219 – Cyborg vs. Holm. The authors apply a sentiment analysis to the multimodal corpus's texts and images, focusing on positive and negative appraisals of action from the sports event as it occurs in real time, as well as 'reaction' images of the poster's personal response to the event or to other user's reactions. Their aim is to characterize the intermodality regarding 4chanposters' juxtaposition of text, images and videos while communicating their reactions to the event and to other posters, which is expected to reveal how the meanings made in one mode are interwoven with the other. Besides, the authors also investigate how hyperlinks serve to direct the sentiment of text and images to other posters on the *4chan* message board. Their findings demonstrate that the general sentiment conveyed in 4chan'sdiscussion is almost entirely negative, with a strong sense of fear and disgust expressed via texts and images. As is mentioned by the authors, the present study represents one of the very few principled multimodal corpus-based analyses of 4chan data by shedding light on the discursive practices within a typically shady corner of the Internet population (as occupied by 4chanusers). It thus contributes to a greater understanding of intermodal patternings as mediated by thousands of anonymous users in (semi-) real-time.

Finally, Chapter 11 'A comparative multimodal corpus study of dislocation structures in live football commentary' (Marcus Callies & Magnus Levin) presents a comparative study of dislocation in live (i.e., play-by-play) TV football commentary, with a focus on the (con-)textual functions of dislocation in association with the live action visible on the screen. The point of departure for this study is the assumption that live TV sports commentary is a specialized register which is characterized by largely unplanned discourse and shaped by the time-critical nature of the action that unfolds on and off the field, but also as to what is visible on TV screen. To test the assumption, a trilingual comparative corpus-based study is conducted based on a corpus of 14,726 words comprising English, German and Swedish transcripts of live TV commentaries of the 2014 men's football FIFA World Cup final between Germany and Argentina. Findings suggest that right dislocation could be considered as a register-specific, functionally motivated discourse feature of live TV sports commentary. In addition, considering there are no major differences in the use of dislocation in sports commentary regarding frequency nor regarding the distributions of the different discourse functions, the authors suggest that future research have to determine to what extent these cross-linguistic similarities hold true in general.

3. General Comments

On the whole, the present volume makes a strong contribution to corpus linguistics and the application of corpus linguistic methods to the discourse and contexts of sports. As the only volume dealing with sports in the *Corpus and Discourse* series, it offers innovative empirical studies that use new corpus resources to show case the structural-linguistic and discourse aspects of a wide range of sports (e.g., football, cycling, motor racing), genres (e.g., live text/commentary, post-match reports, legal texts) and contexts of use (e.g., sports media, in-team communication). Considering the pioneering investigations involved in each chapter, the volume is especially impressive in its exploratory nature and rich implications for future research. In addition, detailed corpus linguistic research methods in each chapter make it easier for both experienced corpus linguists and newcomers to immediately apply these approaches to their research in corpus-based/driven (sports) discourse analysis. For newcomers especially, they will benefit from the thorough literature review of sports discourse research, which is expected to serve as the theoretical foundations while fitting themselves into this field.

In terms of structure, this volume is compiled of ten structural-unified studies. To be more detailed, after the introductory chapter, the remaining ten core chapters all follow the macro-structure of Introduction–Literature Review–Method–Results and/or Discussion–Conclusion (ILM[RD]C) pattern, which is a prominent structural pattern in the empirical research articles published in high-impact applied linguistic journals (Lin & Evans, 2012). With regard to content, these core studies are designed in an interconnected and logical way, though they are independent studies per se. As can be seen obviously, the core feature for each main part sequentially matches the subtitle of this volume.

Among them, Part one is featured by the application of corpus-driven, contrastive methods to the structural-linguistic aspects of football match reports. Although football-related language, especially with a view to specialized terminology and jargon (e.g., Berg & Ohlander, 2012; Liu & Farha, 1996), has been well studied over the past few decades, this line of research is receiving new fuel by the increasing availability of new corpus resources which allow a broader examination of online genres. As mentioned above, the corpus linguistic contrastive methods are particularly suitable for detecting formulaic sequences in phraseological-geared research as they could reveal the lexico-grammatical nature of different varieties. Nevertheless, such research could have but only scratched the surface by setting off from 'decontextualized' angles, such as n-gram frequency lists. As a consequence, given that only some of the most salient features derived from corpora might have been commented on, readers have to beware of the conclusions even though most arguments are based on empirical findings which seem to have been exploited to the full.

Compared with Part one, Part two concentrates on a mono modal (i.e., textual) level but extends the existing research to new media as well as more fruitful sports contexts outside of football. Sports, such as cycling and car racing, have enjoyed an increasing popularity globally due to their association with modern lifestyle, despite the fact that some specific pragmatic contexts of such sports still remain unexplored in the field of linguistics. It also remains largely unknown as for the roleplayed by expletives and metaphors of sports fraud both in sports discursive practices and in the discourse of present-day society. This is why, it is significant for this part to fill the gap as well as inspire readers to be aware of the unexamined and under-examined research topics.

The final part highlights the potential and necessity of examining the language of sports in association with accompanying modes of communication and visualization. From a methodological perspective, adopting the multimodal approach is no doubt the best choice for this part due to the nature of both the corpus data and multimodal analysis. Following Bateman's (2014) framework, multimodality is a way of characterizing communicative situations which rely on combinations of various forms of communication to be effective, and thus a multimodal interpretation can take account of "materiality, language, semiotics and society" (Bateman, Wildfeuer & Hiippala, 2017, p. 26) in both live sports commentary and online anonymous posts. By this means, the final part shows the additional gains of going beyond a single textual dimension towards "complex artefacts" (Bateman et al, 2017, p. 7-9) with a view to throw light on how different modes interact to create synergies in naturally occurring pragmatic situations.

Despite these strong points, the volume is not devoid of certain limitations, a nit-picky note being the bibliographical references which are placed at the end of each chapter itself. The chapter-by-chapter arrangement might do a favor for the reader interested in some article, but for those who are keen to explore a particular topic, it would be even more useful to have a single unified list of references at the end. This is out of the concern that similarities in research foci are shared across different chapters. For instance, for readers interested in football live commentaries, they have to look at least Chapter 2, Chapter 9 and Chapter 11, not to mention the topic of football which is discussed almost throughout the book. Besides, a unified citation list can serve as an up-to-date, comprehensive compilation of relevant research in corpus linguistics and (sports) discourse analysis, which may as well assist undergraduate and graduate students who might have no/limited research experience but would like to discover their potential interests by trying out these fields independently.

In addition, the other issue with the volume is the corpora analyzed. As mentioned, the scale of corpus data in some chapters (e.g., Chapter 4) is relatively limited. Similarly, since the ENMaRC in Chapter 3 is still under construction, the corpus itself could have prevented researchers from offering more in-depth studies of specific linguistic tendencies. Such chapters that do provide some evidences are still limited in the number of subjects, and thus, the generalizability of their results. Nevertheless, this shortcoming does not in essence detract from the strength of the present volume. Actually, it could be said that each chapter, while filling one gap in the literature of sports discourse analysis, simultaneously opens another avenue of academic research in the application of corpus linguistic methods to a wider range of real-word language contexts. As such, the present volume serves as an indispensable step for future research which, if based on larger corpora of sports language, will corroborate the present observations and offer new insights into the specificity of computer-mediated sports language.

Biographical Note

Shuyi Amelia Sun is a student researcher in the School of Languages and Cultures at the University of Queensland. Her areas of research expertise include corpus linguistics, quantitative linguistics, English for Academic Purposes (EAP), and natural language processing (NLP). Her previous experience includes publications in academic journals and conferences, reviewing articles for *Open Linguistics*, managing the Student Research Training Program (SRTP) as project leader, and working as teaching assistants for the undergraduate programs of English Language and Literature as well as Software Engineering.

References

Bateman, J. (2014). Using multimodal corpora for empirical research. In C. Jewitt (Ed.), *The Routledge Handbook of Multimodal Analysis* (pp. 238-252). London and New York: Routledge.

Bateman, J., Wildfeuer, J., & Hiippala, T. (2017). *Multimodality: Foundations, research and analysis–A problem-oriented introduction*. Berlin: De Gruyter Mouton. https://www.degruyter.com/view/title/521291

Bergh, G., & Ohlander, S. (2017). Loan translations versus direct loans: The impact of English on European football lexis. *Nordic Journal of Linguistics*, *40*(1), 5-35. doi: https://doi.org/10.1017/S0332586517000014

Herring, S. C. (1996). *Computer-mediated communication: Linguistic, social, and cross-cultural perspectives*. Amsterdam and Philadelphis: John Benjamins.

Lin, L., & Evans, S. (2012). Structural patterns in empirical research articles: Across-disciplinary study. *English for Specific Purposes*, *31*(3): 150-160. doi: https://doi.org/10.1016/j.esp.2011.10.002

Liu, D. & Farha, B. (1996). Three strikes and you're out: A study of the use of football and baseball jargon in present-day American English. *English Today*, *12*(1): 36-40. doi: https://doi.org/10.1017/S0266078400008828

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